PHYSICAL REVIEW B **90**, 235137 (2014)

Classifying symmetry-protected topological phases through the anomalous action of the symmetry on the edge

Dominic V. Else[1] and Chetan Nayak[1,2]

[1]*Department of Physics, University of California, Santa Barbara, California 93106, USA*

[2]*Microsoft Research, Station Q, Elings Hall, University of California, Santa Barbara, California 93106, USA*

(Received 29 September 2014; revised manuscript received 26 November 2014; published 22 December 2014)

It is well known that $(1+1)$-dimensional [$(1+1)$-D] bosonic symmetry-protected topological (SPT) phases with symmetry group G can be identified by the projective representation of the symmetry at the edge. Here we generalize this result to higher dimensions. We assume that the representation of the symmetry on the spatial edge of a $(d+1)$-D SPT is *local* but not necessarily *on site*, such that there is an obstruction to its implementation on a region with a boundary. We show that such obstructions are classified by the cohomology group $H^{d+1}(G,\mathrm{U}(1))$, in agreement with the classification of bosonic SPT phases proposed in Chen *et al.* [Science **338**, 1604 (2012)]. Our analysis allows for a straightforward calculation of the element of $H^{d+1}(G,\mathrm{U}(1))$ corresponding to physically meaningful models such as nonlinear σ models with a θ term in the action. SPT phases outside the classification of Chen *et al.* are those in which the symmetry cannot be represented locally on the edge. With some modifications, our framework can also be applied to fermionic systems in $(2+1)$-D.

DOI: 10.1103/PhysRevB.90.235137 PACS number(s): 71.10.Pm

The classification of phases of matter in quantum systems at zero temperature has proven to be much richer than in classical statistical mechanical systems. For many such phases, the feature which distinguishes them from other phases is quantum mechanical and not related to the spontaneous breaking of a symmetry. One such family of quantum phases which has been much studied in recent years is the *symmetry-protected topological* (SPT) phases [1–43]. A system with a symmetry is considered to lie in a SPT phase if (a) the symmetry is not spontaneously broken and (b) the system can be connected to one whose ground state is a trivial product state without a phase transition, but *only* if we allow the symmetry to be broken explicitly. In some sense, SPT phases are "trivial" in the bulk, but boundaries between different SPT phases are nontrivial and must either be gapless, break the symmetry (explicitly or spontaneously), or be topologically ordered.

The central problem in the study of SPT phases is classifying the different phases that can occur for a given symmetry. In bosonic systems with an internal symmetry group G, an early result was that in $(1+1)$-dimensional [$(1+1)$-D] systems, the possible SPT phases are classified [9,10] by the *second cohomology group* $H^2(G,\mathrm{U}(1))$. This result has a natural interpretation [11] in terms of the symmetry transformation properties of an edge between the SPT and vacuum (or equivalently, of the entanglement spectrum [12]). Such an edge will, in general, transform *projectively* under the symmetry. The second cohomology group arises naturally from a consideration of these projective representations.

It has been argued [13] that, more generally, the SPT phases in d spatial dimensions are classified by the cohomology group $H^{d+1}(G,\mathrm{U}(1))$. This result was based on an explicit construction of field theories in discrete space-time which are believed to be representative of each SPT phase. However, making a definitive identification between these lattice field theories and other, more physically motivated, descriptions of the corresponding SPT phases [14–16,22] has proved difficult. In this paper, therefore, we propose to recast the cohomological classification in a different, hopefully more intuitive viewpoint, inspired by the original $(1+1)$-D treatment. The central idea is that, just as in the $(1+1)$-D case, the symmetry transformation on the edge of a $(d+1)$-D system will be, in some sense, anomalous [44–47]. Specifically, if we have a system defined on a d-dimensional spatial manifold M_{bulk} with a boundary, the edge symmetry acts on the boundary $\partial M_{\mathrm{bulk}}$, which itself has no boundary [$\partial(\partial M_{\mathrm{bulk}})=0$]. Therefore, there might be an obstruction to implementing the edge symmetry in a consistent way on a $(d-1)$-dimensional manifold M *with* boundary $\partial M \neq 0$. We argue that this obstruction is indeed classified by the cohomology group $H^{d+1}(G,\mathrm{U}(1))$. [For $(2+1)$-D systems, our approach is related to, though more general than, that of Ref. [44], which was based on a tensor-network representation for the edge symmetry.] In fact, in $(2+1)$-D our approach also leads to a classification of SPT phases in interacting fermion systems, as we show.

The remainder of this paper is organized as follows. In Sec. I, we give the general demonstration that the obstruction is classified by $H^{d+1}(G,\mathrm{U}(1))$. For $(2+1)$-D SPTs, this argument can be given in full generality (assuming only that the symmetry acts *locally* on the edge), but in higher dimensions we need to make additional assumptions about the form of the symmetry. In Sec. II, we discuss, by way of illustration, a simple example of an anomalous symmetry that appears on the edge of a $(2+1)$-D SPT. In Sec. III, we use the ideas of this paper to prove that $(2+1)$-D SPT phases characterized by different elements of $H^3(G,\mathrm{U}(1))$ are necessarily separated by a phase transition unless the symmetry is broken explicitly. In Sec. IV, we show how to use our approach to derive the element of the cohomology group corresponding to nonlinear σ models containing a topological term. In Sec. V, we make explicit the connection between our work and the original classification of Ref. [13]. In Sec. VI, we explain why, in the presence of antiunitary symmetries, there exist bosonic SPT phases not captured by our arguments. In Sec. VII, we show how our ideas can be applied also to fermionic systems in $(2+1)$-D.

I. THE GENERAL FORMALISM

Consider a system in a bosonic SPT phase. By definition, this means it is gapped and nondegenerate in the bulk and (disregarding symmetry considerations) can be continuously connected to a product state without a phase transition. However, in a system with a boundary, we can define an effective low-energy theory for the boundary, which may be gapless notwithstanding the gap in the bulk. A key property of SPT phases is that the boundary theory of an SPT phase in d spatial dimensions can always be realized at the microscopic level in a strictly $(d-1)$-dimensional system (see Appendix A for a careful proof of this well-known fact). This is in contrast to, for example, integer quantum Hall states in which the boundary is chiral and cannot be realized as a stand-alone system [48]. For SPT phases, the anomalous nature of the edge arises not from the boundary theory itself but from the way it is acted upon by the symmetry.

We assume that the symmetry in the bulk is unitary and on site; that is, for a lattice system with N sites, the symmetry group G is is represented as a unitary tensor product $U(g) = [u(g)]^{\otimes N}$ of operators acting on each site. (We may need to group several sites together into a single effective site in order to satisfy this condition.) We now consider the low-energy Hilbert space of states with energies below some cutoff that is less than the bulk gap; these states are edge excitations. Projecting the unitary representation of the symmetry group onto this low-energy Hilbert space, we obtain a unitary representation, acting only on the boundary degrees of freedom, that may not be on site. On the contrary, it appears to be a characteristic of nontrivial SPT phases that the symmetry is realized on the boundary in a fundamentally non-on-site way [13,44,45]. Nevertheless, the key assumption that we make in this paper is that the boundary symmetry, albeit not on site, is nevertheless still *local* in the sense of Ref. [8] (e.g., it can be represented as a finite-depth quantum circuit.) This seems to be a natural assumption, but we expect it to be violated by SPT phases not captured by the cohomological classification (see Sec. VI for further discussion).

For a non-on-site symmetry, there is the possibility that there is an obstruction to implementing the symmetry on a manifold with boundary in a consistent way. We intend to show that, by classifying these obstructions, one recovers the cohomological classification of SPT phases. A simple example of this idea is the well-known connection between (1+1)-D SPTs and the projective symmetry transformation of the edge [9–11], which we now review.

A. (1 + 1)-D SPTs

The boundary of a 1D system simply comprises a pair of points a and b (see Fig. 1). Let $U(g)$ be the representation of the symmetry group G on this boundary. Assuming that we chose the system size such that the end points a and b are well separated (i.e., by a distance large compared to all intrinsic length scales), the locality of $U(g)$ simply implies that it must act on a and b separately; that is, it must be a tensor product $U(g) = U_a(g) \otimes U_b(g)$. We can think of $U_a(g)$ as the *restriction* of $U(g)$ to the point a. Importantly, however, this restriction is uniquely defined only modulo

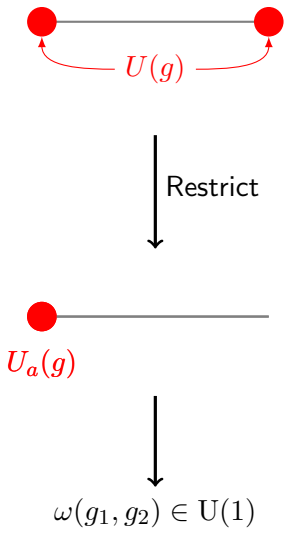

FIG. 1. (Color online) Obtaining a 2-cocycle on the $(0+1)$-D edge of a $(1+1)$-D SPT.

phase factors. Indeed, $U(g)$ is left invariant under $U_a(g) \to \beta(g)U_a(g), U_b(g) \to \beta(g)^{-1}U_b(g)$ for any U(1)-valued function $\beta(g)$. Thus, while $U(g)$ is always a representation of the symmetry group G, that is, $U(g_1)U(g_2) = U(g_1g_2)$, the nonuniqueness of the restriction procedure implies that $U_a(g)$ need only be a *projective* representation of G, which is to say that $U_a(g_1)U_a(g_2) = \omega(g_1,g_2)U_a(g_1g_2)$ for some U(1)-valued function $\omega(g_1,g_2)$. The function ω describes the obstruction to consistently (i.e., nonprojectively) implementing the symmetry on the point a.

Since multiplication of the U_a's must be associative, one can derive a consistency condition on ω by evaluating $U_a(g_1)U_a(g_2)U_a(g_3)$ in two different ways, namely,

$$\omega(g_1,g_2)\omega(g_1g_2,g_3) = \omega(g_2,g_3)\omega(g_1,g_2g_3). \tag{1}$$

A function ω satisfying Eq. (1) is known as a 2-cocycle. Furthermore, due to the fact that $U_a(g)$ is only defined up to a g-dependent phase factor $\beta(g)$, it follows that we have an equivalence relation on 2-cocycles:

$$\omega(g_1,g_2) \sim \omega(g_1,g_2)\beta(g_1)\beta(g_2)\beta(g_1g_2)^{-1}. \tag{2}$$

The group of 2-cocycles quotiented by the above equivalence relation is, by definition, the second cohomology group $H^2(G,\mathrm{U}(1))$. One can then show that two models are in the same SPT phase if and only if they correspond to the same element of $H^2(G,\mathrm{U}(1))$. Therefore, SPT phases in (1+1)-D are classified by $H^2(G,\mathrm{U}(1))$.

B. (2 + 1)-D SPTs

When presented as it was above, the $(1+1)$-D case suggests an obvious generalization to higher dimensions: We consider the symmetry $U(g)$ acting on the boundary C, then *restrict* it to a subregion M, which in general is a manifold with boundary (C itself has no boundary as it is the boundary of a higher-dimensional manifold), to see if the symmetry is implemented consistently or not.

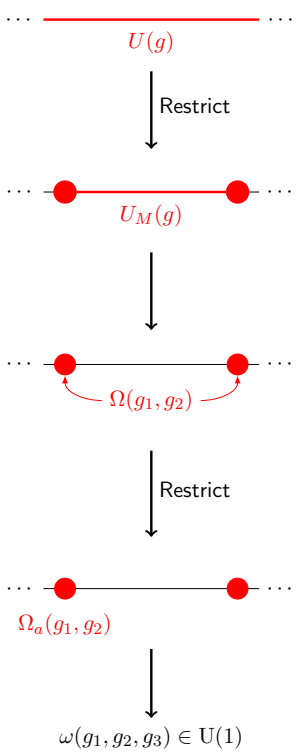

FIG. 2. (Color online) Obtaining a 3-cocycle on the $(1+1)$-D edge of a $(2+1)$-D system.

First, we need to give a more general definition of what it means to restrict a local unitary U acting on a spatial manifold C to a submanifold M, which for the case discussed above was obvious due to the tensor-product structure. Specifically, we say that a local unitary U_M acting on the region M is the restriction of U to the region M if it acts the same as U in the *interior* of M, well away from the boundary ∂M. We observe two properties about this restriction.

(a) It always exists for any local unitary. This can easily be seen from, for example, the quantum circuit description.

(b) It is defined modulo local unitaries acting in the vicinity of the boundary ∂M.

The second property is the higher-dimensional generalization of the restriction being defined only up to phase factors. Thus, in general, if $U(g)$ is a representation of the symmetry group G, then $U_M(g)$ need only satisfy

$$U_M(g_1)U_M(g_2) = \Omega(g_1,g_2)U_M(g_1g_2), \tag{3}$$

where $\Omega(g_1,g_2)$ is a local unitary acting in the vicinity of ∂M, which represents the obstruction to a consistent representation on M due to the fact that it is a manifold with a boundary. Thus, we have reduced the problem of classifying local unitary representations $U(g)$ on a d-dimensional manifold to that of classifying local unitary obstructions $\Omega(g_1,g_2)$ on a $(d-1)$-dimensional manifold. The idea now is to perform more such reductions, each time reducing by 1 the dimensionality of the manifold acted upon, until we get down to the simplest case of 0 dimensions (i.e., points).

For $(2+1)$-D SPTs, this reduction can be completed as follows (see Fig. 2). In this case the boundary has only one spatial dimension, and so $\Omega(g_1,g_2)$ as constructed above already acts on just a pair of points a and b. We observe that Eq. (3), together with the associativity of the operators $U_M(g)$, implies that Ω must satisfy

$$\Omega(g_1,g_2)\Omega(g_1g_2,g_3) = {}^{U_M(g_1)}\Omega(g_2,g_3)\Omega(g_1,g_2g_3), \tag{4}$$

which is a non-Abelian analog of Eq. (1) and where we have introduced the conjugation notation ${}^x y = xyx^{-1}$. Now we perform a second restriction, from $\partial M = \{a,b\}$ to the single point a. The restriction $\Omega \to \Omega_a$ is defined only up to phase factors, and so we conclude that Ω_a satisfies Eq. (4) only up to phase factors

$$\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3) = \omega(g_1,g_2,g_3)\,{}^{U_M(g_1)}\Omega_a(g_2,g_3) \times \Omega_a(g_1,g_2g_3), \tag{5}$$

where $\omega(g_1,g_2,g_3) \in \mathrm{U}(1)$. We show in Appendix B that ω must satisfy the 3-cocycle condition

$$\omega(g_1,g_2,g_3)\omega(g_1g_2,g_3,g_4)^{-1}\omega(g_1,g_2g_3,g_4)\omega(g_1,g_2,g_3g_4)^{-1} \times \omega(g_2,g_3,g_4) = 1. \tag{6}$$

Furthermore, as $\Omega_a(g,g')$ is only defined up to phase factors $\beta(g,g')$, we must identify

$$\omega(g_1,g_2,g_3) \sim \omega(g_1,g_2,g_3)\,\beta(g_1,g_2)\beta(g_1g_2,g_3) \times \beta(g_2,g_3)^{-1}\beta(g_1,g_2g_3)^{-1}. \tag{7}$$

We show in Appendix B that, up to equivalence, the choice of restriction $U(g) \to U_M(g)$ does not affect the 3-cocycle. The group of 3-cocycles quotiented by the equivalence relation Eq. (7) is, by definition, the third cohomology group $H^3(G,\mathrm{U}(1))$. Hence, we recover the cohomological classification of $(2+1)$-D SPTs.

C. Higher dimensions

In higher dimensions it is not clear whether we can still do the reduction procedure in complete generality as in the $(2+1)$-D case. Nevertheless, we can still perform the reduction if we make some simplifying assumptions about the action of the symmetry on the boundary. (The nonlinear σ models discussed in Sec. IV are a nontrivial example in which the symmetry on the edge takes the required form.) Specifically, we consider a symmetry group G acting on a Hilbert space equipped with a set of basis states labeled by the variables $\alpha(x)$ associated with each spatial location in a closed $(d-1)$-dimensional space C_1. We can take the spatial coordinate x to be either discrete (i.e., a lattice) or continuous. The class of symmetry actions that we consider are those that can be written in the form

$$U(g) = N(g)S(g), \tag{8}$$

such that

(a) $S(g)$ is the on-site part of the symmetry which can be written in the form

$$S(g) = \sum_\alpha |g\alpha\rangle\langle\alpha|, \tag{9}$$

where $\alpha \to g\alpha$ is some on-site action of the symmetry on the classical labels α; and

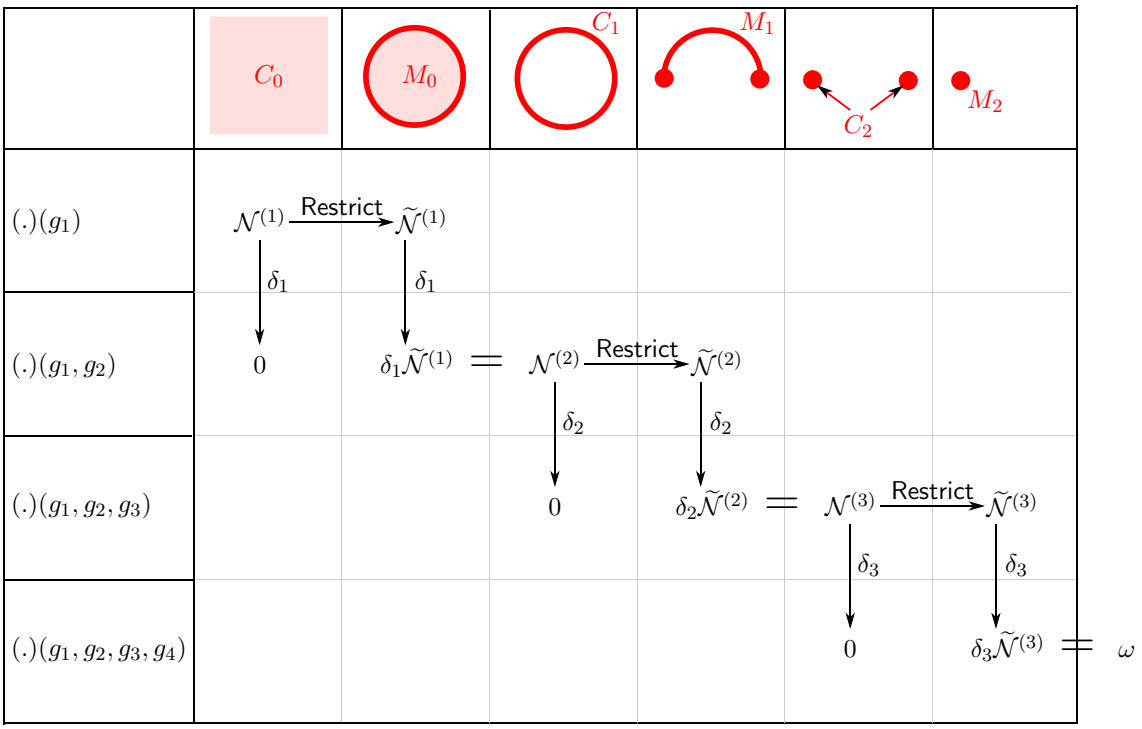

FIG. 3. (Color online) The reduction process to obtain a 4-cocycle ω on the $(2+1)$-D edge of a $(3+1)$-D system, assuming a symmetry representation on the edge of the form Eq. (8).

(b) in the same basis, the non-on-site part $N(g)$ is diagonal, namely,

$$N(g) = \sum_{\alpha} e^{i\mathcal{N}^{(1)}(g)[\alpha]} |\alpha\rangle\langle\alpha|, \tag{10}$$

where $\mathcal{N}(g)$ are functionals of the configuration α. We require these functionals to be sufficiently local that $N(g)$, and hence $U(g)$, are local unitaries.

The requirement that $U(g)$ be a representation, $U(g_1)U(g_2) = U(g_1g_2)$, can be written in terms of the functionals $\mathcal{N}(g)$ as

$$g_1\mathcal{N}^{(1)}(g_2) + \mathcal{N}^{(1)}(g_1) - \mathcal{N}^{(1)}(g_1g_2) = 0 \pmod{2\pi}, \tag{11}$$

where we have defined the action of group elements on functionals in the obvious way: $(g\mathcal{F})[\alpha] = \mathcal{F}[g^{-1}\alpha]$. Henceforth, we take the (mod 2π) to be implied, or, in other words, we consider the functionals to take values in $\mathbb{R}/(2\pi\mathbb{Z})$.

Now as before, we can restrict $U(g)$ to a subregion M_1 with a boundary, which [since $S(g)$ can be trivially restricted] amounts to restricting the functionals $\mathcal{N}^{(1)}(g)$. Then Eq. (11) need be satisfied by the restricted functionals $\widetilde{\mathcal{N}}^{(1)}(g)$ only up to boundary terms,

$$g_1\widetilde{\mathcal{N}}^{(1)}(g_2) + \widetilde{\mathcal{N}}^{(1)}(g_1) - \widetilde{\mathcal{N}}^{(1)}(g_1g_2) = \mathcal{N}^{(2)}(g_1,g_2), \tag{12}$$

where the $\mathcal{N}^{(2)}(g_1,g_2)$ are functionals which depend only on the value of α near the boundary ∂M_1 and describe the obstruction. This corresponds to Eq. (3).

In order to continue the reduction process, we find it useful to define the group coboundary operators δ_k which map functionals depending on k group elements into functionals depending on $k+1$ group elements, as follows:

$$\begin{aligned}(\delta_k\mathcal{N}^{(k)})(g_1,\ldots,g_{k+1}) &= g_1\mathcal{N}^{(k)}(g_2,\ldots,g_n) + (-1)^{k+1}\mathcal{N}^{(k)}(g_1,\ldots,g_k) \\ &\quad + \sum_{i=1}^{k}(-1)^i\mathcal{N}^{(k)}(g_1,\ldots,g_{i-1},g_ig_{i+1},g_{i+2},\ldots,g_{k+1}).\end{aligned} \tag{13}$$

In particular, $(\delta_1\mathcal{N}^{(1)})(g_1,g_2)$ corresponds to the left-hand side of Eq. (11). The important property which the coboundary operators satisfy is that they form a chain complex, i.e., $\delta_{k+1}\circ\delta_k = 0$.

We can now formulate the reduction process for symmetries acting on a manifold of spatial dimension d (see Fig. 3). At the kth step of the process, we have a set of functionals $\mathcal{N}^{(k)}$ acting on a closed $d-k$-dimensional manifold C_k and indexed by k group elements, satisfying $\delta_k\mathcal{N}^{(k)} = 0$. We then consider restrictions $\widetilde{\mathcal{N}}^{(k)}$ of these functionals onto the manifold M_k, where M_k is a submanifold of C_k with boundary. As $\widetilde{\mathcal{N}}^{(k)}$ must act the same as $\mathcal{N}^{(k)}$ in the *interior* of M_k, it follows that $\mathcal{N}^{(k+1)} \equiv \delta_k\widetilde{\mathcal{N}}^{(k)}$ acts on the boundary $\partial M_k \equiv C_{k+1}$. Furthermore, as $\delta_{k+1}\circ\delta_k = 0$, it follows that $\delta_{k+1}\mathcal{N}^{(k+1)} = 0$. Thus, we just iterate these reduction steps, terminating when we reach $\omega = \mathcal{N}^{(d+1)}$, which is simply a mapping from $d+1$ group elements to U(1) satisfying $\delta_{d+1}\mathcal{N}^{(d+1)}$; this the definition of a U(1)$(d+1)$-cocycle. Due to the ambiguity in the choice of restrictions, it follows that ω is only defined up to

$$\omega \sim \omega + \delta_{d+1}\lambda, \tag{14}$$

where λ is some element of U(1) depending on $d+1$ group elements. The group of $(d+1)$-cocycles quotiented by the equivalence relation Eq. (14) is, by definition, the cohomology group $H^{d+1}(G,\mathrm{U}(1))$. Thus, we recover the cohomological classification of SPT phases in arbitrary dimensions.

Finally, let us discuss the case of symmetry groups that contain antiunitary operations. It is perhaps unclear, in general, what is meant by restriction of an antiunitary operation (although see Ref. [49]). Nevertheless, if we consider only symmetries that can be represented as a suitable generalization of Eq. (8), the same arguments as above can be applied with only minor modifications. Specifically, we consider symmetries of the form

$$U(g) = N(g)S(g)K^{n(g)}, \tag{15}$$

where $N(g)$ and $S(g)$ are as before, K is complex conjugation in the $\{|\alpha\rangle\}$ basis, and $n(g)$ is 0 for unitary elements of G and 1 for antiunitary elements. If we define the action of G on functionals as $g\mathcal{F}[\alpha] = (-1)^{n(g)}\mathcal{F}[g^{-1}\alpha]$, all of the steps in the above derivation can be carried through without change, except that there is a residual nontrivial action of G on U(1). Thus, the classification is $H^{d+1}(G,\mathrm{U}(1))$, but with U(1) considered as a nontrivial G module, with antiunitary elements acting by complex conjugation.

II. EXAMPLE: "CHIRAL" SYMMETRY ON THE EDGE OF A (2 + 1)-D SPT

It was shown in Ref. [50] that the action of the symmetry on the gapless edge of some nontrivial (2 + 1)-D SPTs is "chiral," as expressed (for example) in the fact that it acts differently on the left- and right-moving fields. Let us show how this corresponds to a local but not on-site symmetry and calculate the corresponding 3-cocycle. We focus on the simplest case where the symmetry is just $\mathbb{Z}_2$, but similar arguments can be made for $\mathbb{Z}_n$ or U(1) symmetries.

We assume the low-energy theory of the (1 + 1)-D edge is described by a massless boson field φ with compactification radius 2π, i.e., a bosonic Luttinger liquid, with Lagrangian density

$$\mathcal{L} = \frac{g}{2\pi}\left[\frac{1}{v}(\partial_t\varphi)^2 - v(\partial_x\varphi)^2\right]. \tag{16}$$

We introduce the dual boson field θ according to $\partial_x\theta = 2\pi\Pi$, where Π is the canonical momentum conjugate to φ. The commutation relation for θ and φ is, therefore,

$$[\varphi(x),\theta(x')] = -2\pi i\Theta(x-x'), \tag{17}$$

where $\Theta(x)$ is the unit step function. Note that this definition, together with the fact that total angular momentum is quantized to integers, implies that θ is also an angular variable defined modulo 2π.

Now, suppose that the fields φ and θ transform under $\mathbb{Z}_2$ according to

$$\varphi \to \varphi + n\pi, \quad \theta \to \theta + m\pi. \tag{18}$$

Here $(n,m) = (1,0)$ corresponds to a normal on-site π rotation of the boson field. On the other hand, as we shall see, $(n,m) = (1,1)$ is the non-on-site symmetry that we would expect at the edge of a nontrivial $\mathbb{Z}_2$ SPT. Also, $m \neq 0$ corresponds to a superficially "chiral" symmetry in the sense that the left- and right-moving fields $\phi_{L,R} = \varphi \pm \theta$ transform differently under $\mathbb{Z}_2$, but in the $\mathbb{Z}_2$ case [though not for $\mathbb{Z}_n$ or U(1)] this chirality is not physically meaningful because $\theta \sim \theta + 2\pi$ so m is actually only defined modulo 2.

From the commutation relations (17), one can show that Eq. (18) is effected by the unitary operator $U = (-1)^{nL+mW} = N^m S^n$, where L is the total angular momentum and W is the total winding number, and we define

$$N = \exp\left(-\frac{i}{2}\int \partial_x\varphi\, dx\right), \tag{19}$$

$$S = \exp\left(-\frac{i}{2}\int \partial_x\theta dx\right). \tag{20}$$

We now define the restriction $U_{[a,b]} = N^m_{[a,b]}S^n_{[a,b]}$ to a finite interval $[a,b]$, where

$$N_{[a,b]} = \exp\left(-\frac{i}{2}\int_a^b \partial_x\varphi\, dx\right), \tag{21}$$

$$S_{[a,b]} = \exp\left(-\frac{i}{2}\int_{a-\epsilon}^{b+\epsilon} \partial_x\theta\, dx\right), \tag{22}$$

where we have made use of our freedom to redefine the restriction near the boundary of $[a,b]$ to shift the end points of the second integral by some small $\epsilon > 0$. This ensures that $N_{[a,b]}$ and $S_{[a,b]}$ commute. Hence, we find that $U^2_{[a,b]} = N^{2m}_{[a,b]}S^{2n}_{[a,b]}$, where

$$N^2_{[a,b]} = \exp\left(-i\int_a^b \partial_x\varphi dx\right) \tag{23}$$

$$= e^{i\varphi(a)}e^{-i\varphi(b)}, \tag{24}$$

$$S^2_{[a,b]} = \exp\left(-i\int_{a-\epsilon}^{b+\epsilon} \partial_x\theta dx\right) \tag{25}$$

$$= e^{i\theta(a-\epsilon)}e^{-i\theta(b+\epsilon)}. \tag{26}$$

Thus, as expected, we find that $\Omega \equiv U^2_{[a,b]} = [e^{in\theta(a-\epsilon)}e^{im\varphi(a)}][e^{-in\theta(b+\epsilon)}e^{-im\varphi(b)}] \equiv \Omega_a\Omega_b$ still acts nontrivially at the end points a and b even though $U^2 = 1$.

In the present example, Eq. (4) takes the form

$$U_{[a,b]}\Omega U^{-1}_{[a,b]} = \Omega, \tag{27}$$

and this equality can readily be verified directly from the forms of $U_{[a,b]}$ and Ω given above. On the other hand, the restriction Ω_a satisfies this equation, in general, only up to a phase factor. Indeed, we find

$$U_{[a,b]}\Omega_a U^{-1}_{[a,b]} = e^{-in\theta(a-\epsilon)}e^{-im\varphi(a)-imn\pi} = (-1)^{mn}\Omega_a. \tag{28}$$

Hence, we find that the 3-cocycle associated with the realization of $\mathbb{Z}_2$ is given by $\omega(X,X,X) = (-1)^{mn}$ and $\omega(g_1,g_2,g_3) = 1$ for $(g_1,g_2,g_3) \neq (X,X,X)$, where X is the generator of $\mathbb{Z}_2$. For $m = n = 1$ this corresponds to a nontrivial 3-cocycle, and the corresponding representation of $\mathbb{Z}_2$ would appear at the boundary of a nontrivial (2 + 1)-D $\mathbb{Z}_2$ SPT.

III. PROOF OF SEPARATION OF PHASES IN (2 + 1)-D

In this section, we outline how one can use the ideas given above to prove for $(2+1)$-D systems that systems characterized by different elements of the cohomology group $H^3(G,\mathrm{U}(1))$ must be separated by a bulk phase transition; the details are left to the appendixes. (Unfortunately, the proof cannot be applied in higher dimensions due to the lack of a completely general characterization of anomalous symmetry.)

First, as we want to make statements about bulk properties, we need to reformulate the ideas of Sec. I in a slightly different way, in terms of properties of the ground state in the bulk rather than the low-energy physics at the edge. We show in Appendix C that, given a general ground state $|\Psi\rangle$ in some SPT phase in d spatial dimensions ($d \leqslant 2$), and a region A in the bulk, one can find a representation $V_{\partial A}(g)$ of the symmetry group, which acts *inside* A, but only near the boundary ∂A, such that $U_A(g)|\Psi\rangle = V_{\partial A}(g)|\Psi\rangle$. Here $U_A(g)$ is the restriction of the symmetry onto the region A (which can be defined consistently since we are assuming the symmetry is represented on-site in the bulk.) The physical interpretation of this result is simply that, as $|\Psi\rangle$ is invariant under $U(g)$; therefore, $U_A(g)|\Psi\rangle$ can differ from $|\Psi\rangle$ only near the boundary ∂A. This representation $V_{\partial A}(g)$ can be anomalous in the same way as the representation of the symmetry on a physical edge, and the anomaly can be classified using the method of Sec. I.

The final result that we need is that the element of $H^3(G,\mathrm{U}(1))$ is independent of the choice of region A, *even* in the presence of spatial inhomogeneity; this is also proved in Appendix C . (Actually, as discussed in that appendix, we only prove this for certain regions A, but that is sufficient for the following discussion.) This allows us to prove that two systems $\mathcal{S}$ and $\mathcal{S}'$ characterized by different elements of $H^{d+1}(G,\mathrm{U}(1))$ must be separated by a phase transition [51]. Indeed, consider two systems connected without a phase transition. Then, without closing the gap, one can create an interpolated system that looks like $\mathcal{S}$ on some region A and like $\mathcal{S}'$ on another region A' (see Appendix C for a careful proof of this fact). It therefore follows that the same element of $H^{d+1}(G,\mathrm{U}(1))$ must be obtained in both cases. By a similar argument, one also finds that a spatial boundary between two different SPT phases must either be gapless or break the symmetry.

IV. NONLINEAR σ MODELS

It has been found [14,15,52,53] that a quite general way to reproduce the essential features of various SPT phases is through the field theory of a quantum nonlinear σ model (NLσM), where topological properties of the SPT phase arise out of the bulk θ term included in the action. Here we show in such models, the presence of the θ term indeed leads to an obstruction to on-site representation of the symmetry on a spatial edge, in such a way as to allow a straightforward calculation of the corresponding element of the cohomology group.

For example, consider in D space-time dimensions (i.e., $D = d+1$) a NLσM for the $(D+1)$ component vector field $\mathbf{n}$, constrained to have unit norm; i.e., $\mathbf{n}$ lies on a unit D sphere. The (Euclidean) action can be written as the sum of a dynamical contribution S_{dyn} and a topological contribution S_{top},

$$S^{\mathrm{bulk}} = S^{\mathrm{bulk}}_{\mathrm{dyn}} + S^{\mathrm{bulk}}_{\mathrm{top}}, \tag{29}$$

$$S^{\mathrm{bulk}}_{\mathrm{dyn}} = \frac{1}{g}\int d^D x\ \partial^\mu \mathbf{n}\cdot\partial_\mu \mathbf{n}, \tag{30}$$

$$S^{\mathrm{bulk}}_{\mathrm{top}} = i\Theta\frac{1}{V_D}\int \mathbf{n}^*(\omega_V), \tag{31}$$

where V_D is the volume of the unit D sphere and $\mathbf{n}^*(\omega_V)$ is the pullback through the map $\mathbf{n}$ of the volume form on the unit D sphere. Written componentwise, this amounts to

$$S^{\mathrm{bulk}}_{\mathrm{top}} = i\Theta\frac{1}{V_D}\int d^D x\ \epsilon^{a_1,\ldots,a_{D+2}} n^{a_1}\partial_0 n^{a_2}\partial_1 n^{a_2}\cdots\partial_{D-1}n^{a_D}, \tag{32}$$

where $\epsilon^{a_1,\ldots,a_{D+1}}$ is the $(D+1)$-dimensional Levi-Civita symbol. The θ term $S^{\mathrm{bulk}}_{\mathrm{top}}$ measures a topologically invariant "generalized winding number" in $\pi_D(S^D) \cong \mathbb{Z}$ and for space-times without a boundary is quantized to integer multiples of $i\Theta$. Hence, we implement the requirement that SPT phases be trivial in the bulk by setting Θ to be an integer multiple of 2π, thus ensuring that $S^{\mathrm{bulk}}_{\mathrm{top}}$ makes no contribution to the partition function $\int \mathcal{D}[\mathbf{n}]e^{-S}$. In fact, although we have given a specific form of $S^{\mathrm{bulk}}_{\mathrm{dyn}}$ for concreteness, it will not be important for our analysis as the topological features of the system are entirely captured by $S^{\mathrm{bulk}}_{\mathrm{top}}$.

Although the inclusion of $S^{\mathrm{bulk}}_{\mathrm{top}}$ has no effect on the partition function in the bulk, it does play a crucial role once we introduce a spatial edge. In that case $S^{\mathrm{bulk}}_{\mathrm{top}}$ depends (mod $2\pi i$) only on the values of $\mathbf{n}$ on the boundary (to see this, note that any two extensions into the bulk can be connected at the boundary to give a closed surface, on which $e^{-S_{\mathrm{top}}} = 1$); the action on the boundary is referred to as the *Wess-Zumino-Witten* action S_{WZW}. Thus, we can integrate out the gapped bulk to give an effective action for the low-energy excitations on the edge of the form

$$\exp(-S^{\mathrm{edge}}) = \exp\left(-S^{\mathrm{edge}}_{\mathrm{dyn}} - S_{\mathrm{WZW}}\right), \tag{33}$$

where $S^{\mathrm{edge}}_{\mathrm{dyn}} = \int d^d x \mathcal{L}^{\mathrm{edge}}_{\mathrm{dyn}}$ is some unimportant dynamical term derived from $S^{\mathrm{bulk}}_{\mathrm{dyn}}$. Note that one can then write $S_{\mathrm{WZW}} = \int d^d x \mathcal{L}_{\mathrm{WZW}}$ for some local Lagrangian density $\mathcal{L}_{\mathrm{WZW}}$ defined on the edge. However, there is no canonical way to do so.

Now let us consider the symmetry group G in the bulk corresponding to some invertible action $\mathbf{n} \to g\mathbf{n}$ for $g \in G$. We demand that $S^{\mathrm{bulk}}_{\mathrm{dyn}}$ and S_{top} be *locally* invariant under the symmetry, i.e., that the integrands in Eqs. (30) and (31) must be invariant, not just the integral. Then we expect that $S^{\mathrm{edge}}_{\mathrm{dyn}}$ is also locally invariant under the symmetry. S_{WZW} must also be globally invariant (at least, modulo $2\pi i$), but, in general, we do not expect it to be locally invariant. Indeed, because there is no canonical choice for $\mathcal{L}_{\mathrm{WZW}}$, one expects that the symmetry will transform $\mathcal{L}_{\mathrm{WZW}}$ to a different Lagrangian that nevertheless integrates to the same action (modulo $2\pi i$) in a space-time without boundary.

We now show that, after quantization, the lack of local invariance of S_{WZW} implies the non-on-site nature of the

unitary representation of the symmetry on the edge. We assume that after quantization the Hilbert space is spanned by a basis of states labeled by spatial configurations of $\mathbf{n}$ at a fixed time. We can calculate the imaginary-time propagator $e^{-\beta H}$ (or equivalently, the Hamiltonian H) by a path integral

$$\langle \mathbf{n}'|e^{-\beta H}|\mathbf{n}\rangle = \int \mathcal{D}[\mathbf{n}(\tau)] e^{-S^{\text{edge}}\{0,\beta\}}, \tag{34}$$

where

$$S^{\text{edge}}\{0,\beta\} = \int d^{D-2}x \int_0^\beta d\tau \left(\mathcal{L}^{\text{edge}}_{\text{dyn}} + \mathcal{L}_{\text{WZW}}\right) \tag{35}$$

is the action evaluated on a space-time with temporal boundaries at $\tau = 0$ and $\tau = \beta$. Now so far we only know that S_{WZW} is globally invariant (modulo $2\pi i$) on a space-time manifold without boundary. Since S_{WZW} is not *locally* invariant, in the presence of a temporal boundary we can only conclude that it will transform as $S_{\text{WZW}}\{0,\beta\} \to g S_{\text{WZW}}\{0,\beta\}$ ($g \in G$), where the difference can be expressed in terms of the field configurations at the temporal boundaries,

$$\begin{aligned} &g S_{\text{WZW}}\{0,\beta\} - S_{\text{WZW}}\{0,\beta\} \\ &\quad = i\mathcal{N}(g)[\mathbf{n}(\tau)] - i\mathcal{N}(g)[\mathbf{n}(0)] \pmod{2\pi i}, \end{aligned} \tag{36}$$

where $\mathcal{N}(g)$ is a functional of the field configuration at a fixed time.

Equation (36) implies that the edge Hamiltonian is not invariant under the naive on-site implementation of the symmetry, $S(g) = \int \mathcal{D}[\mathbf{n}]\, |g\mathbf{n}\rangle\langle\mathbf{n}|$. Indeed, combined with Eq. (34), we find

$$\langle \mathbf{n}'|S(g)^\dagger e^{-\beta H} S(g)|\mathbf{n}\rangle = e^{i\mathcal{N}(g)[\mathbf{n}'] - i\mathcal{N}(g)[\mathbf{n}]} \langle \mathbf{n}'|e^{-\beta H}|\mathbf{n}\rangle \tag{37}$$

$$= \langle \mathbf{n}'|N(g)^\dagger e^{-\beta H} N(g)|\mathbf{n}\rangle, \tag{38}$$

where

$$N(g) = \int \mathcal{D}[\mathbf{n}] e^{i\mathcal{N}(g)[\mathbf{n}]} |\mathbf{n}\rangle\langle\mathbf{n}|. \tag{39}$$

Hence, we see that the correct implementation of the symmetry on the edge, which does commute with the Hamiltonian, is $U(g) = N(g)S(g)$. In general, there is no reason to expect $N(g)$ to be on-site, as we shall see. However, as we show in Sec. IV A, it is necessarily local. Thus, the symmetry on the edge is a local but non-on-site symmetry precisely of the form considered in Sec. I C, and we can calculate the appropriate element of the cohomology group using the reduction procedure of that section.

We can also consider antiunitary symmetries by a straightforward extension of the above considerations. Specifically, an antiunitary symmetry is implemented in the action by $\mathbf{n} \to g\mathbf{n}$, $i \to (-1)^{n(g)} i$. Then we find that the representation of the symmetry on the edge is $U(g) = N(g)S(g)K^{n(g)}$, with $N(g)$ and $S(g)$ as before and K complex conjugation in the $\mathbf{n}$ basis.

A. Calculating the cocycle in nonlinear σ models using U(1) cochains on the target manifold

A particularly compact and elegant way of calculating the cocycle for NLσMs is by interpreting the θ term in terms of a U(1) cochain defined on the target manifold $T = S^D$. First we need to state some definitions. We refer to k-dimensional oriented integration domains on a manifold T as k chains. Given a k chain A, we denote the opposite orientation by $-A$, and we can also define a sum operation on k chains in the natural way, so that the k chains can be viewed as an additive group. (If one wanted to be rigorous, one would define k chains as formal linear combinations of oriented k simplices with integer coefficients.) A U(1) k *cochain* is a linear mapping from k chains to U(1) [which we here write additively as $\mathbb{R}/(2\pi\mathbb{Z})$]. (Note that we are here referring to *topological* cochains on a manifold; these should be distinguished from the *group* cochains that are used to construct the group cohomology of some group G.) In particular, each differential k-form ω induces a U(1) k cochain by integration,

$$\omega(A) = \left(\int_A \omega\right) \bmod 2\pi, \tag{40}$$

where in an abuse of notation we denote the k form and the U(1) k cochain by the same symbol. Any U(1) k cochain ω on the target manifold T can be used to define a local U(1)-valued functional F_ω for a T-valued field $\mathbf{n}$ on a k-dimensional space(-time) manifold M via

$$F_\omega[\mathbf{n}] = \omega(\mathbf{n}(M)), \tag{41}$$

where $\mathbf{n}(M)$ is the image of M, viewed as a chain, under the mapping $\mathbf{n}$. If ω is derived from a differential k form, this is equivalent to defining F_ω as the integral of the pullback, $F_\omega[\mathbf{n}] = [\int_M \mathbf{n}^*(\omega)] \bmod 2\pi$. In particular, the topological θ term action of Eq. (31) is a special case of Eq. (41).

We define the coboundary operator d which maps k cochains to $(k+1)$ cochains according to

$$(d\omega)(A) = \omega(\partial A), \tag{42}$$

where ∂A is the boundary of A. We call a k cochain ω *exact* if it can be written as $\omega = d\kappa$ for some $(k-1)$ cochain κ. Our central tool is the following result.

Lemma 1. A U(1) k cochain ω on a manifold T is exact if and only if $\omega(C) = 0$ for all closed (i.e., boundaryless) k chains C.

Proof. See Appendix D.

The property that $\omega(C) = 0$ for closed C, in turn, is equivalent to requiring of the induced functional F_ω that it vanish on all closed space-time manifolds. If this is satisfied, then one expects that for a space-time manifold M with boundary, $F_\omega[\mathbf{n}]$ should depend only on the values of $\mathbf{n}$ on the boundary ∂M. Indeed, given $\omega = d\kappa$, one finds that

$$F_\omega[\mathbf{n}] = (d\kappa)(\mathbf{n}(M)) \tag{43}$$

$$= \kappa(\partial \mathbf{n}(M)) \tag{44}$$

$$= \kappa(\mathbf{n}(\partial M)) \tag{45}$$

$$\equiv F_\kappa[\mathbf{n}(\partial M)]. \tag{46}$$

Given the above considerations, one can show that the procedure for obtaining the edge symmetry from the θ term, and then the cocycle from the edge symmetry, can be reduced to a simple prescription in terms of the U(1) cochains defined on the target manifold, with no reference to the space-time manifold at all, which we now describe (see Table I).

We start from a topological action S_{top} on a space-time-manifold M with d-dimensional target manifold T, written as $S_{\text{top}}[\mathbf{n}] = F_{\omega^{(0)}}[\mathbf{n}] = \omega^{(0)}(\mathbf{n}(M))$, where $\omega^{(0)}$ is an exact U(1)

TABLE I. A tabular representation of the reduction process to extract a U(1) group 3-cocycle $\nu = \omega^{(3)}$ starting from a symmetric topological term in (2 + 1)-D represented by a topological U(1) cochain $\omega^{(0)}$. Each cell in the table is specified by a row label l and a column label k and corresponds to a set of k cochains labeled by l group elements. Going left in the table corresponds to applying the *topological* coboundary operator d, whereas going down corresponds to applying the *group* coboundary operator δ defined by Eq. (47). These two operations commute, so the table can be interpreted as a commutative diagram.

		$\leftarrow d$			
		3	2	1	0
δ	0	$\omega^{(0)}$	$\kappa^{(0)}$		
$\downarrow$	1	0	$\omega^{(1)}$	$\kappa^{(1)}$	
	2		0	$\omega^{(2)}$	$\kappa^{(2)}$
	3			0	$\omega^{(3)}$
	4				0

d cochain on T, which is invariant under the action of the symmetry, $g\omega^{(0)} = \omega^{(0)}$. Here we defined the action of the symmetry on a cochain by $g\omega(A) = (-1)^{n(g)}\omega(gA)$, where $n(g)$ is 1 for antiunitary elements and 0 for unitary elements, and the action of g on chains is derived from its action on $\mathbf{n}$. Hence, we have $\delta_0\omega^{(0)} = g\omega^{(0)} - \omega^{(0)} = 0$, where we have introduced the *group* coboundary operators δ_k (*not* the same as the topological coboundary operator d defined above) in the same way as Eq. (13) above, namely,

$$\begin{aligned}(\delta_k\omega^{(k)})&(g_1,\ldots,g_{k+1})\\ &= g_1\omega^{(k)}(g_2,\ldots g_n) + (-1)^{k+1}\omega^{(k)}(g_1,\ldots,g_k)\\ &\quad + \sum_{i=1}^{k}(-1)^i\omega^{(k)}(g_1,\ldots,g_{i-1},g_ig_{i+1},g_{i+2},\ldots,g_{k+1}).\end{aligned} \tag{47}$$

Given a set of exact $(d-k)$ cochains $\omega^{(k)}$ indexed by k group elements which satisfy $\delta_k\omega^{(k)} = 0$, we can write $\omega^{(k)} = d\kappa^{(k)}$ for some set of $(d-k-1)$ cochains $\kappa^{(k)}$. Now $\delta_k\omega^{(k)} = 0$ implies that, for closed chains C, $(\delta_k\kappa^{(k)})(C) = (\delta_k\omega^{(k)})(\partial C) = 0$. Hence, we can define $\omega^{(k+1)} = \delta_k\kappa^{(k)}$ which is exact and satisfies $\delta_{k+1}\omega^{(k+1)} = 0$. The sequence terminates when we reach $\omega^{(D)}$, which is a set of 0 cochains indexed by k group elements. Now a 0 cochain is essentially just a scalar U(1) function defined on the target manifold T. However, the fact that $\omega^{(D)}$ evaluates to zero for the closed 0 chain $a - b$ (where a and b are any two points) implies that the $\omega^{(D)}$ are *constant* U(1) functions. Thus, $\omega^{(D)}$ defines a mapping from D group elements to U(1) satisfying $\delta_D\omega^{(D)} = 0$, which defines an element of the group cohomology group $H^D(G,\mathrm{U}(1))$.

B. Examples

The possible symmetry transformations that leave the Lagrangian of Eq. (29) invariant in space-time dimensions $D = 2,3,4$ were constructed in Ref. [53] for a variety of different symmetry groups. Our framework allows, in principle, for the element of the cohomology group $H^D(G,U(1))$ to be calculated in all of these cases. Let us consider a few examples.

1. Z_2^T in (1+1)-D

We write the symmetry group as $Z_2^T = \{1,\mathbb{T}\}$. The target manifold is S^2 and we work in spherical coordinates $\mathbf{n} = (\cos\theta, \sin\theta\cos\varphi, \sin\theta\sin\varphi)$. The action of $\mathbb{T}$ on $\mathbf{n}$ is $\mathbb{T}\mathbf{n} = -\mathbf{n}$, or in terms of the spherical coordinates, $\theta \to \pi - \theta, \varphi \to \varphi + \pi$. The initial U(1) cochain can be written in terms of a 2 form

$$\omega^{(0)} = \Theta\frac{1}{4\pi}\sin\theta(d\theta \wedge d\varphi). \tag{48}$$

As $\omega^{(0)}$ integrates to 0 (mod 2π) over the whole 2 sphere, it follows that it can be written as $\omega^{(0)} = d\kappa^{(0)}$ for some U(1)1 cochain $\kappa^{(0)}$. We can write $\kappa^{(0)}$ explicitly as

$$\kappa^{(0)} = \Theta\frac{1}{4\pi}(1-\cos\theta)d\varphi. \tag{49}$$

Treating $\kappa^{(0)}$ as a differential 1 form and taking the exterior derivative, one recovers Eq. (48). When written as a 1 form, $\kappa^{(0)}$ appears to have a singularity at $\theta = \pi$. To show that, as a U(1) 1 cochain, $\kappa^{(0)}$ is actually well defined and satisfies $d\kappa^{(0)} = \omega^{(0)}$ globally, it is sufficient to check that $\int_C \kappa^{(0)} = 0$ (mod 2π) for a loop C of infinitesimal size encircling the apparent singularity at $\theta = \pi$, which is indeed the case.

Now, following the general prescription of Sec. IV A, we define $\omega^{(1)} = \delta_0\kappa^{(0)}$. The only nontrivial component is

$$\omega^{(1)}(\mathbb{T}) = \mathbb{T}\kappa^{(0)} - \kappa^{(0)} \tag{50}$$

$$= -\frac{\Theta}{4\pi}(1+\cos\theta) - \frac{\Theta}{4\pi}(1-\cos\theta) \tag{51}$$

$$= -\frac{\Theta}{2\pi}d\varphi, \tag{52}$$

from which we immediately read off that $\omega^{(1)} = d\kappa^{(1)}$, where $\kappa^{(1)} = -\frac{\Theta}{2\pi}\varphi$ [which is well defined as a U(1) 0 cochain because φ is defined modulo 2π]. Thus, we can define the cocycle $\nu = \delta_1\kappa^{(1)}$, and the only nonzero component is

$$\nu(\mathbb{T},\mathbb{T}) = \mathbb{T}\kappa^{(1)} + \kappa^{(1)} \tag{53}$$

$$= \frac{\Theta}{2\pi}\{\varphi + \pi - \varphi\} \tag{54}$$

$$= \frac{\Theta}{2}. \tag{55}$$

Thus, if Θ is an odd multiple of 2π, this 2-cocycle corresponds to a nontrivial SPT phase, with the zero-dimensional boundary transforming projectively under the symmetry, i.e., as a Kramers doublet with $\mathbb{T}^2 = -1$. On the other hand, if Θ is an even multiple of 2π, we have a trivial SPT phase with the edge transforming as $\mathbb{T}^2 = 1$. Thus, by different choices of Θ one recovers both elements of the cohomology group $H^2(Z_2^T,\mathrm{U}(1)) \cong Z_2$.

2. Z_2 in (2+1)-D

We write the symmetry group as $Z_2 = \{1,X\}$. The target manifold is S^3 and we work in generalized spherical coordinates $\mathbf{n} = (\cos\theta, \sin\theta\mathbf{n}_2)$, where $\mathbf{n}_2 \in S^2$. The action of X on $\mathbf{n}$ is $X\mathbf{n} = -\mathbf{n}$, or in terms of the generalized

spherical coordinates $\theta \to \pi - \theta, \mathbf{n}_2 \to -\mathbf{n}_2$. The initial U(1) cochain is

$$\omega^{(0)} = \Theta \frac{1}{V_3} \sin^2\theta (d\theta \wedge \omega_{V,2}), \tag{56}$$

where V_3 is the volume of the 3 sphere, and $\omega_{V,2}$ is the volume form for $\mathbf{n}_2$. We then find that $\omega^{(0)} = d\kappa^{(0)}$, where

$$\kappa^{(0)} = \Theta \frac{1}{V_3} \left(\int_0^\theta \sin^2 x \, dx \right) \omega_{V,2}. \tag{57}$$

We observe that V_3 can be expressed as ($V_2 = 4\pi$),

$$V_3 = V_2 \int_0^\pi \sin^2\theta \, d\theta. \tag{58}$$

From this one can show that $\kappa^{(0)}$ is well defined despite the apparent singularity at $\theta = \pi$. Now the only nontrivial element of $\omega^{(1)} = d\kappa^{(1)}$ is

$$\omega^{(1)}(X) = X\kappa^{(0)} - \kappa^{(0)} \tag{59}$$

$$= \Theta \frac{1}{V_3} \omega_{V,2} \int_0^\pi \sin^2 x dx \tag{60}$$

$$= \Theta \frac{1}{4\pi} \omega_{V,2} \tag{61}$$

[here we used Eq. (58) and the fact that $\omega_{V,2}$ is odd under $\mathbf{n}_2 \to -\mathbf{n}_2$]. In fact, this is identical to Eq. (48). The reduction process then proceeds nearly identically to that in Sec. IV B 1 above and one finds that the only nonzero component of the 3-cocycle is

$$\nu(X,X,X) = \frac{\Theta}{2}. \tag{62}$$

Thus, one recovers both elements of $H^3(Z_2, U(1)) \cong Z_2$ for Θ an odd or even multiple of 2π, respectively.

V. LATTICE MODELS OF SPT PHASES

In Ref. [13], the classification of SPT phases in d spatial dimensions was based on an explicit construction of a field theory for a $(d+1)$-dimensional discrete space-time for each element of the cohomology group $H^{d+1}(G, \mathrm{U}(1))$. Although a discrete space-time is perhaps hard to interpret physically, the construction of Ref. [13] can also be used to derive a ground-state wave function on a spatial lattice; a gapped Hamiltonian with this wave function as its ground state constitutes an (albeit unrealistic) lattice Hamiltonian realizing the SPT phase. Hence, it is worthwhile to show that the symmetry on the edge of such of a lattice model is indeed classified under our scheme by the same element of the cohomology group that was used to construct the wave function. We do this in Appendix E . In particular, this shows that every element of the cohomology group $H^{d+1}(G, \mathrm{U}(1))$ can be realized in an explicit lattice model.

VI. BEYOND THE COHOMOLOGICAL CLASSIFICATION

It is now well established [15,26,28,54,55] that in $(3+1)$-D there exists an SPT phase with respect to time-reversal symmetry that is beyond the standard cohomological classification. The reason why this phase is outside the cohomological classification can be readily understood, as follows. Deriving the cohomological classification using arguments such as those presented in this paper requires at the very least the assumption that the symmetry can be implemented locally on a stand-alone realization of the edge. We now argue that the beyond-cohomology phase violates this assumption.

Indeed, one possible surface termination for the beyond-cohomology phase is a gapped "three-fermion" topological phase $\mathcal{F}$ in which all three nontrivial particle sectors are fermions. Any purely $(2+1)$-D realization of this phase is necessarily chiral; that is, its conjugate $\overline{\mathcal{F}}$ under time reversal cannot be connected to $\mathcal{F}$ without a phase transition. (One way to see this is to note that $\mathcal{F}$ and $\overline{\mathcal{F}}$ have opposite edge chiral central charges $c_- = \pm 4$ and hence a spatial boundary between them must be gapless. If we make the spatial variation from $\mathcal{F}$ to $\overline{\mathcal{F}}$ sufficiently slow, this gapless spatial boundary must be interpreted as a bulk phase transition [56]. Suppose that a state $|\Psi\rangle$ within the phase $\mathcal{F}$ could be invariant under a local antiunitary operation T. Then one can always write $T = U\mathbb{T}$, where $\mathbb{T}$ is the normal on-site representation of time reversal, and U is a local unitary. However, then, since $\mathbb{T}|\Psi\rangle$ is in the conjugate phase $\overline{\mathcal{F}}$, we see that U is a local unitary connecting $\mathcal{F}$ and $\overline{\mathcal{F}}$, which is a contradiction.

VII. FERMIONIC SYSTEMS

The restriction arguments given in Sec. I are quite general and therefore can be equally well applied to fermionic systems, at least in $(2+1)$-D. (Generalizing to higher dimensions would require one to find an appropriate fermionic equivalent of the special form of the symmetry considered in Sec. I C.) Here we discuss, in general terms, the issues arising which result in the fermionic classification differing from the bosonic one, with reference to a particular example of a Fermion SPT protected by a Z_2 symmetry. As the general classification is somewhat complicated, we we leave the details to Appendix F . It would be interesting to see whether it can be related to the "supercohomology" classification proposed in Ref. [27]. We consider only cases in which the symmetry is unitary and on-site. Thus, our classification will not include the well-known cases of topological insulators and superconductors [4–6], which are protected by nonunitary symmetries.

The first issue that needs to be considered is the privileged role of fermion parity. Any local fermionic system must be invariant under the fermion parity $(-1)^F$, where F is the total fermion number. Therefore, the fermionic symmetry group G_f characterizing a fermion SPT always contains fermion parity. This must commute with all the other elements of G_f if they describe local symmetries. If we now consider the $(1+1)$-D edge of a $(2+1)$-D SPT, by assumption it is realizable as a strictly $(1+1)$-D local fermion system. As this $(1+1)$-D system must always be invariant under the fermion parity of the edge, we expect that, in the realization of G_f on the $(1+1)$-D edge, the parity element is represented as the actual fermion parity of the edge. (This can be verified by using the techniques of Appendix A to construct the edge representation.) That is, by contrast to the bosonic case, the fermionic symmetry group contains an element that is *always realized on site on the boundary*. Furthermore, even when we restrict and consider the action of the symmetry on a finite interval, *the restricted operations must be local and therefore must still commute with*

the fermion parity (whereas there is no analogous requirement in the bosonic case.)

The other main difference from the fermionic case occurs when defining the restriction of the obstruction operator $\Omega(g_1,g_2)$, which acts on a pair of points a and b, to a single point a. At this point, one encounters a subtlety that was glossed over in the bosonic treatment. $\Omega(g_1,g_2)$ is clearly local in the sense ("locality preserving") that it maps local operators (including fermion creation and annihilation operators) to local operators under Heisenberg evolution. [We can deduce this from the fact that it is true for the $U_M(g)$s and that the locality preserving property is invariant under multiplication.] This does not necessarily imply [57,58] that it is a local unitary in the sense ("locally generated") that it can be written as the time evolution of a local fermionic Hamiltonian in a domain containing only the two points a and b. In other words, we might not be able to write $\Omega(g_1,g_2) = \Omega_a(g_1,g_2)\Omega_b(g_1,g_2)$, where Ω_a and Ω_b are fermionic local unitaries acting only near the points a and b. For example, the following unitary is locality preserving but not locally generated,

$$\Omega = (c_a + c_a^\dagger)(c_b + c_b^\dagger), \tag{63}$$

where $c_{a,b}$ are the annihilation operators for fermions at points a and b, respectively. If Ω is not locally generated, this presents an obstacle for defining the restriction $\Omega \to \Omega_a$. This problem was not present in the bosonic case because for bosonic systems locality preserving actually does imply locally generated for unitaries acting on a pair of points (for unitaries acting on higher-dimensional regions this is no longer the case [57,58].) Nevertheless, it is clear that, for the Ω given by Eq. (63) there is still a natural definition of "restriction" $\Omega_a = (c_a + c_a^\dagger)$, even though Ω_a is not really local. More generally, it is always true that $\Omega(g_1,g_2) = \Omega_a(g_1,g_2)\Omega_b(g_1,g_2)$, where $\Omega_a(g_1,g_2)$ is either a fermionic local unitary acting near the point a, or it is such a local unitary multiplied by $c_a + c_a^\dagger$ (and similarly for Ω_b). In the latter case, however, the *restricted operations* $\Omega_a(g_1,g_2)$ *and* $\Omega_b(g_1,g_2)$ *will anticommute rather than commute.* This anticommutation leads to fermionic corrections to the 3-cocycle condition [Eq. (6)], to the the equivalence relation [Eq. (7)], and to the product rule for "stacking" SPT phases; see Appendix F for more details.

Example: Fermionic SPT with Z_2 symmetry

In order to illustrate the ideas discussed above, let us consider a $(1+1)$-D field theory which we expect to describe the edge of a $(2+1)$-D fermionic SPT protected by a Z_2 symmetry. (This Z_2 is in addition to the always-present fermion parity; thus, the full fermionic symmetry group is $G_f = Z_2 \times Z_2^f$.) This theory is the fermionic analog of the bosonic edge we considered in Sec. II. The low-energy physics is described by a gapless Dirac point (which can emerge, for example, from a microscopic lattice model of noninteracting electrons with a Fermi surface.) Thus, we define the fermionic fields $\Psi_R(x)$ and $\Psi_L(x)$, corresponding to left- and right-moving fermions (in terms of the original lattice operators, these will be local on a length scale set by the energy cutoff). We can define the corresponding number operators $N_{L,R} = \int \Psi_{L,R}^\dagger \Psi_{L,R} dx$. The Hamiltonian is

$$H = J(N_L + N_R). \tag{64}$$

The fermion parity is $(-1)^{N_R+N_L}$ and sends $\Psi_L \to -\Psi_L, \Psi_R \to -\Psi_R$. We assume that the additional Z_2 symmetry is given by $U = (-1)^{N_R}$; thus, it acts only on the right movers and sends $\Psi_L \to \Psi_L, \Psi_R \to -\Psi_R$. This forbids perturbations like $\Psi_L^\dagger \Psi_R$, which would open up a gap, suggesting that the gapless edge is protected by the symmetry. Indeed, we show that the symmetry corresponds to a nontrivial fermionic cocycle.

We can define the restriction of the Z_2 symmetry to a finite interval $[a,b]$ according to

$$U_{[a,b]} = \exp\left(-i\pi \int_a^b \Psi_R^\dagger \Psi_R dx\right). \tag{65}$$

If we invoke the bosonization correspondences $\Psi_R^\dagger \Psi_R \sim \partial_x \phi_R(x)/(2\pi)$, $\Psi_R(x) \sim e^{i\phi_R}$, we see that $\Omega \equiv U_{[a,b]}^2 \sim \Psi_R(a)\Psi_R^\dagger(b) \equiv \Omega_a\Omega_b$. Thus, $U_{[a,b]}^2$ acts only on the end points as expected. However, this is an example of the possibility discussed above, of the operators Ω_a carrying nontrivial fermion parity.

The parity of Ω_a, which we call σ ($\sigma = -1$ in the current calculation) constitutes one aspect of the nontrivial fermionic cocycle. The other aspect comes from the relation $U_{[a,b]}\Omega U_{[a,b]}^\dagger = \Omega$. The restricted operations Ω_a satisfy this relation only up to a phase factor ω. To calculate this phase factor, we need to regularize the integral Eq. (65) by introducing a soft cutoff; that is, we replace Eq. (65) with

$$U_{[a,b]} = \exp\left(-i\pi \int_{-\infty}^{\infty} f(x)\Psi_R^\dagger \Psi_R dx\right), \tag{66}$$

where f is a smooth function such that $f(x) = 1$ for $x \in [a+\epsilon, b-\epsilon]$, and $f(x) = 0$ for $x < (a-\epsilon)$ or $x > (b+\epsilon)$. Using the bosonization correspondence to express $U_{[a,b]}^2$ in terms of $\partial_x \phi_R$ and integrating by parts gives

$$\Omega_a \sim \exp\left[i \int_{a-\epsilon}^{a+\epsilon} f'(x)\phi_R dx\right]. \tag{67}$$

Using the fact that $\phi_R \to \phi_R + f(x)\pi$ under $U_{[a,b]}$ gives

$$\omega \equiv U_{[a,b]}\Omega_a U_{[a,b]}^\dagger \Omega_a^\dagger = \exp\left[i\pi \int_{a-\epsilon}^{a+\epsilon} f'(x)f(x)dx\right] \tag{68}$$

$$= \exp\left\{\frac{i\pi}{2} \int_{a-\epsilon}^{a+\epsilon} \frac{d}{dx}[f(x)]^2\right\} \tag{69}$$

$$= i. \tag{70}$$

The numbers $(\omega,\sigma) = (i,-1)$ constitute the fermionic 3-cocycle for the Z_2 symmetry. We see that taking four copies of the same edge leads to a trivial fermionic 3-cocycle (in agreement with the results of Ref. [16] showing that four copies of the theory under consideration can be gapped out without breaking the symmetry). Furthermore, if one applies the fermionic 3-cocycle condition (see Appendix F), one sees that the only allowable values of the fermionic 3-cocycle are the ones obtained by taking copies in this way, namely, $(1,1)$, $(i,-1)$, $(-1,1)$, and $(-i,-1)$. Thus, we have recovered all the elements of a $\mathbb{Z}_4$ classification

for fermionic SPTs with $G_f = Z_2 \times Z_2^f$ (which is the same result obtained from supercohomology [27]). By contrast, Refs. [38–40,59] obtained a $\mathbb{Z}_8$ classification for the same G_f. The odd-numbered phases in this classification have an odd number of gapless Majorana modes at the edge, each of which is "half" of the gapless Dirac mode considered here. The explanation for the discrepancy in the classification is that the symmetry in these odd-numbered phases does not act locally on the edge, and hence they are not captured by our approach.

VIII. CONCLUSIONS

Suppose we have a system whose bulk ground state is invariant under a group G of symmetries that commute with the Hamiltonian. Let us further suppose that there is an energy gap to all bulk excitations and a concomitant finite correlation length and that we can solve the Hamiltonian (with a sufficiently powerful computer, for instance) for systems much larger than the correlation length. Armed with this information, we wish to determine if the system is in a SPT phase and, if so, which one. In a 1D system on a finite interval, we can identify an SPT phase by the presence of gapless excitations at the ends of the system that transform under a projective representation of the symmetry (or, alternatively, the presence of such states in the bipartite entanglement spectrum) [9–11,36]. However, how do we identify an SPT phase in higher dimensions? One approach is to gauge the symmetry G [25,30,60,61]. In 2D, the resulting theory has anyonic excitations in the bulk [25]. By determining the statistics of these excitations, one can deduce the SPT phase of the ungauged system. In 3D, the gauged theory has anyonic excitations on its surface [61] and extended excitations (e.g., vortex lines) in its bulk [30]. The topological properties of surface and bulk excitations of the gauged model can be used to identify the underlying SPT phase. However, this approach involves modifying the system drastically, and it cannot be used if all that we are given are the low-energy eigenstates of the original Hamiltonian. Moreover, it may be more difficult, as a practical matter, to solve the gauged model and deduce its quasiparticles' topological properties than it is to solve the original Hamiltonian.

Here we take a different approach, which identifies an SPT directly from the realization of the symmetry group G on its boundary states. We consider d-dimensional SPTs for which the restriction of G to the low-energy Hilbert space has a local action on the $(d-1)$-dimensional boundary of the system. In such a phase, there may be an obstruction to restricting the action of the symmetry to a $(d-1)$-dimensional proper submanifold of the boundary. To analyze such an obstruction, we construct a new functional of two group elements by taking a suitably defined coboundary of the restriction. This localizes the obstruction to the $(d-2)$-dimensional boundary of the $(d-1)$-dimensional proper submanifold of the boundary. We then continue in the same fashion, either restricting a functional of k group elements on a closed $(d-k)$-dimensional manifold to a $(d-k)$-dimensional submanifold with boundary or constructing the coboundary of a functional of k group elements on a $(d-k)$-dimensional submanifold with boundary, thereby obtaining a functional of $k+1$ group elements on a $(d-k-1)$-dimensional closed manifold. These functionals are operators that act on the local Hilbert spaces of the corresponding submanifolds. The resulting sequence of maps between functionals terminates after we reach functionals of d group elements acting on a single point; the coboundary of such a functional must be an ordinary phase. Equivalence classes of such sequences are classified by the cohomology group $H^{d+1}(G,U(1))$ in $d = 1,2$ and, with an additional assumption, in $d \geqslant 3$. Consequently, given the low-energy states of the boundary (or large eigenvalue eigenstates of the reduced density matrix for a bipartition of a system without a real boundary), we can, in principle, determine the corresponding element of $H^{d+1}(G,U(1))$. The Hamiltonian need not take any special form; in fact, it is not even necessary to know the Hamiltonian. As we have shown, this procedure gives the expected results when applied to discrete [13] and continuous [14,15,52,53] nonlinear σ models.

The obstructions classified by these arguments prevent a model from being continuously deformed into a model in which the symmetry is realized on the boundary in an on-site manner. (By assumption, the symmetry can be realized in an on-site manner in the full bulk theory, by grouping multiple sites into a single site, for instance.) As a result of the incorrigibly non-on-site nature of the symmetry, if we try to gauge it, the resulting gauge theory will be anomalous [28,45–47,62,63]. Only the action of the symmetry on the whole system, bulk and edge together, can be gauged in an anomaly-free fashion. A simple example is a $(2+1)$-D U(1) SPT, which is very similar to the $\mathbb{Z}_2$ case discussed in Sec. II. Such a state is a bosonic integer quantum Hall state [52]. If the theory is gauged, the edge effective Lagrangian takes the form $\mathcal{L}_{\text{edge}} = \frac{g}{2\pi}(\partial_\mu\varphi - nA_\mu)^2 + \frac{m}{2\pi}A_\mu\epsilon_{\mu\nu}\partial_\nu\varphi$. Charge is no longer conserved at the edge since an electric field along the edge will cause charge to flow from the bulk to the edge. Following Laughlin [64], we can understand this in an annular geometry. By adiabatically increasing the flux through the center of the annulus by 2π, the charge at the outer edge is increased by $2nm$, the integer (necessarily even in a bosonic SPT) that characterizes the Hall conductance. The 3-cocycle obtained by our construction reflects this charge pumped to the edge, as may be seen by noting that a U(1) transformation applied to a finite interval along the edge is equivalent to applying equal and opposite gauge fields at the ends of the interval [65]. Since they are equal and opposite, such gauge fields cannot increase the total charge on the edge, but if we focus on the charge to the left of an arbitrary point in the middle of the interval, then this increases by $2nm$ when the gauge field winds by 2π. The restriction Ω_a defined in Eq. (3) measures such a charge [66]. Meanwhile, U_M applies gauge fields at the ends of the interval. Then, according to the definition (5), the cocycle measures accumulated charge in response to this change in gauge field. We expect that similar reasoning can relate our constructions to anomalies in higher dimensions and for discrete symmetries.

In this paper, we have confined our attention to "internal" symmetries. It would be interesting to extend them to space-group symmetries. States of free fermions protected by inversion symmetry [67,68], time-reversal symmetry combined with a point group symmetry [69], or a rotational symmetry alone have been classified [70]. With the methods described

here, it might be possible to extend these ideas to interacting fermion systems and to bosonic systems in which a space group symmetry, projected to the low-energy boundary theory, maps sites to sites and then has an additional "internal" action that is non-on-site. However, care must be taken to consider a boundary that respects the space group symmetry and to consider a sequence of submanifolds (which are, presumably, not connected manifolds) that also respects the symmetry.

We have given a prescription that, in principle, allows one to identify an SPT phase, given its ground-state wave function, and we have shown how to apply it to some long-wavelength effective field theories and exactly soluble lattice models. However, how useful can this prescription be in practice, given an arbitrary, perhaps experimentally motivated, model? This remains to be seen. However, ground-state wave functions with tensor-network descriptions are natural candidates for the reduction procedure [44]. A numerical implementation would open an important avenue for future research.

Conversely, we have shown in Sec. V and Appendix E that each SPT phase in the cohomological classification has a representative wave function which is the ground state of some lattice Hamiltonian. However, these Hamiltonians are certainly not expected to describe any experimentally realizable systems; finding more realistic Hamiltonians giving rise to SPT phases is an important open problem.

As noted above, our construction leads to $H^{d+1}(G,U(1))$ in $d \geqslant 3$ provided we make an additional assumption: There exists a local basis for the Hilbert space of the $(d-1)$-dimensional boundary in which the symmetry acts on the boundary in an on-site manner, except for a diagonal part which cannot be made on site. This assumption holds in a system that is described by a d-dimensional nonlinear σ model with a θ term at long wavelengths [14,15,52,53] since the symmetry acts in an on-site manner on all gradient energy terms in the $(d-1)$-dimensional boundary effective action and non-on-site only on the Wess-Zumino term, which only enters the phase of the ground-state wave function. However, it remains an interesting open question whether there are SPT phases in three dimensions that violate this assumption and, consequently, realize the aforementioned nontrivial sequence but in a manner that is not classified by group cohomology. Such an exception to a cohomological classification, if it exists, would be distinct from the so-called "beyond cohomology" SPT phases [15], which occur due to the violation of a different assumption, that the symmetry is realized locally (but not necessarily on-site) at the boundary of the system. In beyond-cohomology SPT phases, the symmetry is realized in an inherently nonlocal manner at the boundary of the system. Our methods do not enable us to classify such phases; once the condition of locality is relaxed, a very different approach may be necessary.

This comment also applies to the most famous SPT phase, the 3D time-reversal-invariant topological insulator [71–73], where time reversal acts in an inherently nonlocal manner at a 2D surface. However, there are fermionic SPT phases in which the symmetry is realized locally on the boundary, and these can be classified along the lines discussed in Sec. VII. Carrying out this classification to completion and relating it to the notion of "supercohomology" [27] is an important goal for future work. We remark that dimensional reductions have previously been employed in the classification of fermionic SPT phases [33,74], and it would be interesting if a connection could be drawn with the reduction procedure described here.

Finally, we note that symmetry-enriched topological (SET) phases [60,75–79] generalize SPT phases to systems with topological order. In SET phases, symmetry realization interacts nontrivially with the fusion and braiding properties of anyons, as already occurs in topological phases at the $(2+1)$-D boundary of a $(3+1)$-D SPT. The possible symmetry fractionalization patterns in $(2+1)$-D correspond to different projective representations of the anyons and are classified by $H^2(G,A)$, where A is the group of Abelian anyons. It is possible that an extension of our methods can also be applied to the analysis of symmetry fractionalization in $(3+1)$-D SET phases which have topological excitations occupying closed loops.

ACKNOWLEDGMENTS

We thank M. Cheng, C. Xu, A.D. Rasmussen, and B. Ware for helpful discussions. D. Else was supported by a gift from the Microsoft Corporation.

APPENDIX A: EXPLICIT CONSTRUCTION OF THE EDGE THEORY

In this appendix we give an explicit proof of the property that the edge theory of an SPT in d spatial dimensions can always be realized in a strictly $(d-1)$-dimensional system and show, given certain assumptions, how to construct the representation of the symmetry in this realization.

By definition, the ground state of an SPT phase is gapped and not topologically ordered, which means it can be connected to a product state by a local unitary. Indeed, let $\mathcal{D}$ be the local unitary which turns the bulk ground state $|\Psi_{\mathrm{gr}}\rangle$ on a boundaryless spatial region product state $|\phi\rangle^{\otimes N}$, and let $\widetilde{\mathcal{D}}$ be the restriction of $\mathcal{D}$ to a spatial region with boundary. Any low-energy state $|\Psi\rangle$ associated with the boundary must be identical to $|\Psi_{\mathrm{gr}}\rangle$ far from the boundary. It follows that $\widetilde{\mathcal{D}}|\Psi\rangle$ must be identical to $|\phi\rangle^{\otimes N}$ far from the boundary. Hence, $\widetilde{\mathcal{D}}|\Psi\rangle$ is simply a direct product with copies of $|\phi\rangle$ in the bulk of some state $|\Psi\rangle_B$ defined on a strip B near the boundary,

$$\widetilde{\mathcal{D}}|\Psi\rangle = |\Psi\rangle_B \otimes |\Phi\rangle_{B^c}, \tag{A1}$$

where B^c is the complement of B and $|\Phi\rangle_{B^c}$ is a product state of $|\phi\rangle$ on every site in B^c. Thus, the states $\{|\Psi\rangle_B : |\Psi\rangle$ a low-energy boundary state$\}$ constitute a $(d-1)$-dimensional realization of the boundary theory. One can also apply the mapping $\widetilde{\mathcal{D}}$ to the original Hamiltonian for the system with boundary in order to obtain a Hamiltonian for this realization of the boundary theory.

Now suppose that the bulk ground state is invariant under an on-site representation $U(g)$ of the symmetry. As the local unitary $\mathcal{D}$ is not required to have any particular properties with respect to the symmetry, in general, it might not be easy to determine how the symmetry acts on the boundary theory. However, the task becomes easier if we make the following simplifying assumption: We assume that $\mathcal{D}$ can be chosen to commute with $U(g)$ in the absence of boundary. (We emphasize that this does *not* necessarily imply that we are considering a trivial SPT phase. That would be only be true if we made the

stronger assumption that $\mathcal{D}$ can be continuously connected to the identity by a path which everywhere commutes with the symmetry.) In fact, this assumption is always true in any SPT phase described by the cohomological classification, because, in particular, it is true for the ground states constructed via the discrete topological term construction of Ref. [13] (see Sec. V above.) This implies that it is also true for any other ground state in the same SPT phase, since, by definition, any two ground states in the same SPT phase can be related by a symmetry-respecting local unitary.

Given this assumption, one can explicitly construct the realization of the symmetry on the edge, as follows. That $U(g)$ and $\mathcal{D}$ commute in the absence of a boundary implies that their restrictions $\widetilde{U}(g)$ and $\widetilde{\mathcal{D}}$ to a region with a boundary must commute up to boundary terms. Thus, $\widetilde{\mathcal{D}}\widetilde{U}(g)\widetilde{\mathcal{D}}^\dagger = W_B(g)U_{B^c}(g)$, where $W_B(g)$ acts only in the strip B, and $U_{B^c}(g)$ is simply the restriction of $U(g)$ into the complement B^c. Hence, we find that

$$\widetilde{\mathcal{D}}\{\widetilde{U}(g)|\Psi\rangle\} = \widetilde{\mathcal{D}}\widetilde{U}(g)\widetilde{\mathcal{D}}^\dagger\widetilde{\mathcal{D}}|\Psi\rangle \tag{A2}$$

$$= W_B(g)U_{B^c}(g)(|\Psi\rangle_B \otimes |\Phi\rangle_{B^c}) \tag{A3}$$

$$= \{W_B(g)|\Psi\rangle_B\} \otimes |\Phi\rangle_{B^c}. \tag{A4}$$

To get to the last line, we used the fact that $U_{B^c}(g)|\Phi\rangle_{B^c} = |\Phi\rangle_{B^c}$. This follows from the fact that, *without* boundary, $|\phi\rangle^{\otimes N}$ is invariant under $U(g)$, since it is obtained from $|\Psi_{\rm gr}\rangle$ [which is certainly invariant under $U(g)$] by $\mathcal{D}$ which by assumption commutes with $U(g)$. Comparing Eq. (A4) with Eq. (A1), we see that $W_B(g)$ is the representation of the symmetry on the stand-alone realization of the boundary.

APPENDIX B: THE (2 + 1)-D REDUCTION PROCEDURE

Here we prove the two key properties of $\omega(g_1,g_2,g_3)$ defined by Eq. (5) in Sec. I B: first, that it must be a 3-cocycle, and second, that up to equivalence it is independent of the choice of restriction $U(g) \to U_M(g)$.

We first make a general remark: The structure described in Sec. I B is known in the mathematics literature as a *crossed module extension*. Recall that a projective representation of a group G corresponds to a *central extension*, which is an exact sequence,

$$1 \to \mathrm{U}(1) \to H \to G \to 1, \tag{B1}$$

such that the image of U(1) is in the center of H. Similarly, a crossed module extension is an exact sequence,

$$1 \to \mathrm{U}(1) \to K \xrightarrow{\varphi} H \to G \to 1, \tag{B2}$$

along with a left action of H on K, represented by $k \mapsto {}^h k$, such that ${}^{\varphi(k)}k' = kk'k^{-1}$ for all $k,k' \in K$. It is a well-known theorem in the mathematics literature [80–83] that the crossed module extensions of G by U(1) are classified by $H^3(G,\mathrm{U}(1))$. The procedure described in Sec. I B for obtaining the 3-cocycle $\omega(g_1,g_2,g_3)$, as well as the proofs of the properties of ω given below, are adapted from the proof of this classification theorem given in Ref. [83]; however, the reader does not need to understand the connection to crossed module extensions in order to follow these proofs.

To prove that ω is a 3-cocycle, we calculate $\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3)\Omega_a(g_1g_2g_3,g_4)$ in two different ways. First,

$$\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3)\Omega_a(g_1g_2g_3,g_4) \tag{B3}$$

$$= \omega(g_1g_2,g_3,g_4) \times \Omega_a(g_1,g_2)^{U_M(g_1g_2)}\Omega_a(g_3,g_4)\Omega_a(g_1g_2,g_3g_4) \tag{B4}$$

$$= \omega(g_1g_2,g_3,g_4) \times {}^{\Omega_a(g_1,g_2)U_M(g_1g_2)}\Omega_a(g_3,g_4)\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3g_4) \tag{B5}$$

$$= \omega(g_1g_2,g_3g_4) \times {}^{\Omega(g_1,g_2)U_M(g_1g_2)}\Omega_a(g_3,g_4)\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3g_4) \tag{B6}$$

$$= \omega(g_1g_2,g_3,g_4) \times {}^{U_M(g_1)U_M(g_2)}\Omega_a(g_3,g_4)\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3g_4) \tag{B7}$$

$$= \omega(g_1g_2,g_3,g_4)\omega(g_1,g_2,g_3g_4) \times {}^{U_M(g_1)U_M(g_2)}\Omega_a(g_3,g_4)^{U_M(g_1)}\Omega_a(g_2,g_3g_4)\Omega_a(g_1,g_2g_3g_4), \tag{B8}$$

where we applied Eq. (5) twice. To get from Eq. (B6) to Eq. (B7) we used Eq. (3). To get from Eq. (B5) to Eq. (B6), we used the fact that $\Omega(g,g')$ can be written as a product of a contributions near a and contributions near b, which commute; it follows that for any operator X_a localized near a,

$${}^{\Omega(g_1,g_2)}X_a = {}^{\Omega_a(g_1,g_2)}X_a. \tag{B9}$$

Proceeding in a different way, we also have

$$\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3)\Omega_a(g_1g_2g_3,g_4) \tag{B10}$$

$$= \omega(g_1,g_2,g_3) \times {}^{U_M(g_1)}\Omega_a(g_2,g_3)\Omega_a(g_1,g_2g_3)\Omega_a(g_1g_2g_3,g_4) \tag{B11}$$

$$= \omega(g_1,g_2,g_3)\omega(g_1,g_2g_3,g_4) \times {}^{U_M(g_1)}\Omega_a(g_2,g_3)^{U_M(g_1)}\Omega_a(g_2g_3,g_4)\Omega_a(g_1,g_2g_3g_4) \tag{B12}$$

$$= \omega(g_1,g_2,g_3)\omega(g_1,g_2g_3,g_4)^{U_M(g_1)}\{\Omega_a(g_2,g_3)\Omega_a(g_2g_3,g_4)\}\Omega_a(g_1,g_2g_3g_4) \tag{B13}$$

$$= \omega(g_1,g_2,g_3)\omega(g_1,g_2g_3,g_4)\omega(g_2,g_3,g_4) \times {}^{U_M(g_1)}\{{}^{U_M(g_2)}\Omega_a(g_3,g_4)\Omega_a(g_2,g_3g_4)\}\Omega_a(g_1,g_2g_3g_4) \tag{B14}$$

$$= \omega(g_1,g_2,g_3)\omega(g_1,g_2g_3,g_4)\omega(g_2,g_3,g_4) \times {}^{U_M(g_1)U_M(g_2)}\Omega_a(g_3,g_4)^{U_M(g_1)}\Omega_a(g_2,g_3g_4)\Omega_a(g_1,g_2g_3g_4). \tag{B15}$$

Comparing Eq. (B15) with Eq. (B8) we see that ω must obey the 3-cocycle condition

$$\omega(g_1g_2,g_3,g_4)\omega(g_1,g_2,g_3g_4) = \omega(g_1,g_2,g_3)\omega(g_1,g_2g_3,g_4)\omega(g_2,g_3,g_4). \tag{B16}$$

Next we prove independence from the choice of restriction $U(g) \to U_M(g)$. Indeed, consider two restrictions, $U_M(g)$ and $\widetilde{U}_M(g) = \Sigma(g)U_M(g)$, where $\Sigma(g)$ is a local unitary acting near $\partial M = \{a,b\}$. Then we find that

$$\widetilde{U}_M(g)\widetilde{U}_M(g') = \widetilde{\Omega}(g,g')\widetilde{U}_M(gg'), \tag{B17}$$

where

$$\widetilde{\Omega}(g,g') = \Sigma(g)^{U_M(g)}\Sigma(g')\Omega(g,g')\Sigma(gg')^{-1}. \tag{B18}$$

It is obvious that the equivalence class of the 3-cocycle is independent of the choice of restriction $\Omega \to \Omega_a$, so we are free to choose a restriction of $\widetilde{\Omega}$ such that

$$\widetilde{\Omega}_a(g,g') = \Sigma_a(g)^{U_M(g)}\Sigma_a(g')\Omega_a(g,g')\Sigma_a(gg')^{-1}, \tag{B19}$$

where $\Sigma_a(g)$ is the restriction of $\Sigma(g)$ to the point a. Now we calculate

$$\begin{aligned}
&\widetilde{\Omega}_a(g_1,g_2)\widetilde{\Omega}_a(g_1g_2,g_3)\Sigma_a(g_1g_2g_3) && \text{(B20)}\\
&= \widetilde{\Omega}_a(g_1,g_2)\Sigma_a(g_1g_2)^{U_M(g_1g_2)}\Sigma_a(g_3)\Omega(g_1g_2,g_3) && \text{(B21)}\\
&= \Sigma_a(g_1)^{U_M(g_1)}\Sigma_a(g_2)\Omega_a(g_1,g_2)^{U_M(g_1g_2)}\Sigma_a(g_3)\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3) && \text{(B22)}\\
&= \Sigma_a(g_1)^{U_M(g_1)}\Sigma_a(g_2)^{\Omega_a(g_1,g_2)U_M(g_1g_2)}\Sigma_a(g_3)\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3) && \text{(B23)}\\
&= \Sigma_a(g_1)^{U_M(g_1)}\Sigma_a(g_2)^{\Omega(g_1,g_2)U_M(g_1g_2)}\Sigma_a(g_3)\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3) && \text{(B24)}\\
&= \Sigma_a(g_1)^{U_M(g_1)}\Sigma_a(g_2)^{U_M(g_1)U_M(g_2)}\Sigma_a(g_3)\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3), && \text{(B25)}
\end{aligned}$$

where we applied Eq. (B19) several times, and we also used, again, Eq. (B9) to go from Eq. (B23) to Eq. (B24). The final line used the definition of ω, Eq. (5). On the other hand,

$$\begin{aligned}
&{}^{\widetilde{U}_M(g_1)}\widetilde{\Omega}_a(g_2,g_3)\widetilde{\Omega}_a(g_1,g_2g_3)\Sigma_a(g_1g_2g_3) && \text{(B26)}\\
&= \Sigma_a(g_1)^{U_M(g_1)}\widetilde{\Omega}_a(g_2,g_3)\Sigma_a(g_1)^{-1}\widetilde{\Omega}_a(g_1,g_2g_3)\Sigma(g_1g_2g_3) && \text{(B27)}\\
&= \Sigma_a(g_1)^{U_M(g_1)}\widetilde{\Omega}_a(g_2,g_3)^{U_M(g_1)}\Sigma_a(g_2g_3)\Omega_a(g_1,g_2g_3) && \text{(B28)}\\
&= \Sigma_a(g_1)^{\,U_M(g_1)}\{\widetilde{\Omega}_a(g_2,g_3)\Sigma_a(g_2g_3)\}\Omega_a(g_1,g_2g_3) && \text{(B29)}\\
&= \Sigma_a(g_1)^{\,U_M(g_1)}\{\Sigma_a(g_2)^{U_M(g_2)}\Sigma_a(g_3)\Omega_a(g_2,g_3)\}\Omega_a(g_1,g_2g_3) && \text{(B30)}\\
&= \Sigma_a(g_1)^{\,U_M(g_1)}\Sigma_a(g_2)^{U_M(g_1)U_M(g_2)}\Sigma_a(g_3)^{U_M(g_1)}\Omega_a(g_2,g_3)\Omega_a(g_1,g_2g_3) && \text{(B31)}\\
&= \omega(g_1,g_2,g_3)\Sigma_a(g_1)^{\,U_M(g_1)}\Sigma_a(g_2)^{U_M(g_1)U_M(g_2)}\Sigma_a(g_3)\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3). && \text{(B32)}
\end{aligned}$$

Comparing Eqs. (B32) and (B25), we find that

$${}^{\widetilde{U}_M(g_1)}\widetilde{\Omega}_a(g_2,g_3)\widetilde{\Omega}_a(g_1,g_2g_3) = \omega(g_1,g_2,g_3)\widetilde{\Omega}_a(g_1,g_2)\widetilde{\Omega}_a(g_1g_2,g_3). \tag{B33}$$

On the other hand, by definition, ω also satisfies Eq. (B33) with $\widetilde{\Omega}_a$ and $\widetilde{U}_M$ replaced by Ω_a and U_M. Thus, it does not matter whether we use the restriction U_M or $\widetilde{U}_M$; one obtains the same 3-cocycle ω.

APPENDIX C: COMPLETING THE PROOF OF SEPARATION OF PHASES IN (2 + 1)-D

In this section, we fill in the details of the proof outlined in Sec. III showing that (2 + 1)-D SPT phases corresponding to different elements of the cohomology group $H^3(G,\mathrm{U}(1))$ are necessarily separated by a phase transition. Although we write the arguments in lemma-proof form, we emphasize that we do not aim for mathematical rigor in our treatment of locality; rigorous proofs could potentially be constructed based on the arguments sketched here, but would require much more careful estimates of how relevant quantities decay at large distances.

The first result that needed to be proved was as follows.

Lemma 2. Let $|\Psi\rangle$ be the ground state of an SPT phase in d spatial dimensions captured by the cohomological classification, and let $U(g)$ be a local unitary representation of a group G which leaves $|\Psi\rangle$ invariant. Let $U_A(g)$ be the restriction of $U(g)$ to a region A. Then there exists a representation $V_{\partial A}(g)$ acting only in a strip near the boundary ∂A (and only within the region A) such that $U_A(g)|\Psi\rangle = V_{\partial A}(g)|\Psi\rangle$.

Proof. We use the fact, discussed in Appendix A above, that the state $|\Psi\rangle$ can be transformed into a product state $|\Phi\rangle = |\phi\rangle^{\otimes N}$ through a local unitary $\mathcal{D}$ that commutes with $U(g)$. Hence, as in Appendix A, if we define the restriction $\mathcal{D}_A$, it follows that $\mathcal{D}_A U_A(g)\mathcal{D}_A^\dagger = W_B(g)U_{A\setminus B}(g)$, where $W_B(g)$ acts within a strip B near the boundary, and $U_A(g)$ and $U_{A\setminus B}(g)$ are the restriction of $U(g)$ to the respective regions (and $A \setminus B$ is the region A with the strip B excluded.) It is clear (by a similar argument to the one given above in Appendix A) that $U_{A\setminus B}(g)|\Phi\rangle = |\Phi\rangle$, and hence we find that

$$\begin{aligned}
U_A(g)|\Psi\rangle &= \mathcal{D}_A^\dagger\mathcal{D}_A U_A(g)\mathcal{D}_A^\dagger\mathcal{D}_A|\Psi\rangle && \text{(C1)}\\
&= \mathcal{D}_A^\dagger W_B(g)U_{A\setminus B}(g)\mathcal{D}_A|\Psi\rangle && \text{(C2)}\\
&= \mathcal{D}_A^\dagger W_B(g)\mathcal{D}_A|\Psi\rangle, && \text{(C3)}
\end{aligned}$$

where in going from Eq. (C2) to Eq. (C3) we used the fact that $\mathcal{D}_A|\Psi\rangle$ looks like $|\Phi\rangle$ on $A \setminus B$, and therefore $U_{A\setminus B}(g)\mathcal{D}_A|\Psi\rangle = \mathcal{D}_A|\Psi\rangle$. Hence, defining $V_{\partial A}(g) = \mathcal{D}_A^\dagger W_B(g)\mathcal{D}_A$, we have the desired result. ■

To proceed further, we also require the following lemma, which states that a "trivial" state cannot be invariant under an "anomalous" symmetry representation.

Lemma 3. Let $|\Psi\rangle$ be some short-ranged entangled state, and let $U(g)$ be some local unitary representation of a symmetry group G on a closed 1D subregion C of the space on which $|\Psi\rangle$ is defined, such that $U(g)|\Psi\rangle = |\Psi\rangle$. Then the element of the cohomology group $H^3(G,\mathrm{U}(1))$ obtained via the reduction procedure of Sec. I B is necessarily trivial.

Proof. Consider the restriction $U_M(g)$ to a subregion M of C with boundary. Then $U_M(g)|\Psi\rangle$ looks like $|\Psi\rangle$ away from the boundary ∂M. This implies that, for *short-ranged entangled* states $|\Psi\rangle$ it must be the case that $U_M(g)|\Psi\rangle = \Sigma_{\partial M}(g)|\Psi\rangle$ for some set of local unitaries $\Sigma_{\partial M}(g)$ acting near the boundary [25]. (To show this, one first establishes it for a product state, from which one can show that it must also apply for any state that can be turned into a product state by a local unitary.) However, the restriction $U(g) \to U_M(g)$ was only defined up to unitaries at the boundary anyway, so we are free to set $U_M(g)|\Psi\rangle = |\Psi\rangle$. Then, defining $\Omega(g_1,g_2)$ via

$$U_M(g_1)U_M(g_2) = \Omega(g_1,g_2)U_M(g_1g_2), \tag{C4}$$

we can deduce that $\Omega(g_1,g_2)|\Psi\rangle = |\Psi\rangle$. Assuming that $\partial M = \{a,b\}$ where a and b are two points, we can choose the restriction $\Omega_a(g_1,g_2)$ such that $\Omega_a(g_1,g_2)|\Psi\rangle = |\Psi\rangle$. Then, given the definition [Eq. (5) in Sec. I B] of the 3-cocycle $\omega(g_1,g_2,g_3)$, one finds that $\omega(g_1,g_2,g_3)|\Psi\rangle = |\Psi\rangle$, and hence $\omega(g_1,g_2,g_3) = 1$. ■

Now, let us consider a (possibly spatially inhomogeneous) ground state $|\Psi\rangle$ in an SPT phase, invariant under an on-site symmetry representation $U(g)$ of G. We choose two regions, A and A', separated by a quasi-1D buffer region K, as depicted in Fig. 4. We assume that the combined region $A \cup K \cup A'$ has no boundary. We define $V_{\partial A}(g)$ and $V_{\partial A'}(g)$ according to $U_A(g)|\Psi\rangle = V_{\partial A}|\Psi\rangle$ as per Lemma 2, and denote the corresponding classes of 3-cocycles as $[\omega],[\omega'] \in H^3(G,\mathrm{U}(1))$. We want to show that $[\omega] = [\omega']$.

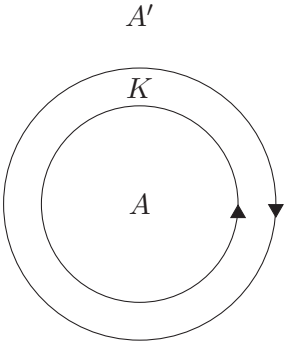

FIG. 4. The regions A and A' on which we can prove that the anomalous symmetry on the boundary is classified by the same element of $H^3(G,\mathrm{U}(1))$. The orientations of the boundaries ∂A and $\partial A'$ are depicted with arrows.

First we need to discuss an important subtlety involved in the definitions of $[\omega]$ and $[\omega']$. Specifically, the reduction procedure of Sec. I implicitly depends on an orientation for the 1D space in which the local unitaries act, in order to provide a consistent convention for reducing from $\partial M = \{a,b\}$ to a single point a. Opposite orientations give rise to inverse cocycles. We take the orientations of ∂A and $\partial A'$ to be derived from that of A and A'; if we choose A and A' to have the same orientation (e.g., both specified by normal vectors pointing out of the page), then ∂A and $\partial A'$ have *opposite* orientations, as depicted by the arrows in Fig. 4.

We now observe that $|\Psi\rangle$ is invariant under $V_{\mathrm{sum}}(g) = U_K(g)V_{\partial A}(g)V_{\partial A'}(g)$. Now it can readily be verified [using the fact that the three components of $V_{\mathrm{sum}}(g)$ all commute with each other] that the element of $H^3(G,\mathrm{U}(1))$ characterizing $V_{\mathrm{sum}}(g)$ is equal to $[\omega][\omega']^{-1}$. (The inverse comes from the fact that we have defined the orientations of ∂A and $\partial A'$ to be opposite.) On the other hand, by Lemma 3, this product must be trivial; hence, we have established that $[\omega] = [\omega']$.

Finally, let us justify the following claim we made in Sec. III.

Lemma 4. Consider two gapped systems, $\mathcal{S}$ and $\mathcal{S}'$, connected without a phase transition and two well-separated regions, A and A'. Then one can construct an interpolating system such that the ground state looks like that of $\mathcal{S}$ on A and like that of $\mathcal{S}'$ on A'.

Proof. Let $|\Psi\rangle$ and $|\Psi'\rangle$ be the corresponding ground states. Then there must exist a local unitary $\mathcal{D}$ such that $\mathcal{D}|\Psi\rangle = |\Psi'\rangle$. Define the restriction $\mathcal{D}_{\widetilde{A}'}$ of $\mathcal{D}$ to some region $\widetilde{A}'$ that contains A' well inside, but is also well separated from A. Then applying $\mathcal{D}_{\widetilde{A}'}$ to $\mathcal{S}$ gives a system with the desired properties. ■

APPENDIX D: PROOF OF LEMMA 1

Here we give a proof of Lemma 1, which we stated in Sec. IV A. The proof is based on a result called the universal coefficient theorem [84]. Let us first state some definitions. For a manifold T, we define $C^k(T,\mathrm{U}(1))$ to be the group of *closed* U(1) k cochains, i.e., those cochains ω for which $d\omega = 0$, and $B^k(T,\mathrm{U}(1))$ to be the group of *exact* cochains, i.e., those which can be written as $\omega = d\kappa$ for some κ. We define the *cohomology group* $H^k(T,\mathrm{U}(1)) \equiv C^k(T,\mathrm{U}(1))/B^k(T,\mathrm{U}(1))$. Similarly, we define $C_k(T)$ and $M_k(T)$ to be the group of closed (i.e., boundaryless) and exact (can be expressed as a boundary) k chains, respectively; and the homology group $H_k(T) = C_k(T)/M_k(T)$.

We observe that there is a natural homomorphism $\gamma : H_k(T,\mathrm{U}(1)) \to \hom(H^k(T),\mathrm{U}(1))$ defined according to $[\omega] \mapsto ([\sigma] \mapsto \omega(\sigma))$ (where $[\cdots]$ denotes equivalence classes in cohomology or homology.) The universal coefficient theorem states that γ is in fact an isomorphism. [In general, replacing U(1) with an arbitrary Abelian group A, the universal coefficient theorem states that the homomorphism is surjective and its kernel is isomorphic to $\mathrm{ext}(G,A)$. However, U(1) is divisible and it follows that $\mathrm{ext}(G,\mathrm{U}(1)) = 0$.]

Hence, we can prove the following.

Lemma 1. A U(1) k cochain ω on a manifold T is exact if and only if $\omega(C) = 0$ for all closed (i.e., boundaryless) k chains C.

Proof. If ω is exact, then $\omega = d\kappa$, and hence $\omega(C) = \kappa(\partial C) = 0$ for any closed k chain C.

Conversely, let ω be some k cochain such that $\omega(C) = 0$ for all closed A. Then ω is closed because $(d\omega)(A) = \omega(\partial A) = 0$. Also, $\gamma([\omega]) = 0$. Since γ is an isomorphism, it follows that the equivalence class $[\omega] = 0$. Hence, ω is exact. ■

APPENDIX E: CALCULATING THE ELEMENT OF THE COHOMOLOGY GROUP FOR DISCRETE NONLINEAR σ MODELS

The action for the field theories of Ref. [13] is a discrete analog of the topological θ term that appeared in the continuous NLσMs in Sec. IV. Recall from Sec. IV A that the θ term is derived from a U(1) cochain defined on the target manifold T. The same is true in the discrete case, except that the target space T is now discrete, and so the interpretation of the "chains" which are the arguments of the cochains needs to be revised. Specifically, we define a k chain on T to be a formal linear combination (with integer coefficients) of "k simplices," which are simply ordered k-tuples $\Delta = (\Delta_1, \ldots, \Delta_k) \in T^{\times k}$. Then we can define the "boundary" operator ∂ acting linearly on k chains by specifying its action on k simplices:

$$\partial(\Delta_1, \ldots, \Delta_k) = \sum_{j=1}^{k} (-1)^{j-1} (\Delta_1, \ldots, \Delta_{j-1}, \Delta_{j+1}, \ldots, \Delta_k). \tag{E1}$$

To construct the discrete topological term corresponding to a U(1) cochain on T, one considers a triangulation of a D-dimensional space time manifold M; that is, we build M up out of D simplices. The degrees of freedom of the field theory will live on the vertices of the simplices. We can represent the D simplices in space time in terms of their vertices $(x_1, \ldots, x_D)$ [the abstract definition of boundary given in Eq. (E1) then agrees with the geometrical definition]. Thus, we interpret the manifold M as a formal D chain $M = \sum_\Delta (\Delta_1, \ldots, \Delta_D)$. Given a U(1) D cochain ω on M, and a function α assigning a value of the target space to each vertex, we can define the action

$$S_{\text{top}} = \omega(\alpha(M)) \tag{E2}$$

$$\equiv \sum_{\Delta} \omega(\alpha(\Delta_1), \ldots, \alpha(\Delta_D)). \tag{E3}$$

We ensure that this action will vanish for *closed* space-time manifolds M by requiring that $\omega(C) = 0$ for any closed chain C. By Lemma 1 in Sec. IV A above (which holds equally well for discrete cochains), this is equivalent to requiring that ω be exact.

If we have an action of a group G on the target space T, then for each symmetric D cochain ω one can derive an element of the group cohomology group $H^D(G, \mathrm{U}(1))$ by following the exact same reduction procedure as we did in the continuous case in Sec. IV A. One might, however, object that the physical significance of this is not clear unless we specify some way to quantize a field theory defined in discrete time. For this reason, we want to reinterpret discrete field theories like those of Sec. IV A as prescriptions for constructing lattice models.

Specifically, consider a triangulation of a closed d-dimensional *spatial* manifold M ($d = D - 1$). At each vertex, we put a quantum particle whose basis states are labeled by the elements of the target space T. (Hence, each basis state of the whole system is labeled by a function α mapping vertices into T.) We can define a quantum state for the system by "imaginary-time evolution" of a $(d+1)$-dimensional discrete topological action derived from an exact U(1) $(d+1)$ cochain on T. This state is given by

$$|\Psi\rangle = \sum_{\alpha} \Psi(\alpha) |\alpha\rangle, \tag{E4}$$

$$\Psi(\alpha) = \exp[i\kappa(\alpha(M))], \tag{E5}$$

where κ is a U(1) d cochain such that $d\kappa = \omega$, and we define $\kappa(\alpha(M))$ in the analogous way to Eq. (E2) above. This wave function is invariant under the on-site representation of the symmetry,

$$S(g) = \sum_{\alpha} |g\alpha\rangle\langle\alpha|. \tag{E6}$$

Once we have the wave function, it is easy to construct a corresponding local Hamiltonian for which it is the gapped ground state. For example, if we let $\mathcal{V} = \sum_\alpha \Psi(\alpha)|\alpha\rangle\langle\alpha|$ be the local unitary which creates $|\Psi\rangle$ from the trivial product state $|\Psi_{\text{prod}}\rangle = \sum_\alpha |\alpha\rangle$, then, starting from a local Hamiltonian H_{prod} which has $|\Psi_{\text{prod}}\rangle$ as its gapped ground state, we can define $H = \mathcal{V} H_{\text{prod}} \mathcal{V}^\dagger$.

In Appendix A, we give a general discussion of how to decouple a bulk theory from its boundary in order to find the form of the edge symmetry. Applying the method of Appendix A to the situation at hand, one finds that the edge symmetry takes the form $U_{\text{edge}}(g) = N(g)S(g)$, where $S(g)$ is as in Eq. (E6) (but acting only on the degrees of freedom at the edge), and

$$N(g) = \sum_{\alpha} \exp\left\{ i\kappa_g^{(2)}(\alpha(\partial M)) \right\} |\alpha\rangle\langle\alpha|, \tag{E7}$$

where $\kappa_g^{(2)}$ is the U(1) $d-1$ cochain such that $d\kappa_g^{(2)} = g\kappa - \kappa$. This is precisely the discrete version of Eq. (39). It is now easy to see that the general reduction procedure of Sec. I C will produce the same result as naively applying the method of Sec. IV A for the discrete case.

It turns out that the reduction procedure can actually be done explicitly starting from an arbitrary exact U(1) $(d+1)$ cochain ω on a target space T symmetric under the action of a group G. To see this, choose some arbitrary fixed $t_* \in T$, and define

$$\kappa^{(k)}(g_1, \ldots, g_k)(\Delta_1, \ldots, \Delta_{d-k}) = \omega\left(g_k^{-1} \ldots g_1^{-1} t_*, \ldots, g_1^{-1} t_*, t_*, \Delta_1, \ldots, \Delta_{d-k}\right) \tag{E8}$$

and $\omega^{(k)} = d\kappa^{(k)}$. Using the fact that ω is invariant under the symmetry and $\omega(C) = 0$ for closed chains C, one can show that (a) $\omega^{(0)} = \omega$; and (b) $\delta_k \kappa^{(k)} = \omega^{(k+1)}$. Thus, we have explicitly constructed the reduction sequence of Sec. IV A, and we find that the resulting element of the group cohomology group $H^{d+1}(G, \mathrm{U}(1))$ is the equivalence class of the following U(1) $(d+1)$ (group cocycle):

$$\nu(g_1, \ldots, g_{d+1}) = \omega\left(g_{d+1}^{-1} \ldots g_1^{-1} t_*, \ldots, g_1^{-1} t_*, t_*\right). \tag{E9}$$

In particular, following Ref. [13], we can consider the case where the target space T is the symmetry group G itself, with G acting on itself by left multiplication. In that case,

it is easy to see that Eq. (E9) actually defines a one-to-one mapping between symmetric exact "topological" cochains on the right-hand side and *group* cocycles on the left-hand side. Thus, for every element of the group cohomology group $H^{d+1}(G,\mathrm{U}(1))$, one can construct a discrete topological term in $d+1$ space-time dimensions via Eq. (E9), and applying our general reduction procedure returns the same element of $H^{d+1}(G,\mathrm{U}(1))$.

APPENDIX F: CLASSIFICATION OF (2 + 1)-D FERMIONIC SPTS

Here we implement the ideas discussed in Sec. VII in order to give a classification of (2 + 1)-D fermion SPT's. Consider a (2 + 1)-D fermionic SPT with fermionic symmetry group G_f (represented on site), including an element π corresponding to the fermion parity. All the symmetries are assumed to be local, so π must commute with all the elements of G_f. The fermion parity plays a such a key role in the following argument that we find it convenient to write elements of G_f in the form $\varpi(g)\pi^m$, where $m = 0$ or 1, $g \in G_b \equiv G_f/Z_2^f$, and ϖ is an arbitrary identification of elements of G_b with coset representatives in G_f, such that $\varpi(g_1)\varpi(g_2) = \pi^{\lambda(g_1,g_2)}\varpi(g_1,g_2)$. Here $\lambda(g_1,g_2)$ takes values 0 or 1, and associativity implies that it must be a $\mathbb{Z}_2$ 2-cocycle, i.e., $\delta\lambda = 0$ [mod 2], where δ is the coboundary operator,

$$(\delta\lambda)(g_1,g_2) = \lambda(g_1,g_2) + \lambda(g_1g_2,g_3) + \lambda(g_2,g_3) + \lambda(g_1,g_2g_3). \tag{F1}$$

Now we assume that the edge of this SPT can be realized in a strictly (1 + 1)-D local fermionic system and be invariant under a local unitary (but not necessarily on site) representation of G_f. The fermion parity is still represented as $\Pi \equiv (-1)^F$ on the edge. We write the fermionic local unitary operator implementing $\varpi(g)\pi^m$ on the edge as $U(g)\Pi^m$. Then $U(g)$ must satisfy

$$U(g_1)U(g_2) = \Pi^{\lambda(g_1,g_2)}U(g_1g_2). \tag{F2}$$

If we restrict the symmetry action to an interval $M = [a,b]$, then the restricted unitaries must satisfy Eq. (F2) up to a boundary term $\Omega(g_1,g_2)$:

$$U_M(g_1)U_M(g_2) = \Omega(g_1,g_2)\Pi^{\lambda(g_1,g_2)}U_M(g_1g_2). \tag{F3}$$

Using the associativity of the U_M's, combined with $\delta\lambda = 0$ and the fact that the U_M's commute with Π, we see that the Ω's must satisfy an identical equation to the bosonic case:

$$\Omega(g_1,g_2)\Omega(g_1g_2,g_3) = {}^{U_M(g_1)}\Omega(g_2,g_3)\Omega(g_1,g_2g_3). \tag{F4}$$

As discussed in Sec. VII, in defining the restriction $\Omega \to \Omega_a$ we might obtain an operator carrying nontrivial charge under fermion parity. We define the function $\sigma(g_1,g_2)$ to be 0 if Ω_a is a fermionic local unitary (no charge under fermion parity) acting at the point a, and 1 if it is equal to such a local unitary, multiplied by $(c_a + c_a^\dagger)$.

The restriction $\Omega_a(g_1,g_2)$ must satisfy Eq. (F4) up to a phase factor

$$\begin{aligned}&{}^{U_M(g_1)}\Omega_a(g_2,g_3)\Omega_a(g_1,g_2g_3)\\ &\quad= \omega(g_1,g_2,g_3)\Omega_a(g_1,g_2)\Omega_a(g_1g_2,g_3),\end{aligned} \tag{F5}$$

where ω is a U(1)-valued function. The pair of functions (ω,σ) constitutes the fermionic 3-cocycle. From Eq. (F5) we immediately see that σ must be a $\mathbb{Z}_2$ cocycle, i.e.,

$$\delta\sigma = 0. \tag{F6}$$

Following a similar derivation to the one in Appendix B that gave the bosonic 3-cocycle condition [85], we also also find that ω must obey

$$(\delta\omega)(g_1,g_2,g_3,g_4) = (-1)^{[\sigma(g_1,g_2)+\lambda(g_1,g_2)]\sigma(g_3,g_4)}, \tag{F7}$$

where

$$\begin{aligned}(\delta\omega)(g_1,g_2,g_3,g_4) = {}&\omega(g_1,g_2,g_3)\omega(g_1g_2,g_3,g_4)^{-1}\\ &\times\omega(g_1,g_2g_3,g_4)\omega(g_1,g_2,g_3g_4)^{-1}\\ &\times\omega(g_2,g_3,g_4).\end{aligned} \tag{F8}$$

Equations (F6) and (F7) constitute the condition for (ω,σ) to be a fermionic 3-cocycle.

Furthermore, the freedom to redefine the restriction $U \to U_M$ and $\Omega \to \Omega_a$ implies (again following similar arguments as in Appendix B) that we must identify fermionic 3-cocycles that differ by the transformation

$$\sigma(g_1,g_2) \to \sigma(g_1,g_2) + (\delta\mu)(g_1,g_2) \quad [\mathrm{mod}\ 2], \tag{F9}$$

$$\omega(g_1,g_2,g_3) \to \omega(g_1,g_2,g_3)(-1)^{[\sigma(g_1,g_2)+\lambda(g_1,g_2)]\mu(g_3)+\mu(g_1)[\sigma(g_2,g_3)+(d\mu)(g_2,g_3)]}(\delta\beta)(g_1,g_2,g_3), \tag{F10}$$

where

$$(\delta\beta)(g_1,g_2,g_3) = \beta(g_1,g_2)\beta(g_1g_2,g_3)\beta(g_2,g_3)^{-1}\beta(g_1,g_2g_3)^{-1}, \tag{F11}$$

$$(\delta\mu)(g_1,g_2) = \mu(g_1) + \mu(g_2) + \mu(g_1g_2), \tag{F12}$$

and β and μ take values in U(1) and $\{0,1\}$, respectively. [The numbers $\mu(g)$ correspond to the fermion parity of the restriction $\Sigma_a(g)$ of the operator $\Sigma(g)$ that implements the redefinition $U_M(g) \to \Sigma(g)U_M(g)$.] If (ω,σ) is not equivalent to the trivial fermionic 3-cocycle according to the above equivalence relation, then we expect that the edge must correspond to the boundary of a nontrivial SPT phase. This is because such an anomalous nontrivial symmetry precludes a gapped ground state unless the symmetry is spontaneously broken. (This can be derived in a similar way to the equivalent bosonic result, Lemma 3 in Appendix C .) Similarly, two SPT phases characterized by fermionic 3-cocycles *not* related by

the above equivalence relation must be separated by a phase transition.

We also can define a product rule for fermionic 3-cocycles. Physically, the product rule corresponds to "stacking" two SPT phases on top of each other. If the edges of the two systems are characterized by (σ,ω) and (σ',ω'), then one can show that the edge of the combined system will be described by the fermionic 3-cocycle $(\sigma_{\text{prod}},\omega_{\text{prod}})$, where

$$\sigma_{\text{prod}} = \sigma + \sigma', \tag{F13}$$

$$\omega_{\text{prod}}(g_1,g_2,g_3) = (-1)^{\sigma'(g_2,g_3)\sigma(g_1,g_2g_3)+\sigma'(g_1,g_2)\sigma(g_1g_2,g_3)} \times \omega(g_1,g_2,g_3)\omega'(g_1,g_2,g_3). \tag{F14}$$

Finally, let us remark that if we set $\sigma = 0$, then the fermionic 3-cocycles reduce to ordinary 3-cocycles for the "bosonic" symmetry group $G_b = G_f/Z_2^f$. This reflects the fact that bosonic SPTs can be realized in a fermion system by pairing fermions to form bosons. However, according to the equivalence relation Eq. (F9), when $\lambda \neq 0$ (i.e., the fermionic symmetry group is not simply a direct product $G_f = G_b \times Z_2^f$), a nontrivial bosonic 3-cocycle might still be trivial as a fermionic 3-cocycle. Thus, there is the possibility that a bosonic SPT phase could become trivial in the presence of fermions. Examples of this phenomenon (albeit for symmetry groups including antiunitary symmetries, which we have not considered here) can be found in Ref. [16].

[1] F. D. M. Haldane, Phys. Lett. A **93**, 464 (1983); Phys. Rev. Lett. **50**, 1153 (1983).
[2] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Commun. Math. Phys. **115**, 477 (1988); Phys. Rev. Lett. **59**, 799 (1987).
[3] F. Verstraete, M. A. Martín-Delgado, and J. I. Cirac, Phys. Rev. Lett. **92**, 087201 (2004).
[4] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
[5] J. E. Moore, Nature (London) **464**, 194 (2010).
[6] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
[7] Z.-C. Gu and X.-G. Wen, Phys. Rev. B **80**, 155131 (2009).
[8] X. Chen, Z.-C. Gu, and X.-G. Wen, Phys. Rev. B **82**, 155138 (2010).
[9] X. Chen, Z.-C. Gu, and X.-G. Wen, Phys. Rev. B **83**, 035107 (2011).
[10] N. Schuch, D. Pérez-García, and I. Cirac, Phys. Rev. B **84**, 165139 (2011).
[11] F. Pollmann, E. Berg, A. M. Turner, and M. Oshikawa, Phys. Rev. B **85**, 075125 (2012).
[12] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, Phys. Rev. B **81**, 064439 (2010).
[13] X. Chen, Z.-C. Gu, Z.-X. Liu, and X.-G. Wen, Science **338**, 1604 (2012); Phys. Rev. B **87**, 155114 (2013).
[14] C. Xu, Phys. Rev. B **87**, 144421 (2013).
[15] A. Vishwanath and T. Senthil, Phys. Rev. X **3**, 011016 (2013).
[16] Y.-M. Lu and A. Vishwanath, Phys. Rev. B **86**, 125119 (2012).
[17] C. Xu and T. Senthil, Phys. Rev. B **87**, 174412 (2013).
[18] P. Ye and X.-G. Wen, Phys. Rev. B **87**, 195128 (2013).
[19] Z.-X. Liu and X.-G. Wen, Phys. Rev. Lett. **110**, 067205 (2013).
[20] L. H. Santos and J. Wang, Phys. Rev. B **89**, 195122 (2014).
[21] J. Wang, L. H. Santos, and X.-G. Wen, arXiv:1403.5256.
[22] T. Senthil and M. Levin, Phys. Rev. Lett. **110**, 046801 (2013).
[23] M. A. Metlitski, C. L. Kane, and M. P. A. Fisher, Phys. Rev. B **88**, 035131 (2013).
[24] N. Regnault and T. Senthil, Phys. Rev. B **88**, 161106 (2013).
[25] M. Levin and Z.-C. Gu, Phys. Rev. B **86**, 115109 (2012).
[26] C. Wang and T. Senthil, Phys. Rev. B **87**, 235122 (2013).
[27] Z.-C. Gu and X.-G. Wen, Phys. Rev. B **90**, 115141 (2014).
[28] A. Kapustin, arXiv:1403.1467.
[29] O. M. Sule, X. Chen, and S. Ryu, Phys. Rev. B **88**, 075125 (2013).
[30] C. Wang and M. Levin, Phys. Rev. Lett. **113**, 080403 (2014).
[31] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).
[32] A. Kitaev, arXiv:0901.2686.
[33] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. W. Ludwig, New J. Phys. **12**, 065010 (2010).
[34] X.-G. Wen, Phys. Rev. B **85**, 085103 (2012).
[35] E. Tang and X.-G. Wen, Phys. Rev. Lett. **109**, 096403 (2012).
[36] L. Fidkowski and A. Kitaev, Phys. Rev. B **83**, 075103 (2011).
[37] C. Wang, A. C. Potter, and T. Senthil, Science **343**, 629 (2014).
[38] X.-L. Qi, New J. Phys. **15**, 065002 (2013).
[39] H. Yao and S. Ryu, Phys. Rev. B **88**, 064507 (2013).
[40] Z.-C. Gu and M. Levin, Phys. Rev. B **89**, 201113(R) (2014).
[41] Z. Bi, M. Cheng, and Z.-C. Gu (unpublished).
[42] T. Neupert, C. Chamon, C. Mudry, and R. Thomale, Phys. Rev. B **90**, 205101 (2014).
[43] A. Kapustin, R. Thorngren, A. Turzillo, and Z. Wang, arXiv:1406.7329.
[44] X. Chen, Z.-X. Liu, and X.-G. Wen, Phys. Rev. B **84**, 235141 (2011).
[45] X.-G. Wen, Phys. Rev. D **88**, 045013 (2013).
[46] A. Kapustin and R. Thorngren, Phys. Rev. Lett. **112**, 231602 (2014).
[47] A. Kapustin and R. Thorngren, arXiv:1404.3230.
[48] X. G. Wen, Phys. Rev. B **43**, 11025 (1991).
[49] X. Chen and A. Vishwanath, arXiv:1401.3736.
[50] X. Chen and X.-G. Wen, Phys. Rev. B **86**, 235135 (2012).
[51] A. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006).
[52] T. Senthil and M. Levin, arXiv:1206.1604.
[53] Z. Bi, A. Rasmussen, and C. Xu, arXiv:1309.0515.
[54] F. J. Burnell, X. Chen, L. Fidkowski, and A. Vishwanath, arXiv:1302.7072.
[55] R. Thorngren, arXiv:1404.4385.
[56] X. Chen, Fa Wang, Y.-M. Lu, and D.-H. Lee, Nucl. Phys. B **873**, 248 (2013).
[57] D. Gross, V. Nesme, H. Vogts, and R. Werner, Commun. Math. Phys. **310**, 419 (2012).
[58] M. B. Hastings, arXiv:1305.6625.
[59] S. Ryu and S.-C. Zhang, Phys. Rev. B **85**, 245132 (2012).
[60] P. Etingof, D. Nikshych, and V. Ostrik, Quantum Topol. **1**, 209 (2010).
[61] X. Chen, F. J. Burnell, A. Vishwanath, and L. Fidkowski, arXiv:1403.6491.

[62] S. Ryu, J. E. Moore, and A. W. W. Ludwig, Phys. Rev. B **85**, 045104 (2012).
[63] J. Wang, Z.-C. Gu, and X.-G. Wen, arXiv:1405.7689.
[64] R. B. Laughlin, Phys. Rev. B **23**, 5632 (1981).
[65] By restricting the rotation to an interval, we cannot introduce a net twist, but we can separate equal and opposite twists and focus on the vicinity of just one of them.
[66] The precise relation is slightly subtle, however, because Ω_a measures the charge to the left of the end point a, the very place at which charge is accumulating due to the Hall effect. The precise manner in which Ω_a takes into account the charge located at the end point itself depends on the restriction $U \to U_M$. However, as long as we use the same restriction both to define Ω_a and to implement the winding, the value of the measured "charge" is independent of the choice of restriction. Perhaps the easiest case to interpret is where U_M is defined so that the winding of the phase interpolates linearly over a transition region between no winding outside the interval and the desired winding in the interior of the interval. Then Ω_a measures the charge to the left of a point x, averaged over all x in the transition region. This is nm, rather than the charge $2nm$ measured to the left of a point in the interior of the interval.
[67] A. M. Turner, Y. Zhang, and A. Vishwanath, Phys. Rev. B **82**, 241102(R) (2010).
[68] T. L. Hughes, E. Prodan, and B. A. Bernevig, Phys. Rev. B **83**, 245132 (2011).
[69] L. Fu, Phys. Rev. Lett. **106**, 106802 (2011).
[70] P. Jadaun, D. Xiao, Q. Niu, and S. K. Banerjee, Phys. Rev. B **88**, 085110 (2013).
[71] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306(R) (2007).
[72] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).
[73] R. Roy, Phys. Rev. B **79**, 195322 (2009).
[74] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, Phys. Rev. B **78**, 195424 (2008).
[75] A. Mesaros and Y. Ran, Phys. Rev. B **87**, 155115 (2013).
[76] L.-Y. Hung and X.-G. Wen, Phys. Rev. B **87**, 165107 (2013).
[77] Y.-M. Lu and A. Vishwanath, arXiv:1302.2634.
[78] C. Xu, Phys. Rev. B **88**, 205137 (2013).
[79] A. M. Essin and M. Hermele, Phys. Rev. B **87**, 104406 (2013).
[80] D. Holt, J. Algebra **60**, 307 (1979).
[81] J. Huebschmann, Comment. Math. Helv. **55**, 302 (1980).
[82] K. S. Brown, *Cohomology of Groups* (Springer, Berlin, 1982).
[83] S. Thomas, arXiv:0911.2861.
[84] C. T. Dodson and P. E. Parker, *A User's Guide to Algebraic Topology* (Kluwer Academic, Dordrecht, 1997).
[85] Specifically, the difference from the bosonic case comes from going from Eq. (B5) to Eq. (B6) [due to the potential for $\Omega_b(g_1,g_2)$ and $\Omega_a(g_3,g_4)$ to anticommute] and from going from Eq. (B6) to Eq. (B7) [due to the factor of $\Pi^{\lambda(g_1,g_2)}$ appearing in Eq. (F3).].