doi:10.1088/1751-8113/48/11/11FT01

J. Phys. A: Math. Theor. 48 (2015) 11FT01 (12pp)

Journal of Physics A: Mathematical and Theoretical

doi.10.1000/1731-0110/40/11/111

Fast Track Communication

Delocalization of boundary states in disordered topological insulators

Andrew M Essin¹ and Victor Gurarie²

¹ Institute for Quantum Information and Matter and Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA

² Department of Physics, CB390, University of Colorado, Boulder, CO 80309, USA

E-mail: victor.gurarie@colorado.edu

Received 17 October 2014, revised 28 January 2015 Accepted for publication 3 February 2015 Published 23 February 2015



Abstract

We use the method of bulk-boundary correspondence of topological invariants to show that disordered topological insulators have at least one delocalized state at their boundary at zero energy. Those insulators which do not have chiral (sublattice) symmetry have in addition the whole band of delocalized states at their boundary, with the zero energy state lying in the middle of the band. This result was previously conjectured based on the anticipated properties of the supersymmetric (or replicated) sigma models with WZW-type terms, as well as verified in some cases using numerical simulations and a variety of other arguments. Here we derive this result generally, in arbitrary number of dimensions, and without relying on the description in the language of sigma models.

Keywords: topological insulators, quenched disorder, Anderson localization

1. Introduction

Topological insulators are non-interacting fermionic systems which are bulk insulators, have gapless excitations at their boundary, and which are characterized by topological invariants. As free fermion systems, they are described by the Hamiltonian

$$\hat{H} = \sum_{\alpha\beta} \mathcal{H}_{\alpha\beta} \hat{a}^{\dagger}_{\alpha} \hat{a}_{\beta}.$$
⁽¹⁾

Here α , β label points in space, spin and flavor of the fermions.

A procedure developed by Volovik in the 80s [1] to relate the edge states of twodimensional (2D) integer Hall systems to the bulk topological invariants in a direct way was recently generalized to all topological insulators [2]. In effect, that procedure showed that an edge of a topological insulator is a topological metal, characterized by its own topological invariant whose value must be equal to the value of the bulk invariant. All of this was done in the absence of disorder.

Here we show that a suitable modification of this formalism extends it to the case when the topological insulators are disordered. It immediately follows from this formalism that the edge of topological insulators cannot be fully localized by disorder. Furthermore, it is well known that topological insulators can be split into those without chiral (sublattice) symmetry and with chiral symmetry [3]. It can then be shown that the edge of non-chiral disordered topological insulators are characterized by a band of delocalized state spanning the energy interval between the delocalized bulk states. They are delocalized along the edge while exponentially decaying into the bulk, as is expected from the proper edge states. At the same time, the edge of chiral topological insulators has at least one delocalized state at zero energy while the rest of the states may be localized by disorder.

These statements generally match what is expected from the topological insulators from the study of sigma models or by using other methods. In particular, the localization of the one-dimensional edge of 2D topological insulators is very well understood. There is no doubt, for example, that the edge of an integer Hall state is delocalized regardless of disorder, thanks to its chiral nature (absence of backscattering). Similar arguments can be made in case of other 2D topological insulators.

The three-dimensional (3D) topological insulators were also studied in recent years. Their 2D boundaries, in the presence of disorder, can be analyzed using sigma models. Those can belong to one of five symmetry classes. Of those, arguably the most important is the AII topological insulator, or the standard strong time-reversal invariant topological insulators with spin–orbit coupling [4, 5]. It has no sublattice symmetry. That insulator is known to have a fully delocalized edge, as confirmed in a variety of studies [6, 7]. This agrees with our claim that all non-chiral insulators have a band of delocalized states at the boundary.

The remaining four insulators in 3D are all chiral. Among them, the simplest is the insulator with sublattice symmetry only, known as AIII topological insulator. Its edge is described by 2D Dirac fermions with random gauge potential [8]. Surprisingly, it was shown only a few years ago that when the random gauge potential is zero on the average these insulators have either a fully delocalized edge or an edge with localized states with localization length which diverges as energy is taken to zero, depending on whether their topological invariant is even or odd [9]. This results is obtained by mapping the problem at finite energy into a Pruisken-type sigma model with a topological term which corresponds to exactly the point of the integer quantum Hall transition it describes if the invariant is odd integer, and to the localized quantum Hall plateau if the invariant is even integer. However, this result is not completely robust: adding a constant magnetic field to the random gauge potential shifts the coefficient of the Pruisken-type sigma model, potentially localizing all states except those at zero energy, in all cases [10].

The next is the insulator in class DIII represented by a superfluid ³He in its phase B. By mapping it into a sigma model, one finds that, similarly to the AIII case, this insulator has an edge which is either fully delocalized if the invariant is odd (like for ³He where it is exactly 1) or with just one delocalized state at zero energy if the invariant is even. Technically this occurs because at finite energy an edge of such an insulator crosses over to the symmetry class AII which has a \mathbb{Z}_2 structure.

The insulator in class CI (represented by an exotic spin-singlet superconductor [11]) is known to have a fully localized edge except one state in the middle of the band, since at finite energy its edge crosses over to class AI, time-reversal invariant spin rotation invariant systems which were known for a very long time to be localized in two-dimensions [12]. Finally, the topological insulator in class CII, the most exotic of the five topological insulators in three-dimensions, has a fully localized edge except one state, as can be argued based on its mapping to the trivial (non-topological) AII insulator at finite energy.

All of these examples match what follows from the arguments which are presented in this paper. However, we note that a variety of chiral insulators, at least in three-dimensions, have a fully delocalized edge going beyond the prediction of at least one delocalized state given here. Whether our method can be generalized to explain these additional features is not known.

On the other hand, our method works not only in two or three-dimensions, but also for all topological insulators of arbitrarily large spacial dimension where sigma models might be difficult to analyze.

Finally we would like to point out that there exists an alternative method of studying topological insulators with disorder, based on the non-commutative Chern number and its generalizations [13–17]. These methods provide yet another way to look at the boundary of topological insulators with disorder [13], which may well prove to be more powerful than the methods discussed here.

2. Topological invariants of disordered insulators

We start with non-chiral topological insulators in even-dimensional space d. When disorder is absent they are characterized by the topological invariant which can be constructed out of its Green's function

$$G = [i\omega - \mathcal{H}]^{-1}.$$
(2)

Assuming translational invariance, the Green's function takes a form of a matrix $G_{ab}(\omega, \mathbf{k})$ which depends on the frequency ω and the *d*-dimensional lattice momentum \mathbf{k} with indices *a*, *b* labeling bands as well as spin and flavor. The topological invariant is known to take the form

$$N_d = C_d \epsilon_{\alpha_0 \dots \alpha_d} \operatorname{tr} \int \mathrm{d}\omega \mathrm{d}^d k \ G^{-1} \partial_{\alpha_0} G \dots G^{-1} \partial_{\alpha_d} G.$$
(3)

Here C_d is a constant which makes N_d an integer, known to be given by

$$C_d = -\frac{\left(\frac{d}{2}\right)!}{(2\pi i)^{\frac{d}{2}+1}(d+1)!}.$$
(4)

Each of the indices $\alpha_0, \alpha_1, ..., \alpha_d$ in equation (3) is actually implicitly summed over the values 0, 1, ..., d, and a convenient shorthand notation is introduced

$$\partial_{\alpha_i} \equiv \frac{\partial}{\partial k_{\alpha_i}},$$

where furthermore k_1, k_2, \dots, k_d are the cartesian components of the *d*-dimensional vector \mathbf{k} , with an additional notation $k_0 \equiv \omega$. This type of a summation will occur throughout the paper. To simplify notations, from now on to describe it we will simply say that each of the indices α is being summed over ω , k_1, \dots, k_d . Other notations in equation (3) include ϵ as the Levi–Civita symbol and tr as the trace over the matrix indices of G_{ab} . The integration over ω is taken from $-\infty$ to ∞ , while the integration over \mathbf{k} is over the first Brillouin zone of the lattice where equation (1) is defined. This invariant, if equation (2) is taken into account, is nothing but the Chern number of negative energy (filled) bands at d = 2, second Chern number at d = 4, etc.

The fact that equation (3) is an integer valued topological invariant relies on the existence of the homotopy group

$$\pi_{d+1}(\mathrm{GL}(\mathcal{N},\mathbb{C})) = \mathbb{Z}, d \text{ even}, \tag{5}$$

and on the fact that Green's functions given in equation (2) are not singular (have neither zero nor infinite eigenvalues; the latter follows from the system being an insulator and \mathcal{H} having no zero energy eigenvalues). The equivalence of the topological invariant equation (3) with the Chern number was first discussed in [18]. More generally, the invariants of this form, inducing the precise coefficient given in equation (4) in front to make them integer, are introduced and discussed in [1, 3]. It is actually very straightforward to check that equation (3) is an invariant as any perturbation of the form $G \rightarrow G + \delta G$ keeps equation (3) unchanged. It is a little more difficult to check that it always produces integer values [3].

Once the insulator is disordered, it is no longer translationally invariant and equation (3) loses any meaning. An alternative form for the topological invariant can be introduced in the following way.

Following [18] (see also [19]), introduce the finite size system such that the wave function for each particle satisfies periodic boundary conditions with an additional phases θ_i , where i = 1, ..., d labels d directions in space. In other words

$$\psi(x_i + L_i) = e^{i\theta_i}\psi(x) \tag{6}$$

for a particular coordinate x_i , where L_i is the size of the system in the *i*th direction. Then one introduces a Green's function $G_{\alpha\beta}(\theta)$ which is no longer Fourier transformed, but which depends on the *d* angles θ . Here α , β label not only spin and flavor of the fermions but also the sites of the lattice. The topological invariant is given by essentially exactly the same formula, (3), but just interpreted in a slightly different way. The indices $\alpha_0, \dots, \alpha_d$ are now summed over $\omega, \theta_1, \dots, \theta_d$, and the symbol tr implies summation over all the matrix indices, while the integral is performed over $d\omega d^d \theta$

$$N_d = C_d \ \epsilon_{\alpha_0...\alpha_d} \ \text{tr} \ \int d\omega d^d \theta \ G^{-1} \partial_{\alpha_0} G ... G^{-1} \partial_{\alpha_d} G.$$
(7)

Here the integration over each θ extends from $-\pi$ to π and the trace is over the matrix indices of $G_{\alpha\beta}$.

In case when there is no disorder, the invariant introduced in this way coincides with the invariant defined in terms of momenta. Indeed, we can take advantage of translational invariance and reintroduce the momenta in equation (7). Due to the periodic boundary conditions with the phases, the momenta are restricted to the values (for each of the *d* directions) $k_i^{(n)} = (2\pi n_i + \theta_i)/L_i$ with n_i being integers. Integration over θ_i and summation over n_i together are now equivalent to an unrestricted integration over all the values of *k*, as in equation (3). Thus in the absence of disorder, equations (3) and (7) are equivalent. Yet unlike equation (3), the expression in terms of phases equation (7) can be used even in the presence of disorder when translational invariance is broken.

If d is odd, then equation (3) is zero. Instead we follow [3] and consider insulators with chiral symmetry, such that there is a matrix Σ

$$\Sigma \mathcal{H} \Sigma = -\mathcal{H}.$$
 (8)

As well known, only the insulators with this symmetry have invariants of the integer type in odd spatial dimensions. The invariant itself, without disorder, can be written as follows



Figure 1. A *d*-dimensional system with a domain wall separating two phases, one with a topological invariant N^L at large negative coordinate *s* and one with a topological invariant N^R at large positive coordinate *s*. The insulator is translationally invariant in the direction perpendicular to *s* so that direction can be spanned by the d - 1 momenta *k*.

$$N_d = \frac{C_{d-1}}{2} \epsilon_{\alpha_1 \dots \alpha_d} \operatorname{tr} \int \mathrm{d}^d k \ \Sigma \ V^{-1} \partial_{\alpha_1} V \dots V^{-1} \partial_{\alpha_d} V.$$
(9)

Here $\alpha_1, \ldots, \alpha_d$ are summed over k_1 to k_d each, and

$$V = G|_{\omega=0} = -\mathcal{H}^{-1}.$$
 (10)

(We could equally well use \mathcal{H} instead of V in the definition of the topological invariant, but use V to smoothen the difference between even d and odd d cases.) Again, in case if there is disorder present (which preserves symmetry equation (8), we can rewrite the invariant in terms of phases

$$N_d = \frac{C_{d-1}}{2} \epsilon_{\alpha_1 \alpha_2 \dots \alpha_d} \operatorname{tr} \int d^d \theta \ \Sigma \ V^{-1} \partial_{\alpha_1} V \dots V^{-1} \partial_{\alpha_d} V.$$
(11)

Here $\alpha_1, \ldots, \alpha_d$ are summed over values θ_1 to θ_d each.

Again, simple arguments can be given that these two expressions for N_d in case when there is no disorder coincide. Once the disorder is switched on, equation (9) loses any meaning, while equation (11) can still be used.

We will not separately study the invariants of the type \mathbb{Z}_2 , as those can be obtained by dimensional reduction from the invariants of the type \mathbb{Z} introduced above. This will be briefly discussed at the end of the paper.

3. Boundary of topological insulators

3.1. Boundary of a disorder-free topological insulator

Following our prior work [2], we would like to consider a situation where a domain wall is present such that the topological insulator is characterized by one value of the topological invariant on the one side of the domain wall and another value on another side. We would like to examine the nature of the edge states forming at the boundary.

Let us first review the approach taken in [2] in case when there is no disorder. It is possible to introduce the Green's function of the entire system $G_{ab}(\omega; k_1, ..., k_{d-1}; s, s')$. Here k_1, \dots, k_{d-1} span the boundary separating two insulators, and s is the coordinate perpendicular to the boundary where the system is not translationally invariant, see figure 1. With its help it was furthermore possible to introduce the Wigner transformed Green's function J. Phys. A: Math. Theor. 48 (2015) 11FT01

$$G_{ab}^{W}(\omega, \mathbf{k}, R) = \int dr \ e^{-ik_{d}r} \ G_{ab}\left(\omega; k_{1}, ..., k_{d-1}; R + \frac{r}{2}, R - \frac{r}{2}\right).$$
(12)

Here **k** now denotes all *d* momenta k_1, \ldots, k_d . We can now introduce the concept of a Green's function on the far left of the boundary and the far right of the boundary

$$G^{R}(\omega, \mathbf{k}) = \lim_{R \to \infty} G^{W}(\omega, \mathbf{k}, R),$$

$$G^{L}(\omega, \mathbf{k}) = \lim_{R \to -\infty} G^{W}(\omega, \mathbf{k}, R).$$
(13)

These can now be used to calculate the topological invariant on the right and on the left of the boundary, N_d^R and N_d^L respectively, according to equations (3) or (9), depending on whether d is even or odd, with G^R and G^L substituted for G. Since we are considering a situation where the boundary separate two topologically distinct states, these two values are distinct, $N_d^R \neq N_d^L$.

Furthermore, as it was discussed in [2], there is also a boundary topological invariant, which can be defined with the help of the original Green's function $G_{ab}(\omega; k_1, ..., k_{d-1}; s, s')$ by

$$N_d^B = C_{d-2} \epsilon_{\alpha_0 \alpha_1 \dots \alpha_{d-1}} \int dn^{\alpha_0} X_{\alpha_1 \dots \alpha_{d-1}},$$

$$X_{\alpha_1 \alpha_2 \dots \alpha_{d-1}} = \operatorname{Tr} \left[G^{-1} \circ \partial_{\alpha_1} G \circ \dots \circ \partial_{\alpha_{d-1}} G \right].$$
(14)

Here we introduced convenient notations $A \circ B$ and Tr A, following [2], where

$$(A \circ B)_{ac}(s, s'') = \sum_{b} \int ds' A_{ab}(s, s') B_{bc}(s', s''),$$

Tr $A = \sum_{a} \int ds A_{aa}(s, s).$ (15)

The integral in equation (14) is over a d - 1-dimensional surface in the *d*-dimensional space formed by ω , k_1 , ..., k_{d-1} . n^{α_0} is a vector in this *d*-dimensional space normal to the surface which is being integrated over. This surface is closed and surrounds the singularities of *G* in this space (those are present because the boundary is not an insulator and has gapless excitations; thus at $\omega = 0$ the zero eigenvalues of \mathcal{H} , related to *G* via equation (2), make *G* singular). In fact, it is the presence of these singularities which makes equation (14) to be nonzero [2]; in their absence it can be shown that *X* is a divergence-free vector and thus the boundary invariant vanishes.

It was shown in [2] that

$$N_d^R - N_d^L = N_d^B. aga{16}$$

(A much earlier work [1] showed this for d = 2.)

If the space is of odd-dimensions, closely similar definition of N_d^B can be given, with equation (16) still valid. For completeness, let us give them here. We now have a system with a chiral symmetry, implying that

$$\Sigma G(\omega)\Sigma = -G(-\omega). \tag{17}$$



Figure 2. A *d*-dimensional system periodically repeated. Each rectangle represents a system with disorder. That disorder is identical in every spatial replica of the system. However, a certain parameter in the Hamiltonian varies from replica to replica, as well as within replicas, along the direction of the discrete coordinate *s* so that at some point the system has a boundary to a topologically distinct phase, indicated here by a think dashed line, whose direction is spanned by the phases θ_1 to θ_{d-1} . Far from the boundary that parameter asymptotically reaches the two values corresponding to the left phase with the topological invariant N^R .

The boundary invariant can now be defined with

$$N_d^B = \frac{C_{d-3}}{2} \epsilon_{\alpha_1,\dots,\alpha_{d-1}} \int dn^{\alpha_1} X_{\alpha_2\dots\alpha_{d-1}},$$

$$X_{\alpha_2\dots\alpha_{d-1}} = \operatorname{Tr} \Big[\Sigma V^{-1} \circ \partial_{\alpha_2} V \circ \dots \circ V^{-1} \partial_{\alpha_{d-1}} V \Big].$$
(18)

Here

$$V(k_1, k_2, ..., k_{d-1}; s, s') = G(\omega, k_1, ..., k_{d-1}; s, s')\Big|_{\omega=0}.$$
 (19)

Indices α_1 to α_{d-1} are summed over $k_1, k_2, ..., k_{d-1}$. The integral is over a d-2-dimensional surface in the d-1-dimensional space formed by $k_1, ..., k_{d-1}$. The bulk invariants can still be computed using equation (9), with $V^R = G^R \Big|_{\omega=0}$ substituted for \mathcal{H} , and similarly for V^L . The boundary invariant N_d^B can be useful in analyzing boundary theories of particular

The boundary invariant N_d^p can be useful in analyzing boundary theories of particular topological insulators. These boundary theories must have non-zero boundary invariants, which in case when the boundary is between a non-trivial insulator and an empty space (whose invariant is zero) must be equal, up to a sign, to the bulk invariant.

3.2. Boundary with disorder

We would like to generalize N_d^B to the case when disorder is present. It is clear that we need to replace the momenta k_1, \dots, k_{d-1} along the boundary with d - 1 phases across the boundary $\theta_1, \dots, \theta_{d-1}$. However, it is not immediately clear what to do in the direction perpendicular to the boundary.

In order to deal with this direction, we use the following trick. Imposing phases across the system is tantamount to periodically replicating our system in space (with disorder being exactly the same in each replica), with the phase θ becoming equivalent to the usual crystalline quasi-momentum. Recall that in the absence of disorder we work with variable *s* conjugate to the momentum k_d . A variable conjugate to the quasi-momentum is the discrete variable $m = 0, \pm 1, \pm 2, \ldots$ which labels replicas of the system.

Therefore, to maintain the continuity with the previous approach taken in the absence of disorder, we also periodically repeat the system in the direction perpendicular to the boundary. However, in that direction we should also ensure that the Hamiltonian changes close to the boundary, in such a way that once we move past the boundary the system goes into a different topological class with a different invariant. We can make the Hamiltonian change from a replica to a replica as well as within replicas (by varying some appropriate parameter in it) until the system goes through a transition somewhere within a particular

replica where the boundary between two topologically distinct states resides. Far away from the boundary on either side of it the Hamiltonian no longer changes from replica to replica. It is even possible, if desired, to make this parameter change across just the single replica, and staying constant in the rest of the replicas (although it we shall see below, it is advantageous to work in the limit when the parameter changes very little across each replica and it takes many replicas for it to reach its asymptotic far from the boundary value). This is schematically shown in figure 2.

Although this may appear to be a rather special kind of a setup with systems with periodically repeated disorder, it should be clear that this constitutes just a convenient approach which allows to probe the boundary, which really occurs within a particular replica of the system, as shown in figure 2. The rest of the replicated systems are there just for allowing the formalism presented here to describe simultaneously the boundary and the insulator far away from the boundary. Assuming that there are only two phases in the replicated systems, the one in replicas on the left of the boundary, and the one in the replicas on the right of the boundary, there is only one boundary here which separates topologically distinct systems (the boundaries between replicas do not separate systems which are topologically distinct).

Labeling the replicas of the system in the direction perpendicular to the boundary by the (discrete) variable m, we can define the Green's function of the replicated system as $G(\omega; \theta_1, ..., \theta_{d-1}; m, m')$. Here $\theta_i, i = 1, ..., d - 1$ phases are quasimomenta (or phases) across the d – 1-dimensional boundary, and the remaining variables m, m' label copies of the system in the direction perpendicular to the boundary.

Given this Green's function we define the Wigner transformed function by Fourier series (compare with equation (12))

$$G^{W}(\omega; \theta_{1}, ..., \theta_{d}; M) = \sum_{s} e^{-is\theta_{d}} G\left(\omega; \theta_{1}, ..., \theta_{d-1}; M + \frac{m}{2}, M - \frac{m}{2}\right).$$
(20)

Here M can be either integer or half-integer, and the summation over m goes over either even integers or odd integers respectively. Given this function, we can again find the far left and far right Green's function

$$G^{R}(\omega; \theta_{1}, ..., \theta_{d}) = \lim_{M \to \infty} G^{W}(\omega; \theta_{1}, ..., \theta_{d}, M),$$

$$G^{L}(\omega; \theta_{1}, ..., \theta_{d}) = \lim_{M \to -\infty} G^{W}(\omega; \theta_{1}, ..., \theta_{d}, M).$$
(21)

These can be used to define the left and right invariants according to equation (7) for even *d*, with $G^{L,R}$ substituted for *G*, or according to equation (11) for odd *d*, with $G^{L,R}\Big|_{\omega=0}$ substituted for *V*.

Now the difference $N^R - N^L$ is again expected to be equal to N^B , the boundary invariant. This invariant is defined analogously with equations (14) and (18) for *d* even and odd respectively, with the momenta replaced by the phases. If *d* is even

$$N_{d}^{B} = C_{d-2}\epsilon_{\alpha_{0}\alpha_{1}...\alpha_{d-1}} \int \mathrm{d}n^{\alpha_{0}} X_{\alpha_{1}...\alpha_{d-1}},$$

$$X_{\alpha_{1}...\alpha_{d-1}} = \mathrm{Tr} \Big[G^{-1} \circ \partial_{\alpha_{1}} G \circ \cdots \circ \partial_{\alpha_{d-1}} G \Big],$$
(22)

where the indices α_i are summed over ω , and $\theta_1, ..., \theta_{d-1}$, and the integration over *s* (implicit in the definitions of Tr and $A \circ B$ in equation (15)) is replaced by summation over *m*.

Similarly for d odd we can use construct the boundary invariant by using

$$N_d^B = \frac{C_{d-3}}{2} \epsilon_{\alpha_1,\dots,\alpha_{d-1}} \int \mathrm{d}n^{\alpha_1} X_{\alpha_2\dots\alpha_{d-1}},$$

$$X_{\alpha_2\dots\alpha_{d-1}} = \operatorname{Tr} \Big[\Sigma V^{-1} \circ \partial_{\alpha_2} V \circ \dots \circ V^{-1} \partial_{\alpha_{d-1}} V \Big],$$
(23)

where $V = G \mid_{\omega=0}$, and α_i are summed over $\theta_1, ..., \theta_{d-1}$.

The integrals in equations (22) and (23) are taken over closed surfaces in the *d*-dimensional (if *d* even) or d - 1-dimensional (if *d* is odd) space.

The derivation of equation (16) for this disordered case is essentially the same as in the disorderless case presented in [2]. The only (minor) difference is that here M and m are discrete variables. The derivation of [2] relies on gradient expansion in powers of derivatives with respect to the variable M and it seems to be essential in this technique for M to be a continuous variable. In order to work with M as if it were continuous variable we need to ensure that our system changes little as M is increased by 1, or in other words the parameter of the system whose change results in the topological phase transition changes just slightly as one goes from replica to replica. Therefore, at this stage of the derivation we have to assume that the system changes just slightly from replica to replica as one moves in the direction perpendicular to the boundary (in particular, a setup where the parameter controlling which phase the system is in changes only within one replica should not be used here). However, as always in the gradient expansion of integer valued topological invariants, small corrections to them coming from terms neglected as M is made a continuous variable must vanish to ensure that the result is still an integer. Similar issues are discussed in a somewhat related context in [2]. As a result, we expect that even as the variations in the parameters from replica to replica become larger, the approximation of continuous M remains exact. Thus we come to the conclusion that when properly defined equation (16) holds also in the presence of disorder.

3.3. Analysis of the boundary invariants in the presence of disorder

We would now like to analyze the boundary topological invariants. Take the invariant from the equation (22) (applicable when *d* is even). The integral in equation (22) is computed over a closed surface in the *d*-dimensional space formed by the frequency ω and d - 1 phases spanning the boundary of the system. The choice of the surface is arbitrary, as long as it encloses the points or surfaces where *G* is singular (see [2] for the discussion concerning the surface choice).

It is then natural to choose as a surface to be integrated over the two planes at two values of the phase $\theta_{d-1} = \pm \Lambda$, as it is done in our prior work [2] (the choice of θ_{d-1} is arbitrary; any of θ_i can be chosen for this construction). Here Λ is such that all the singularities of G lie between these two planes in the θ -space. Then the boundary invariant can be rewritten as essentially a bulk invariant in the d - 1-dimensional space formed by frequency and d - 2phases across the boundary, $\theta_1, \ldots, \theta_{d-2}$, with θ_{d-1} fixed. More precisely, we can define

$$N_{d-2}(\theta_{d-1}) = C_{d-2} \epsilon_{\alpha_0...\alpha_{d-2}} \int d\omega d^{d-2}\theta X_{\alpha_0...\alpha_{d-2}},$$

$$X_{\alpha_0...\alpha_{d-2}} = \operatorname{Tr} \Big[G^{-1} \circ \partial_{\alpha_0} G \circ \dots G^{-1} \circ \partial_{\alpha_{d-2}} G \Big].$$
(24)

Here $\alpha_0, \dots, \alpha_{d-2}$ are summed over $\omega, \theta_1, \dots, \theta_{d-2}$.

Then we can rewrite N_d^B as

$$N_d^B = N_{d-2}(\Lambda) - N_{d-2}(-\Lambda).$$
(25)

Therefore, if N_d^B is not zero, N_{d-2} is a function of θ_{d-1} in such a way that as θ_{d-1} changes from $-\Lambda$ to Λ , N_{d-2} changes by an amount equal to N_d^B . And as we discussed if the boundary we study is the boundary of a topological insulator with non-zero bulk invariant N_d , $N_d^B = N_d \neq 0$.

Now N_{d-2} is a topological invariant. The only way for it to change as a function of θ_{d-1} is if there is some special value $-\Lambda < \theta_{d-1}^c < \Lambda$ such that at θ_{d-1} equal to this value, *G* becomes singular. Then at that value N_{d-2} is not well defined, and the difference $N_{d-2}(\Lambda) - N_{d-2}(-\Lambda)$ can be non-zero.

G is related to the Hamiltonian by $G = [i\omega - \mathcal{H}]^{-1}$. The only way for it to be singular if $\omega = 0$ and \mathcal{H} has a zero eigenvalue. That means, there is a single particle energy level at the boundary whose energy is zero at $\theta_{d-1} = \theta_{d-1}^c$. At the same time, when $\theta_{d-1} = \pm \Lambda$, that energy level is not zero.

We can now invoke a well known criterion of localization [20] which states that a localized level's energy cannot shift as a function of the phase imposed across a disordered system. Therefore, in order for N_d^B to be non-zero, there has to be at least one energy level whose energy depends on θ_{d-1} . That level must be delocalized. (To be even more precise, this argument gives support to having at least one energy level which is not exponentially localized, as one can perhaps argue that some levels which are not exponentially localized are sometimes not fully delocalized either.) For this argument to work, we must also keep the lengths of the system L_i , introduced in equation (6), to be much larger than the localization lengths encountered in the localized states.

Furthermore, the system we study must have more than one delocalized energy level. Indeed, zero was an arbitrary reference point for the energy. We can always consider a Hamiltonian shifted by some chemical potential μ (a position-independent constant)

$$\mathcal{H}' = \mathcal{H} - \mu. \tag{26}$$

We can repeat all the arguments for this shifted Hamiltonian. As long as this shift does not change N_d (the bulk invariant), there should be a delocalized state at new zero energy, or at energy μ of the original model. Now we can anticipate that the bulk system has states at all energies, but states in the energy interval $-\Delta$ to Δ are all localized. This concept of Δ generalizes the concept of a gap in case of a disorder-free system. Then for any $-\Delta < \mu < \Delta$, the bulk invariant N_d is insensitive to μ . Then the system has a delocalized state at any energy μ which spans the interval $-\Delta$ to Δ . This concludes the argument about the delocalized states at the edge of any even-dimensional insulator.

The situation with odd-dimensional insulators is somewhat more restrictive. Their boundary invariant given in equation (23) can still be rewritten in a way equivalent to equation (25), with $N_{d-2}(\theta_{d-1})$ given by

$$N_{d-2}(\theta_{d-1}) = \frac{C_{d-3}}{2} \epsilon_{\alpha_1 \dots \alpha_{d-2}} \int d^{d-2}\theta \ X_{\alpha_1 \dots \alpha_{d-2}},$$

$$X_{\alpha_1 \dots \alpha_{d-2}} = \operatorname{Tr} \Big[\Sigma V^{-1} \circ \partial_{\alpha_1} V \circ \dots \circ V^{-1} \partial_{\alpha_{d-2}} V \Big],$$
(27)

where as before $V = G|_{\omega=0}$. Just as before, at the boundary of a topologic insulator $N_d = N_d^B = N_{d-2}(\Lambda) - N_{d-2}(-\Lambda) \neq 0$. It again follows that there is an energy level which crosses zero as a function of θ_{d-1} at some value θ_{d-1}^c . That level must be delocalized.

However, no other delocalized levels can be generally expected. Indeed, the Hamiltonian cannot be shifted by a chemical potential, as before, because we must ensure the symmetry

represented by equation (8). An arbitrary chemical potential added to the Hamiltonian breaks it. Therefore, we expect only one delocalized state close to zero energy (in the limit of the infinitely large system, exactly at zero energy). All other states are generally localized.

This is indeed what is expected from the chiral systems as explained in the introduction. Some of the chiral systems have only one delocalized state at zero energy. Yet others have many delocalized states spanning some energy interval centered around zero, similarly to the non-chiral disordered insulators. The arguments given here cannot establish which of the chiral systems will have a fully delocalized edge. All one can establish is that at least one state at zero energy must be delocalized.

3.4. Edge of a topological insulator with \mathbb{Z}_2 invariant

Finally, let us address the rest of topological insulators described by a invariant \mathbb{Z}_2 taking just two values, 0 and 1. All topological insulators are described by either an integer invariant of the types discussed here earlier or the invariant of the type \mathbb{Z}_2 as is well summarized in [3].

Systems with \mathbb{Z}_2 topological invariant can be understood as a dimensional reduction of the system with an invariant described in this paper by residing in higher dimensions [3, 21]. This construction can easily be incorporated into the formalism used here, as discussed in [2]. Relying on the arguments from this work, we can imagine that higher dimensional system have disorder which does not vary in the spatial direction we plan to eliminate. Those dimensions can be spanned by momenta in the Green's functions, while other dimensional are still spanned by phases. Setting those momenta to zero we obtain the dimensionally reduced system with the invariant \mathbb{Z}_2 . The boundary states of both higher dimensional \mathbb{Z} 'parent' and lower dimensional \mathbb{Z}_2 'descendant' system must be delocalized in the same way. Thus we find that the boundary of \mathbb{Z}_2 topological insulators have a fully delocalized band if they do not have chiral symmetry or at least a single delocalized state at zero energy if they do have chiral symmetry.

4. Conclusions

We examined the boundary states of disordered topological insulators and were able to understand their localization properties by directly examining the topological invariants of these disordered systems. In doing so, we reproduced what should be considered a widely anticipated answer. Nevertheless it was only for two and some 3D insulators that this answer has been derived. Further elaborations were based on the approach of the sigma models with WZW-type terms, and on their mostly conjectured behavior (although in some low dimensional cases this can be derived). Here we derived this answer without any conjectures in an arbitrary number of dimensions.

Finally in view of the existence of other methods to look at boundaries of topological insulators [17], it would be interesting to further explore the connection between these methods and the one discussed here and see if this could shed additional light on the structure of the boundary states in disordered topological insulators.

Acknowledgments

The authors are grateful to P Ostrovsky for sharing insights concerning the boundaries of three-dimensional disordered topological insulators. This work was supported by the NSF grants DMR-1205303 and PHY-1211914 (VG), and by the Institute for Quantum Information



and Matter, an NSF Physics Frontiers Center with support of the Gordon and Betty Moore Foundation through Grant GBMF1250 (AE).

References

- [1] Volovik G E 2003 The Universe in a Helium Droplet (Oxford: Oxford University Press) pp 275-81
- [2] Essin A and Gurarie V 2011 Phys. Rev. B 84 125132
- [3] Ryu S, Schnyder A P, Furusaki A and Ludwig A W W 2010 New J. Phys. 12 065010
- [4] Moore J E and Balents L 2007 Phys. Rev. B 75 121306
- [5] Fu L, Kane C L and Mele E J 2007 Phys. Rev. Lett. 98 106803
- [6] Bardarson J H, Tworzydło J, Brouwer P W and Beenakker C W J 2007 Phys. Rev. Lett. 99 106801
- [7] Nomura K, Koshino M and Ryu S 2007 Phys. Rev. Lett. 99 146806
- [8] Ludwig A W W, Fisher M P A, Shankar R and Grinstein G 1994 Phys. Rev. B 50 7526
- [9] Ostrovsky P M, Gornyi I V and Mirlin A D 2007 Phys. Rev. Lett. 98 256801
- [10] Mirlin A 2013 private communication
- [11] Schnyder A P, Ryu S and Ludwig A W W 2009 Phys. Rev. Lett. 102 196804
- [12] Abrahams E, Anderson P W, Licciardello D C and Ramakrishnan T V 1979 Phys. Rev. Lett. 42 673
- [13] Kellendonk J, Richter T and Schulz-Baldes H 2002 Rev. Math. Phys. 14 87
- [14] Kellendonk J and Schulz-Baldes H 2003 Commun. Math. Phys. 249 611
- [15] Loring T A and Hastings M B 2010 Europhys. Lett. 92 67004
- [16] Loring T A and Hastings M B 2011 Ann. Phys. 326 1699
- [17] Prodan E 2011 J. Phys. A: Math. Theor. 44 113001
- [18] Niu Q, Thouless D J and Wu Y S 1985 *Phys. Rev.* B **31** 3372
- [19] Essin A M and Moore J E 2007 Phys. Rev. B 76 165307
- [20] Edwards J T and Thouless D J 1972 J. Phys. C: Solid State Phys. 5 807
- [21] Qi X-L, Hughes T L and Zhang S-C 2008 Phys. Rev. B 78 195424