Generalized Hall currents in topological insulators and superconductors

David B. Kaplan[1,*] and Srimoyee Sen[2,†]
[1]Institute for Nuclear Theory, University of Washington, Seattle Washington 98195 USA
[2]Department of Physics and Astronomy, Iowa State University, Ames Iowa 50011 USA

(Received 5 April 2023; accepted 6 August 2023; published 30 August 2023)

We generalize the idea of the quantized Hall current to count gapless edge states in topological materials, applying equally well to theories in different dimensions, with or without continuous symmetries in the bulk or chiral anomalies on the boundaries. This current is related to the index of the Euclidean fermion operator and can be calculated via one-loop Feynman diagrams. Quantization of the current is shown to be governed by topology in phase space, and the procedure can be applied to topological classes governed by either $\mathbb{Z}$ or $\mathbb{Z}_2$ invariants. We analyze several explicit examples of free fermions in relativistic field theories. We speculate that it may be possible to extend the technique to interacting theories as well, such as the interesting cases where interactions gap the edge states.

DOI: 10.1103/PhysRevD.108.045019

I. INTRODUCTION

Topological insulators and superconductors, often well described by free fermion theories [1–5], generally have gapless fermion excitations confined to the boundary between two different topological phases. For $d+1$ dimensional defects, where d is odd, these massless fermions can be Weyl fermions. In that case the boundary theory has a chiral anomaly, and the currents for classical axial symmetries can have nonzero divergence in the presence of background gauge fields. These currents are conserved in the higher dimensional bulk theory, and the violation of axial charge in the boundary theory can be understood as the result of a fermion number current flowing in from the bulk, as in the integer quantum Hall effect [6]. The argument can be turned around: the inflow of current from the bulk indicates a boundary theory with a chiral anomaly, which in turn requires massless particles to exist at the boundary.

It is appealing that one can deduce the existence of gapless edge states from the inflow of current from infinity, but it is not general. For example, gapless edge states exist for superconducting systems for which there is no conserved fermion number current, as well as for boundaries with even spatial dimension d where there are no chiral anomalies. The anomaly inflow observed by Callan and Harvey has been shown to be an example of a more general phenomenon operative with all topological insulators and superconductors (see Ref. [7] and references therein). In general, the fermion operators in the bulk for such systems are not self-adjoint when a boundary is present, and the partition function for the bulk has a phase canceled by the anomaly of the edge states. This phenomenon still goes by the title "anomaly inflow" even though for discrete and global anomalies there is no actual current flowing onto or off of the boundary.

In a recent paper we showed, however, that in the cases controlled by perturbative anomalies, one can generalize the idea of Hall currents and detect the existence gapless edge states by this current's divergence [8], even when the anomaly is for a discrete symmetry. That theory makes use of the index of the fermion operator in the Euclidian action in the presence of external "diagnostic fields." A very simple procedure for calculating the current flow in terms of a one-loop Feynman diagram allows one to detect the existence of gapless edge states. One benefit of this complementary construction is that it makes the topological protection of the gapless modes manifest, arising from a winding number in momentum space in direct analogy to the [9] explanation for current quantization in the integer quantum Hall effect [9]. This feature is obscured in the anomaly inflow analysis in [7]. Also, by presenting the calculation in the familiar language of Feynman diagrams, we hope to develop a framework for computing the effects of interactions.

In the present paper we work through the examples given in Ref. [8] in greater detail, including the construction of the generalized Hall current, the topological origins of its nonzero divergence, and the role of regularization. We also extend the previous work with a discussion of the effect of interactions. Our discussion throughout is in the context of relativistic quantum field theory.

II. GENERAL FRAMEWORK

Our starting point is a relativistic quantum field theory of free fermions in infinite $d+1$ dimensional Minkowski

*dbkaplan@uw.edu
†srimoyee08@gmail.com

spacetime. The action is $S_{\mathrm{M}} = \int d^{d+1}x\bar{\psi}\mathcal{D}_{\mathrm{M}}\psi$, where the fermion operator $\mathcal{D}_{\mathrm{M}}$ can include a position dependent mass term, e.g. one that exhibits defects like a domain wall or a vortex singularity, which will serve as proxies for a boundary. If the fermion operator $\mathcal{D}_{\mathrm{M}}$ hosts massless fermion states trapped in these defects, and the defects are space and time translation invariant in the codimensions, then the spectrum will include a state with zero momentum that is constant in the spacetime coordinates of the defect, and is localized on the defect in the transverse direction(s). We wish to have a simple way to detect the existence of such states which makes manifest the underlying topology. The approach we use is to note that when $\mathcal{D}_{\mathrm{M}}$ is analytically continued to Euclidian time, the Euclidian fermion operator $\mathcal{D}$ is an elliptic operator and the propagator $1/\mathcal{D}$ will have poles corresponding to gapless Minkowski edge states with zero momentum. We detect these states by computing the index of $\mathcal{D}$—the number of zeromodes of $\mathcal{D}$ minus that of $\mathcal{D}^\dagger$. One obstacle to this program is that the index only counts normalizable zeromodes, while the states we are discussing are constant in the noncompact boundary dimensions; another is that $\mathcal{D}^\dagger$ also exhibits poles corresponding to these states. Both can be circumvented by adding "diagnostic" background fields that localize these states within the boundary. These diagnostic fields are instantons, where we use that term loosely to describe background field configurations in Euclidian spacetime which have nontrivial topology which localize to a spacetime point massless fermions that they couple to and give the edge state fermion operator a nonzero index. An interesting feature of this approach is that the diagnostic fields are typically not fields that could have been introduced into the original Minkowski theory, as we will explain below. In the presence of these diagnostic fields, solutions to $\mathcal{D}\psi = 0$ can be localized, while those to $\mathcal{D}^\dagger\psi = 0$ are delocalized, resulting in a nonzero index for $\mathcal{D}$. This effect is seen with the toy operator $\mathcal{D} = \partial_x + \epsilon(x)$ [where $\epsilon(x) \equiv x/|x|$ plays the role of the diagnostic field] which has a localized solution to $\mathcal{D}\psi = 0$ while the conjugate operator $\mathcal{D}^\dagger = -\partial_x + \epsilon(x)$ does not. The index then indicates the presence of a massless edge state in the original Minkowski theory, provided that it persists in the limit that the energy density in the diagnostic fields tends to zero so that it cannot significantly affect the bulk gap and change its topological phase. The index being the difference between the number of zeromodes of $\mathcal{D}$ and of $\mathcal{D}^\dagger$ it follows that a nonzero index necessarily implies the presence of a boundary zeromode. The converse does not hold: an index of zero does not rule out the existence of an equal number of zeromodes, but such an occurrence would not be topologically protected and would presumably require a fine-tuning of parameters.

Computing the index then simply requires computing the divergence of an in-flowing current by means of a 1-loop diagram, which generalizes the inflow picture from the quantum Hall system in any dimensions and whether the system has a conserved current or not. Furthermore, the index is shown to be the product of the winding number of the diagnostic fields in coordinate space times the winding number of the fermion dispersion relation in momentum space, making manifest the topological nature of the gapless modes. We will see that the Euclidian momentum space topology is simpler than the topological invariants of the Minkowski systems would suggest: in each case we examine it is governed by the homotopy group $\pi_n(S^n)$, where n is the number of spacetime dimensions. The index we compute reflects the correct invariant of the Minkowski system, $\mathbb{Z}_2$ for example, due to the interplay of momentum space and coordinate space topology in its definition.

Our procedure is (i) start with a Minkowski theory of interest; (ii) analytically continue to Euclidian spacetime; (iii) introduce diagnostic fields to localize edge states; (iv) compute the index of the Euclidian fermion operator in the limit of vanishing diagnostic fields, which involves computing the inflow of a generalized Hall current. In the cases where the fermion number is not conserved, relativistic systems analogous to topological superconductors, the Euclidian action takes the form $S_{\mathrm{E}} = \frac{1}{2}\int \psi^T C\mathcal{D}\psi$, where C is the charge conjugation matrix. In these cases at step (ii) we consider instead the Dirac action $S_E = \int \bar{\psi}\mathcal{D}\psi$ and proceed from there, since solutions to $\mathcal{D}\psi = 0$ are also solutions to $C\mathcal{D}\psi = 0$. The Dirac theory allows for the addition of diagnostic fields not possible in the original theory, such as a $U(1)$ gauge field coupled to fermion number, or fields in $\mathcal{D}$ which appear symmetrically in $C\mathcal{D}$, and hence would not couple to the fermions in $\psi^T C\mathcal{D}\psi$.

The procedure for computing the index of an elliptic operator $\mathcal{D}$ is to first define

$$\mathcal{I}(M) = \mathrm{Tr}\left(\frac{M^2}{\mathcal{D}^\dagger\mathcal{D} + M^2} - \frac{M^2}{\mathcal{D}\mathcal{D}^\dagger + M^2}\right) \tag{2.1}$$

with

$$\mathrm{ind}(\mathcal{D}) = \mathcal{I}(0) \equiv \lim_{M\to 0}\mathcal{I}(M). \tag{2.2}$$

Note that $\mathcal{I}(M)$ can generally take noninteger values for arbitrary M. However, it must be an integer when $M \to 0$. The function $\mathcal{I}(M)$ can be computed by means of Feynman diagrams by first defining

$$K = \begin{pmatrix} 0 & -\mathcal{D}^\dagger \\ \mathcal{D} & 0 \end{pmatrix}, \qquad \Gamma_\chi = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.3}$$

in terms of which $\mathcal{I}(M)$ is expressed as

$$\mathcal{I}(M) = \mathrm{Tr}\left(\Gamma_\chi \frac{M}{K + M}\right). \tag{2.4}$$

The matrix K and Γ have twice the dimension of $\mathcal{D}$. The operator $1/(K+M)$ looks like a fermion propagator in a theory whose action is

$$S_K = \int d^{d+1}x \bar{\Psi}(K+M)\Psi, \tag{2.5}$$

where Ψ has twice the number of components as the original ψ field off our Minkowski theory, but the integration is still over $d+1$ spacetime coordinates. In this extended theory, the quantity $\mathcal{I}(M)$ in Eq. (2.4) can be expressed as the matrix element of the pseudoscalar density,

$$\mathcal{I}(M) = -M \int d^{d+1}x \langle \bar{\Psi}\Gamma_\chi \Psi \rangle, \tag{2.6}$$

where $\langle \bar{\Psi}\Gamma_\chi \Psi \rangle$ is computed using a path integral with weight e^{-S_K}. We can also define an axial current

$$\mathcal{J}^\chi_\mu = \bar{\Psi}\Gamma_\mu \Gamma_\chi \Psi, \tag{2.7}$$

where $\Gamma_\mu = i\partial\tilde{K}(p)/\partial p^\mu$, where we use the tilde to indicate the operator has been Fourier transformed to momentum space. Assuming that $\mathcal{D} = \partial\!\!\!/ + \cdots$, where the ellipsis denotes nonderivative terms, then $\Gamma_\mu = \sigma_1 \otimes \gamma_\mu$, in the notation where $\Gamma_\chi = \sigma_3 \otimes 1$. Applying the axial transformation $\Psi \to e^{i\Gamma_\chi \theta(x)}\Psi$, we can derive the Ward-Takahashi identity for this axial current as

$$\partial^\mu \mathcal{J}^\chi_\mu = 2M\bar{\Psi}\Gamma_\chi \Psi - \mathcal{A}, \tag{2.8}$$

where the first term on the right is the classical divergence due to the mass M, and the second term is the anomaly, which can be computed using the method of Ref. [10],

$$\mathcal{A} = -2 \lim_{\Lambda\to\infty} \mathrm{Tr}(\Gamma_\chi e^{K^2/\Lambda^2}) = -2\mathcal{I}(\infty). \tag{2.9}$$

As we show below the anomaly vanishes in the cases we consider. With $\mathcal{A} = 0$, Eq. (2.8) can then be used to express the index $\mathrm{ind}(\mathcal{D})$ as

$$\mathrm{ind}(\mathcal{D}) = -\frac{1}{2}\lim_{M\to 0} \int d^{d+1}x \partial_\mu \langle \mathcal{J}^\chi_\mu \rangle. \tag{2.10}$$

The current $\mathcal{J}^\chi_\mu$ is what we refer to as the generalized Hall current. With the anomaly $\mathcal{A}$ vanishing, the nonzero divergence of this current in the $M \to 0$ limit arises from infrared divergences in the theory, and its inflow counts gapless edge states in the original Minkowski theory. The diagnostic fields we will add to the $d+1$ dimensional bulk theory are chosen to contribute a term to $\mathcal{J}_\mu$ proportional to a $(d+1)$-index epsilon tensor, where one index is saturated by the derivative acting on the domain wall mass, while the remaining $(d-2)$ indices are saturated by derivatives or gauge fields. With the limited variety of marginal interactions for the fermions, these requirements seem to uniquely specify the choice of fields.

The generalized Hall current in Eq. (2.10) in the $M \to 0$ limit can be computed from a one-loop Feynman diagram $\mathrm{Tr}\Gamma_\mu \Gamma_\chi K^{-1}$, which may need to regulated. The UV regulator will in general contribute to the current in odd spacetime dimensions, but not to its divergence, which arises from infrared physics. However, in these cases the topological meaning of the current is obscured if regulator contributions are neglected, as the Feynman diagram can be interpreted as being proportional to the winding number of a map from momentum space to an n sphere—but only if the momentum space of the fermion is compact. This makes the winding number sensitive to the ultraviolet behavior of the fermion propagator. Here we use Pauli-Villars regularization when required, computing the index of $\mathcal{D}_{\mathrm{reg}} = \frac{\mathcal{D}(m)}{\mathcal{D}(\Lambda)}$, and then sending the regulator mass Λ to infinity. The regulated current then is given by $\mathrm{Tr}\Gamma_\mu \Gamma_\chi K^{-1}_{\mathrm{reg}}$, where K_{reg} is given by Eq. (2.3) with $\mathcal{D}$ replaced by $\mathcal{D}_{\mathrm{reg}}$. We do this in any spacetime dimension $d+1$, but find that for even $d+1$, the regulator does not contribute to the index.

In the next sections we work out in detail the examples of Ref. [8], starting with the continuum example of a single Majorana fermion chain in $1+1$ dimensions with $0+1$ dimensional defect hosting a Majorana zero mode. This is followed by the example of multiple flavors of Majorana fermions in $1+1$ dimension with time-reversal symmetry violation. In both cases, the generalized Hall current correctly produces the index of the fermion operator and counts the number of massless fermions in Minkowski space. We then discuss Dirac fermion in $2+1$ dimensions, which exhibits the integer quantum Hall effect, with the inflow anomaly current described in Ref. [6]. We then add a fermion number violating Majorana mass term to this theory, modeling a topological superconductor in $2+1$ dimensions. Again we show how the divergence of the generalized Hall current correctly detects edge states, even though there is no conventional Hall current in such a system, and we discuss the topological origins of the index. After discussing the $2+1$ dimensional case, we work through an analogous example in $3+1$ dimensions: that of Dirac fermion with a domain wall in its mass which describes a three-dimensional topological insulator and its edge states. We again compute the generalized Hall current, show how it reproduces the index, and display it topological origins. The final section provides a qualitative argument in the context of $1+1$ dimensional interacting Majorana chains that the divergence of the generalized Hall current detects zeromodes even in the presence of interactions, even though the index of the free fermion propagator is no longer relevant.

III. MAJORANA FERMION 1+1 DIMENSIONS

A. One flavor of Majorana fermion

For our first example we consider a single Majorana fermion in $1+1$ dimension with domain wall profile for the Majorana mass, with the $0+1$ dimensional defect hosting a 1-component real massless fermion. Majorana edge states were first discussed in Refs. [11] and [12]. In this theory there is no continuous symmetry, and hence no conserved current in the bulk, and no chirality or chiral anomaly on the $0+1$ dimensional defect. Therefore this system cannot exhibit the conventional Hall response. As we will show, however, we can compute generalized Hall currents which converge on the defect when it hosts gapless states.

The Minkowski Lagrangian in this case is

$$\mathcal{L}_M = \frac{1}{2}\psi^T C(i\partial\!\!\!/ - m)\psi, \tag{3.1}$$

which is the continuum version of the 1-flavor theory considered in Ref. [13]. Here ψ is a real, two-component Grassmann spinor. The γ matrices we take are

$$\gamma^0 = C = \sigma_2, \qquad \gamma^1 = -i\sigma_1, \qquad \gamma_\chi = \sigma_3, \tag{3.2}$$

where σ_i are the Pauli matrices. This model of Majorana fermions has a rich phase structure when interactions are included [13], but in this section we restrict ourselves to analyzing the free theory. To construct the generalized Hall current for this model we first define the Euclidian theory[1]

$$\mathcal{L}_E = \frac{1}{2}\psi^T C\mathcal{D}\psi, \tag{3.3}$$

where in Euclidean spacetime

$$\mathcal{D} = \partial\!\!\!/ + m, \quad \gamma_0 = C = \sigma_2, \quad \gamma_1 = -\sigma_1, \quad \gamma_\chi = \sigma_3. \tag{3.4}$$

We now wish to study the index of $\mathcal{D}$, somewhat modified (there being a one-to-one correspondence between zeromodes of $\mathcal{D}$ and $C\mathcal{D}$). The modifications involve replacing the step function mass with a general scalar field $\phi_1(x)$, and adding a pseudoscalar field $\phi_2(x)$ as our diagnostic field, so that the more general operator we consider is

$$\mathcal{D} = \partial\!\!\!/ + \phi_1 + i\phi_2\gamma_\chi = \begin{pmatrix} \phi_1 + i\phi_2 & -i\partial_0 - \partial_1 \\ i\partial_0 - \partial_1 & \phi_1 - i\phi_2 \end{pmatrix},$$
$$\mathcal{D}^\dagger = -\partial\!\!\!/ + \phi_1 - i\phi_2\gamma_\chi = \begin{pmatrix} \phi_1 - i\phi_2 & i\partial_0 + \partial_1 \\ -i\partial_0 + \partial_1 & \phi_1 + i\phi_2 \end{pmatrix}. \tag{3.5}$$

To understand the logic of this construction, first consider the operator we are ultimately interested in, where $\phi_2 = 0$ and $\phi_1(x) = m_0\epsilon(x_1)$ with $m_0 > 0$. In this case we find two solutions each to the equations $\mathcal{D}\psi = 0$ and $\mathcal{D}^\dagger\chi = 0$, namely

$$\psi_-(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{m_0|x_1|}, \quad \psi_+(x) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-m_0|x_1|},$$
$$\chi_-(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-m_0|x_1|}, \quad \chi_+(x) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{m_0|x_1|}, \tag{3.6}$$

where the $\pm$ subscript refers to the eigenvalue of γ_1. None are normalizable since they are constant in x_0, whether or not they are localized in x_1 about $x_1 = 0$, and therefore the index calculation will give us $\mathrm{ind}(\mathcal{D}) = 0$. However, if we now add $\phi_2 = \mu_0\epsilon(x_0)$ with $\mu_0 > 0$, a domain wall in Euclidian time, we find the four solutions to $\mathcal{D}\psi = 0$ and $\mathcal{D}^\dagger\chi = 0$ are

$$\psi_-(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{m_0|x_1|+\mu_0|x_0|}, \quad \psi_+(x) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-m_0|x_1|-\mu_0|x_0|},$$
$$\chi_-(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-m_0|x_1|+\mu_0|x_0|}, \quad \chi_+(x) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{m_0|x_1|-\mu_0|x_0|},$$
$$(x_2 \to x_0). \tag{3.7}$$

$\mathcal{D}$ now has a single normalizable zeromode ψ_+ localized at $x_0 = x_1 = 0$ while $\mathcal{D}^\dagger$ has none. Therefore the index is given by $\mathrm{ind}(\mathcal{D}) = 1$ in this background scalar field. By considering the other signs for m_0 and μ_0 one sees that more generally we get $\mathrm{ind}(\mathcal{D}) = -\nu_\phi$, where ν_ϕ is the winding number of $\phi = \phi_1 + i\phi_2$ in the $x_0 - x_1$ plane.[2] As we shall see, this expectation is born out by the explicit calculation we give below. Furthermore, our results persist even no matter how small we take $|\mu_0|$. We also see how the zeromodes we are counting would not have existed if there had not been a solution to $\mathcal{D}\psi = 0$ for the case of interest, $\phi_2 = 0$ and $\phi_1 = m_0\epsilon(x_1)$ (for either sign of m_0), and so the nonzero value for $\mathrm{ind}(\mathcal{D})$ informs us that the original Minkowski theory has massless edge states.

The reason this procedure works is that the background ϕ_2 field breaks symmetries of the original theory (in this case, time-reversal symmetry), and a constant field would

[1] We use the mostly minus convention for the Minkowski metric, and in $d+1$ dimensions we denote both Minkowski and Euclidean time by x^0. The relation between Euclidian and Minkowski γ matrices is $\gamma^0_M = \gamma^0_E$, $\gamma^i_M = i\gamma^i_E$, and γ_χ is the same in both Minkowski and Euclidean spacetime.

[2] This is easier to see if one replaces the singular step functions with something smoother, such as a tanh function. In fact, it has to be smoothed out to justify the derivative expansion we perform next.

gap the spectrum entirely. However, by giving a domain wall structure to ϕ_2, which by itself would lead to gapless edge states, it cannot gap fermions along the domain wall surface at $x_0 = 0$. Thus theory with both $\phi_{1,2}$ domain walls will be left with a massless, normalizable fermion at the origin where the two domain walls cross.

Following the procedure laid out in the previous section, we compute the index of $\mathcal{D}$ by constructing the fermion operator K which serves as the fermion operator in a theory with twice the number of fermion degrees of freedom,

$$K = \begin{pmatrix} 0 & -\mathcal{D}^\dagger \\ \mathcal{D} & 0 \end{pmatrix} = \Gamma_\mu \partial_\mu + i(\phi_2 \Gamma_2 + \phi_1 \Gamma_3), \qquad (3.8)$$

where we have defined the five 4×4 matrices

$$\Gamma_i = \sigma_1 \otimes \gamma_i, \qquad \Gamma_2 = \sigma_1 \otimes \gamma_\chi, \qquad \Gamma_3 = -\sigma_2 \otimes 1, \qquad \Gamma_\chi = \sigma_3 \otimes 1 \qquad (3.9)$$

with $i = 0, 1$.

Our task is to compute the part of the chiral current $\mathcal{J}_\mu = \bar{\Psi}\Gamma_\mu \Gamma_\chi \Psi$ that contributes to the index, where $\mu = 0, 1$ and Ψ is a fermion with action $S = \bar{\Psi} K \Psi$. To do this we first write the scalars as

$$\phi = \phi_1 + i\phi_2 = (v + \rho(x)) e^{i\theta(x)} \qquad (3.10)$$

assuming constant v with ρ and θ slowly varying about $\rho = \theta = 0$, their gradients falling off at infinity. To compute the part of the current that contributes to the index we need the leading term in a $1/v$ expansion, since higher order terms will drop off too fast at infinity to contribute to the integral $\int \partial_\mu \mathcal{J}_\mu$. Thus we can write

$$K = K_0 + \delta K, \qquad (3.11)$$

with

$$K_0 = \partial_\mu \Gamma_\mu + iv\Gamma_3, \qquad \delta K = iv\theta(x)\Gamma_2 + i\rho(x)\Gamma_3. \qquad (3.12)$$

Then K_0^{-1} will be our free fermion propagator, while we perturb in δK.

When expanding $\mathcal{J}_\mu$ in δK, note that because of the insertion of Γ_χ in the fermion loop we require that the rest of the graph supplies one each of the other four Γ matrices in order to get a nonzero contribution from the trace. The matrices $\Gamma_{0,1,3}$ can be supplied by the fermion propagators $1/K_0$, while Γ_2 must arise from an insertion of θ. Thus the leading contribution is given by the graph in Fig. 1 expanded to linear order in the momentum carried by θ.

Using $\tilde{K}_0$ to denote the Fourier transform of K_0 in momentum space, the diagram can be computed as

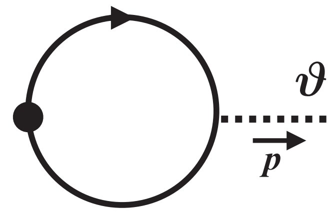

FIG. 1. The loop diagram for computing the generalized Hall current for the $1+1$-dimension Dirac fermion. The black dot is an insertion of the chiral current $\Gamma_\mu \Gamma_\chi$ with incoming momentum p_ν. The outgoing field α is the spatially varying part of the phase of the complex field $\phi_1 + i\phi_2$, and the fermion propagator is given by K_0^{-1}.

$$\begin{aligned} \mathcal{J}_\mu &= v \frac{\partial \theta}{\partial x_\nu} \int \frac{d^2 q}{(2\pi)^2} \mathrm{Tr}\left[\Gamma_\mu \Gamma_\chi \left(\frac{\partial \tilde{K}_0^{-1}}{\partial q_\nu}\right) \Gamma_2 \tilde{K}_0^{-1}\right] \\ &= \epsilon_{\mu\nu} \partial_\nu \theta \int \frac{d^2 q}{(2\pi)^2} \frac{4v^2}{(q^2 + v^2)^2} \\ &= \frac{1}{\pi} \epsilon_{\mu\nu} \partial_\nu \theta. \end{aligned} \qquad (3.13)$$

In this expression, the derivative ∂_ν acting on $\theta(x)$ is with respect to x_ν, while the derivative ∂_ν acting on $\tilde{K}_0$ is with respect to q_ν. The positive sign in the first line above arises from (-1) from the fermion loop, a $(-i)$ from Fourier transforming $p_\nu \to -i\partial/(\partial x_\nu)$, and another $(-i)$ from the θ vertex factor $-iv\Gamma_2$.

From Eq. (2.10) it follows that the index of $\mathcal{D}$ is given by

$$\mathrm{ind}(\mathcal{D}) = -\frac{1}{2} \int d^2 x \partial_\mu \mathcal{J}_\mu = -\frac{1}{2\pi} \oint \frac{\partial \theta}{\partial x_\mu} d\ell_\mu = -\nu_\phi, \quad (3.14)$$

where ν_ϕ is the winding number of ϕ. This result agrees with what was predicted from our earlier heuristic argument with a configuration of crossed domain walls.

Before turning to the question of topology we address two issues about the way we handled the Ward-Takahashi identity for the generalized Hall current, given in Eq. (2.8). First of all, we set $M = 0$ from the outset, rather than performing the computation at nonzero M and then taking the limit $M \to 0$ at the end. This is discussed further in Sec. VI, but in brief, the role of M was to serve as an IR regulator for the calculation, but instead we used our background field ϕ to serve as the IR regulator. Note that the generalized Hall current we found in Eq. (3.13) is proportional to $\partial_\mu \theta = i\phi^* \overleftrightarrow{\partial}_\mu \phi / 2|\phi|^2$, and that the inverse dependence on $|\phi|$ indicates its role as an IR regulator. Secondly, we ignored the anomaly term, $\mathcal{A}$. This quantity can be computed using the methods of Fujikawa [10], where

$$\mathcal{A} = \lim_{\Lambda \to \infty} \mathrm{Tr} \Gamma_\chi e^{K^2/\Lambda^2}. \qquad (3.15)$$

[Note that K, defined in Eq. (3.12), is anti-Hermitean.] The two-derivative term in K^2 gives rise to the Gaussian integral

$$\int \frac{d^2k}{(2\pi)^2} e^{-k^2/\Lambda^2} \propto \Lambda^2. \tag{3.16}$$

On the other hand, since the Γ matrices obey the Clifford algebra for $SO(4)$, the Γ-matrix trace with Γ_χ requires that the expansion of e^{K^2/Λ^2} supplies at least four different Γ_μ to the trace before one obtains a nonzero value. As K/Λ is linear in the Γ matrices, these four Γ matrices will be accompanied by a factor of $1/\Lambda^4$ which overwhelms the factor of Λ^2 from the momentum integral and causes the anomaly term $\mathcal{A}$ to vanish as $\Lambda \to \infty$. This will occur in all of our examples, since the doubled theory always has the same $d+1$ spacetime dimension as the original theory, while the Γ matrices belong to the Clifford algebra for $SO(d+3)$.

The calculation we have performed does not make explicit the topological quantization of the index. To display the topology underpinning of the calculation explicitly we define $\tilde{\mathcal{D}}_0$ to equal $\tilde{\mathcal{D}}$ with $\phi_1 \to v$ and $\phi_2 \to 0$ and use the identities $\Gamma_\mu = -i\partial_\mu \tilde{K}_0$ and $\{\gamma_\chi, \tilde{\mathcal{D}}_0\} = 2v\gamma_\chi$ to rewrite Eq. (3.13) as

$$\begin{aligned}\mathcal{J}_\mu &= -iv\partial_\nu\theta \int \frac{d^2q}{(2\pi)^2} \mathrm{Tr}[\Gamma_\chi(\partial_\nu \tilde{K}_0^{-1})\Gamma_2 \tilde{K}_0^{-1}\partial_\mu \tilde{K}_0] \\ &= -iv\partial_\nu\theta \int \frac{d^2q}{(2\pi)^2} \mathrm{Tr}[\gamma_\chi(\tilde{D}_0^{-1}\partial_\mu \tilde{D}_0 \partial_\nu \tilde{D}_0^{-1} + \tilde{D}_0 \to \tilde{D}_0^\dagger)] \\ &= \frac{i}{4}\epsilon_{\mu\nu}\partial_\nu\theta\epsilon_{\sigma\tau} \int \frac{d^2q}{(2\pi)^2} \mathrm{Tr}[\gamma_\chi(\tilde{D}_0^{-1}\partial_\sigma \tilde{D}_0 \tilde{D}_0^{-1}\partial_\tau \tilde{D}_0 \\ &\quad + \tilde{D}_0\partial_\sigma \tilde{D}_0^{-1}\tilde{D}_0\partial_\tau \tilde{D}_0^{-1} + \tilde{D}_0 \to \tilde{D}_0^\dagger)].\end{aligned} \tag{3.17}$$

We can then define the special unitary matrix

$$\xi = \frac{\tilde{\mathcal{D}}_0}{\sqrt{\det \tilde{\mathcal{D}}_0}} = \frac{v}{\sqrt{q^2+v^2}} + i\hat{q}_\mu\gamma_\mu \frac{q}{\sqrt{q^2+v^2}}, \tag{3.18}$$

in terms of which Eq. (3.17) can be rewritten as

$$\begin{aligned}\mathcal{J}_\mu &= \frac{i}{2}\epsilon_{\mu\nu}\partial_\nu\theta\epsilon_{\sigma\tau} \int \frac{d^2q}{(2\pi)^2} \mathrm{Tr}\gamma_\chi[(\xi^\dagger\partial_\sigma\xi)(\xi^\dagger\partial_\tau\xi) \\ &\quad + (\xi\partial_\sigma\xi^\dagger)(\xi\partial_\tau\xi^\dagger)],\end{aligned} \tag{3.19}$$

unaffected by the normalizing determinant factor (which cannot vanish for nonzero v). As a final manipulation, we can define the "axial current,"

$$A_j = \frac{i}{2}(\xi^\dagger\partial_j\xi - \xi\partial_j\xi^\dagger), \tag{3.20}$$

and using the identity $\gamma_\chi\xi\gamma_\chi = \xi^\dagger$ we can write the generalized Hall current as

$$\mathcal{J}_\mu = -i\epsilon_{\mu\nu}\partial_\nu\theta\epsilon_{\sigma\tau} \int \frac{d^2q}{(2\pi)^2} \mathrm{Tr}[\gamma_\chi A_\sigma A_\tau]. \tag{3.21}$$

In Appendix A we show that the above integral is proportional to the winding number of the map provided by the Dirac propagator from momentum space compactified to S^2, to "Dirac space" (the analog of the Bloch sphere), also S^2 in two spacetime dimensions. That winding number ν_q is an element of $\pi_2(S^2) = \mathbb{Z}$, and for this particular map we have $\nu_q = 1$. Making use of the normalization given in Eq. (A.11) we arrive at

$$\mathcal{J}_\mu = \frac{\epsilon_{\mu\nu}\partial_\nu\theta}{\pi}\nu_q, \qquad \nu_q = 1. \tag{3.22}$$

The topological quantization in momentum space is analogous to the topological origin of the quantization of the integer quantum Hall effect discovered in the celebrated TKNN paper and related work [9,14,15].[3]

Our final result for the index is then given by

$$\mathrm{ind}(\mathcal{D}) = -\nu_\phi\nu_q, \tag{3.23}$$

where ν_ϕ is the winding number of our diagnostic field in coordinate space, obtained by integrating $\epsilon_{\mu\nu}\partial_\nu\theta$ and determined by the homotopy group $\pi_1[U(1)] = \mathbb{Z}$. We see that the index, which counts edge states, is manifestly quantized by simultaneous nontrivial topology in both momentum and coordinate space. This result confirms our heuristic argument about localizing the zeromode with crossed domain walls, and agrees with Eq. (3.14).

Relativistic quantum field theories can be assigned to the same ten topological classes as used for condensed matter systems; for example see Ref. [5]. The translation is simple, with what are called time-reversal symmetry, particle hole symmetry, and sublattice symmetry replaced by the conventional T, C and CT symmetries respectively in the relativistic theory. The ten categories are constructed from (i) time-reversal symmetry with $T^T = T$, or T symmetry with $T^T = -T$, or T violation; (ii) charge conjugation symmetry with $C^T = C$, or C symmetry with $C^T = -C$, or

[3]In contrast with the physics in $2+1$ dimensions for which the topology is dependent on the UV regulator (as we demonstrate in the next section), the calculation here and in $3+1$ dimensions is insensitive to UV physics. That is because the current in Eq. (3.21) is proportional to $\partial_\mu\theta = i\phi^*\overleftrightarrow{\partial}_\mu\phi/2|\phi|^2$ which cannot be obtained by the variation of a local relevant or marginal operator, and hence cannot be sensitive to the UV. In contrast, the current we will find in $2+1$ dimensions [Eq. (4.9)] is the variation of the Chern-Simons operator which is a local marginal operator and receives contributions from the UV regulator.

C violation; (iii) both C and T violation, but good CT symmetry. The only caveat in comparing using the tables from condensed matter papers is that the symmetry of C is opposite that of the conventional particle-hole symmetry "P," since in the relativistic case charge conjugation is the transformation $\psi \to C\bar{\psi}^T$ for a Dirac fermion, while particle hole symmetry is $\psi \to P\psi^*$, without the extra γ^0 matrix. In the present example we have both T and C with symmetric T and antisymmetric C, which puts the system in the BDI [16] class, with topological invariant $\mathbb{Z}$. If we had considered a N_f -flavor version of the theory considered in this section without special global flavor symmetries, we would have trivially found $\mathrm{ind}(\mathcal{D}) = -N_f \nu_\phi \nu_q$, which can take on any value in $\mathbb{Z}$ as one would expect for a BDI topological class in one spatial dimension.

B. Multiple flavors of $d=1+1$ Majorana fermions with time-reversal symmetry violation

With N_f flavors of free fermions in our $1+1$ dimensional model, it is possible to include both scalar and pseudoscalar mass terms:

$$\mathcal{L}_M = \frac{1}{2}\psi_i^T C(i\partial\!\!\!/ \delta_{ij} - m_{ij} - i\gamma_\chi \mu_{ij})\psi_j, \qquad (3.24)$$

where m must be a real and symmetric matrix, while μ is imaginary and antisymmetric. Without loss of generality it is possible to take m_{ij} to be diagonal.

The theory with $\mu_{ij} = 0$ is invariant under the antiunitary time-reversal symmetry, $\psi_i(x,t) \to \sigma_1\psi_i(x,-t)$. When the masses have domain wall profiles, such as $m_i = m_0\epsilon(x)$ with $m_0 > 0$, there will be N_f massless Majorana modes localized at the mass defect with wave function

$$\eta_i(t)\begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad (3.25)$$

in the $0+1$-dimension theory at the defect. In order to gap these modes in the free theory, one might add a term $\mu_{ij}\eta_i\sigma_1\eta_j$ with μ imaginary and antisymmetric, and it is easy to see that this term lifts into the bulk theory as the μ term in Eq. (3.24). However, the pseudoscalar μ term is odd under the time-reversal symmetry we identified (since μ is imaginary), and therefore time-reversal symmetry requires $\mu = 0$ and ensures that the edge states remain gapless. Such a system is in the BDI topological class, whose gapless states are characterized by the group $\mathbb{Z}$ [17].

For $N_f \geq 3$ it is possible to show that one can choose m and μ such that time-reversal symmetry is broken, and generically the edge states will be gapped pairwise, so that there will be one massless edge state for N_f odd, and none for N_f even. Such a system is in the D class, with topology characterized by the group $\mathbb{Z}_2$ [17].

For $N_f = 2$, time reversal is not actually broken when $\mu \neq 0$, since the theory is invariant under the simultaneous antiunitary transformations $\psi_1(x,t) \to +\sigma_1\psi_1(x,-t)$ and $\psi_2(x,t) \to -\sigma_1\psi_2(x,-t)$. In this case edge states can still be gapped because the fermions transform with opposite signs under time reversal and the system is topologically trivial. So the system is indistinguishable from the D class as far as edge states are concerned.

We focus on this simplest $N_f = 2$ case to show how the index calculation procedure we have developed here gets the correct answer, that there are no gapless edge states. The Minkowski theory we consider is

$$\mathcal{L}_M = \frac{1}{2}\psi_i^T C(i\partial\!\!\!/ - m_0\epsilon(x) - i\gamma_\chi\mu\tau_2)\psi_i, \qquad (3.26)$$

where we have suppressed the two flavor indices. The $(i\partial\!\!\!/ - m_0\epsilon(x))$ operator is diagonal in flavor, while τ_2 is the y-Pauli matrix acting in flavor space, and μ is real and constant in spacetime.

Our Euclidean operator $\mathcal{D}$ with a diagnostic field $\phi(x)$ in this case is given by

$$\mathcal{D} = \partial\!\!\!/ + \phi_1(x) + i(\phi_2(x) + \mu\tau_2)\gamma_\chi. \qquad (3.27)$$

This can be diagonalized in flavor to give two 1-flavor Dirac operators

$$\mathcal{D}_\pm = \partial\!\!\!/ + \phi_1(x) + i(\phi_2(x) \pm \mu)\gamma_\chi. \qquad (3.28)$$

Now we expect the sum of two contributions to $\mathcal{J}_\mu$ proportional to $\epsilon_{\mu\nu}\partial_\nu(\theta_+ + \theta_-)$ where

$$\theta_\pm = \arctan\frac{\phi_2(x) \pm \mu}{\phi_1(x)}. \qquad (3.29)$$

Now we see that even if $\phi(x) = \phi_1(x) + i\phi_2(x)$ has winding number, as we take the limit $\phi_2 \to 0$, that winding number vanishes. For example, if we take $\phi_1 = m\epsilon(x)$ and $\phi_2(x) = m'\epsilon(\tau)$, the two contributions to the index will both equal 1 for $|m'| > |\mu|$, but will jump to zero for $|m'| < |\mu|$ as we remove our diagnostic field—this is possible because the bulk goes gapless at the critical value $|m'| = |\mu|$. So the index in this case would give the correct answer of zero, as it will for any even number of flavors.

If we have an odd number of flavors with a constant μ matrix for $N_f = 2n+1$, we will find n pairs of fermions coupling to $\phi_2(x)$ with $\pm\mu_i$ shifts, each contributing zero to the index; however there would be one flavor with no shift, and we would find an index of 1 then. Thus we find that $\mathrm{ind}(\mathcal{D})$ takes values in $\mathbb{Z}_2$. That is the correct answer since having no T symmetry while having C symmetry with antisymmetric C puts the model in the D topological class, whose topological invariant in one spatial dimension is $\mathbb{Z}_2$.

We emphasize though that in our calculation the reduction of the topological invariant from $\mathbb{Z}$ in the time-reversal symmetric case to $\mathbb{Z}_2$ in the case with broken time-reversal symmetry manifests itself in the change in the spacetime coordinate topology that the fermion sees, and not due to a change in momentum space. The index we compute sees the topology in phase space, and is sensitive to both.

IV. DIRAC FERMION AND MAJORANA FERMIONS IN 2+1 DIMENSIONS

A. A $d=2+1$ Dirac fermion with $U(1)$ fermion number symmetry

Our next example is a massive Dirac fermion in $2+1$ (Minkowski) dimensions, which is directly analogous to the integer quantum Hall effect and is quite familiar. The coordinates are $\{x^0, x^1, x^2\} = \{t, x, y\}$, and the fermion mass $m(y)$ which has a "domain wall" structure—a monotonically increasing function of y with $m(0) = 0$ [18]. For convenience we choose the particular basis for the γ matrices

$$\gamma^0 = \sigma_2, \qquad \gamma^1 = -i\sigma_1, \qquad \gamma^2 = i\sigma_3, \tag{4.1}$$

in which case the Dirac equation $[i\partial\!\!\!/ - m(y)]\psi = 0$ has the two special solutions,

$$\psi_\pm = e^{-i\omega(t\pm x)} e^{\mp\int_0^y ds m(s)} \chi_\pm, \tag{4.2}$$

where $\chi_\pm$ are constant 2-component spinors satisfying $\sigma_3\chi_\pm = \pm\chi_\pm$. The solution ψ_+ is localized on the $1+1$ dimensional domain wall and corresponds to a massless Weyl fermion that travels at the speed of light in the $-x$ direction. However, since $e^{+\int_0^y ds m(s)}$ is not normalizable, the ψ_- solution does not correspond to a state in the Hilbert space. Therefore the spectrum on the domain wall is chiral, and if the fermion number is gauged, this $1+1$ dimensional theory on the domain wall is anomalous.

We must next determine how to localize the massless edge state with diagnostic fields and produce a nonzero index. This is readily accomplished with help from the Atiyah-Singer index theorem, which states that in $d = 2$ the index of the massless Dirac operator in the presence of gauge fields is given by $\frac{1}{2\pi}\int d^2x\epsilon_{ij}F_{ij} = \frac{1}{2\pi}\oint \vec{A}\cdot d\vec{\ell}$, where the loop integral is computed at infinity. Therefore we can add a 3D gauge field and be assured that for some background field configurations a nonzero index will result whenever a gapless edge state existed in the absence of those gauge fields.

All other eigenstates of the Dirac operator are gapped and are not localized. One might therefore expect that these heavy states could be integrated out, leaving an effective $1+1$ dimensional theory of a Weyl fermion at $y = 0$, along with irrelevant operators. However, when fermion number symmetry is gauged the heavy fermions do not decouple entirely, giving rise to a marginal Chern-Simons operator after being integrated out of the theory, as pointed out by a number of authors [6,19,20]. For a constant mass m, this contribution is[4]

$$\mathcal{L}_{\rm CS} = \frac{1}{4\pi}\frac{m}{|m|}\epsilon^{\mu\nu\rho}A_\mu\partial_\nu A_\rho. \tag{4.3}$$

The coefficient of this operator depends on the sign of m but not its magnitude; the dependence on the sign is required because m changes sign under T or P transformations (time-reversal and space reflection), as does the Chern-Simons operator. Variation of $\mathcal{L}_{\rm CS}$ with respect to the gauge field gives rise to a current,

$$J^\mu_{\rm CS} = \frac{1}{2\pi}\frac{m}{|m|}\epsilon^{\mu\nu\rho}\partial_\nu A_\rho. \tag{4.4}$$

Callan and Harvey [6] discussed the effects of this operator in the presence of a domain wall profile, substituting the step function $\epsilon(y)$ for $m/|m|$ in the above expression, so that $J^\mu_{\rm CS}$ corresponds to a current flowing on or off of the domain wall from both sides, with a divergence proportional to $\delta(y)$ that exactly cancels the anomalous divergence of the chiral current on the domain wall. There are a few problems with their analysis. For one, the substitution of spatially varying mass for a constant mass in the coefficient of the Chern Simons term is not correct where the fermion mass is varying appreciably, i.e. near the domain wall; however to demonstrate overall charge conservation, one need only look at the spatial integral of the divergence of the current, which involves the Chern Simons coefficient at infinity, where the substitution is valid if the mass is not rapidly changing there. A second problem is that while the 1-loop integral for deriving $\mathcal{L}_{\rm CS}$ is finite, the full theory requires a regulator which contributes to the Chern-Simons operator as well. For example, with a Pauli-Villars regulator of mass Λ, the coefficient in Eqs. (4.3) and (4.4) becomes proportional to $(m/|m| - \Lambda/|\Lambda|)$. With the Pauli-Villars mass Λ being independent of y, the effect is to double the inflowing current on one side of the domain wall, and cancel it on the other. This has no effect on the divergence of the current, but corrects the coefficient of the Chern-Simons term and makes a topological interpretation possible by compactifying momentum space. Other regularization schemes, such as the lattice, can have richer topological phase structure.

Even with the correct expression for the current, the connection with the integer quantum Hall effect is obscure, beyond the fact that an electric field in the x direction gives

[4]Our notation is that the covariant derivative is $D_\mu = (\partial_\mu - iA_\mu)$, where A_μ has mass dimension 1, setting the electric charge to $e = 1$.

rise to a current in the y direction; missing is an analog of the famous plot of the resistivity ρ_{xy} versus magnetic field in the condensed matter system, with its characteristic steps. In the Dirac fermion case, the domain wall plays the role of one of the boundaries of the quantum Hall system, and the Dirac mass—which is time-reversal violating in $2+1$ dimensions—plays the role of the magnetic field, while the resistivity is the ratio of the applied electric field to the Chern Simons current. However, one does not see a stepwise increase in the current as a function of the Dirac mass. Once again, that is a result of how the short-distance physics of the theory is being regulated. For example, when one uses a lattice and Wilson terms to regulate the UV, one does see quantized jumps in the Chern-Simons current as a function of the fermion mass [21–23].

The Euclidean fermion operator for this fermion is

$$\mathcal{D}=[\not{D}+m(y)],\quad \gamma_0=\sigma_2,\quad \gamma_1=-\sigma_1,\quad \gamma_2=\sigma_3, \tag{4.5}$$

with $D_\mu=\partial_\mu+iA_\mu$. From Eq. (2.3) we have

$$K=\begin{pmatrix}0 & -\mathcal{D}^\dagger\\ \mathcal{D} & 0\end{pmatrix}=D_\mu\Gamma_\mu-im(y)\Gamma_3, \tag{4.6}$$

where we define the matrices [satisfying the $SO(5)$ Clifford algebra]

$$\Gamma_\mu=\sigma_1\otimes\gamma_\mu,\quad \Gamma_3=\sigma_2\otimes 1,\quad \Gamma_\chi=\sigma_3\otimes 1 \tag{4.7}$$

for $\mu=0, 1, 2$. Our task now is to compute the generalized Hall current $\mathcal{J}_\mu$ in Eq. (2.7). Since the bulk mass $|m|$ serves to regulate the IR divergences of the theory, we can take the limit $M\to 0$ in Eq. (2.10) from the start. Another simplification is that since we only need the current inflow from infinity in the limit that the gauge field strength is weak, we can compute $\mathcal{J}_\mu$ to leading order in a $1/m$ expansion which is given by the Feynman diagram in Fig. 2 expanded to first order in the incoming momentum.[5]

We first perform the calculation naively, without a Pauli-Villars regulator, with the result

$$\mathcal{J}_\alpha=\partial_\gamma A_\beta\int\frac{d^3q}{(2\pi)^3}\mathrm{Tr}\Gamma_\alpha\Gamma_\chi(\partial_\gamma\tilde{K}_0^{-1}(q))\Gamma_\beta\tilde{K}_0^{-1}(q), \tag{4.8}$$

where as before, the tilde indicates a Fourier transform to momentum space, and $\tilde{K}_0=\tilde{K}|_{A_\mu=0}$. The factor of $\Gamma_\alpha\Gamma_\chi$ in the trace comes from the insertion of the generalized Hall current, while the photon vertex gives $i\Gamma_\beta$; the two factors of $\tilde{K}_0^{-1}$ are the two fermion propagators, and the derivative $\partial_\gamma=\partial/\partial q_\gamma$ arises from expanding the graph to first order in the external momentum p. With $\Gamma_\alpha=-i\partial_\alpha\tilde{K}_0$, we can rewrite this as

$$\begin{aligned}\mathcal{J}_\alpha(p)&=\partial_\gamma A_\beta\int\frac{d^3q}{(2\pi)^3}\\&\quad\times\mathrm{Tr}[\Gamma_\chi(\tilde{K}_0^{-1}\partial_\gamma\tilde{K}_0)(\tilde{K}_0^{-1}\partial_\beta\tilde{K}_0)(\tilde{K}_0^{-1}\partial_\alpha\tilde{K}_0)]\\&=-\epsilon_{\alpha\beta\gamma}\partial_\gamma A_\beta\int\frac{d^3q}{(2\pi)^3}\frac{4m}{(m^2+q^2)^2}\\&=-\frac{1}{2\pi}\frac{m}{|m|}\epsilon_{\alpha\beta\gamma}\partial_\gamma A_\beta.\end{aligned} \tag{4.9}$$

With a domain wall profile for $m(y)$, the above expression exhibits a generalized Hall current flow converging on (or diverging from) the wall similar to the electromagnetic current found by Callan and Harvey [6].

To better understand the topology behind the loop integral, it is convenient to rewrite Eq. (4.9) in terms of the fermion operator $\tilde{\mathcal{D}}_0$ of the undoubled theory,

$$\begin{aligned}\mathcal{J}_\alpha&=-\epsilon_{\alpha\beta\gamma}\partial_\gamma A_\beta\bigg(\frac{1}{3}\epsilon_{ijk}\int\frac{d^3q}{(2\pi)^3}\\&\quad\times\mathrm{Tr}[(\tilde{\mathcal{D}}_0^{-1}\partial_i\tilde{\mathcal{D}}_0)(\tilde{\mathcal{D}}_0^{-1}\partial_j\tilde{\mathcal{D}}_0)(\tilde{\mathcal{D}}_0^{-1}\partial_k\tilde{\mathcal{D}}_0)]\bigg).\end{aligned} \tag{4.10}$$

As in the example in the previous section, for constant m we can define the $SU(2)$ matrix

$$U(\mathbf{q})=\frac{\tilde{\mathcal{D}}_0(\mathbf{q})}{\sqrt{\det\tilde{\mathcal{D}}_0(\mathbf{q})}}\equiv\cos\frac{\theta}{2}+i\hat{\boldsymbol{\theta}}\cdot\boldsymbol{\gamma}\sin\frac{\theta}{2} \tag{4.11}$$

where $\boldsymbol{\theta}$ is a real 3-vector with $\theta=|\boldsymbol{\theta}|$ and

$$\cos\frac{\theta}{2}=\frac{m}{\sqrt{m^2+q^2}},\quad \sin\frac{\theta}{2}=\frac{q}{\sqrt{m^2+q^2}},\quad \hat{\boldsymbol{\theta}}=\hat{\mathbf{q}}. \tag{4.12}$$

The coordinates θ_i parametrize $SU(2)\cong S^3$ as a three-dimensional ball of radius 2π, where we can identify $U=1$ and $|\boldsymbol{\theta}|=0$ as the "north pole" of S^3, while $U=-1$ and $|\boldsymbol{\theta}|=2\pi$ corresponds to the "south pole." However, as the magnitude of the momentum q ranges from $q=0$ to

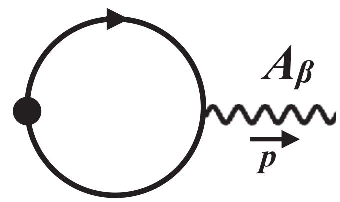

FIG. 2. Loop diagram for computing the generalized Hall current for the $2+1$-dimension Dirac fermion. The black dot is an insertion of the chiral current $\Gamma_\alpha\Gamma_\chi$ with incoming momentum p, and the propagators are given by K^{-1}.

[5]The term we keep will give a finite contribution to the index, while higher order operators fall off too fast at infinity to contribute to the integral $\int\partial_\mu\mathcal{J}_\mu$.

$q = \infty$, we see that $U(\mathbf{q})$ only visits half of S^3. In particular, for $m > 0$ we see that $q = 0$ corresponds to $|\boldsymbol{\theta}| = 0$, the north pole, while $q \to \infty$ corresponds to $|\boldsymbol{\theta}| = \pi$, the equator, so the mapping only covers the northern hemisphere of S^3. On the other hand, when $m < 0$ one can see that the mapping just covers the southern hemisphere. We can rewrite our expression [Eq. (4.10)] in terms of the θ_i variables as

$$\begin{aligned}\mathcal{J}_\alpha &= -\frac{1}{3}\epsilon_{\alpha\beta\gamma}\partial_\gamma A_\beta \epsilon_{ijk}\int \frac{d^3q}{(2\pi)^3} \\ &\quad\times \mathrm{Tr}\left[\left(U^\dagger \frac{\partial U}{\partial q_i}\right)\left(U^\dagger \frac{\partial U}{\partial q_j}\right)\left(U^\dagger \frac{\partial U}{\partial q_k}\right)\right] \\ &= -\frac{1}{\pi}\epsilon_{\alpha\beta\gamma}\partial_\gamma A_\beta\left(\frac{1}{24\pi^2}\epsilon_{ijk}\int_{V_{1/2}} d^3\theta \right.\\ &\quad\left.\times \mathrm{Tr}\left[\left(U^\dagger \frac{\partial U}{\partial \theta_i}\right)\left(U^\dagger \frac{\partial U}{\partial \theta_j}\right)\left(U^\dagger \frac{\partial U}{\partial \theta_k}\right)\right]\right) \qquad (4.13)\end{aligned}$$

where as discussed above, the volume of integration $V_{1/2}$ only covers half of the ball, corresponding to the northern hemisphere of S^3 for $m > 0$, or the southern hemisphere for $m < 0$. As such, the integral cannot be considered to be the winding number of a map from momentum space to S^3. The situation is remedied when a regulator is included. Here we consider a Pauli-Villars regulator, which corresponds to replacing $\mathcal{D} \to \mathcal{D}_{\mathrm{reg}} = \mathcal{D}/\mathcal{D}_{\mathrm{PV}} = \mathcal{D}(m)/\mathcal{D}(\Lambda)$ everywhere, where in $\mathcal{D}_{\mathrm{PV}}$ we have simply replaced m by the Pauli-Villars mass Λ, which we will take to $+\infty$ after the calculation. This has the effect of compactifying momentum space, since $\lim_{q\to\infty}\tilde{\mathcal{D}}_{\mathrm{reg}}(\mathbf{q}) = 1$, independent of the orientation of the momentum vector $\mathbf{q}$. The calculation proceeds as before, but now the $SU(2)$ matrix in Eq. (4.11) gets replaced by

$$\begin{aligned}U_{\mathrm{reg}}(\mathbf{q}) &= \frac{\tilde{\mathcal{D}}_{\mathrm{reg}}(\mathbf{q})}{\sqrt{\det \tilde{\mathcal{D}}_{\mathrm{reg}}(\mathbf{q})}} \\ &\equiv \cos\frac{\theta_{\mathrm{reg}}}{2} + i\hat{\boldsymbol{\theta}}_{\mathrm{reg}}\cdot\boldsymbol{\gamma}\sin\frac{\theta_{\mathrm{reg}}}{2} \qquad (4.14)\end{aligned}$$

where $\hat{\boldsymbol{\theta}}_{\mathrm{reg}} = \hat{\mathbf{q}}$ as before, and

$$\begin{aligned}\cos\frac{\theta_{\mathrm{reg}}}{2} &= \frac{\Lambda m + q^2}{\sqrt{(m^2+q^2)(\Lambda^2+q^2)}}, \\ \sin\frac{\theta_{\mathrm{reg}}}{2} &= \frac{q(\Lambda - m)}{\sqrt{(m^2+q^2)(\Lambda^2+q^2)}}. \qquad (4.15)\end{aligned}$$

We note that at $q \to \infty$ we have $U_{\mathrm{reg}}(\mathbf{q}) \to 1$, corresponding to compactifying momentum space to S^3 and mapping the point at $q = \infty$ to the north pole of the S^3 parametrized by θ. [This is in contrast to U in Eq. (4.11) which maps the 2-sphere at $q \to \infty$ onto the equator of S^3.] At $q = 0$, however, we find that $U_{\mathrm{reg}}(0) = \pm 1$ depending on the relative sign of m and Λ. When Λ and m have the same sign, $\cos\frac{\theta_{\mathrm{reg}}}{2} \geq 0$ for all values of q, with $\theta_{\mathrm{reg}} = 0$ for both $q = 0$ and $q = \infty$. In this case $U_{\mathrm{reg}}(q)$ describes a topologically trivial map from our compact momentum space to the northern hemisphere of S^3 $(0 \leq \theta_{\mathrm{reg}} \leq \pi/2)$. However, when m and Λ have opposite signs $U_{\mathrm{reg}}(q)$ is a nontrivial map from momentum space to S^3 with winding number equal to 1. It is no surprise then that we find that for the regulated theory

$$\begin{aligned}\mathcal{J}_\alpha &= -\frac{1}{\pi}\epsilon_{\alpha\beta\gamma}\partial_\gamma A_\beta(p)\left(\frac{1}{24\pi^2}\epsilon_{ijk}\int_V d^3\theta_{\mathrm{reg}}\right.\\ &\quad\left.\times \mathrm{Tr}\left[\left(U^\dagger_{\mathrm{reg}}\frac{\partial U_{\mathrm{reg}}}{\partial\theta_i}\right)\left(U^\dagger_{\mathrm{reg}}\frac{\partial U_{\mathrm{reg}}}{\partial\theta_j}\right)\left(U^\dagger_{\mathrm{reg}}\frac{\partial U_{\mathrm{reg}}}{\partial\theta_k}\right)\right]\right) \\ &= \frac{\nu_q}{\pi}\epsilon_{\alpha\beta\gamma}\partial_\beta A_\gamma, \qquad (4.16)\end{aligned}$$

where

$$\nu_q = \frac{1}{2}\left(\frac{\Lambda}{|\Lambda|} - \frac{m}{|m|}\right) \qquad (4.17)$$

is the winding number of the map from compact momentum space to S^3, where $\nu_q = 0$ when Λ and m have the same signs, and $\nu_q = \pm 1$ when Λ and m have the opposite signs. This relation between the topology in momentum space of the fermion dispersion relation and the quantization of the Hall current in $2+1$ dimensions has been remarked on previously in connection with the Ward-Takahashi identity in Refs. [23,24].

We arrive at the index of $\mathcal{D}$ as the surface integral at infinity in Euclidian 3-space

$$\mathrm{ind}(\mathcal{D}) = -\frac{1}{2}\int_S \mathcal{J}_\alpha dS_\alpha = -\frac{\nu_q}{2\pi}\int_S \epsilon_{\alpha\beta\gamma}\partial_\beta A_\gamma dS_\alpha, \qquad (4.18)$$

and it remains for us to show that what is multiplying ν_q is an integer winding number in coordinate space. For our diagnostic gauge field, we choose $A_{0,1}$ to be independent of y, while $A_2 = 0$, and for our integration region we take the volume to be a cylinder with its axis perpendicular to the domain wall, which we then take to infinite size in every direction. The surface integral only gets contributions from the end caps of the cylinder, ν_q being different at the two ends when we assume that Λ (with $\Lambda > 0$) and m have the same sign for $y > 0$ and the opposite signs for $y < 0$. The expression for the index then becomes

$$\begin{aligned}\mathrm{ind}(\mathcal{D}) &= -\left(\oint \vec{A}\cdot d\vec{\ell}\right)\nu_q(y)\Big|_{y=-\infty}^{y=\infty} = -\nu_A\nu_q(y)\Big|_{y=-\infty}^{y=\infty}\\ &= -\frac{\nu_A}{2}\left[\frac{\Lambda}{|\Lambda|}-\frac{m(y)}{|m(y)|}\right]_{y=-\infty}^{y=\infty} = \nu_A,\end{aligned} \tag{4.19}$$

where ν_A is the winding number of our Abelian gauge field integrated over the circle at infinity in the $\tau - x$ plane (where $\tau = x_0$ and $x = x_1$), evaluated at $y = x_2 = \pm\infty$. We see that a diagnostic gauge field with winding number $\nu_A = 1$ yields $\mathrm{ind}(\mathcal{D}) = 1$, revealing the existence of the gapless edge state in the corresponding Minkowski spacetime theory through a combined topological configuration in both coordinate and momentum space. Again, with more flavors we could get any integer value for the index, consistent with the fact that this model is in the D class (antisymmetric C and broken T symmetry) for which the topological invariant is known to be $\mathbb{Z}$ in two spatial dimensions.

B. A $d=2+1$ Majorana fermion with only $\mathbb{Z}_2$ fermion number symmetry

The model becomes more interesting when we explicitly break the fermion number symmetry down to $\mathbb{Z}_2$ by adding a Majorana mass (with the domain wall profile still in the Dirac mass). This system is analogous to a topological superconductor with the Majorana mass playing the role of the condensation of Cooper pairs. As it has no conserved fermion number, there is no conventional Hall current in this model, even though, as we shall see, for some parameters there exist gapless edge states. The Minkowski theory is

$$\mathcal{L}_{\mathrm{M}} = \bar{\psi}(i\partial\!\!\!/ - m)\psi + \frac{i\mu}{2}\psi^T C\psi + \frac{i\mu}{2}\bar{\psi}C\bar{\psi}^T, \tag{4.20}$$

where m is real, we can take μ to be real and positive, and C is the charge conjugation matrix satisfying

$$C^\dagger = C^{-1} = C, \qquad C\gamma^\mu C^{-1} = -(\gamma^\mu)^T. \tag{4.21}$$

Once again we will assume a domain wall profile $m(y)$, while keeping the Majorana mass μ constant. Because of the lack of continuous symmetry in this model, there are no conserved currents and hence no anomaly current inflow picture at nonzero μ. Nevertheless, we will show that one can still detect massless edge states in this model by computing the generalized Hall current, which does have inflow onto the defect when massless edge states exist, and we show that we can use this inflow to count such states.

Since Dirac notation is cumbersome when the fermion number is violated, our first step is to rewrite $\mathcal{L}_{\mathrm{M}}$ in terms of two real spinor fields $\chi_{1,2}$, where $\psi = (\chi_1 + i\chi_2)/\sqrt{2}$. The Lagrangian can then be expressed in terms of a 4-component spinor

$$\chi = \begin{pmatrix}\chi_1\\ \chi_2\end{pmatrix} \tag{4.22}$$

as

$$\mathcal{L}_{\mathrm{M}} = \frac{1}{2}\chi^T\left[\begin{pmatrix}1 & i\\ -i & 1\end{pmatrix}\otimes\gamma_M^0(i\partial\!\!\!/ - m) - \mu\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\otimes C\right]_{\mathcal{A}}\chi, \tag{4.23}$$

where the subscript "$\mathcal{A}$" means "antisymmetric part," derivatives being antisymmetric. We can now Wick rotate to Euclidean space and write the Euclidean Lagrangian as

$$\mathcal{L}_{\mathrm{E}} = \frac{1}{2}\chi^T\left[\begin{pmatrix}1 & i\\ -i & 1\end{pmatrix}\otimes\gamma_0(\partial\!\!\!/ + m) + \mu\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\otimes C\right]_{\mathcal{A}}\chi. \tag{4.24}$$

In the particular γ-matrix basis of Eq. (4.5)

$$\gamma_0 = C = \sigma_2, \qquad \gamma_1 = -\sigma_1, \qquad \gamma_2 = \sigma_3, \tag{4.25}$$

we can write $\mathcal{L}_{\mathrm{E}}$ in terms of the fermion fields

$$\zeta = \frac{1}{\sqrt{2}}\begin{pmatrix}\chi_1 + \chi_2\\ \chi_1 - \chi_2\end{pmatrix} \equiv \begin{pmatrix}\zeta_+\\ \zeta_-\end{pmatrix} \tag{4.26}$$

as

$$\begin{aligned}\mathcal{L}_E &= \frac{1}{2}[\zeta_+^T C\mathcal{D}_+\zeta_+ + \zeta_-^T C\mathcal{D}_-\zeta_-],\\ \mathcal{D}_\pm &= \partial\!\!\!/ + (m \pm \mu).\end{aligned} \tag{4.27}$$

As in the $1+1$ dimensional case we can drop the C matrix and use $[\mathrm{ind}(\mathcal{D}_+) + \mathrm{ind}(\mathcal{D}_-)]$ to detect edge states in this case, with $m(y)$ having a domain wall form and μ being constant. We can immediately adapt the result [Eq. (4.19)] in the previous section and evaluate this index as

$$\begin{aligned}\mathrm{ind}(\mathcal{D}) &\equiv [\mathrm{ind}(\mathcal{D}_+) + \mathrm{ind}(\mathcal{D}_-)]\\ &= -\left(\oint \vec{A}\cdot d\vec{\ell}\right)(\nu_q^+(y) + \nu_q^-(y))\Big|_{y=-\infty}^{y=\infty}\\ &= -\nu_A(\nu_q^+(y) + \nu_q^-(y))|_{y=-\infty}^{y=\infty},\end{aligned} \tag{4.28}$$

where

$$\nu_q^{(\pm)} = \frac{1}{2}\left(\frac{\Lambda}{|\Lambda|} - \frac{m\pm\mu}{|m\pm\mu|}\right). \tag{4.29}$$

Assuming positive Λ, we have

$$(\nu_q^{(+)} + \nu_q^{(-)}) = \begin{cases} 2 & m < -|\mu|, \\ 1 & -|\mu| < m < |\mu|, \\ 0 & |\mu| < m. \end{cases} \quad (4.30)$$

The analog of Eq. (4.19) then follows for the index in this theory,

$$\mathrm{ind}(\mathcal{D}) = -\nu_A[\nu_q^{(+)}(y) + \nu_q^{(-)}(y)]|_{y=-\infty}^{y=\infty}, \quad (4.31)$$

where ν_A is the winding number in the diagnostic gauge field. For $\nu_A = 1$ it follows that the index takes on one of the values $0, \pm 1, \pm 2$ depending on which the cases in Eq. (4.30) pertain for the asymptotic values of m at the two sides of the domain wall. If we denote the asymptotic values $m(\pm\infty) = m_\pm$, then our result for the index for various cases is given in Table I.

We now show that the above results agree with what one finds when constructing explicit domain wall solutions for this model. We return the Minkowski Lagrangian

$$\mathcal{L}_M = \bar{\psi}(i\partial\!\!\!/ - m)\psi + i\frac{\mu}{2}\psi^T C\psi + i\frac{\mu}{2}\bar{\psi}C\bar{\psi}^T, \text{sign of } \mu \quad (4.32)$$

and consider the equation of motion in our basis

$$\gamma^0 = \sigma_2, \qquad \gamma^1 = -i\sigma_1, \qquad \gamma^2 = i\sigma_3, \qquad C = \sigma_2. \quad (4.33)$$

To obtain the solutions to the equations of motion we redefine the fermion field $\psi = e^{i\frac{\pi}{4}}\varphi$ in which case we can write the solutions as

$$\varphi(x) = \begin{pmatrix} \alpha_1 e^{(-m-\mu)y} + i\beta_1 e^{(-m+\mu)y} \\ \alpha_2 e^{(m+\mu)y} + i\beta_2 e^{(m-\mu)y} \end{pmatrix}. \quad (4.34)$$

We take μ to be spatially constant, and without a loss of generality we assume $\mu > 0$, while the mass m depends on the coordinate y and takes the asymptotic values $m \to m_\pm$ as $y \to \pm\infty$. For $y > 0$, a localized solution requires that the coefficient of y in the exponent in Eq. (4.34) be negative when m is replaced by m_+. Out of the four exponentials in Eq. (4.34), there are always two that meet this criterion. Labeling them by their coefficients, they are

$$\begin{aligned} m_+ > |\mu|: &\qquad \alpha_1, \beta_1, \\ -|\mu| < m_+ < |\mu|: &\qquad \alpha_1, \beta_2, \\ m_+ < -|\mu|: &\qquad \alpha_2, \beta_2. \end{aligned} \quad (4.35)$$

We can do the same thing for solutions at $y < 0$ and $m \to m_-$; in this case the coefficient of y must be positive, and again there are always two solutions, the same as the above but with subscripts 1,2 reversed:

$$\begin{aligned} m_- > |\mu|: &\qquad \alpha_2, \beta_2, \\ -|\mu| < m_- < |\mu|: &\qquad \alpha_2, \beta_1, \\ m_- < -|\mu|: &\qquad \alpha_1, \beta_1. \end{aligned} \quad (4.36)$$

When we match solutions at $y = 0$ there must be localized solutions of the same chirality on both sides, and they must be both real or both imaginary. That means that the must be the same α_i or β_i solution for positive and negative y. It is evident then that depending on the two values $m_\pm$ relative to $|\mu|$ there can be 2, 1, or 0 solutions. For example, if $m_+ > |\mu|$ while $-|\mu| < m_- < |\mu|$, then there is one localized positive chirality β_1 solution that can be matched across $y = 0$. The number and chirality of edge state solutions are given in Table II, where the R, L entry tells us whether we have upper or lower component solutions respectively.

Note that the index in Table I is giving us the number of positive chirality massless edge states minus the number of negative chirality ones. The reason for that is that the equation $\mathcal{D}^\dagger\psi = 0$ has the same solutions as $\mathcal{D}\psi = 0$ except for a parity flip, exchanging $R \leftrightarrow L$. When the gauge field is turned on with $\nu_A = 1$, then it localizes the R solutions of $\mathcal{D}$ while delocalizing the L solutions. Therefore what the index is counting is the number of R solutions for $\mathcal{D}\psi = 0$, minus the number of R solutions for $\mathcal{D}^\dagger\psi = 0$, which is equivalent to the number of positive chirality massless edge states minus the number of negative chirality ones for the operator $\mathcal{D}$.

The index we computed still takes values in $\mathbb{Z}$, which is appropriate since the Majorana mass μ does not break T symmetry and the system remains in the D topological class.

TABLE I. The index $\mathrm{ind}(\mathcal{D})$ for spatial topology $\nu_A = 1$ as a function of $m_\pm$ and the asymptotic values of the fermion mass on the two sides of the domain wall, relative to the constant Majorana mass μ.

	$m_+ < -\lvert\mu\rvert$	$-\lvert\mu\rvert < m_+ < \lvert\mu\rvert$	$\lvert\mu\rvert < m_+$
$m_- < -\lvert\mu\rvert$	0	1	2
$-\lvert\mu\rvert < m_- < \lvert\mu\rvert$	−1	0	1
$\lvert\mu\rvert < m_-$	−2	−1	0

TABLE II. The number and chirality of edge state solutions to $\mathcal{D}\psi = 0$ before introducing a gauge field, where the R,L indicates chirality. The table for solutions to $\mathcal{D}^\dagger\psi = 0$ would be the same with substitution L $\leftrightarrow$ R.

	$m_+ < -\lvert\mu\rvert$	$-\lvert\mu\rvert < m_+ < \lvert\mu\rvert$	$\lvert\mu\rvert < m_+$
$m_- < -\lvert\mu\rvert$	0	R	2R
$-\lvert\mu\rvert < m_- < \lvert\mu\rvert$	L	0	R
$\lvert\mu\rvert < m_-$	2L	L	0

V. DIRAC FERMION IN 3+1 DIMENSIONS

Our last example of a Dirac fermion $3+1$ dimension shares many features with the $1+1$-dimension example. When the mass m has a domain wall profile, the Dirac equation in Minkowski spacetime has exact solutions corresponding to 2-component massless edge states localized on the $2+1$ dimensional wall. Such a domain wall describes the physics of a topological insulator, where the region with $m<0$ is considered the interior of the topological insulator and the region with $m>0$ is considered to be the exterior. There is no analog of chirality in $2+1$ dimensions, and hence no charge violation on the wall in the presence of background gauge fields and no inflowing current from the bulk to the wall maintaining current conservation. This makes the theory of three-dimensional topological insulator another interesting example to which one could imagine applying our construction for which to compute the generalized Hall current and index.

The Euclidean Dirac operator is simply $\not{\partial}+m(x_3)$ which has a static and unnormalizable solution localized at $x_3=0$ and constant in all the other coordinates. The edge state has two nonzero spinor components and is an eigenstate of γ_3. To fully localize this state we add as diagnostic fields a four-dimensional gauge field and a pseudoscalar, and again we consider the mass to be an arbitrary scalar field for now. Thus the operator we consider is

$$\mathcal{D}=D_\mu\gamma_\mu+\phi_1+i\phi_2\gamma_\chi, \tag{5.1}$$

where $D_\mu=(\partial_\mu+iA_\mu)$ is the $d=4$ gauge covariant derivative and γ_μ, γ_χ are our 4D Dirac matrices.

To compute the index of $\mathcal{D}$ we once again construct the K operator,

$$K=\begin{pmatrix}0 & -\mathcal{D}^\dagger\\ \mathcal{D} & 0\end{pmatrix}=D_\mu\Gamma_\mu+i\phi_2\Gamma_4+i\phi_1\Gamma_5, \tag{5.2}$$

where $\mu=0,\ldots 3$ and the Γ_a are the eight-dimensional matrices

$$\begin{aligned}\Gamma_\mu&=\sigma_1\otimes\gamma_\mu, \qquad \mu=0,\ldots,3,\\ \Gamma_4&=\sigma_1\otimes\gamma_\chi,\\ \Gamma_5&=-\sigma_2\otimes 1,\\ \Gamma_\chi&=\sigma_3\otimes 1, \qquad \text{index starts at } 0.\end{aligned} \tag{5.3}$$

Our task is to compute the part of the chiral current $\mathcal{J}_a=\bar{\Psi}\Gamma_a\Gamma_\chi\Psi$ that contributes to the index, where Ψ is a fermion with action $S=\bar{\Psi}K\Psi$. As in the $1+1$ dimensional example in Eq. (3.10), we first write the scalars as

$$\phi=\phi_1+i\phi_2=(v+\rho(x))e^{i\theta(x)}, \tag{5.4}$$

assuming constant v and slowly varying ρ and θ where $\rho=\theta=0$. We compute the leading contribution to the chiral current in a $1/v$ expansion, since higher order terms will drop off too fast at infinity to contribute to the integral $\int\partial_\mu\mathcal{J}_\mu$. To this end we write

$$K=K_0+\delta K, \tag{5.5}$$

with

$$\begin{aligned}K_0&=\partial_\mu\Gamma_\mu+iv\Gamma_5,\\ \delta K&=iA_\mu\Gamma_\mu+i\theta v\Gamma_4+i\rho(x)\Gamma_5.\end{aligned} \tag{5.6}$$

Then K_0^{-1} will be the free fermion propagator, while we perturb in δK. To compute the part of the current that contributes to the index we need the leading term in a $1/v$ expansion, since higher order terms will drop off too fast at infinity to contribute to the integral $\int\partial_\mu\mathcal{J}_\mu$.

When expanding $\mathcal{J}_\mu$ in δK, the Γ_χ insertion in the fermion loop requires that the rest of the graph supplies one each of the other six Γ_a matrices in order to get a nonzero contribution from the trace. First consider the source of $\Gamma_{4,5}$ in the graph. We see the Γ_5 can come from one of the fermion propagators K_0^{-1}, but to obtain Γ_4 we require an insertion of θ in the graph, while to lowest order the ρ contribution will vanish. To obtain the other four Γ_μ we note that the result for $\mathcal{J}_\mu$ will be proportional to an epsilon tensor $\epsilon_{\mu\alpha\beta\gamma}$ and that we can only contract the α,β,γ indices with a gauge field and two derivatives—one acting on the gauge field, the other on θ. Thus we must expand the graphs in Fig. 3 to linear order in the momenta carried by the gauge field and by θ.

It is straightforward to evaluate this loop integral, with the result

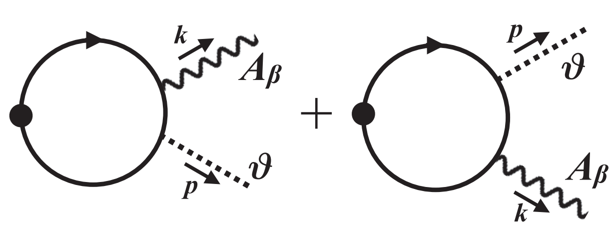

FIG. 3. Loop diagrams for computing the generalized Hall current for the $3+1$-dimension Dirac fermion. The black dot is an insertion of the chiral current $\Gamma_\alpha\Gamma_\chi$ with incoming momentum $p+k$. The outgoing fields are the gauge field and the phase θ of the complex field $\phi_1+i\phi_2$, and the fermion propagator is given by K_0^{-1}.

$$\begin{aligned}\mathcal{J}_\mu &= v\partial_\gamma\theta\partial_\alpha A_\beta \int \frac{d^4q}{(2\pi)^4}\mathrm{Tr}(\Gamma_\mu\Gamma_\chi[(\partial_\alpha\tilde{K}_0^{-1})\Gamma_\beta\tilde{K}_0^{-1}\Gamma_4(\partial_\gamma\tilde{K}_0^{-1}) \\ &\quad + (\partial_\gamma\tilde{K}_0^{-1})\Gamma_4\tilde{K}_0^{-1}\Gamma_\beta(\partial_\alpha\tilde{K}_0^{-1})]) \\ &= -\epsilon_{\mu\alpha\beta\gamma}\partial_\gamma\theta\partial_\alpha A_\beta\int\frac{d^4q}{(2\pi)^4}\frac{16v^2}{(q^2+v^2)^3} \\ &= -\frac{1}{2\pi^2}\epsilon_{\mu\alpha\beta\gamma}\partial_\gamma\theta\partial_\alpha A_\beta, \end{aligned} \tag{5.7}$$

where the derivatives inside the integral are with respect to q. The prefactor in the first line includes a (-1) for the fermion loop, (-1) from the Fourier transform $p^2 \to -\partial_x^2$, (-1) from the two derivatives with respect to momentum with opposite signs due to momentum flow, and factors $(-iv)$ and $-i\Gamma_\beta$ for θ and A_β vertices respectively.

To highlight the momentum space topology underlying the above calculation, we follow similar steps as in the $1+1$ dimensional example of Sec. III to rewrite the result as the analog of Eq. (3.19):

$$\begin{aligned}\mathcal{J}_\mu &= \frac{1}{12}\epsilon_{\mu\alpha\beta\gamma}\partial_\gamma\theta\partial_\alpha A_\beta \\ &\quad\times\epsilon_{\rho\sigma\tau\omega}\int\frac{d^4q}{(2\pi)^4}\mathrm{Tr}\gamma_\chi[(\xi^\dagger\partial_\rho\xi)(\xi^\dagger\partial_\sigma\xi)(\xi^\dagger\partial_\tau\xi)(\xi^\dagger\partial_\omega\xi) \\ &\quad + (\xi\partial_\rho\xi^\dagger)(\xi\partial_\sigma\xi^\dagger)(\xi\partial_\tau\xi^\dagger)(\xi\partial_\omega\xi^\dagger)],\end{aligned} \tag{5.8}$$

where

$$\xi = \frac{\tilde{\mathcal{D}}_0}{(\det\tilde{\mathcal{D}}_0)^{\frac{1}{4}}} = \frac{v}{\sqrt{q^2+v^2}} + i\frac{q_\mu\gamma_\mu}{\sqrt{q^2+v^2}}. \tag{5.9}$$

Making use of the currents A_μ and V_μ defined as

$$\begin{aligned}A_\mu &= \frac{i}{2}(\xi^\dagger\partial_\mu\xi - \xi\partial_\mu\xi^\dagger), \\ V_\mu &= \frac{1}{2}(\xi^\dagger\partial_\mu\xi + \xi\partial_\mu\xi^\dagger),\end{aligned} \tag{5.10}$$

the integrand in Eq. (5.8) can be related to the volume form derived for four spacetime dimensions in Eq. (A11) using the relation that one can derive

$$\begin{aligned}&-\frac{\epsilon_{\rho\sigma\tau\omega}}{96\pi^2}[(\xi^\dagger\partial_\rho\xi)(\xi^\dagger\partial_\sigma\xi)(\xi^\dagger\partial_\tau\xi)(\xi^\dagger\partial_\omega\xi) \\ &\quad + (\xi\partial_\rho\xi^\dagger)(\xi\partial_\sigma\xi^\dagger)(\xi\partial_\tau\xi^\dagger)(\xi\partial_\omega\xi^\dagger)] \\ &= \epsilon_{\rho\sigma\tau\omega}\left[-\frac{1}{16\pi^2}A_\rho A_\sigma A_\tau A_\omega + \partial_\sigma\left(\frac{1}{24\pi^2}A_\sigma A_\tau V_\omega\right)\right].\end{aligned} \tag{5.11}$$

On taking the trace of the above equation multiplied by γ_χ, the first term on the right is recognized from Eq. (A11) as the volume measure for S^4, while the second term is the divergence of a current which is well defined without singularities on S^4, and hence integrates to zero. Therefore the generalized Hall current in Eq. (5.8) can be written as

$$\mathcal{J}_\mu = -\frac{\nu_q}{2\pi^2}\epsilon_{\mu\alpha\beta\gamma}\partial_\gamma\theta\partial_\alpha A_\beta, \tag{5.12}$$

where ν_q is the element of $\pi_4(S^4)$:

$$\nu_q = -\frac{1}{16\pi^2}\epsilon_{abcd}\int d^4q\mathrm{Tr}\gamma_\chi A_a A_b A_c A_d = 1. \tag{5.13}$$

The index is given by

$$\begin{aligned}\mathrm{ind}(\mathcal{D}) &= -\frac{1}{2}\int d^4x\partial_\mu\mathcal{J}_\mu \\ &= \frac{\nu_q}{4\pi^2}\int_S \epsilon_{\mu\alpha\beta\gamma}\partial_\gamma\theta\partial_\alpha A_\beta dS_\mu.\end{aligned} \tag{5.14}$$

To evaluate the spatial integral we first must choose a topologically nontrivial set of diagnostic fields to localize the solutions to $\mathcal{D}\psi = 0$. That can be done by making use of the Bogomolny-Prasad-Sommerfeld monopole field configuration discussed in Ref. [25] which considers a massless Dirac fermion in three Euclidean dimensions interacting with a scalar field and a gauge field,

$$\Phi = \frac{g}{2\pi}\left(a - \frac{1}{2r}\right), \qquad \mathbf{A} = -\frac{g(1+\cos\theta)}{4\pi r\sin\theta}\hat{\mathbf{e}}_\varphi, \tag{5.15}$$

and discuss how when the couplings obey the minimal Dirac quantization relation, $eg = 2\pi$, the $d = 3$ conjugate Dirac operator $\mathcal{D}_3^\dagger$ has a zeromode proportional to $\exp(-ar)$ while $\mathcal{D}_3$ has none. This configuration can be lifted into our four-dimensional Euclidean theory by taking $\mathbf{A}$ to be the first three components of our four-component gauge field, independent of the fourth coordinate, while

$$\phi = m(x_4) + i\Phi(\mathbf{x}), \quad A_4 = 0, \quad F_{ij} = \epsilon_{ijk4}\frac{g\hat{r}_k}{4\pi r^2}, \tag{5.16}$$

where for notational convenience we have relabeled our coordinates so that the domain wall mass is a function of x_4 and coordinates on the mass defect are labeled by $x_{1,2,3}$. Since we have set the electric charge $e = 1$ in our covariant derivative, we take $g = 2\pi$ for the magnetic charge. With these background fields we then expect that the existence of a massless edge state in the original $d = 3+1$ Minkowski theory implies one zeromode for $\mathcal{D}^\dagger$ only if $a > 0$ and no zeromode for $\mathcal{D}$, so that $\mathrm{ind}(\mathcal{D}) = -\theta(a)$.

This is indeed what we find. Plugging in these fields into Eq. (5.12), we get the generalized Hall current

$$\mathcal{J}_i = \frac{\hat{\mathbf{r}}_i}{2\pi^2 r} \frac{(2ar-1)\frac{\mathrm{d}m(x_3)}{\mathrm{d}x_3}}{(2ar-1)^2 + 4r^2 m(x_3)^2},$$
$$\mathcal{J}_3 = \frac{m(x_3)}{2\pi^2 r^2((2ar-1)^2 + 4r^2 m(x_3)^2)}, \tag{5.17}$$

for $i = 0, 1, 2$. We then compute the integral of its divergence,

$$\int d^4x \partial_\mu J_\mu = \lim_{L_3, R\to\infty} \left[\int^R d^3r J_4 \Big|_{x_4=-L_3}^{x_3=L_3} + \int_{-L_3}^{L_3} dx_3 \int d\Omega R^2 \hat{R}\cdot\vec{J} \right]. \tag{5.18}$$

Specializing to the mass profile $m(\pm\infty) = \pm m_0$, and using the expression for current given in Eq. (5.17) we find

$$\lim_{L_3, R\to\infty} \int_{-L_4}^{L_4} dx_4 \int d\Omega R^2 \hat{r}\cdot\vec{J} = \frac{2\tan^{-1}(\frac{m_0}{a})}{\pi},$$
$$\lim_{L_3, R\to\infty} \int^R d^3r J_3 \Big|_{x_3=-L_3}^{x_3=L_3} = \left(1 + \frac{2\tan^{-1}\frac{a}{m_0}}{\pi}\right). \tag{5.19}$$

Summing these two terms gives us

$$\int d^4x \partial_\mu \mathcal{J}_\mu = 2\theta(a) \tag{5.20}$$

and yields the anticipated result for the index,

$$\mathrm{ind}(\mathcal{D}) = -\frac{1}{2}\int d^4x \partial_\mu \mathcal{J}_\mu = -\theta(a). \tag{5.21}$$

So we see again that with diagnostic fields to localize massless edge states, the index of the Euclidean fermion operator counts these edge states and is quantized because of both the spacetime topology of the background fields, and the momentum space topology of the fermion dispersion relation. As in the $d = 1+1$ case and unlike the $d = 2+1$ example, we find that the $d = 3+1$ result is not affected by regulator fields, which decouple.

This model is in the DIII class, symmetric under both C and T with both C and T matrices antisymmetric [5]. The topological invariant in this case for three spatial dimensions is $\mathbb{Z}$, consistent with what we would find if we generalized this theory to more flavors without any flavor symmetry.

VI. INTERACTIONS

Our interest in understanding edge states via the index of the Euclidean fermion operator led us to computing the incoming flux at infinity of the generalized Hall current. This has all been for free fermions, but interacting systems are more interesting [26–34], and we know that interactions can change the topological classification [13]. The ramifications go beyond condensed matter systems and have been applied to lattice models for chiral gauge theories, where vectorlike fermions appear as chiral edge states until interactions are turned on, gapping some of them and leaving behind a chiral representation [35–38]. For an analogous discussion in continuum chiral gauge theories, see [39,40]. Clearly the index of the free fermion operator cannot capture this physics. The generalized Hall current, on the other hand, is well defined even in the presence of interactions, and so it is reasonable to ask whether its divergence still tells us about massless edge states in an interacting theory. In this section we speculate that that is plausible, in the context of the same $1+1$ model described in Ref. [13].

First we examine more closely how the calculations for free fermions were done. Our method followed the work of Callan and Harvey [6], which in turn used the methods developed by Goldstone and Wilczek [41]. We computed the generalized Hall current in a derivative expansion, integrating out the fermions in a background field. In the $d = 1+1$ example, this background field was a complex scalar field, and we obtained a contribution to the current proportional to $\partial_\mu\theta = i\phi^*\overleftrightarrow{\partial}_\mu\phi/|\phi|^2$. Recall that the index was defined in Eq. (2.10) as the integral of the divergence of the generalized Hall current in the limit that the doubled fermion's mass M tended to zero. As seen in Eq. (2.1), the mass M was introduced as an infrared cutoff, and in the Ward-Takahashi identity for the current in Eq. (2.8),

$$\partial^\mu \mathcal{J}_\mu^\chi = 2M\bar{\Psi}\Gamma_\chi\Psi - \mathcal{A}, \tag{6.1}$$

where $\mathcal{A} = 0$. The nonzero divergence indicating the existence of the massless edge state comes from the $2M\bar{\Psi}\Gamma_\chi\Psi$ term on the right-hand side, in the limit that $M \to 0$, where M serves as an IR regulator. In our calculation, the inverse dependence on $|\phi|^2$ is the sign of an infrared divergence regulated by $\langle\phi\rangle$—it was because the background field served as an IR regulator that we could set the parameter M in Eq. (2.10) to zero before computing the Feynman diagrams. We replaced M by a spatially varying the ϕ field as the IR regulator so that a nonzero index of the fermion operator $\mathcal{D}$ would indicate there was a massless edge state when ϕ was removed. The calculation was performed as if the fermions were fully gapped by ϕ. This is clearly false, since we were studying systems with an exact zeromode. The current we computed cannot be valid in the region where the zeromode wave function is appreciable since the fermion is not gapped there and the derivative expansion in the background field breaks down. However, since the index is proportional to $\int \partial_\mu \mathcal{J}_\mu$, it only depends on the current asymptotically far away from the localized zeromode, For that reason we were able to treat the fermion as gapped, and justify the derivative expansion. This is the same justification as for the Callan-Harvey calculation [6].

We will argue that in theories where interactions fully gap the fermions, this spatially constant gap will serve as the dominant infrared regulator in calculating the generalized Hall current, and since it cannot localize the fermions, it will lead to a vanishing divergence of the current. We envision a mechanism similar to what we saw in Sec. III where we discussed the example of $2N$ flavors of free, $d = 1 + 1$ Majorana fermions in the presence of time-reversal violation. There we found that if we explicitly broke time-reversal invariance via a spatially constant $i\mu\psi^T C\gamma_\chi\psi$ term, the index vanished. The reason was that in this case, as we removed the diagnostic field $\phi_2(x) \to 0$, the spatial topology experienced by the fermion abruptly became trivial at the point where μ dominated over ϕ_2 as the infrared regulator. The question then is whether the same phenomenon can occur when interactions gap the fermions as we remove the diagnostic fields. This seems plausible, making the procedure described here for detecting massless edge states relevant when interactions are introduced, even though the original motivation of looking at zeromodes of the free fermion operator is no longer applicable. To understand this better, we look at the model of Ref. [13] in greater detail.

Consider $2N$ copies of the $1 + 1$-dimension Majorana fermion model discussed in Sec. III. As shown in Eq. (3.6) the Minkowski domain wall solutions in the free theory with a step function mass take the form

$$\psi_i(x,t) = \eta_i(t)e^{-m|x|}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \tag{6.2}$$

and when quantized, the Hermitian $\hat{\eta}_i$ operators obey the Clifford algebra

$$\{\hat{\eta}_i, \hat{\eta}_j\} = 2\delta_{ij}, \quad i = 1, \ldots, 2N. \tag{6.3}$$

From these we can construct the ladder operators

$$\hat{c}_a = \frac{\hat{\eta}_a + i\hat{\eta}_{a+N}}{2}, \quad \hat{c}_a^\dagger = \frac{\hat{\eta}_a - i\hat{\eta}_{a+N}}{2}, \quad a = 1, \ldots N, \tag{6.4}$$

which obey the usual fermion anticommutation relations

$$\{\hat{c}_a, \hat{c}_b\} = \{\hat{c}_a^\dagger, \hat{c}_b^\dagger\} = 0, \qquad \{\hat{c}_a, \hat{c}_b^\dagger\} = \delta_{ab}. \tag{6.5}$$

The 2^N-fold degenerate edge states can then be constructed by acting with $c_a^\dagger$ operators on a state $|0\rangle$, which is defined to be the state annihilated by all of the $\hat{c}_a$ operators.

Since the $\hat{\eta}_i$ operators obey a Clifford algebra, they define an $\mathfrak{so}(2N)$ Lie algebra, under which the degenerate ground states transform as the 2^N-dimensional reducible spinor representation. When considering interactions between the edge states, it is convenient to represent the $\hat{\eta}_i$ operators as $2^N \times 2^N$ Hermitian Dirac gamma matrices which act on these states. Interactions between these states can then be represented as a matrix consisting of sums of totally antisymmetrized products of even numbers of gamma matrices, which we will call $H_{\rm int}$.

One constraint we will impose on the interactions is that they preserve time-reversal symmetry, since the gapless edge states in the free theory owe their existence to that symmetry in the first place, as discussed in Sec. III B. The action of time reversal on the bulk states is $\psi \to \sigma_1\psi$ which takes $\hat{\eta}_i \to -\hat{\eta}_i$. This sign is not interesting since we will only be considering products of an even number of the $\hat{\eta}_i$ operators; however, in order for an interaction represented as an antisymmetrized product of $2k$ antisymmetrized gamma matrices to be Hermitian it must be proportional to i^k, which means that operators with an odd k flip sign under the antiunitary time-reversal transformation. So we restrict the interaction to operators involving products of multiples of *four* fermion fields.

One of the results of Ref. [13] is that this time-reversal invariant $H_{\rm int}$ can gap all of the edge states in this model if and only if the number of flavors is a multiple of eight. Completely gapping the edge states means that there is a unique, nondegenerate ground state—so to prove this result we need to show that a sum of totally antisymmetrized products of $4k$ $SO(2N)$ gamma matrices can only have a unique lowest eigenvalue when $2N = 0 \bmod 8$. This is easy to show, and the argument is given in Appendix B.

Consider the case with eight flavors of fermions and consider the interaction defined in Eqs. (B10) and (B11). On the domain wall,

$$H_{\rm int} = \omega(\hat{c}_1\hat{c}_2\hat{c}_3\hat{c}_4 + \text{H.c.} + 1), \tag{6.6}$$

which for $\omega > 0$ has the unique ground state $|\Omega\rangle = (|0000\rangle - |1111\rangle)/\sqrt{2}$ with energy $E = 0$, fourteen degenerate states with $E = \omega$, and an isolated state with energy $E = 2\omega$. In the effective Euclidean $0 + 1$-dimension theory one can compute the 1-particle propagator and find

$$\langle\Omega|\eta_i e^{-\hat{H}_{\rm int}\tau}\eta_j|\Omega\rangle = e^{-\omega\tau}\delta_{ij}. \tag{6.7}$$

How do these interactions affect the actual calculation of the divergence of the generalized Hall current in the doubled Euclidean version of the $1 + 1$-dimension Minkowski theory? We do not have a quantitative answer, but believe that the gapping of the single particle propagator in the fully interacting theory will render the generalized Hall current divergenceless.[6]

In the $1 + 1$-dimension bulk theory we expect the interactions act like the 't Hooft operator induced by

[6]For work on the Greens function for theories with gapped edge states, see Refs. [42–45].

instantons in Ref. [46], saturating the zeromodes and serving as topologically trivial infrared cutoff. In the absence of interactions and diagnostic fields, the fermion propagator we use to compute the generalized Hall current in the $N_f = 8$ model defined as

$$\frac{\int\prod_{i=1}^{8} d\Psi_i e^{-S}\Psi_j(y_1)\bar{\Psi}_k(y_2)}{\int\prod_{i=1}^{8} d\Psi_i e^{-S}}, \qquad S = \int d^2x\bar{\Psi}\mathcal{D}\Psi \tag{6.8}$$

is divergent since the denominator vanishes due to the integration over zeromodes of $\mathcal{D}$ which do not appear in the action.

Suppose we now add a fully gapping interaction, such as the one given in (Eq. (B11),

$$\mathcal{L}_{\text{int}} = \omega[(\psi_1 + i\psi_5)^T C\gamma_\chi(\psi_2 + i\psi_6)] \times [(\psi_3 + i\psi_7)^T C\gamma_\chi(\psi_4 + i\psi_8)] + \text{H.c.} \tag{6.9}$$

The natural extension to add in the doubled Euclidian theory is then

$$\mathcal{L}_{\text{int}} = \omega[(\bar{\Psi}_1 + i\bar{\Psi}_5)\Gamma_3(\Psi_2 + i\Psi_6)][(\bar{\Psi}_3 + i\bar{\Psi}_7)\Gamma_3(\Psi_4 + i\Psi_8)] + \text{H.c.}, \tag{6.10}$$

where $\Gamma_3 = \sigma_1 \times \gamma_\chi$ is the doubled version of γ_χ given in Eq. (3.9). When this term is added to the action in Eq. (6.8) the fermion propagator is no longer IR divergent since the zeromodes of $\mathcal{D}$ now appear in the action. As a result, one should still find a well defined index as the diagnostic fields are removed—but as the infrared cutoff arising from the interaction presumably has trivial spatial topology, we expect a topological phase transition to a trivial phase with vanishing divergence for the generalized Hall current, similar to that seen in Sec. III B. It would be interesting to develop a quantitative method to perform the calculation in the presence of interactions, but this is beyond the scope of this paper.

VII. DISCUSSION

We have shown that the presence of the massless edge states in topological matter manifests itself by the inflow of a current—which we call a generalized Hall current—in a related system in Euclidian spacetime. The divergence of this current indicates the existence of massless edge states just as the Hall current inflow does for the integer quantum Hall effect. But this current appears in all topological classes in Minkowski spacetime, including those that do not have conserved currents because of a lack of continuous symmetries (such as a topological superconductor), or whose edges do not suffer from chiral anomalies because they lack chiral symmetry (such as topological insulators in $1+1$ and $3+1$ dimensions). In this sense one arrives at a unified picture for disparate manifestations of topological matter.

Furthermore, while the original motivation was to study the index of the free Euclidean Dirac operator, the generalized Hall currents can be computed for interacting systems as well, and we gave qualitative arguments for why we expect the utility of such currents to persist. Whether this idea can be put on a firmer foundation is an open question. It is an attractive proposition to be able to compute analytically how interactions affect topological properties in different systems in various dimensions. If a general theory for gapping massless chiral edge states in $3+1$ dimensions can be derived, that might shed light on what restrictions there are on the matter content of chiral gauge theories regulated on a lattice; this is of obvious interest given that the Standard Model is a chiral gauge theory.

ACKNOWLEDGMENTS

We thank Lukasz Fidkowski and Michael Clancy for useful conversations. D. B. K. is supported in part by DOE Grant No. DE-FG02-00ER41132 and by the DOE QuantISED program through the theory consortium "Intersections of QIS and Theoretical Particle Physics" at Fermilab. S. S. acknowledges support from the U.S. Department of Energy, Nuclear Physics Quantum Horizons program through the Early Career Award No. DE-SC0021892.

APPENDIX A: TOPOLOGY IN MOMENTUM SPACE FOR DIRAC OPERATORS

We have seen repeatedly in this paper that the Feynman diagram for the generalized Hall current can be expressed as a trace involving a unitary matrix ξ and its derivatives, contracted with an epsilon tensor, where ξ is the unitarized Euclidiean fermion operator in $d+1$ dimensions, $\xi = \tilde{D}/(\det\tilde{D})^{2^{-k}}$ where $d+1 = 2k$ for even dimensions and $d = 2k$ for odd dimensions. ξ in general takes the form

$$\xi = a(p) + i\not{b}(p), \qquad a(p)^2 + b_\mu(p)b_\mu(p) = 1. \tag{A1}$$

For an ordinary Dirac fermion, we have

$$a = \frac{m}{\sqrt{m^2+p^2}}, \qquad b_\mu = \frac{p_\mu}{\sqrt{m^2+p^2}}. \tag{A2}$$

We see that the $d+2$-component unit vector $\{a, b_\mu\}$ represents a map from $(d+1)$-dimensional momentum space to the sphere S^{d+1}. Note, however, that ξ maps all of momentum space onto half of the sphere, as $p \to \infty$. At $p = 0$ we have $\xi = 1$, while for infinite momentum, $\xi \to i\hat{p}_\mu\gamma_\mu$ which can be thought of as the equator of the sphere, where the poles are represented by the matrices ± 1. A map onto all of S^{d+1} is described by $U = \xi^2$, where for infinite momentum $U \to -1$. Evidently the topology for a Dirac fermion in $(d+1)$-dimensional momentum space is described by the $O(d+1)$ nonlinear sigma model in $(d+1)$ dimensions.

A convenient way to describe this nonlinear sigma model is as an $SO(d+2)/SO(d+1)$ sigma model, which is what we do here. By computing $\sqrt{g}$, where $g_{\mu\nu}$ is the metric of the sigma model and g is its determinant, we can simply express the volume form in terms of the matrix ξ, in a form easy to relate to the Feynman diagram calculations.

We define

$$\xi = e^{i\theta_\mu\gamma_\mu/4} = \cos\frac{\theta}{4} + i\hat{\theta\!\!/}\sin\frac{\theta}{4},$$
$$U = \xi^2 = e^{i\theta_\mu\gamma_\mu/2} = \cos\frac{\theta}{2} + i\hat{\theta\!\!/}\sin\frac{\theta}{2},$$
$$A_a = \frac{i}{2}(\xi^\dagger\partial_a\xi - \xi\partial_a\xi^\dagger) = \mathcal{A}_{a\mu}\gamma_\mu. \tag{A3}$$

For even $d+1$, it is easy to show that U parametrizes an $SO(d+2)/SO(d+1)$ sigma model. We can write U as

$$U = e^{i\theta_\mu\sigma'_{\mu,d+2}}, \qquad \mu = 1,\ldots,d+1, \tag{A4}$$

where

$$\sigma'_{\mu\nu} = \frac{i}{4}[\gamma'_\mu,\gamma'_\nu], \quad \gamma'_\mu = i\gamma_{d+2}\gamma_\mu, \quad \gamma'_{d+2} = \gamma_{d+2}, \tag{A5}$$

since $\sigma'_{\mu,d+2} = \gamma_\mu/2$. In this form we see that $U(x)$ would describe the Goldstone bosons of the spontaneous symmetry breaking pattern $SO(d+2) \to SO(d+1)$, where the $\sigma'_{\mu\nu}$ generate $SO(d+2)$, and the subset $\sigma'_{\mu,d+2}$ are the "broken generators" which are not also generators of the $SO(d+1)$ subalgebra. A similar construction can be made for odd $d+1$.

The metric for this sigma model is given by

$$g_{\mu\nu} \propto \mathrm{Tr}\partial_\mu U^\dagger\partial_\nu U \propto \mathrm{Tr}A_\mu A_\nu = \mathcal{A}_{\mu\alpha}\mathcal{A}_{\nu\beta}\mathrm{Tr}\gamma_\alpha\gamma_\beta \propto (\mathcal{A}\mathcal{A}^T)_{\mu\nu}, \tag{A6}$$

where the proportionalities are all constant. Thus we have

$$\sqrt{g} = N\det\mathcal{A}. \tag{A7}$$

With the convention for both $SO(2k)$ and $SO(2k+1)$

$$\mathrm{Tr}[\gamma_{2k+1}\gamma_{\mu_1}\ldots\gamma_{\mu_{2k}}] = (2i)^k\epsilon_{\mu_1\ldots\mu_{2k}}, \tag{A8}$$

we can rewrite this as

$$\sqrt{g} = N\epsilon_{\mu_1\ldots\mu_{d+1}} \times \begin{cases} \mathrm{Tr}\gamma_{d+2}A_{\mu_1}\ldots A_{\mu_{d+1}} & \text{even } d+1 \\ \mathrm{Tr}A_{\mu_1}\ldots A_{\mu_{d+1}} & \text{odd } d+1 \end{cases}. \tag{A9}$$

With the normalization condition

$$\int d^{d+1}\theta\sqrt{g} = 1, \tag{A10}$$

we have

$$\begin{aligned}
d+1=2\colon\ \sqrt{g} &= -\frac{i}{4\pi}\epsilon_{ij}\mathrm{Tr}\gamma_3A_iA_j \\
&= \frac{i}{16\pi}\epsilon_{ij}\mathrm{Tr}[\gamma_\chi(\partial_iU)(U^\dagger\partial_jU) \\
&= \frac{1}{8\pi}\left(\frac{\sin\theta/2}{\theta}\right), \\
d+1=3\colon\ \sqrt{g} &= \frac{i}{3\pi^2}\epsilon_{ijk}\mathrm{Tr}A_iA_jA_k \\
&= \frac{1}{24\pi^2}\epsilon_{ijk}\mathrm{Tr}(U^\dagger\partial_iU)(U^\dagger\partial_jU)(U^\dagger\partial_kU) \\
&= \frac{1}{4\pi^2}\left(\frac{\sin\theta/2}{\theta}\right)^2, \\
d+1=4\colon\ \sqrt{g} &= -\frac{1}{16\pi^2}\epsilon_{ijk\ell}\mathrm{Tr}\gamma_5A_iA_jA_kA_\ell \\
&= -\frac{1}{256\pi^2}\epsilon_{ijk\ell}\mathrm{Tr}[\gamma_5(\partial_iU)(U^\dagger\partial_jU) \\
&\quad\times(U^\dagger\partial_kU)(U^\dagger\partial_\ell U) \\
&= \frac{3}{16\pi^2}\left(\frac{\sin\theta/2}{\theta}\right)^3, \\
d+1=5\colon\ \sqrt{g} &= \frac{1}{15\pi^3}\epsilon_{ijk\ell m}\mathrm{Tr}A_iA_jA_kA_\ell A_m \\
&= \frac{i}{480\pi^3}\epsilon_{ijk\ell m}\mathrm{Tr}[(U^\dagger\partial_iU)(U^\dagger\partial_jU) \\
&\quad\times(U^\dagger\partial_kU)(U^\dagger\partial_\ell U)(U^\dagger\partial_mU) \\
&= i\frac{1}{2\pi^3}\left(\frac{\sin\theta/2}{\theta}\right)^4.
\end{aligned} \tag{A11}$$

The integral in Eq. (A10) has an interpretation other than as the normalized volume integral: with the definitions in Eqs. (A1) and (A3) it is the winding number of a map $U(p)$ from momentum space to $SO(d+2)/SO(d+1) \cong S^{d+1}$. Since $U(p) \to -1$ as $|\mathbf{p}| \to \infty$ in any direction, momentum space is effectively compactified to S^{d+1}, and so the winding number is an element of the homotopy group $\pi_n(S^n)$ with $n = d+1$.

APPENDIX B: GAPPING THE MAJORANA EDGE STATES WITH INTERACTIONS IN 1+1 DIMENSIONS

Here we give a simple argument for a result of [13], that one can fully gap the edge states in the $1+1$ dimensional Majorana model when the number of flavors is

$2N = 0 \bmod 8$, and not for other numbers of flavors. For a pedagogical review, see [47].

We have seen in Sec. VI that for $2N$ flavors, the most general time-reversal invariant interactions can be written as sums of totally antisymmetrized products of $4k$ $SO(2N)$ gamma matrices acting on the 2^N-dimensional Hilbert space of degenerate ground states. Call this interaction matrix $H_{\rm int}$, and consider the effect on $H_{\rm int}$ of two special matrices: γ_{2N+1}, which anticommutes with all the other gamma matrices, and $\tilde{C}$ which has the following properties[7]:

$$\begin{aligned} &\tilde{C} = \tilde{C}^\dagger = \tilde{C}^{-1}, \\ &\tilde{C}\gamma_\mu\tilde{C} = \begin{cases} -\gamma_\mu^T & 2N = 8k+2, \quad 8k+6 \\ +\gamma_\mu^T & 2N = 8k, \quad 8k+4. \end{cases}, \\ &\tilde{C}^T = \begin{cases} +\tilde{C} & 2N = 8k, \quad 8k+6 \\ -\tilde{C} & 2N = 8k+2, \quad 8k+4. \end{cases}, \\ &\tilde{C}\gamma_{2N+1}\tilde{C} = \begin{cases} +\gamma_{2N+1} & 2N = 4k, \\ -\gamma_{2N+1} & 2N = 4k+2. \end{cases}. \end{aligned} \tag{B1}$$

An explicit representation of $\tilde{C}$ can be easily found using the following recursive definition for the gamma matrices of $SO(2N)$. For $SO(2)$ we take the Pauli matrices as our gamma matrices:

$$SO(2){:}\ \gamma_i^{(2)} = \sigma_i, \quad i = 1, \ldots, 3, \tag{B2}$$

and then for $SO(2N)$ for $N \geq 2$ we take

$$\begin{aligned} \gamma_i^{(2N)} &= \sigma_1 \otimes \gamma_i^{(2N-2)}, \quad i = 1, \ldots, 2N-1, \\ \gamma_{2N}^{(2N)} &= \sigma_2 \otimes 1, \\ \gamma_{2N+1}^{(2N)} &= \sigma_3 \otimes 1. \end{aligned} \tag{B3}$$

In this basis the matrix $\tilde{C}$ is found to be a direct product of matrices alternating between σ_2 and σ_3:

$$\begin{aligned} &\tilde{C}^{(2)} = \sigma_2, \qquad \tilde{C}^{(4)} = \sigma_3 \otimes \sigma_2, \qquad \tilde{C}^{(6)} = \sigma_2 \otimes \sigma_3 \otimes \sigma_2, \\ &\tilde{C}^{(8)} = \sigma_3 \otimes \sigma_2 \otimes \sigma_3 \otimes \sigma_2, \end{aligned} \tag{B4}$$

and so on. It can be easily verified in this basis that $\tilde{C}$ possesses the properties in Eq. (B1).

Now it is easy to prove that the eigenvalues of $H_{\rm int}$ have to be at least doubly degenerate for all $SO(2N)$ groups except $2N = 8k$, where k is an integer. $H_{\rm int}$ consists of sums of products of $4k$ γ matrices totally antisymmetrized in their indices, with real coefficients; it follows that

$$H_{\rm int} = H_{\rm int}^\dagger, \quad [H_{\rm int}, \gamma_{2N+1}] = 0, \quad \tilde{C}H_{\rm int}\tilde{C} = H_{\rm int}^T. \tag{B5}$$

Since $[H_{\rm int}, \gamma_{2N+1}] = 0$ we can simultaneously diagonalize $H_{\rm int}$ and γ_{2N+1} with eigenvectors $\psi_{n,\sigma}$ where

$$\begin{aligned} &H_{\rm int}\psi_{n,\sigma} = \lambda_n\psi_{n,\sigma}, \\ &\gamma_{2N+1}\psi_{n,\sigma} = \sigma\psi_{n,\sigma}, \qquad \lambda_n \in \mathbb{R}, \qquad \sigma = \pm 1. \end{aligned} \tag{B6}$$

Now consider their action on the vector $\chi_{n,\sigma} \equiv \tilde{C}\psi_{n,\sigma}^*$:

$$\begin{aligned} H_{\rm int}\chi_{n,\sigma} &= \tilde{C}(\tilde{C}H_{\rm int}\tilde{C})\psi_{n,\sigma}^* = \tilde{C}H_{\rm int}^T\psi_n^* = \tilde{C}H_{\rm int}^*\psi_n^* \\ &= \lambda_n\tilde{C}\psi_n^* = \lambda_n\chi_{n,\sigma}, \end{aligned} \tag{B7}$$

and

$$\begin{aligned} \gamma_{2N+1}\chi_{n,\sigma} &= \tilde{C}(\tilde{C}\gamma_{2N+1}\tilde{C})\psi_{n,\sigma}^* \\ &= \begin{cases} \tilde{C}\gamma_{2N+1}\psi_{n,\sigma}^* = +\sigma\chi_{n,\sigma} & 2N = 4k \\ -\tilde{C}\gamma_{2N+1}\psi_{n,\sigma}^* = -\sigma\chi_{n,\sigma} & 2N = 4k+2. \end{cases} \end{aligned} \tag{B8}$$

Thus $\chi_{n,\sigma}$ is an eigenstate of $H_{\rm int}$ with eigenvalue λ_n, and now we would like to know if it is proportional to $\psi_{n,\sigma}$, in which case we have learned nothing, or orthogonal to $\psi_{n,\sigma}$, in which case we have shown that the eigenvalue λ is at least doubly degenerate.

First of all, we see that while $\gamma_{2N+1}\psi_{n,\sigma} = \sigma\psi_{n,\sigma}$ we have $\gamma_{2N+1}\chi_{n,\sigma} = -\sigma\chi_{n,\sigma}$ for $2N = 4k'+2 = 8k+2, 8k+6$, proving that $\psi_{n,\sigma}$ and $\chi_{n,\sigma}$ are indeed orthogonal in these cases and the spectrum is doubly degenerate. We can also directly compute their inner product and find

$$\chi_{n,\sigma}^\dagger\psi_{n,\sigma} = \psi_{n,\sigma}^T\tilde{C}\psi_{n,\sigma}, \tag{B9}$$

which equals zero whenever $\tilde{C}$ is antisymmetric, which occurs for $2N = 8k+2, 8k+4$—and so the spectrum is doubly degenerate in these cases also.

Putting the two results together we see double degeneracy for $2N = 8k+2, 8k+4, 8k+6$, leaving only $2N = 8k$ as a possible candidate for $H_{\rm int}$ to have unique eigenvalues.

Now we can ask: for $SO(8k)$ do we need to go beyond the γ^4 terms in $H_{\rm int}$ in order to gap all of the edge states? For this we will just count parameters. $H_{\rm int}$ is 2^N-dimensional, so we should be able to obtain 2^N different eigenvalues if we have at least 2^N parameters in $H_{\rm int}$. (This is overkill, since we only need the lowest eigenvalue to be unique). The number of independent antisymmetric 4-index tensors, whose indices can take $2N$ values is $2N!/[4!(2N-4)!]$ which is greater than 2^N for $N \geq 3$. Therefore a purely 4-fermion interaction can gap all the fermions for every $SO(8k)$.

[7]Note that the conventional definition of the charge conjugation matrix C satisfies $C\gamma_\mu C = -\gamma_\mu^T$ whereas $\tilde{C}$ defined here satisfies that equation for $SO(4k+2)$ but $\tilde{C}\gamma_\mu\tilde{C} = +\gamma_\mu^T$ for $SO(4k)$, both C and $\tilde{C}$ to conjugate the generators, $\sigma_{\mu\nu} \to -\sigma_{\mu\nu}^T$, and the alternating sign better serves our purpose here. For $SO(4k)$ the conventional C is given by multiplying $\tilde{C}$ by the chiral matrix γ_{4k+1}.

A simple example in the eight flavor model of a time-reversal invariant interaction that gaps all of the edge states is

$$H_{\rm int} = \omega(\hat{c}_1\hat{c}_2\hat{c}_3\hat{c}_4 + \text{H.c.} + 1), \tag{B10}$$

where the $\hat{c}_i$ ladder operators were defined in Eq. (6.4), and we assume $\omega > 0$. (Under time reversal, $\hat{c}_i \leftrightarrow -\hat{c}_i^\dagger$.) The eigenstates of $H_{\rm int}$ include a unique ground state with eigenvalue $E = 0$, fourteen degenerate states with eigenvalue $E = \omega$, and a unique maximal state with eigenvalue 2ω. The eigenstates corresponding to the minimum and maximum energy are linear combinations of the empty and fully occupied states, $(|0000\rangle \mp |1111\rangle)/\sqrt{2}$, where $\hat{c}_i|0000\rangle = 0$, and $|1111\rangle = \hat{c}_4^\dagger\hat{c}_3^\dagger\hat{c}_2^\dagger\hat{c}_1^\dagger|0000\rangle$.

The interaction in Eq. (B10) can be realized in our $1+1$-dimension Lagrangian by the term

$$\mathcal{L}_{\rm int} = \omega[(\psi_1 + i\psi_5)^T C\gamma_\chi(\psi_2 + i\psi_6)] \times [(\psi_3 + i\psi_7)^T C\gamma_\chi(\psi_4 + i\psi_8)] + \text{H.c.} \tag{B11}$$

[1] Andreas W. W. Ludwig, Topological phases: Classification of topological insulators and superconductors of non-interacting fermions, and beyond, Phys. Scr. T **168**, 014001 (2015).

[2] Xiao-Gang Wen, Symmetry-protected topological phases in noninteracting fermion systems, Phys. Rev. B **85**, 085103 (2012).

[3] Shinsei Ryu, Andreas P. Schnyder, Akira Furusaki, and Andreas W. W. Ludwig, Topological insulators and superconductors: Tenfold way and dimensional hierarchy, New J. Phys. **12**, 065010 (2010).

[4] Alexei Kitaev, Periodic table for topological insulators and superconductors, AIP Conf. Proc. **1134**, 22 (2009).

[5] Andreas P. Schnyder, Shinsei Ryu, Akira Furusaki, and Andreas W. W. Ludwig, Classification of topological insulators and superconductors in three spatial dimensions, Phys. Rev. B **78**, 195125 (2008).

[6] Curtis G. Callan, Jr. and Jeffrey A. Harvey, Anomalies and fermion zero modes on strings and domain walls, Nucl. Phys. **B250**, 427 (1985).

[7] Edward Witten and Kazuya Yonekura, Anomaly inflow and the η-invariant, In *Memorial Volume for Shoucheng Zhang* (World Scientific, Singapore, 2022), pp. 283–352.

[8] David B. Kaplan and Srimoyee Sen, Index Theorems, Generalized Hall Currents, and Topology for Gapless Defect Fermions, Phys. Rev. Lett. **128**, 251601 (2022).

[9] David J. Thouless, Mahito Kohmoto, M. Peter Nightingale, and Marcel den Nijs, Quantized Hall Conductance in a Two-Dimensional Periodic Potential, Phys. Rev. Lett. **49**, 405 (1982).

[10] Kazuo Fujikawa, Path-Integral Measure for Gauge-Invariant Fermion Theories, Phys. Rev. Lett. **42**, 1195 (1979).

[11] David B. Kaplan and Martin Schmaltz, Supersymmetric Yang-Mills theories from domain wall fermions, Chin. J. Phys. (Taipei) **38**, 543 (2000).

[12] A. Yu Kitaev, Unpaired Majorana fermions in quantum wires, Phys. Usp. **44**, 131 (2001).

[13] Lukasz Fidkowski and Alexei Kitaev, The effects of interactions on the topological classification of free fermion systems, Phys. Rev. B **81**, 134509 (2010).

[14] Qian Niu and D. J. Thouless, Quantised adiabatic charge transport in the presence of substrate disorder and many-body interaction, J. Phys. A **17**, 2453 (1984).

[15] Qian Niu, D. J. Thouless, and Yong-Shi Wu, Quantized Hall conductance as a topological invariant, Phys. Rev. B **31**, 3372 (1985).

[16] Alexander Altland and Martin R. Zirnbauer, Novel Symmetry Classes in Mesoscopic Normal-Superconducting Hybrid Structures, Phys. Rev. B **55**, 1142 (1997).

[17] Wikipedia contributors, Periodic table of topological invariants (2018) (Online; accessed 24-March-2022).

[18] R. Jackiw and C. Rebbi, Solitons with fermion number 1/2, Phys. Rev. D **13**, 3398 (1976).

[19] A. Norman Redlich, Gauge Noninvariance and Parity Nonconservation of Three-Dimensional Fermions, Phys. Rev. Lett. **52**, 18 (1984).

[20] K. Ishikawa, Chiral Anomaly and Quantized Hall Effect, Phys. Rev. Lett. **53**, 1615 (1984).

[21] David B. Kaplan, A method for simulating chiral fermions on the lattice, Phys. Lett. B **288**, 342 (1992).

[22] Karl Jansen and Martin Schmaltz, Critical momenta of lattice chiral fermions, Phys. Lett. B **296**, 374 (1992).

[23] Maarten F. L. Golterman, Karl Jansen, and David B. Kaplan, Chern-Simons currents and chiral fermions on the lattice, Phys. Lett. B **301**, 219 (1993).

[24] Kenzo Ishikawa and Toyoki Matsuyama, A microscopic theory of the quantum Hall effect, Nucl. Phys. **B280**, 523 (1987).

[25] Bobby Cheng and Chris Ford, Fermion zero modes for Abelian BPS monopoles, Phys. Lett. B **720**, 262 (2013).

[26] Shinsei Ryu and Shou-Cheng Zhang, Interacting topological phases and modular invariance, Phys. Rev. B **85**, 245132 (2012).

[27] Hong Yao and Shinsei Ryu, Interaction effect on topological classification of superconductors in two dimensions, Phys. Rev. B **88**, 064507 (2013).

[28] Zheng-Cheng Gu and Michael Levin, The effect of interactions on 2D fermionic symmetry-protected topological phases with Z2 symmetry, Phys. Rev. B **89**, 201113 (2014).

[29] Tsuneya Yoshida and Akira Furusaki, Correlation effects on topological crystalline insulators, Phys. Rev. B **92**, 085114 (2015).
[30] Yingfei Gu and Xiao-Liang Qi, Axion field theory approach and the classification of interacting topological superconductors, arXiv:1512.04919.
[31] Yi-Zhuang You and Cenke Xu, Symmetry-protected topological states of interacting fermions and bosons, Phys. Rev. B **90**, 245120 (2014).
[32] Xue-Yang Song and Andreas P. Schnyder, Interaction effects on the classification of crystalline topological insulators and superconductors, Phys. Rev. B **95**, 195108 (2017).
[33] Raquel Queiroz, Eslam Khalaf, and Ady Stern, Dimensional Hierarchy of Fermionic Interacting Topological Phases, Phys. Rev. Lett. **117**, 206405 (2016).
[34] Juven Wang and Yi-Zhuang You, Symmetric mass generation, Symmetry **14**, 1475 (2022).
[35] Juven Wang and Xiao-Gang Wen, Non-perturbative regularization of $1+1d$ anomaly-free chiral fermions and bosons: On the equivalence of anomaly matching conditions and boundary gapping rules, Phys. Rev. B **107**, 014311 (2023).
[36] Yi-Zhuang You and Cenke Xu, Interacting topological insulator and emergent grand unified theory, Phys. Rev. B **91**, 125147 (2015).
[37] Juven Wang and Xiao-Gang Wen, A solution to the $1+1D$ gauged chiral fermion problem, Phys. Rev. D **99**, 111501 (2018).
[38] Meng Zeng, Zheng Zhu, Juven Wang, and Yi-Zhuang You, Symmetric Mass Generation in the $1+1$ Dimensional Chiral Fermion 3-4-5-0 Model, Phys. Rev. Lett. **128**, 185301 (2022).
[39] David Tong, Comments on symmetric mass generation in 2d and 4d, J. High Energy Phys. 01, 07 (2022.
[40] Shlomo S. Razamat and David Tong, Gapped Chiral Fermions, Phys. Rev. X **11**, 011063 (2021).
[41] Jeffrey Goldstone and Frank Wilczek, Fractional Quantum Numbers on Solitons, Phys. Rev. Lett. **47**, 986 (1981).
[42] Yi-Zhuang You, Zhong Wang, Jeremy Oon, and Cenke Xu, Topological number and fermion Green's function for strongly interacting topological superconductors, Phys. Rev. B **90**, 060502 (2014).
[43] Simon Catterall and David Schaich, Novel phases in strongly coupled four-fermion theories, Phys. Rev. D **96**, 034506 (2017).
[44] Yi-Zhuang You, Yin-Chen He, Cenke Xu, and Ashvin Vishwanath, Symmetric Fermion Mass Generation as Deconfined Quantum Criticality, Phys. Rev. X **8**, 011026 (2018).
[45] Yichen Xu and Cenke Xu, Green's function zero and symmetric mass generation, arXiv:2103.15865.
[46] Gerard 't Hooft, Symmetry Breaking through Bell-Jackiw Anomalies, Phys. Rev. Lett. **37**, 8 (1976).
[47] Andrei Bernevig and Titus Neupert, Topological superconductors and category theory, *Lecture Notes of the Les Houches Summer School: Topological Aspects of Condensed Matter Physics* (2017), pp. 63–121.