PHYSICAL REVIEW B **92**, 085105 (2015)

# Patterns of electromagnetic response in topological semimetals

Srinidhi T. Ramamurthy and Taylor L. Hughes
*Department of Physics and Institute for Condensed Matter Theory, University of Illinois at Urbana-Champaign, Illinois 61801, USA*

(Received 17 July 2014; revised manuscript received 28 May 2015; published 3 August 2015)

Topological semimetals are gapless states of matter which have robust and unique electromagnetic responses and surface states. In this paper, we consider semimetals which have pointlike Fermi surfaces in various spatial dimensions $D = 1,2,3$ which naturally occur in the transition between a weak topological insulator and a trivial insulating phase. These semimetals include those of Dirac and Weyl types. We construct these phases by layering strong topological insulator phases in one dimension lower. This perspective helps us understand their effective response field theory that is generally characterized by a 1-form $b$ which represents a source of Lorentz violation and can be read off from the location of the nodes in momentum space and the helicities/chiralities of the nodes. We derive effective response actions for the two-dimensional (2D) and 3D Dirac semimetals and extensively discuss the response of the Weyl semimetal. We also show how our work can be used to describe semimetals with Fermi surfaces with lower codimension as well as to describe the topological response of 3D topological crystalline insulators.

DOI: 10.1103/PhysRevB.92.085105 PACS number(s): 71.90.+q, 75.47.−m, 73.43.−f

The discovery of topological-band insulators (TIs) and their novel electronic properties has led to a reexamination and search for robust topological features of the electronic structure of many different material types [1]. Some notable properties of TIs include a gapped, insulating bulk interior, protected boundary modes that are robust even in the presence of disorder, and quantized electromagnetic transport. A full (periodic) classification table of noninteracting fermionic states of matter that are protected by time-reversal ($T$), chiral, and/or particle-hole ($C$) symmetries has been established [2–4]. Recent work has further augmented the initial periodic table by including the classification of states protected by spatial symmetries such as translation, reflection, and rotation [5–20]. While these symmetry-protected topological phases are theoretically interesting in their own right, this field would not have attracted so much attention if it were not for the prediction and confirmation of candidate materials for many different topological classes. A few examples are the two-dimensional (2D) quantum spin Hall insulator (e.g., CdTe/HgTe quantum wells [21–23]), the 3D $T$-invariant strong TI (e.g., BiSb [24], $\mathrm{Bi_2Se_3}$ [25–27]), the 2D quantum anomalous Hall (Chern) insulator (e.g., Cr-doped $\mathrm{(Bi,Sb)_2Te_3}$ [28,29]), and the 3D $T$-invariant topological superfluid state (e.g., the B phase of He-3 [2,4,30]).

All of the above work pertains to gapped systems; however, recent theoretical predictions have shown that even materials that are not bulk insulators can harbor robust topological electronic responses, transport properties, and conducting surface/boundary states [31–40]. This class of materials falls under the name topological semimetals and represents another type of noninteracting electronic structure with a topological imprint. The most well-studied examples of topological semimetals (TSMs) are the 2D Dirac semimetal (e.g., graphene [41]), the 3D Weyl semimetal (possibly in pyrochlore irridates [32], inversion-breaking superlattices [42], or optical lattices [43]), and the 3D Dirac semimetal [44–49]. While there are yet to be any confirmed experimental candidates for 3D Weyl semimetals, their unique phenomenology, including incomplete Fermi-arc surface modes, an anomalous Hall effect (AHE), and a chiral magnetic effect, has drawn theoretical and experimental attention to these materials. Two types of 3D Dirac semimetals, i.e., a Dirac semimetal type with nodes at the time-reversal invariant momenta [44,47,48], and one with nodes away from those special momenta [45,46] has been reported to be found. In addition to these TSMs there is a large set of symmetry-protected TSMs which rely on additional symmetries for their stability [35]. Finally, we also note that there are superconducting relatives of these semimetal phases called topological nodal superconductors, or Weyl superconductor phases, that await experimental discovery [35,50,51], though we do not consider them further.

In this article we explore the quasitopological response properties of TSMs in the presence of external electromagnetic fields. We present a generic construction of TSMs that can be adapted to model almost any type of TSM. This construction allows us to manifestly determine the electromagnetic response properties of the TSMs in question. It also enables us to uncover clear patterns in the quasitopological electromagnetic response terms exhibited by the various semimetal types and in different spatial dimensions. In addition, our work nicely complements the extensive recent work studying the topological response properties of Weyl semimetals [32–35,40,52].

The previous field-theoretic calculations of the response of Weyl semimetals have predicted a novel electromagnetic response for the 3D TSMs, but not without some subtlety [34,53–58]. Thus, another goal of this article is to address the electromagnetic (EM) response for various topological semimetals and to show the validity and limitations of the field-theory results. To this end, we provide explicit numerical simulations using simple lattice models to complement our transparent analytic discussion. In addition to the discussion of the 3D Weyl semimetals, we carefully illustrate the pattern of TSM response actions that exist in 1D metallic wires and 2D Dirac semimetals to establish a unified framework of the EM response of TSMs. We discuss the influence of and, in some cases, the necessity of antiunitary and/or spatial symmetries for the stability of the semimetal phase and the resultant implications for the EM response. Furthermore, we provide an analytic solution for the boundary modes of the TSMs

1098-0121/2015/92(8)/085105(42)

in our simple lattice models, derive a topological effective response action for the 2D and 3D Dirac semimetals, calculate the EM response at interfaces between different TSMs, and, where possible, emphasize the important physical quantities of TSMs that can be observed.

The article is organized as follows. In Sec. I we discuss the preliminaries and motivation for the work. This section provides our approach to the characterization of TSMs and further reviews previous work on the response of Weyl semimetals. After this we begin by discussing 1D semimetals in Sec. II as a warm-up problem for the rest of the article. From this, we move on to 2D Dirac semimetals in Sec. III. We discuss the connection between 1D TIs and 2D Dirac semimetals and discuss the low-energy boundary states of the Dirac semimetal. We calculate the "topological" contribution to the EM response for a TSM with two Dirac points using a field-theoretical calculation and then go on to generalize the picture to a generic number of Dirac points. We also discuss the microscopic origin and subtleties of the response using lattice model realizations. For the 2D Dirac semimetal some parts of the quasitopological EM response have been discussed quite extensively for graphene in Refs. [59,60]. Additionally, the time-reversal-breaking quantum anomalous Hall (Chern insulator) response in a gapped Dirac semimetal has been studied quite carefully as well [28]. We generalize these results to the case of an arbitrary number of nodes, including fewer nodes than the four in graphene, and consider subtleties which arise from the $\mathbb{Z}_2$ nature of the edge modes which arise in these models. Based on the work of Ref. [61], we are also able to give a valid definition of the electric charge polarization for Dirac semimetals, a quantity usually reserved for bulk insulators, which we use to define the quasitopological EM response of the semimetal.

After this we proceed to 3D, where we study both Weyl semimetals and 3D Dirac semimetals (we also comment on the possible response of 3D topological crystalline insulators in the Discussion and Conclusions). First, in Sec. IV A we discuss the response properties of a Weyl semimetal. While some of the results in this section are already known, we present the material from a different perspective and include lattice-regularized numerical calculations of the response, which show precisely under what conditions there is a nonzero current due to the chiral magnetic effect (CME). We also connect the numerical result with our earlier discussion on 1D systems since, when placed in a uniform magnetic field, one can map the 3D Weyl semimetal to identical copies of the 1D model. This may help resolve some of the controversy surrounding questions raised over whether the CME exists in a lattice model. We also provide many new results including an analytic description of the boundary modes for a lattice model of the Weyl semimetal, the response behavior of a heterojunction between two different Weyl semimetals, and a discussion of the anomaly cancellation which connects the bulk and surface responses. We finish our discussion of 3D materials with Sec. IV B on 3D Dirac semimetals. The interesting quasitopological response of 3D Dirac semimetals is present when the Dirac nodes are located at generic points in momentum space away from the zone boundaries [45,46]. This type of Dirac semimetal (DSM) was shown to be stable in Ref. [49] and has been experimentally observed in $\mathrm{Na_3Bi}$ [46]. We find a type of EM response which appears when the surface of the 3D DSM is in contact with a magnetic layer and is related to the response predicted for the 2D quantum spin Hall insulator [3,62]. Finally, in Sec. V we summarize our results and briefly discuss an application of our work to the topological response properties of 3D topological crystalline insulators and also how to consider semimetals with Fermi surfaces with different codimensions, both of which will be discussed more in future work.

## I. PRELIMINARIES AND MOTIVATION

### A. Electromagnetic response

One of the primary goals of this work is to produce valuable intuition for understanding the response properties of generic topological semimetals with point nodes (linelike nodes will be considered in future work). In this section we begin with a simple physical construction that is applicable to different types of topological semimetals and provides a basis for understanding the EM response of a wide class of TSMs in a unified manner. In this context we discuss some of the previous work on the EM response of Weyl semimetals as an explicit example. Finally, before we move on to more technical calculations, we illustrate the pattern followed by the EM response of TSMs in various spatial dimensions.

An insightful way to view a topological semimetal is as a stable gapless phase that separates a trivial insulator phase from a weak TI phase. A trivial insulator is essentially a band insulator that is adiabatically connected to the decoupled atomic limit. The electronic structure of trivial insulators does not exhibit any nonvanishing topological properties. On the other hand, weak topological insulators (WTIs) are anisotropic, gapped topological phases that are protected by translation symmetry and characterized by a vector topological invariant $\vec{\nu}$. The fact that the topological invariant is a vector, and not a scalar, is an indication that they are essentially anisotropic. This anisotropy can be made more apparent because each WTI phase in $d$ spatial dimensions can be adiabatically connected to a limit of decoupled $(d-1)$-dimensional systems that are layered perpendicular to $\vec{\nu}$. The $(d-1)$-dimensional building blocks that make up the $d$-dimensional WTI must each be in a $(d-1)$-dimensional TI phase to generate the higher-dimensional WTI phase. Of course, one can also construct a $d$-dimensional WTI from $(d-q)$-dimensional $(1 < q < d)$ topological phases, which will lead to Fermi surfaces with lower codimension although we save their consideration to future work.

The most well-known example of a WTI is a stack of planes of 2D integer quantum Hall states (or 2D Chern insulators) that create the so-called 3D quantum Hall effect (QHE) [63–66]. If the 2D planes are parallel to the $xy$ plane, then the vector invariant $\vec{\nu} \propto \hat{z}$. If the coupling between the planes is weak, then the bulk gap, arising from the initial bulk gaps of the 2D planes, will not be closed by the dispersion in the stacking direction. However, when the interlayer tunneling becomes strong enough, the system will become gapless and exhibit the so-called Weyl semimetal phase. Eventually, as the tunneling strength increases, the system will transition to another gapped phase that will either be a different WTI phase or a trivial insulator. Thus, in the simplest case, the Weyl semimetal is an intermediate gapless phase separating a WTI from a trivial insulator. As we discuss later, a similar picture can be developed for the 2D Dirac semimetal which can be adiabatically connected to an array of 1D TI wires that are stacked into 2D. Ultimately, this type of description of TSMs will be very useful since the relevant EM response properties of the lower-dimensional TI building blocks are known [3], and the problem of the TSM response is transformed into understanding how the interlayer coupling affects the EM responses of the TI constituents.

While it is well-known that TIs and WTIs exhibit topological EM response properties, at the transition between trivial and topological phases the relevant topological response coefficients are no longer well-defined, i.e., not sharply quantized. In fact, there is usually a jump from a quantized nonzero value in the topological phase to a vanishing value of the response coefficient in the trivial phase. Therefore, it is a bit surprising that the semimetal phases intermediate between trivial and topological insulators retain an imprint of the topological response.

This is illustrated beautifully in the case of the Weyl semimetal as we now discuss. A trivial insulator has no topological component to its EM response; it obeys Maxwell's equations with the conventional insulator constituent relations for polarization and magnetization. On the other hand, the nontrivial WTI represented by the 3D quantum Hall insulator produces a topological response term in the effective action

$$S_{\text{eff}}[A_\mu] = -\frac{e^2}{2\pi h}\int d^3x dt\; \nu_\mu \epsilon^{\mu\sigma\rho\tau} A_\sigma \partial_\rho A_\tau, \qquad (1)$$

where $\nu_0 = 0$, $\nu_i = \frac{n}{2} G_i$ are the components of a half-integer multiple $n/2$ of a reciprocal lattice vector $\vec{G}$, and $A_\mu$ are external EM fields. This action implies that spatial planes perpendicular to $\vec{\nu}$ will have a Hall effect, and the 3D Hall conductance is $\sigma_{xy} = -ne^2/ha_G$, where $a_G$ is the lattice spacing along $\vec{G}$ ($a_G = 2\pi/|\vec{G}|$). Note that we have chosen the global negative sign to match the convention of Ref. [54]. The trivial insulator phase can be thought of as the case when $\vec{\nu} = \vec{0}$. It is clear that the topological response is anisotropic, as the particular $\vec{\nu} = \frac{\vec{G}}{2}$ breaks rotation invariance (and as a consequence Lorentz invariance if we are considering relativistic theories which are a common low-energy description of a TSM; see Ref. [67]).

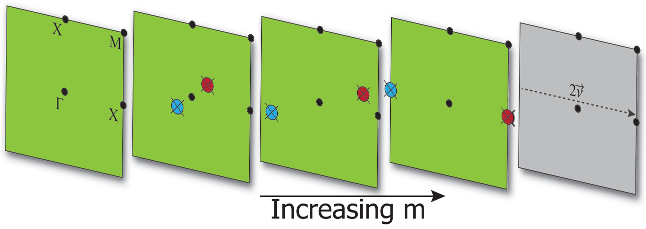

FIG. 1. (Color online) Schematic illustration of the motion of point nodes in the $k_z = 0$ plane of a cubic, 3D BZ as a parameter $m$ is adjusted. As $m$ increases, two Weyl nodes with opposite chirality (as represented by the color shading) are created in the 2D subspace (i.e., $k_z = 0$) of a full 3D BZ. As $m$ increases further, the nodes move throughout the BZ, meet at the boundary, and then finally annihilate to create a gapped phase with a weak topological invariant proportional to the reciprocal lattice vector separation $\vec{G} = 2\vec{\nu}$ of the Weyl nodes before annihilation. The far left BZ represents a trivial insulator, the far right represents a WTI, and the intermediate slices represent the Weyl semimetal phase.

Now that we understand the topological response of the two phases that straddle the Weyl semimetal phase, we can try to understand the response of the simplest type of Weyl semimetal, i.e., the kind with only two Weyl nodes (the minimal number). Let us imagine the following process, where we begin with a trivial insulator and nucleate two Weyl nodes at the Γ point in the 3D Brillouin zone (BZ) by tuning a parameter $m$ (see Fig. 1). The low-energy $k \cdot P$ Hamiltonian near each Weyl node is of the form $H_{\text{Weyl}}(\text{p}) = p_1\sigma^1 + p_2\sigma^2 + p_3\sigma^3$, where $\sigma^a$ are Pauli matrices and we have set the velocity to unity. As $m$ is further changed, the Weyl nodes will move through the BZ but cannot be gapped (assuming translation invariance) unless they meet each other again or another node with opposite chirality. The reason is that if the Weyl nodes are separated, then there is no matrix which anticommutes with $H_{\text{Weyl}}(\text{p})$, and thus no perturbation can be added that will open a gap. If the two Weyl nodes (with opposite chirality) meet and become degenerate, then the resulting $4 \times 4$ Hamiltonian $H_{\text{Weyl}} \oplus \widetilde{H}_{\text{Weyl}}$ has the Dirac form. In this case one can find an anticommuting matrix to add that will perturbatively open a gap and annihilate the nodes. If the Weyl nodes meet at the boundary of the BZ at points which differ by a reciprocal lattice vector $\vec{G}$, then upon annihilation the system will undergo a change of its weak invariant, i.e., $\Delta\vec{\nu} = \frac{\vec{G}}{2}$. Thus, if the system starts with $\vec{\nu} = 0$ then it will have a transition to a nontrivial WTI during this process.

During the process of tuning $m$ we see that before we nucleate the Weyl nodes, there is no topological response, and after they annihilate at the BZ boundary there will be a nontrivial Hall response. We now can ask the following: What is the response in the gapless semimetal phase? The answer turns out to be simple, we just have the response of Eq. (1), with $\vec{\nu} = \vec{b}$, where $2\vec{b}$ is the difference in momentum between the two Weyl nodes [54]. Interestingly, the response coefficient smoothly interpolates between the two insulating end points. This remarkable result can be extended even further because we also have a notion of a relative energy between the Weyl nodes. Because of this, we can generate a coefficient $\nu_0 = b_0$ in Eq. (1), where $2b_0$ is the energy difference between the two Weyl nodes. This enhances the response as now we can have a Lorentz-invariance-violating 4-vector response coefficient $\nu_\mu$ (again, see Ref. [67] for what is meant by "Lorentz violating").

The addition of a response proportional to $\nu_0$ is a new feature of the semimetal since one cannot define a notion of $\nu_0$ in the pure WTI because the low-energy theory is gapped. The reason one can have a spatial vector in the gapped WTI is because of the translation-symmetry (and continuous-rotation-symmetry)-breaking lattice structure which gives rise to the reciprocal lattice vector(s) $\vec{G}$. On the other hand, if we had a periodically driven system, i.e., a system evolving according to Floquet dynamics, then, even in the insulating case, we could have a nonzero $\nu_0$ which would be proportional to the driving frequency of the time-dependent field, i.e., the reciprocal lattice vector for time. In the Weyl semimetal phase, the existence of nondegenerate Weyl nodes immediately gives rise to a Lorentz-breaking 4-vector similar to the kind anticipated by Refs. [68,69] for Lorentz violations in high-energy physics.

The resulting response from Eq. (1) generates an anomalous Hall effect along with a CME. The CME occurs when $b_0 \neq 0$ and is anticipated to give rise to a current when a magnetic field is applied to the system, but in the absence of any electric field. For a translation-invariant 3D material with an even number of Weyl nodes, we can determine

$$\vec{b} = \frac{1}{2}\sum_a \chi_a \vec{K}_a, \quad b_0 = \frac{1}{2}\sum_a \chi_a \epsilon_a, \tag{2}$$

where the sum runs over all of the Weyl nodes and $\chi_a$, $\vec{K}_a$, and $\epsilon_a$ are the chirality, momentum location, and energy of the $a$th node, respectively. Additionally, from the Nielsen-Ninomiya no-go/fermion-doubling theorem we know there is also the constraint that the total chirality $\sum_a \chi_a = 0$ must vanish [31]. With these definitions, the resulting charge density and current in the semimetal are given in terms of $b_\mu$ and the applied EM fields as

$$j^0 = \frac{e^2}{2\pi h}(2\vec{b}) \cdot \vec{B}, \tag{3}$$

$$\vec{j} = \frac{e^2}{2\pi h}[(2\vec{b}) \times \vec{E} - (2b_0)\vec{B}]. \tag{4}$$

As an aside, we note that while the origin and detection of the anomalous Hall current is well understood, there have been some disagreements in the recent literature about the possibility of a nonvanishing CME. To summarize the results so far, the field-theoretical results are somewhat ambiguous because of the dependence on a regularization [53,70]: A tight-binding lattice calculation has shown a vanishing result [55], while a more recent calculation has indicated the need for a slowly varying magnetic field that eventually tends toward a uniform/constant field [56]. In Sec. IV A we comment on these results and note that having an explicit source of Lorentz violation is a necessity for a nonvanishing CME effect. We also discuss the interpretation of the CME effect from a quasi-1D perspective generated from applying a uniform magnetic field to a Weyl semimetal. This allows us to map the 3D problem onto degenerate copies of the 1D system, which can be more easily analyzed.

### *General pattern of quasitopological electromagnetic response in topological semimetals*

While we have seen that it is the case for the Weyl semimetal, it is generically true that the general pattern of EM response for TSMs with point nodes in any spatial dimension stems from the existence of the Lorentz-violating vector response coefficient $b_\mu$. In systems with translation symmetry, the vector is connected to the momentum and energy difference between nondegenerate point nodes (e.g., Dirac nodes in 2D and Weyl nodes in 3D). In general, the vector represents a source of Lorentz violation in the system because it chooses a preferred direction or frame in the system, and its time and space components can both be nonvanishing [67]. For example, the spatial part of $b_\mu$ represents an anisotropic "stacking direction" similar to the case of the weak TI. Now let us denote the external EM gauge field by $A_a$ and its field strength by $F_{ab} = \partial_a A_b - \partial_b A_a$. In odd-dimensional space-time ($D+1$ is odd), the effective EM response action for point-node semimetals is

$$S[A] = \mathcal{A}_D \int d^{D+1}x\, \epsilon^{a_1 a_2 \cdots a_{D+1}} b_{a_1} F_{a_2 a_3} \cdots F_{a_D a_{D+1}}, \tag{5}$$

where the ellipses in the above equation represent further factors of the field strength and $\mathcal{A}_D$ is a dimension-dependent normalization coefficient. We see from this equation that if one calculates the current $j^\mu = \delta S[A]/\delta A_\mu$, then the result always depends on derivatives of $b_\mu$. This is important because it immediately implies that the response of semimetals in odd-dimensional space-time depends crucially on the properties of boundaries or interfaces where $b_\mu$ is changing.

In contrast, in even space-time dimensions ($D+1$ is even), the effective action for the quasitopological EM response of point-node semimetals is

$$S[A] = \mathcal{A}_D \int d^{D+1}x\, \epsilon^{a_1 a_2 \cdots a_{D+1}} b_{a_1} A_{a_2} F_{a_3 a_4} \cdots F_{a_D a_{D+1}}, \tag{6}$$

where the ellipses in the above equation represent further factors of the field strength. For example, in $(1+1)$ dimensions, we have just $S[A] = \mathcal{A}_1 \int d^2x\, \epsilon^{\mu\nu} b_\mu A_\nu$. Now if one calculates the current, the result depends on the value of $b_\mu$ itself (as well as possible derivatives in some cases). This hints that at least part of the response is determined by bulk effects alone and does not involve the properties of surfaces and boundaries. In even-dimensional space-times the literature differs on the convention for the choice of the action and some sources use

$$S[A] = \frac{\mathcal{A}_D}{2} \int d^{D+1}x\, \epsilon^{a_1 a_2 \cdots a_{D+1}} \theta F_{a_1 a_2} F_{a_3 a_4} \cdots F_{a_D a_{D+1}},$$

where $\theta \equiv 2b_\mu x^\mu$. However, this second form, while it looks somewhat nicer as far as gauge invariance is concerned, has an implicit breaking of translation symmetry. This comes from the freedom of the choice of origin in the definition of $\theta$ as we could have alternatively defined $\theta$ to be $\theta \equiv 2b_\mu(x^\mu + x_0^\mu)$ with some constant 4-vector $x_0^\mu$. Because of this, we always choose the form Eq. (6) to avoid the translation symmetry ambiguity. In fact, using the $\theta$-term version of the action leads to spurious effects when the system is not homogeneous, e.g., in the presence of boundaries.

In general, the pattern of response actions for TSMs with nodal (pointlike) Fermi surfaces is attached to an intrinsic 1-form $b = b_\mu dx^\mu$, which is determined from the electronic structure. This type of 1-form indicates some inherent anisotropy in the electronic structure and can appear in any dimension. We also note that because of the lattice periodicity, the vector $2\vec{b}$ is only determined up to a reciprocal lattice vector. Thus, the response of a TSM is only determined up to a quantum determined by the addition of a filled band. For Weyl semimetals this indeterminacy is due to the possibility of a contribution of an integer Hall conductance (per layer) from filled bands; the low-energy Fermi-surface physics does not contain information about the Hall conductance of the filled bands [71]. For cases where the response coefficients are connected to $Z_2$ invariants instead of integers, the ambiguity of contributions from filled bands must be carefully considered, as we do below for the 2D Dirac semimetal. We also note that, more generally, we can have terms in the effective action which involve an $n$ form in space-time dimensions greater than or equal to $n$ when a $d$-dimensional system has a Fermi surface with codimension less than $d$, some cases of which will be discussed elsewhere [72].

The most important feature of the quasitopological response coefficients of TSMs is that the response coefficients *continuously* change throughout the gapless TSM phase from the quantized values in the insulating phases on either side of the gapless phase. One might expect that when the gap closes there might be some complicated singular behavior in the response coefficients; however, what is special about the TSM phases is that we can continuously track the coefficient through the weak TI-TSM-trivial insulator phase diagram. We note that there could be other transport coefficients that do have more complicated singular behavior during the insulator-to-TSM transitions, but the restricted set on which we focus has this important property.

### B. Boundary degrees of freedom

The other generic feature of TSM phases is the existence of low-energy boundary modes. It is well known that TIs have robust, gapless boundary modes that exist in the bulk energy gap. A (strong) TI will contain topological boundary states on any surface, while a WTI only harbors topologically protected boundary states on surfaces where $\vec{\nu}$ does not project to zero in the surface BZ [73]. This is another clear signature of the anisotropy, and it gets passed on to the TSMs that interpolate between the WTI and trivial insulator phases.

TSMs themselves will have low-energy boundary modes, but again, only on surfaces where $b_i$ does not project to zero in the surface BZ. That is, there will be surface states on surfaces where the normal vectors are not parallel to the node separation vector $\vec{b}$. We note that even in cases where $b_i = 0$ (or $b_i$ projects to zero on a surface) there can still be surface states because $b_i$ is only well defined modulo a reciprocal lattice vector. We note that surface states that exist when $b_i = 0$ come from fully filled bands and will exist over the entire BZ (if the ground state does not carry a strong topological invariant). These surface states are not related to the properties of the semimetal and will not crucially depend on the locations of the nodes as they are continuously deformed.

The existence of boundary modes in TSMs is most easily illustrated with a simple example. Let us again resort to the picture of a Weyl semimetal arising out of a stack of identical 2D Chern (quantum anomalous Hall) insulators and, for simplicity, assume that the layers are stacked in the $z$ direction. Then, for the WTI phase in the completely decoupled limit, each Chern insulator layer contributes one set of chiral edge modes on surfaces with normal vectors in the $\hat{x}$ and/or $\hat{y}$ directions [65]. This is the simple picture of a WTI, and if each layer has a first Chern number $C_1 = 1$, then the vector invariant $\vec{\nu} = (0,0,\pi/a)$, where $a$ is the spacing between the Chern insulator layers. If the system has length $L_z = Na$ in the $z$ direction, then the total Hall conductance is $\sigma_{ij} = -\epsilon_{ijk}\frac{e^2}{\pi h}\nu^k L_z = -N\frac{e^2}{h}$, i.e., an amount $e^2/h$ per stacked layer. When the coupling between layers is turned on, then the bulk and edge states will disperse in the $z$ direction, but as long as the interlayer coupling does not close the bulk gap, then the system will remain in the WTI phase with the same Hall conductance.

To further discuss the boundary modes of the topological semimetal, it useful to illustrate with an explicit lattice model. We can represent this system as a tight-binding model on a cubic lattice where each site contains a single electronic orbital with spin-up and spin-down degrees of freedom. A representative Bloch Hamiltonian is

$$H(\vec{k}) = A\sin k_x\sigma^x + A\sin k_y\sigma^y + (2B - m - B\cos k_x - B\cos k_y - C\cos k_z)\sigma^z, \tag{7}$$

where $A$, $B$, $C$, and $m$ are parameters, $\sigma^a$ represents spin, and we have set the lattice constant $a = 1$. If we choose the parameters $A = B = 2m = 1$ and $C = 0$, this will represent a WTI phase built from decoupled layers of Chern insulator states as discussed above. We can see this from the fact that when $C = 0$ there is no dispersion in the $z$ direction, and thus we have many copies of a 2D system, one for each allowed $k_z$, i.e., one for each layer. The important point is that when $A$, $B$, and $m$ are tuned as above, then, ignoring the $z$ direction, the resulting 2D system is in a Chern insulator phase with $C_1 = 1$ [3], and thus we have decoupled copies of a nontrivial Chern insulator. When the tunneling between the layers is activated, the parameter $C$ will be nonvanishing. With $A$, $B$, and $m$ fixed as above, then for $-1/2 < C < 1/2$ the model will remain in the WTI phase. At $C = 1/2$ the bulk energy gap closes at $\vec{k} = (0,0,\pi)$. If $C$ is further increased, then there will be two points where the gap vanishes, i.e., two Weyl nodes, and they will occur at $\vec{k} = (0,0,\cos^{-1}(-\frac{m}{C}))$, where we added the dependence for a variable $m$ parameter back in. Accordingly, when $|m/C| < 1$ the system will exhibit a Weyl semimetal phase if $A = B = 1$.

As was shown in Ref. [74], we can use a model like Eq. (7) to create a nice description of the Weyl semimetal phase. For this picture, it is useful to think about the system as a family of 2D insulators $H_{k_z}(k_x,k_y) \equiv H(k_x,k_y,k_z)$, parametrized by $k_z$. For parameters representing the fully gapped WTI phase (e.g., $A = B = 2m = 1$, $C = 0$), then for each value of $k_z$ the 2D insulator $H_{k_z}(k_x,k_y)$ is in the Chern insulator phase.

Now when we tune the $C$ parameter into the Weyl semimetal phase, then the model will contain gapless Weyl nodes at $\vec{k} = (0,0,\pm k_c)$, and a separation vector $\vec{b} = (0,0,k_c)$. To understand the existence of surface states in the semimetal phase it is again helpful to think of each 2D insulator at fixed $k_z$ being in a trivial $C_1 = 0$ phase when $|k_z| > |k_c|$ and a Chern insulator phase with $C_1 = 1$ when $|k_z| < |k_c|$. Exactly at $k_z = \pm k_c$ there is a gapless "transition" as a function of $k_z$ between the trivial 2D insulator with $C_1 = 0$ and the nontrivial 2D insulator with $C_1 = 1$.

This illustration shows that in the Weyl semimetal phase we should only expect boundary states to exist over a finite range of $k_z$, i.e., $|k_z| < |k_c|$ for this particular example. For each $k_z$ in the topological range, the 2D insulator $H_{k_z}(k_x,k_y)$ contributes one propagating chiral fermion mode to the boundary degrees of freedom. These chiral boundary states manifest as incomplete surface Fermi arcs that connect Weyl points in the surface BZ for surfaces with normal vectors which are not parallel with $\vec{b}$. The picture of a TSM as a momentum-space transition in a family of lower-dimensional gapped insulators is helpful because similar concepts can be applied to understand the properties of all topological semimetals.

This completes the basic review and motivation. To summarize, we have introduced some important physical intuition and concepts pertaining to 3D Weyl semimetals, and during this process reviewed some of the previous work describing the EM response and boundary states of these systems. Now we begin a more in-depth discussion of the response and boundary states of semimetals in 1D, 2D, and 3D following the outline presented above.

## II. SEMIMETAL IN 1 + 1 DIMENSIONS

We begin with a careful study of the properties of a 1D TSM, which in this case is just an ordinary 1D metal, as noted in Ref. [33]. As a representative model we can choose a spinless one-band tight-binding model of the form

$$H_{1\mathrm{D}} = -\alpha \sum_n [c^\dagger_{n+1} c_n + c^\dagger_n c_{n+1}], \tag{8}$$

where the sum over $n$ runs over all of the lattice sites, and we let the lattice constant be $a$. This familiar model is easy to diagonalize and the energy spectrum is

$$E(k) = -2\alpha \cos ka, \tag{9}$$

where $k \in [-\pi/a,\pi/a)$. In the momentum basis the Hamiltonian is just $H_{1\mathrm{D}} = \sum_k E(k) c^\dagger_k c_k$.

Establishing a chemical potential $\mu$ that lies within the band will fill the system with a finite density of electrons. If we keep translation symmetry, we can calculate the number of particles by counting the number of occupied momentum states

$$N = \sum_{k\in\mathrm{occ.}} 1 = \frac{L}{2\pi}\int_{-k_F}^{k_F} dk = \frac{Lk_F}{\pi}, \tag{10}$$

which implies a charge density $\rho = e\frac{k_F}{\pi}$, where $k_F$ is the Fermi wave vector and $e$ is the electron charge. In the language of the previous section we note that this density breaks Lorentz invariance because it establishes a preferred frame, i.e., the rest frame of the fermion density. Thus, we should expect a Lorentz-violating contribution to the effective action. In fact, we can easily write this contribution since a background charge density just couples to the scalar EM potential $A_0$ to give a potential energy term

$$S[A_0] = -\int dx dt \rho A_0. \tag{11}$$

In addition to the density, there is the possibility of introducing an electric current that will also break Lorentz invariance. For a moment, let us consider a generic 1D lattice model with translation invariance and in the momentum basis. When minimally coupled to an EM field (e.g., through Peierls substitution) we find

$$H = \sum_k c^\dagger_k H\left(k - \frac{e}{\hbar}A_1\right) c_k, \tag{12}$$

where $H(k)$ is a Bloch Hamiltonian. The current for this system in the limit $A_1 \to 0$ is given by

$$j = \lim_{A\to 0} \frac{\partial H}{\partial A_1} = -\frac{e}{\hbar}\sum_k \left[\frac{\partial H(k)}{\partial k} n_F\right], \tag{13}$$

where $n_F$ is the Fermi-Dirac distribution, which will be a step function at $T = 0$. This can be rewritten at zero temperature as

$$j = -e \sum_{n\in\mathrm{occ}} \int_{BZ} \frac{dk}{2\pi\hbar}\frac{\partial E_n(k)}{\partial k}, \tag{14}$$

where $n$ runs over the occupied bands. Specializing to the case of our single-band model, the current is equal to $j = -\frac{e}{2\pi\hbar}[E(k_F) - E(-k_F)]$ which is nonzero only if $E(k_F) \neq E(-k_F)$. We discuss two different mechanisms for generating a current in Secs. II A and II B.

### A. 1D model in an electric field

One way to generate a nonzero electric current is to apply an external electric field. We apply an electric field by adiabatically threading magnetic flux through the hole of the periodic lattice ring via Faraday's law. This is equivalent to introducing twisted boundary conditions on the wave functions

$$\Psi(x + L) = e^{i\Phi(t)L}\Psi(x), \tag{15}$$

where

$$\Phi(t) = \frac{eEt}{\hbar} \tag{16}$$

for an electric field $E$ at time $t$. Using Eq. (14) we can easily calculate the electric current to be

$$j = \frac{2\alpha e}{\pi\hbar}\sin(k_F a)\sin[\Phi(t)a]. \tag{17}$$

For comparison, we numerically calculate the charge density and current for the case when the single band is half filled. At half filling $k_F = \pi/2a$, and thus the density should be uniform, time independent, and equal to $\rho = \frac{e}{2a}$, i.e., half an electron per site. At half filling, the current reduces to $j = \frac{2\alpha e}{\pi\hbar}\sin[\Phi(t)a]$. The numerical calculations are shown in Fig. 2, and they agree with the analytic results.

We note in passing that for finite-size lattice models some care must be taken to correctly calculate a smooth electric current response. We have intended to calculate the current of a metallic/gapless system, but there are finite-size gaps in the energy spectrum between each state separated by $\Delta k = 2\pi/L$. Thus, if we want the system to behave as a gapless system should, we must apply a minimum threshold electric field. If too small of an electric field is applied at a given system size, the model will behave like a gapped insulator instead. To avoid this we can simply enforce the canonical momentum $\Pi_x = p_x - eA_1$ to be a multiple of $2\pi\hbar/L$ so that the system remains gapless at each time step. If this is not done, then the system will behave as gapped insulator and we will see steps in the current response. Ensuring that $\frac{e}{\hbar}A_1 = \frac{2\pi m}{L}$ at every time step saves us this trouble, and in our simulations for this section we have always taken $\Phi(t) = eEt/\hbar$ to be a multiple of $2\pi/L$ and never smaller than this value. Physically we understand that, for a system with these finite-size gaps,

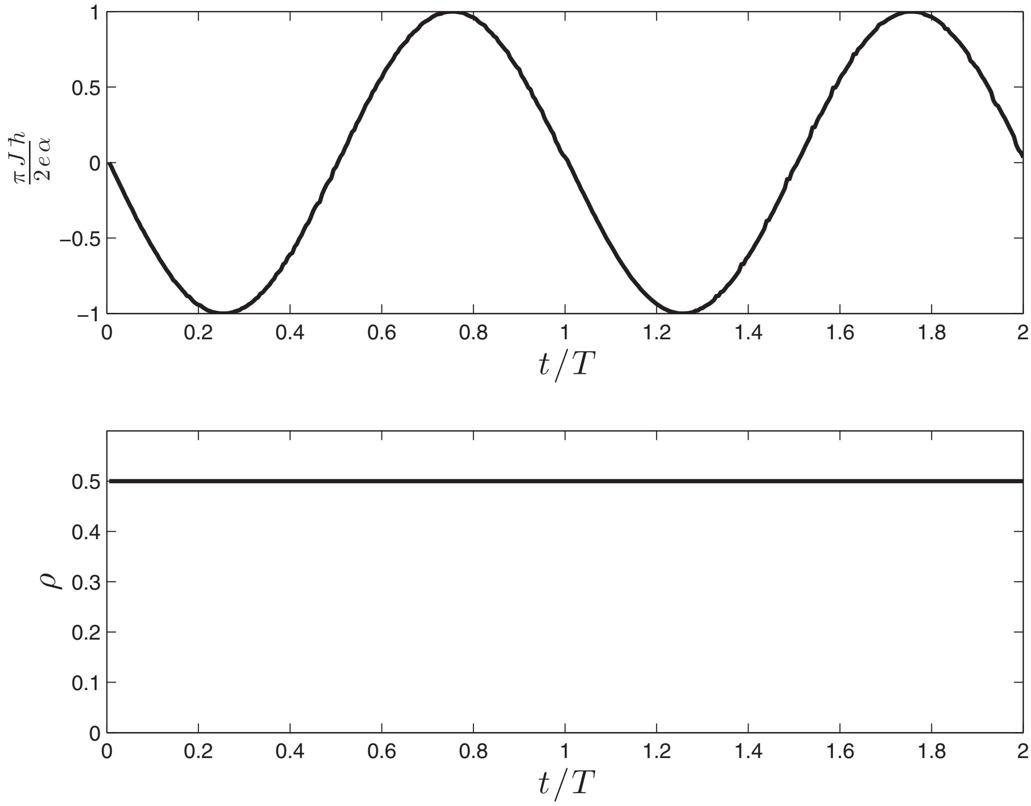

FIG. 2. The current and charge density of the 1D (semi-)metal are plotted vs time for half filling and for nearest-neighbor hopping $\alpha = 1$. The current has a periodic response, as expected, with a period of 200 time slices for an electric field of strength $E = \frac{\hbar}{eT(200a)}$ for some time scale $T$ that is long. The charge density is given by $\rho = e\frac{k_F}{\pi} = e/2a$, as expected, and shows no time-dependent behavior.

an infinitesimal adiabatic current generation will not work. Instead we must turn on a large enough electric field so that there is some nonadiabaticity so that the finite-size gaps can be overcome.

Although we do not present the results here, we have carried out numerical calculations for various filling factors and electric field strengths, and the analytic results match the numerical simulations. If we change the boundary conditions from periodic to open, then the charge density remains the same (possibly up to some damped density oscillations near the ends of the wire), but the current vanishes, as expected. Hence, we see that in the presence of an electric field with periodic boundary conditions the response action of the 1D (semi-)metal is

$$S[A_\mu] = \int dx dt[-\rho A_0 + jA_1] = \int dx dt j^\mu A_\mu, \tag{18}$$

where $j^\mu = (\rho, j)$, which in our convention already has the electric charge factored in. Other than the presentation, most of what we have done here is elementary; we are just using these results to set the stage for the later sections.

Now, we can rewrite the action in a few suggestive ways. First we can define a new 2-vector,

$$b_\mu = \frac{\pi}{e}(j,\rho), \tag{19}$$

such that the action can be rewritten

$$S[A_\mu] = \frac{e}{\pi}\int dx dt \epsilon^{\mu\nu} b_\mu A_\nu. \tag{20}$$

This is to be compared with Eq. (6). Alternatively, we can define an axionlike field,

$$\begin{aligned}\theta(x,t) &\equiv 2b_\mu x^\mu = \frac{2\pi}{e}(\rho x - jt)\\ &= 2k_F x - \frac{4\alpha}{\hbar}\sin(k_F a)\sin[\Phi(t)a]t,\end{aligned} \tag{21}$$

and if the system is homogeneous with no boundaries, we can use $\theta(x,t)$ to rewrite Eq. (18) as

$$S[A_\mu,\theta] = -\frac{e}{4\pi}\int dx dt \theta(x,t)\epsilon^{\mu\nu}F_{\mu\nu}. \tag{22}$$

As mentioned in Sec. I, using $\theta(x,t)$ breaks space-time translation symmetry due to the arbitrary choice of origin, and thus we must be careful to specify that the system is translation invariant when writing down Eq. (22), otherwise spurious response terms will be generated at boundaries and interfaces. Physically, we can interpret $\frac{e\theta}{2\pi}$ as the charge polarization since its space and time derivatives are proportional to the charge density and current, respectively.

While this method of generating an electric current came from an external effect, i.e., an externally applied electric field, we now move on to a discussion of an intrinsic effect that can produce a current in the absence of an external electric field.

### B. 1D model with next-nearest-neighbor (NNN) hopping

In this section we illustrate another way to generate a nonvanishing current. For energies near the Fermi points, the dispersion of our model is linear, and the modes near each Fermi-point are $(1+1)$-dimensional chiral fermions. In fact, it is well known that there is a close connection between the physical electric current for a 1D metallic band in an electric field and the compensating chiral anomalies of the fermion modes near each Fermi point. The previous section explicitly dealt with these issues, albeit using a less elegant perspective, and in that case an electric current was generated by an external source of Lorentz breaking, i.e., the applied electric field. Here we would like to consider an intrinsic source of Lorentz breaking that will lead to a current as well. By considering this effect, we are trying to make an analogy to the 3D CME in Weyl semimetals, where it has been predicted that a current can appear in the presence of an applied magnetic field, but in the absence of an electric field.

The basic idea is that, for the 1D model we have chosen, the chiral fermions near the Fermi points both have the same velocity, except for the sign, and we want to deform the velocities so that each chiral fermion has a different "speed of light." This is an obvious way to break Lorentz invariance. If the velocities are different (and the spectra were linear for all energies), then it is clear that we should have $E(-k_F) = v_L k_F \neq v_R k_F = E(k_F)$, which suggests the presence of a current. Physically, this just means that if we have $(1+1)$-dimensional chiral fermions with the same nonzero density, but different velocities, then there will be a nonvanishing current. Since we are in 1D, the analog of the 3D CME predicts that we should find an intrinsic current without the application of a magnetic field or an electric field, and it should be proportional to the intrinsic quantity $b_0$. In the 3D Weyl semimetal, the number $b_0$ represents the energy difference between Weyl nodes and has units of frequency. A simple interpretation of the effect seen here in 1D is that a nonvanishing frequency scale $b_0$ will be generated by the combination of $\Delta v_F$, i.e., the velocity difference at the two Fermi points, and a length scale. In our system we have two important length scales: the lattice constant $a$ and the inverse of the Fermi-wave vector $k_F$. To see which one enters the result, we perform an explicit calculation.

To generate the velocity modification effect, we deform the tight-binding model in Eq. (8) above to include imaginary next-nearest-neighbor hopping terms,

$$H_{1Dv} = H_{1D} + i\beta \sum_n [c^\dagger_{n+2} c_n - c^\dagger_n c_{n+2}]. \tag{23}$$

The Fourier transform of the Hamiltonian is given by

$$H_{1Dv} = -2 \sum_k [\alpha \cos(ka) - \beta \sin(2ka)] c^\dagger_k c_k. \tag{24}$$

For $\beta \neq 0$, inversion symmetry is broken in the model and subsequently we should consider two Fermi wave vectors $k_{FL}$ and $k_{FR}$, where $k_{FL} \leqslant k_{FR}$ by definition. Exactly at half-filling $k_{FL} = -k_{FR} = \pi/2a$ for all $\beta$ (as shown in Fig. 3). Thus, the electric current is vanishing at half filling (since $\beta \sin[2a(\pi/2a)] = 0$), and the charge density will be $\rho = \frac{e}{2a}$; i.e., the same as was found when no electric field was applied to the model $H_{1D}$ at half filling.

Half filling is just a special point of this model where $\beta$ has no effect because of our choice of next-nearest-neighbor hopping. Instead, let us consider the case where $\mu$ is tuned slightly away from half filling, i.e., $\mu = 0 - \delta\mu$ with $|\delta\mu| \ll \alpha$, and we also take $|\beta| \ll \alpha$ as we want to consider the perturbative effect of turning on this term. We can define $k_{FL} = -\frac{\pi}{2a} + \epsilon_L$ and $k_{FR} = \frac{\pi}{2a} + \epsilon_R$. By expanding Eq. (24) around the Fermi points, we find that consistency requires

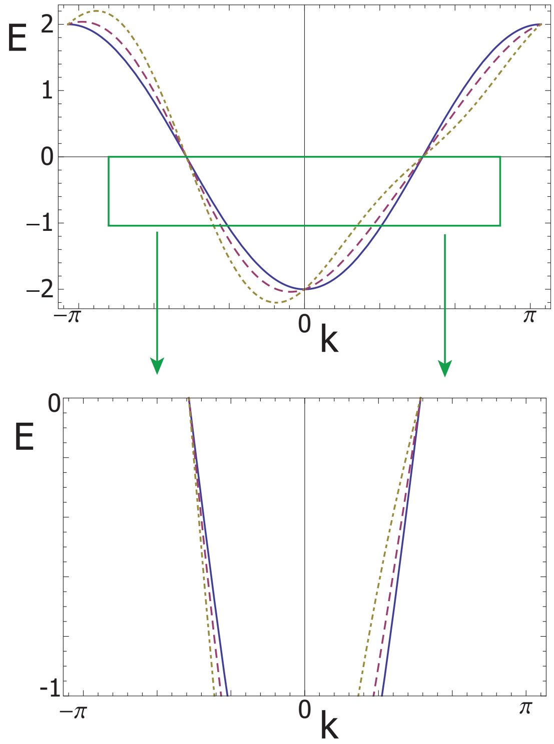

FIG. 3. (Color online) (Top) Energy spectrum of the Hamiltonian $H_{1Dv}$, where each curve represents a different value of $\beta$. The solid blue line is $\beta = 0$, the magenta dashed line is $\beta = 0.1$, and the dash-dotted tan line is $\beta = 0.25$. All curves have $\alpha = 1$. (Bottom) This is a magnified region of the top figure slightly below half-filling, which is the regime for our calculation. Exactly at half filling $\beta$ has no effect, and the stronger $\beta$ is, the more the Fermi-wave vectors and velocities are modified at a fixed $\mu$.

$$\epsilon_{L/R} = \pm \frac{\delta\mu}{2a(\alpha \pm 2\beta)} \approx \pm \frac{1}{2a}\frac{\delta\mu}{\alpha}\left[1 \mp \frac{2\beta}{\alpha}\right]. \tag{25}$$

Thus, we can determine that

$$k_{FL/R} \approx \frac{\pi}{2a}\left[\mp 1 \pm \frac{\delta\mu}{\pi\alpha}\left(1 \mp \frac{2\beta}{\alpha}\right)\right] \tag{26}$$

and can subsequently define $\kappa_F \equiv \frac{\pi}{2a}(1 - \delta\mu/\pi\alpha)$, which would be the Fermi wave vector if $\beta = 0$. Note that the signs in the previous two equations are correlated. From Fig. 3 we can see that as $\beta$ is increased the Fermi-wave vector at a fixed $\mu$ (different than half-filling) changes, as well as the velocity of the low-energy fermions. From Eq. (14), the response should be

$$\rho = e\frac{k_{FR} - k_{FL}}{2\pi} = \frac{e\kappa_F}{\pi} = \frac{e}{2a}\left(1 - \frac{\delta\mu}{\pi\alpha}\right), \tag{27}$$

$$j = \frac{2e\beta}{\pi\hbar}\sin(2\kappa_F a). \tag{28}$$

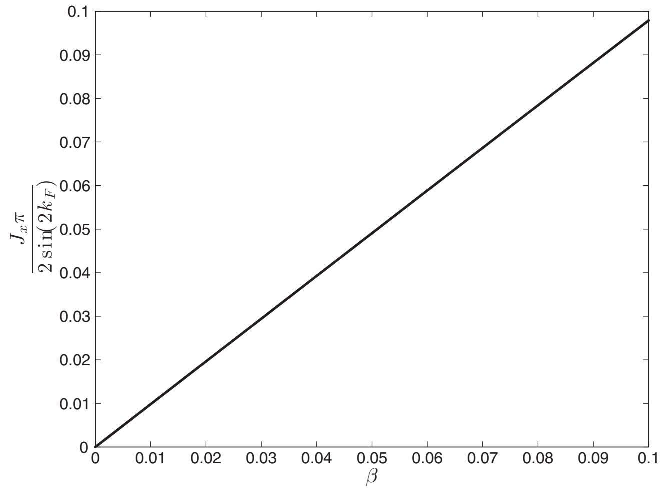

FIG. 4. The current of $H_{1\mathrm{D}v}$ is plotted vs next-nearest-neighbor hopping strength $\beta$ near half filling. $\kappa_F = \pi/2a - \pi/100a$ was chosen and the nearest-neighbor hopping $\alpha = 1$ with periodic boundary conditions. The current increases linearly as a function of $\beta$, as expected from Eq. (28). Note that we have let $a = \hbar = 1$.

This result shows that we find a nonzero electric current even in the absence of an applied electric field, and its magnitude is proportional to the inversion-breaking parameter $\beta$. This effect, while simple in origin, is the 1D analog of the 3D CME. It represents a current proportional to an intrinsic frequency scale, but does not require the application of any external electric or magnetic fields. We do note that the definition of the frequency scale does require a nonvanishing Fermi wave vector, i.e., a nonvanishing background density which cannot arise from a completely empty or filled band. As shown in Fig. 4, the numerical calculation of the electric current matches the analytic formula. The response is linear in $\beta$, as expected from Eq. (28) and, although we do not show the charge density, it matches as well. The numerical calculations were done for slightly less than half filling at $\kappa_F = \pi/2a - \pi/100a$.

Let us take a closer look at the generation of the electric current. The velocity of the chiral fermions at $\pm\kappa_F$ is given by $\hbar v_\pm = \pm[2\alpha a \sin(\kappa_F a) \mp 4\beta a \cos(2\kappa_F a)]$, and, thus,

$$\Delta v_F = \frac{8\beta a}{\hbar}\cos(2\kappa_F a). \tag{29}$$

For our choice of the chemical potential, $\kappa_F = \pi/2a + \delta\kappa_F$, and the current from Eq. (28) is approximately

$$j \approx -\frac{8e\beta}{2\pi\hbar}\delta\kappa_F a = -\frac{e}{2\pi}\frac{8\beta a}{\hbar}\delta\kappa_F = \frac{e}{2\pi}\Delta v_F\delta\kappa_F, \tag{30}$$

where we used that near $\kappa_F = \pi/2a$ we have $\Delta v_F \approx -\frac{8\beta a}{\hbar}$. Thus, $\Delta v_F\delta\kappa_F$ gives a Lorentz-breaking frequency scale that will give rise to a nonvanishing $b_0$ term in the effective response. In fact, the density and current give us the 2-vector $b_\mu = (\frac{1}{2}\Delta v_F\delta\kappa_F, \kappa_F)$, which determines the response action

$$S[A_\mu] = \frac{e}{\pi}\int dx dt \epsilon^{\mu\nu} b_\mu A_\nu. \tag{31}$$

To draw an analogy with the previous literature on the Weyl semimetal response, we could also define a $\theta(x,t)$ by

$$\begin{aligned}\theta(x,t) &= \frac{2\pi}{e}(\rho x - jt) = 2\kappa_F x - \frac{4\beta}{\hbar}\sin(2\kappa_F a)t\\ &\approx 2\kappa_F x - \Delta v_F\delta\kappa_F t,\end{aligned} \tag{32}$$

which couples into the action

$$S[A_\mu,\theta] = -\frac{e}{4\pi}\int dx dt \theta(x,t)\epsilon^{\mu\nu}F_{\mu\nu}. \tag{33}$$

### C. Derivation of the effective response

After our explicit discussion of the different EM responses of the 1D metallic wire, let us elevate our discussion to a field-theoretic calculation. In this section, we use the Fujikawa method to derive the effective response of the low-energy continuum field theory description of the 1D metal in the presence of intrinsic sources of Lorentz invariance violation, e.g., external EM fields, momentum and velocity shifts of the nodes, and nonzero chemical potential. The derivation is similar to that for 3D Weyl semimetals found in Ref. [54].

To carry out the calculation, let us expand the lattice Bloch Hamiltonian given by $H(k) = -2\alpha\cos ka + 2\beta\sin 2ka$ around the chemical potential $\mu = 0 - \delta\mu$ for $|\delta\mu| \ll \alpha$ and $|\beta| \ll \alpha$, as in the previous section. If we expand the right- and left-handed chiral branches around $\pm\kappa_F$, respectively, we find the approximate continuum Hamiltonian

$$H_{\mathrm{cont}} = \left(-\delta\mu + \tfrac{1}{2}\hbar\Delta v_F q\right)\mathbb{I} + \left(\hbar v_F q + \tfrac{1}{2}\hbar\Delta v_F\delta\kappa_F\right)\sigma^z, \tag{34}$$

where the upper component represents the fermions near $k_{FR}$, the lower component represents the fermions near $k_{FL}$, $q$ represents a small wave-vector deviation from $k_{FL/R}$, $\hbar v_F \equiv 2a\alpha$, $\Delta v_F \equiv -\frac{8\beta a}{\hbar}$, and $\delta\kappa_F = -\frac{\delta\mu}{2a\alpha}$. The definitions of the parameters are easy to understand by looking at the lattice model in the previous section when expanded around $\kappa_F$.

Since we know the behavior of the full lattice model, i.e., the high-energy regularization of the continuum model, we can see that our expansion effectively normal orders the current and density with respect to half filling. Since the current vanishes exactly at half filling, the total current is simply $j = \delta j$. The current change away from half filling is simply given by $\delta j = \frac{e}{2\pi\hbar}[E_R(q=0) - E_L(q=0)] = \frac{e}{2\pi}\Delta v_F\delta\kappa_F$, which matches that given in the previous section, which does not include the effective normal ordering. On the other hand, the charge density does not vanish at half filling. The density change away from half filling is given by $\delta\rho = e\frac{-\delta\mu}{2\pi\alpha a} = e\frac{\delta\kappa_F}{\pi}$, and the full density includes the additional amount $\rho_0 = \frac{e}{2a}$ that arises from all the occupied states up to half filling. This makes the total density $\rho = \rho_0 + \delta\rho = e\frac{k_{FR}-k_{FL}}{\pi}$, as expected. However, if we are just given the continuum model, without reference to an initial lattice model, it only has information about $\delta\rho$ and $\delta j$. We note that neither the current nor the density depend on the dispersion term $\frac{1}{2}\hbar\Delta v_F q\mathbb{I}$ and so we drop it from further discussion as it is also higher order in the expansion around the Fermi points.

From this Hamiltonian it is simple to construct the Lagrangian now using the Dirac matrices $\gamma^0 = i\sigma^x, \gamma^1 = \sigma^y$

and the chirality matrix $\gamma^3 = \sigma^z$. We find

$$\mathcal{L} = \overline{\psi}(i\not\partial - \not b\gamma^3)\psi, \tag{35}$$

where $\not b = b_\mu\gamma^\mu$ for $b_\mu = (\frac{1}{2}\Delta v_F \delta\kappa_F, \delta\kappa_F)$. If we included the EM gauge field, this Lagrangian would be analogous to the Lagrangian derived for the Weyl semimetal in Ref. [54], except this is in 1 + 1 dimensions. We can now get rid of the $b_\mu$-dependent term by doing a chiral gauge transformation. As is well known, this transformation can change the measure of the path integral and lead to anomalous terms in the effective action.

We use the Fujikawa method to derive the effective response due to this change of measure. Performing a series of infinitesimal chiral transformations parametrized by the infinitesimal $ds$, we can get rid of the $b_\mu$-dependent term,

$$\psi \to e^{-ids\theta(x)\gamma^3/2}\psi, \tag{36}$$

$$\overline{\psi} \to \overline{\psi}e^{-ids\theta(x)\gamma^3/2}, \tag{37}$$

where $\theta(x) \equiv 2b_\mu x^\mu$. Note that using this choice of $\theta(x)$ we have made an arbitrary choice of origin which is folded into the calculation. To avoid spurious response terms, we need to constrain the system to be homogeneous in space-time so that each choice of space-time origin is equivalent. The Dirac operator $\not D$ acts as

$$\not D = i\not\partial - \not A - \not b\gamma^3(1-s), \tag{38}$$

$$\not D\phi_n(x) = \epsilon_n\phi_n(x), \tag{39}$$

where $A_\mu$ is the EM gauge field, and $\phi_n$ are a complete set of eigenstates of the Dirac operator. Let us write

$$\psi(x) = \sum_n c_n\phi_n(x), \quad \overline{\psi}(x) = \sum_n \overline{c}_n\phi_n^*(x), \tag{40}$$

where $c_n$ are Grassman variables, and we can expand $\psi$ in terms of $\phi_n$ because they are complete. Considering what the infinitesimal chiral transformation does to the $c_n$'s, from Eq. (36), we see that

$$c'_n = \sum_m U_{nm}c_m, \quad \overline{c}'_n = \sum_m U_{nm}\overline{c}_m, \tag{41}$$

$$U_{nm} = \delta_{nm} - \frac{ids}{2}\int d^2x\,\phi_n^*(x)\theta(x)\gamma^3\phi_m(x). \tag{42}$$

The Jacobian of this transformation is $J = \det(U^{-2})$. Using the identity that $\det(U) = e^{\mathrm{Tr}\,\log(U)}$, we see that

$$J = e^{ids\sum_n\int d^2x\,\phi_n^*(x)\theta(x)\gamma^3\phi_n(x)}. \tag{43}$$

The Jacobian due to the chiral rotation thus induces a term in the effective action given by

$$S_{\mathrm{eff}} = \int_0^1 ds\int dx\,dt\,\theta(x)I(x), \tag{44}$$

$$I(x) = \sum_n \phi_n^*(x)\gamma^3\phi_n(x). \tag{45}$$

To evaluate $I(x)$, we can use the heat kernel regularization,

$$I(x) = \lim_{M\to\infty}\sum_n \phi_n^*(x)\gamma^3 e^{-\not D^2/M^2}\phi_n(x), \tag{46}$$

to arrive at the well-known result that

$$I(x) = -\frac{e}{4\pi}\epsilon^{\mu\nu}F_{\mu\nu}. \tag{47}$$

So, the effective action is given by

$$S_{\mathrm{eff}}[A_\mu] = -\frac{e}{4\pi}\int d^2x\,\theta(x)\epsilon^{\mu\nu}F_{\mu\nu}. \tag{48}$$

To remove the dependence on the arbitrary origin, we can rewrite the action as

$$S_{\mathrm{eff}}[A_\mu] = \frac{e}{\pi}\int \epsilon^{\mu\nu}b_\mu A_\nu. \tag{49}$$

This expression matches the result we determined from simpler calculations of the lattice model in Secs. II A and II B if we replace $\rho$ with $\delta\rho$ and $j$ with $\delta j$.

### D. Interfaces

Now that we have derived the EM response via two separate methods, we put it to use in this section where we calculate the properties of interfaces across which $b_\mu$ varies. We show that the response action in Eq. (49) predicts results that match numerical simulations, while the $\theta$-term version in Eq. (48) gives spurious results due to boundary terms that depend on the arbitrary choice of origin embedded in $\theta(x)$. We want to emphasize that this also happens in the case of the 3D Weyl semimetal and is a generic feature. One might think that one could remove these spurious terms by adding boundary degrees of freedom; however, the spurious results to which we refer do not seem to be connected to any anomalies as they can appear on surfaces which do not exhibit gapless boundary modes.

The form of the action to use when studying inhomogeneous systems (i.e., with relaxed translation invariance) is

$$S[A] = \frac{e}{\pi}\int d^2x\,\epsilon^{\mu\nu}b_\mu A_\nu.$$

One might complain that this action appears gauge variant; however, it is not. We note that we can define a current $j^\mu_{(b)} = \frac{e}{\pi}\epsilon^{\mu\nu}b_\nu$. Therefore, the action itself can be written $S = \int d^2x\,j^\mu_{(b)}A_\mu$. If the current is conserved, then the action is gauge invariant due to the continuity equation. For the 1D metal, the current $j^\mu_{(b)}$ is exactly the EM charge current and thus is conserved, yielding a gauge-invariant response functional.

Now, for the first example of an interface, suppose our 1D metal lies in the spatial region $x > x_0$, and there is only vacuum for $x < x_0$. We model this by choosing $b_\mu(x) = b_\mu\Theta(x - x_0)$, where $\Theta(x)$ is the step function, and for simplicity we only turn on a nonvanishing $b_1$. If we look at the charge density the response action would predict, we find

$$\rho(x) = \frac{e}{\pi}b_1\Theta(x - x_0), \tag{50}$$

which is physically correct since the metallic region will have a density equal to this value, and the vacuum will have no density. If we had used the axion action with $\theta(x,t) = 2b_1(x - x_1)$ for some arbitrary constant value $x_1$, we would have obtained the density

$$\begin{aligned}\bar{\rho}(x) &= \frac{e}{2\pi}\partial_x\theta(x,t) = \frac{e}{\pi}b_1\partial_x[(x-x_1)\Theta(x-x_0)]\\ &= \frac{eb_1}{\pi}[(x_0-x_1)\delta(x-x_0)+\Theta(x-x_0)]. \qquad (51)\end{aligned}$$

This predicts a spurious boundary charge located at the interface $x_0$ and proportional to the distance between the boundary point and our *arbitrary* choice of $x_1$. This term is clearly unphysical, and simulations show that there is nothing special happening at the interface. Thus, the first action reproduces the correct response and matches numerics for the one-band lattice metal.

For a more complicated illustration, consider an interface between two different systems such that $b_1$ is nonvanishing in both and varies in the $x$ direction. This will give an $x$-dependent charge density. A simple way to implement an $x$-dependent $b_1$ is to introduce an on-site energy term which is $x$ dependent. If we had a translationally invariant 1D lattice model with a fixed chemical potential $\mu$, then shifting the on-site energy up or down will decrease or increase the electron density, respectively. Let us consider two 1D segments which have a common boundary. Suppose the on-site energies are constant within each region, but are offset between the two regions by $\epsilon_0$. To simplify the description we assume that they are glued periodically so, in fact, there are two interfaces.

For an analytically tractable limit, let us study the case when the offset is not too big when compared to the bandwidth of the system and with the chemical potential fixed at $\mu = 0$. The Hamiltonian is given by

$$H = -t\sum_n[c^\dagger_{n+1}c_n + c^\dagger_n c_{n+1}] + \sum_n \epsilon(n)c^\dagger_n c_n, \qquad (52)$$

where $\epsilon(n) = \pm\epsilon_0/2$, when $n \leqslant N/2$ or $n > N/2$, respectively, for a system with an even number of sites $N$. We want to understand what happens to the charge density in the system and compare it to what is predicted by the EM response action. With this Hamiltonian the system consists of two segments (labeled by $\ell$ and $r$), each of length $L_s = Na/2$, where $a$ is the lattice constant.

We can now compute what $b_{1(\ell)}$ and $b_{1(r)}$ are for each segment since there is a simple relation between charge density and $b_1$. As the length of the segments approaches the thermodynamic limit, the average charge density will not depend on whether we calculate it with open or periodic boundary conditions, so for simplicity we can calculate the density with periodic boundary conditions for each segment separately. With $\mu = 0$ fixed, the Fermi momentum for the segment $\ell$ with the offset $+\epsilon_0/2$ is given by

$$0 = \epsilon_0/2 - 2t\cos(k_{F,\ell}a) \Rightarrow k_{F,\ell} = \frac{1}{a}\cos^{-1}\left(\frac{\epsilon_0}{4t}\right). \qquad (53)$$

The Fermi momentum for system $r$ is given by

$$0 = -\epsilon_0/2 - 2t\cos(k_{F,r}a) \Rightarrow k_{F,r} = \frac{1}{a}\cos^{-1}\left(\frac{-\epsilon_0}{4t}\right). \qquad (54)$$

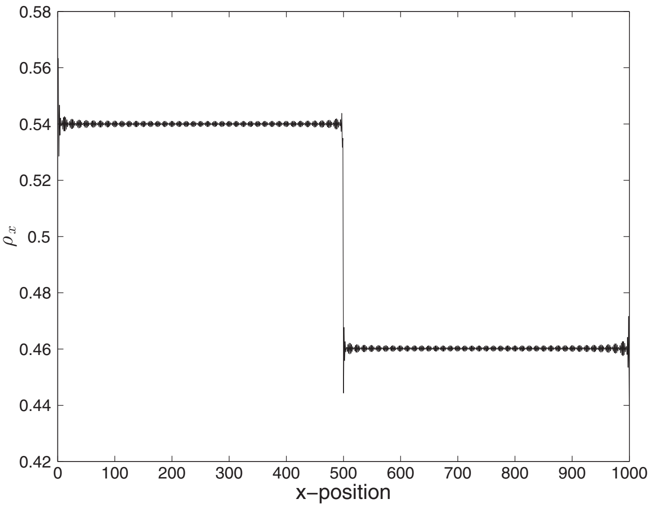

FIG. 5. Charge density in units of $e/a$ as a function of position for an inhomogeneous system with $N = 1000$ lattice sites, where each segment has $L_s = 500$ sites. The chemical potential is $\mu = 0$, and if $\epsilon_0$ was tuned to zero the density would be $\rho = e/2a$. For our choice of $\epsilon_0 = 0.5t$ we have $b_{1(\ell)} = (\pi/a)0.46$, $b_{1(r)} = (\pi/a)0.54$. Away from the interfaces the values match the calculation from the effective response action. Near the interfaces there are damped oscillations due to finite-size effects that are not captured by the analytic calculation. Note that the finite-size boundary effects have nothing to do with the spurious "interface" terms in Eq. (51).

So we have $\rho_r = e\frac{k_{F,r}}{\pi}$ and $\rho_\ell = e\frac{k_{F,\ell}}{\pi}$, which by definition implies that $b_{1(r/\ell)} = (\pi/e)\rho_{r/\ell} = k_{Fr/\ell}$. Explicitly, we have

$$b_{1(\ell)} = \frac{1}{a}\cos^{-1}\left(\frac{\epsilon_0}{4t}\right), \qquad (55)$$

$$b_{1(r)} = \frac{1}{a}\cos^{-1}\left(\frac{-\epsilon_0}{4t}\right). \qquad (56)$$

In our geometry we have interfaces at $x = L_s$ and $x = 2L_s \equiv 0$ and $b_1$ varies across the interfaces. The EM response action predicts

$$\begin{aligned}\rho = \frac{e}{\pi}\{&b_{1(\ell)}[\theta(x) - \theta(x - L_s)]\\ &+ b_{1(r)}[\theta(x - L_s) - \theta(x - 2L_s)]\}. \qquad (57)\end{aligned}$$

This result matches what is found numerically, as shown in Fig. 5.

### E. General comments

Before we move on to discuss the more interesting higher-dimensional semimetals, we pause to make a few important comments.

*(i) Response action without translation invariance.* Initially, we parametrized the EM response of the 1D metal through quantities such as the Fermi-wave vector, and the velocity at the Fermi points, which can only be clearly defined when there is translation symmetry. That is, when the system is homogeneous we can precisely define momentum space and these two quantities. What we have found is that the response is actually more general because we can define it in terms of the sources of Lorentz violation [67], i.e., an intrinsic charge

density and charge current. These two physical quantities can be defined, and measured, without reference to momentum space and thus we can drop all reference to a Fermi-wave vector and a Fermi velocity by using the density and current, respectively. The fact that the EM response is accurate even without translation invariance is clearly shown when we have an interface as shown in the previous section.

This physical definition of the response is special to 1D because the semimetal EM response action is just $\int d^2x j^\mu A_\mu$. This type of term will appear in every dimension, but in higher dimensions there are more interesting anisotropic response terms that appear and which we discuss later. For $d$-dimensional space-time we can introduce a $(d-1)$ form $b_{\mu_1\mu_2\cdots\mu_{d-1}}$ representing a source of Lorentz breaking. We can furthermore take the dual to generate a current $j^\mu_{(b)} = \epsilon^{\mu\mu_1\cdots\mu_{d-1}} b_{\mu_1\mu_2\cdots\mu_{d-1}}$, which represents an intrinsic charge density or charge current which couples to $A_\mu$ minimally. This term yields the higher-dimensional analog of the 1D semimetal EM response. We comment later on the possibility to represent higher-dimensional response actions without reference to momentum space.

*(ii) Response of filled bands.* As is well-known from elementary solid-state physics, a filled band of electrons in a crystal carries no current. Each filled band also contributes a charge density $\rho_{\text{band}} = \frac{e}{a}$ or $e/\Omega$, where $\Omega$ is the size of a unit cell in higher dimensions. The EM response actions of topological semimetals do not capture density or current contributions from filled bands and thus the response coefficients are ambiguous by a finite quantized amount; i.e., $b_\mu$ is ambiguous by the addition of half of a reciprocal lattice vector.

*(iii) Symmetries of $b_\mu$ in 1D.* Let us discuss the transformation properties of $b_\mu$ under time reversal ($T$), charge conjugation ($C$), and inversion symmetry ($P$). Since in 1D we know that $b_0$ is proportional to a current and $b_1$ is proportional to a density we can easily determine their symmetry properties,

$$T: b_0 \to -b_0, \quad C: b_0 \to b_0, \quad P: b_0 \to -b_0, \tag{58}$$

and

$$T: b_1 \to b_1, \quad C: b_1 \to b_1, \quad P: b_1 \to b_1. \tag{59}$$

Note that they are both even under $C$, which is due to the fact that our convention for $b_\mu$ defined in terms of the density and current has the electric charge factored out. Subsequently, the response actions will have factors of electric charge in their normalization coefficients. Note that these symmetry properties only hold in 1D because the transformation properties of $b_\mu$ under these discrete symmetries are dimension dependent.

*(iv) Connection between 1D and 3D semimetals.* As mentioned in Sec. I, the effective response for a 3D Weyl semimetal is

$$S[A_\mu] = -\frac{e^2}{2\pi h}\int d^4x \epsilon^{\mu\nu\rho\sigma} b_\mu A_\nu \partial_\rho A_\sigma.$$

To be explicit, consider a system where $b_\mu = (b_0,0,0,b_z)$ in the presence of a uniform magnetic field $F_{xy} = -B_0$. In this case the action reduces to

$$\frac{e^2\Phi}{\pi h}\int dtdz \epsilon^{ab} b_a A_b = N_\Phi \frac{e}{\pi}\int dtdz \epsilon^{ab} b_a A_b, \tag{60}$$

where $\Phi = -B_0 L_x L_y$ is the magnetic flux and $a,b = 0,z$. From this we see that the 3D action, for this arrangement of $b_\mu$ and $F_{\mu\nu}$, reduces to $N_\Phi = |\Phi/(h/e)|$ copies of the 1D action. This connection hints that it could be possible to define the response of the 3D Weyl semimetal without reference to momentum space and instead only using physical quantities, e.g., the charge density and current in a uniform magnetic field. It also shows why the symmetry transformation properties of $b_\mu$ in 1D are different than those of $b_\mu$ in 3D because of the additional factor of $\Phi$ in 3D which is odd under time reversal. We discuss this more in the section on 3D semimetals.

## III. DIRAC SEMIMETAL IN (2 + 1) DIMENSIONS

After our discussion of the simple one-band metal, we now move on to a discussion of the 2D Dirac semimetal that has become widely recognized with the experimental discovery of graphene [41]. Graphene is a honeycomb lattice of carbon atoms with a low-energy electronic structure consisting of four Dirac points. These four Dirac points are located in spin-degenerate pairs at the special points $K$ and $K'$ in the hexagonal BZ. For models like graphene, with both time-reversal and inversion symmetry, the minimum number of Dirac points that can appear in a 2D lattice model is two. Graphene has twice this amount because of the spin-$\frac{1}{2}$ degeneracy of the electrons due to the time-reversal symmetry with $T^2 = -1$. For our purposes, we focus on a reduced case of spinless (or spin polarized) electrons for which (effectively) $T^2 = +1$. To recover results for graphene, one could trivially add in the degenerate spin degree of freedom. Later in this section we discuss a general (even) number of Dirac nodes, but we always assume they are nondegenerate for simplicity.

This section is organized as follows. (i) We first discuss the construction of WTIs and, subsequently, Dirac semimetals from wire arrays of 1D TIs; (ii) using the connection to the weak TI state we conjecture a form for the EM response of a Dirac semimetal with two nodes, discuss the required symmetries for the robustness of this response, and show that a simple model yields the predicted physical properties; (iii) we derive the conjectured quasitopological response effective action in the continuum limit using two Dirac nodes; (iv) we provide a physical interpretation of the response action in terms of known EM quantities, and we discuss the general measurable properties; (v) we discuss the generalization of the continuum calculation to lattice models and an arbitrary (even) number of Dirac nodes; (vi) finally, we make some general comments on the nature of the quasitopological response, the similarities and differences between responses in even and odd space-time dimensions, and connect the results to the properties of Chern insulator with nonzero charge polarization.

### A. Dirac semimetal from layered topological insulators

#### *1. Topological insulator in 1D protected by C or P symmetry*

As discussed in Sec. I, each TSM can be constructed from a collection of lower-dimensional TIs which are stacked and then coupled; the DSM is no different. To generate a DSM this way, we must begin with 1D TI wires. From the classification of 1D TIs we know that to have a robust, nontrivial 1D topological phase we must require the presence of a symmetry to protect the state [2–4]. This is inherently different than the 3D Weyl semimetal, which is constructed from stacks of 2D Chern insulators that require no symmetry to have a protected topological phase [75]. There are two possibilities for an appropriate 1D TI symmetry: (i) charge-conjugation symmetry ($C$) or (ii) inversion/reflection symmetry ($P$). For $C$ symmetry the 1D topological wire lies in class D of the Altland-Zirnbauer classification [2,3,76], and there is a $\mathbb{Z}_2$ topological invariant that controls the EM response. While, in principle, there is no problem with considering insulators with $C$ symmetry, in practice, such a symmetry is approximate and/or fine tuned. For $P$ symmetry the wire belongs to the set of inversion-symmetric insulators and also has a $\mathbb{Z}_2$ topological invariant [8,9,77,78]. In both cases we call the invariant $\mathcal{Z}_{1\mathrm{D}}$. If $\mathcal{Z}_{1\mathrm{D}}$ takes its trivial (nontrivial) value $\mathcal{Z}_{1\mathrm{D}} = +1(\mathcal{Z}_{1\mathrm{D}} = -1)$, then the insulator will have a bulk charge polarization of $P_1 = ne \bmod \mathbb{Z}e [P_1 = (n + 1/2)e \bmod \mathbb{Z}e]$ and will exhibit an even (odd) number of low-energy fermion bound states on each boundary point. Let us note that we use $P$ to label reflection symmetries (inverting a single coordinate) and $\mathcal{I}$ to represent inversion symmetry (reflection in all coordinates). Of course, in 1D they are the same, so we simply use $P$ for 1D systems.

Since it will become important, let us review the EM response of the 1D TI. The response is captured by the effective action

$$S_{\mathrm{eff}}[A_\mu] = \frac{1}{2}\int d^2x\, P_1 \epsilon^{\mu\nu} F_{\mu\nu}, \tag{61}$$

where $P_1$ depends on the insulating phase as given above. The requirement of either $C$ or $P$ symmetry enforces a quantization of the polarization in units of half an electron charge [3,8,9,77]. Naively, these symmetries should forbid a nonzero $P_1$ since $P_1 \to -P_1$ under $C$ or under $P$. However, since the polarization in 1D crystalline insulators is only well defined modulo integer charge, the allowed values of $P_1$ are 0 and $e/2$, which both satisfy $P_1 = -P_1$ modulo integer electron charges [3,79]. Another way to say this is that 1D insulators with polarizations that differ by an integer electron charge are topologically equivalent (or stable topologically equivalent).

It will be very useful to have an explicit system in mind when discussing the features of the 1D TI and the subsequent weak TI and 2D Dirac semimetal generated by stacking the 1D TIs. Thus, let us choose a simple model which exhibits a 1D TI phase: the 1D lattice Dirac model. For translationally invariant systems, this model has a Bloch Hamiltonian

$$H_{1\mathrm{DTI}}(k) = (A \sin ka)\sigma^y + (B - m - B\cos ka)\sigma^z, \tag{62}$$

where $A$, $B$, and $m$ are model parameters (we set $A = B = 1$ from now on), $a$ is the lattice constant, and $\sigma^\alpha$ are the Pauli matrices representing some degrees of freedom within the unit cell. The phases of this model are controlled by the parameter $m$, and for $m < 0$, or $m > 2$, the system is a trivial insulator with $\mathcal{Z}_{1\mathrm{D}} = +1$. For $0 < m < 2$ the system is in a TI phase with $\mathcal{Z}_{1\mathrm{D}} = -1$. A benefit of this model is that we can judiciously choose a $C$ operator and a $P$ operator such that the Hamiltonian has that symmetry. For example, if we pick $C = \sigma^y$, then $CH_{1\mathrm{DTI}}(k)C^{-1} = -H^*_{1\mathrm{DTI}}(-k)$, and if we pick $P = \sigma^z$, then $PH_{1\mathrm{DTI}}(k)P^{-1} = H_{1\mathrm{DTI}}(-k)$. So, as written, this model is simple enough to have both $C$ and $P$ symmetry and thus can exhibit a protected topological phase. If we add perturbations to the model that break one of the symmetries, but preserve the other, then the topological phase will remain stable. It is only if we break both symmetries that we can destabilize the 1D TI phase.

Usually, for insulators, a $C$ symmetry only exists when the model is fine tuned, but inversion/reflection symmetry can be approximately preserved in real materials. In what follows we emphasize the inversion- or reflection-symmetric cases as it is more relevant when considering semimetal phases that might be realized in materials. We note that this model also has time-reversal symmetry with $T = K$ ($T^2 = +1$). Although this symmetry is not important for the 1D classification, it will become important when we discuss the 2D semimetal phase.

### 2. *Weak topological insulator in 2D protected by $C$, $P$, or $\mathcal{I}$ symmetry*

Before we approach the DSM, let us consider the 2D WTI phase generated by stacking a weakly coupled set of 1D TI wires. To be explicit, suppose that the wires are oriented parallel to the $x$ axis and stacked perpendicularly to spread into the $y$ direction. In the limit of decoupled wires, we can determine that the system will have a charge polarization in the $x$ direction, and will have low-energy boundary states on boundaries with a normal vector in the $x$ direction (or, in general, on boundaries not parallel to the $y$ axis). In this limit, a 2D Hamiltonian representing this phase is just multiple copies of $H_{1\mathrm{DTI}}$ with a fixed value of $0 < m < 2$ for each wire. These distinguishing topological characteristics remain as long as the coupling between the wires does not close the bulk gap, and as long as the relevant symmetries of the 1D TI are preserved.

We can model this using a square-lattice Bloch Hamiltonian,

$$\begin{aligned} &H_{2\mathrm{DWTI}}(\vec{k}) \\ &\quad = \sin(k_x a)\sigma^y + [1 - m - \cos(k_x a) - t_y \cos(k_y a)]\sigma^z, \end{aligned} \tag{63}$$

for a lattice constant $a$ and a new tunneling parameter $t_y$. Again, this model has both $C$ and $P_x$ symmetry (reflection with $x \to -x$), with the same operators as above, since the interwire tunneling term $-t_y \cos(k_y a)\sigma^z$ preserves both. It also has time-reversal symmetry $T = K$, reflection symmetry in the $y$ direction with $P_y = \mathbb{I}$, and inversion symmetry with $\mathcal{I} = \sigma^z$. If we pick $0 < m < 2$, then the model remains in the WTI phase as long as no solutions for at least one of

$$\cos(k_y a) = -\frac{m}{t_y}, \quad \cos(k_y a) = \frac{2 - m}{t_y}, \tag{64}$$

can be found. We immediately see that as long as $|t_y| < |m|$ and $|t_y| < |(2 - m)|$, then the system will be gapped, and if additionally $0 < m < 2$, the model will be in the WTI phase.

This WTI is characterized by a 2D topological vector invariant $\vec{\nu} = (0, \frac{\pi}{a})$, which is a half-reciprocal lattice vector. The EM response of the 2D WTI depends on this vector and is given by

$$S_{\mathrm{eff}}[A_\mu] = \frac{e}{4\pi}\int d^3x\, \nu_\mu \epsilon^{\mu\nu\rho} F_{\nu\rho}, \tag{65}$$

where $\nu_0 = 0$. This response represents the contribution of a charge polarization $\vec{P}_1$ to the action where

$P_1^i=\frac{e}{2\pi}\epsilon^{ij}\nu_j=(\frac{e}{2a},0)$. The magnitude of the polarization is due to a contribution of a 1D polarization of $e/2$ (and $e/2$ boundary charge) per wire, as expected, and the total charge on a boundary with normal vector $\hat{x}$ will be $N_y\frac{e}{2}$, where $N_y$ is the number of wire layers. As discussed in Sec. I, the WTI phase does not give rise to $\nu_0$ because there is no effective Lorentz breaking in the time direction for a filled band. One could generate a $\nu_0$ in an insulator by applying a time-dependent periodic field to generate Floquet dynamics, or perhaps by coupling the system to a varying adiabatic parameter that will drive cyclic adiabatic charge pumping [80]. For the latter, this will drive a constant, quantized current along the wires which will result in a nonvanishing $\nu_0$ proportional to the charge pumping frequency. We prove below that, just as $\nu_i$ is connected to the intrinsic charge polarization, $\nu_0$ is related to the intrinsic magnetization, which is why producing currents will generate such a term.

#### 3. *From 2D weak topological insulator to Dirac semimetal*

We now give an explicit example of a Dirac semimetal, and in Sec. III B we discuss its physical response properties and characteristics. We then move on to deriving the results for a generic Dirac semimetal in the subsequent sections.

It is easy to construct an explicit example of a DSM phase from the WTI model we have been using by choosing $m$ and $t_y$ such that at least one of Eq. (64) has a solution. To be concrete, let $m=1/2$, $t_y=-1$, and $a=1$, for which $\cos k_y=-m/t_y$ has two solutions, $\pm k_{yc}=\pm\pi/3$, which implies there are Dirac points at $\vec{k}=(0,\pm k_{yc})$. If we expand the Hamiltonian in Eq. (63) around these points, we find the continuum Hamiltonians,

$$H_{\text{2Dcon}}=\delta k_x\sigma^x\pm\frac{\sqrt{3}}{2}\delta k_y\sigma^z, \tag{66}$$

which are anisotropic Dirac points with $\delta k_x$ the deviation from $k_x=0$ and $\delta k_y$ the deviation from $k_y=\pm k_{yc}$. If we tuned the velocity parameter $A$ in Eq. (62) to be $\sqrt{3}/2$, we would find isotropic Dirac points. In Fig. 6 we show the energy spectrum of this model, at the parameter values given above, in a strip/cylinder geometry with open boundary conditions in the $x$ direction and periodic boundary conditions in the $y$ direction. We see the Dirac points at the predicted values and also a flat band of midgap states which are exponentially localized on the edges of the strip.

Despite some superficial differences, the square-lattice model for the DSM captures the same physics as the honeycomb-lattice graphene model. In fact, in Appendix A we show that our square-lattice model for the DSM can be continuously deformed to the honeycomb graphene model, and thus we can easily consider graphene to be constructed from layers of 1D TIs if we trivially add spin degeneracy. This matches the well-known result that graphene has anisotropic boundary states that appear only on zigzag edges and not armchair edges, which is a consequence of this layered structure, and the close connection to the WTI model of stacked 1D TIs [41].

### B. Motivation of quasitopological response of 2D Dirac semimetals

Following the general discussion in Sec. I, when the DSM is formed, we expect the quasitopological EM response to be dependent on the momentum and energy differences between the Dirac nodes. In this section we present a form of the quasitopological response that is analogous to Eq. (65) for the weak TI, and we provide physical evidence that our conjecture is correct. In Sec. III C we derive the result more systematically.

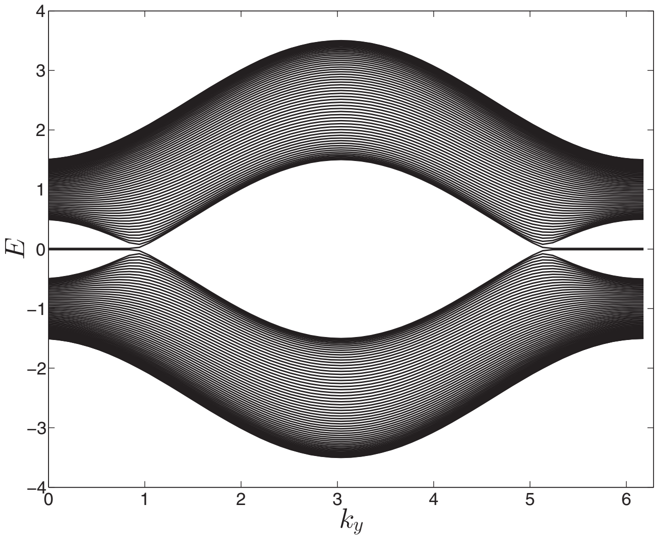

FIG. 6. The energy spectrum for the Hamiltonian in Eq. (63) tuned into the 2D Dirac semimetal. The figure shows exact diagonalization of this model in a strip geometry ($x$ direction with open boundaries and $y$ direction with periodic boundaries) with $\pm k_{yc}=\pm\pi/3$ and $b_y=\pi/3$. The flat band of states stretched between the Dirac nodes are edge modes.

For our explicit choice of parameters we should have a separation vector $b_\mu=(0,0,\pi/3)$. As we prove in Sec. III C, one contribution to the EM response is the analog of Eq. (65) for the WTI phase, that is,

$$S_{\text{eff}}[A_\mu]=\frac{e}{4\pi}\int d^3x\, b_\mu\epsilon^{\mu\nu\rho}F_{\nu\rho}. \tag{67}$$

From the interpretation of the 2D WTI response above, this implies a nonzero charge polarization,

$$P_1^i=-\frac{e}{2\pi}\epsilon^{ij}b_j. \tag{68}$$

We warn that when there are multiple pairs of nodes, one must be careful when constructing the value of $b_i$ that enters the response due to the $\mathbb{Z}_2$ nature of the polarization. We discuss this in detail in Sec. III E, but for now we continue analyzing the simplest case with only two nodes.

#### 1. *Charge polarization in a 2D Dirac semimetal*

Let us now try to understand the origin of the polarization. To illustrate this, we should heuristically view the DSM model Hamiltonian as representing a family of 1D insulators, parametrized by the values of $k_y$. That is, each value of $k_y$ (except $k_y=\pm k_{yc}$) represents a 1D insulating wire; in the model we have picked the wires are effectively oriented in the $x$ direction. From our model we see that the 1D wires with $k_y$ values on opposite sides of a Dirac point have opposite values of $\mathcal{Z}_{1\text{D}}$, and thus their contributions to the overall charge polarization differ by a quantized amount. We already

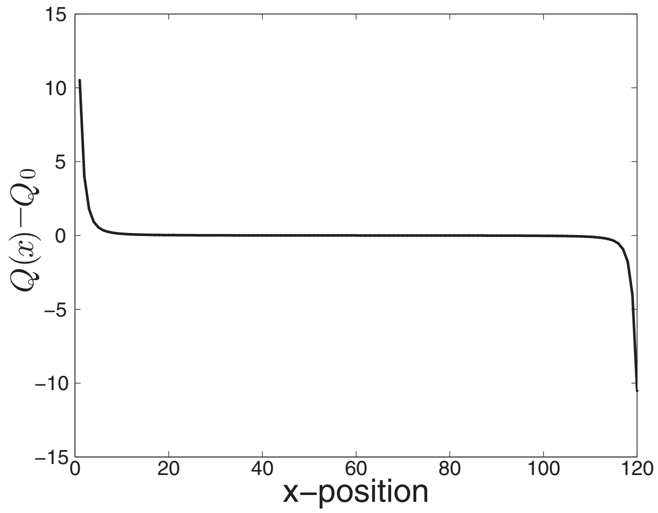

FIG. 7. We have plotted the deviation of the charge density from the average for $L_x = L_y = 120$ at half-filling in a 2D Dirac semimetal with $b_y = \pi/3$ (i.e., same parameters as in the previous figure). The average background charge per site is $Q_0 = 120e$. We notice peaks at the boundaries of the system due to the charge carried by localized midgap modes. The charge density exponentially decays to the value of $Q_0 = 120e$ within a few lattice sites. The total charge at the boundary calculated from summing the boundary charge near the right edge is $Q_b = -19.6e$, which matches the expected result, $Q_b = P_1^x L_y = -\frac{e}{6a} 120a = -20e$. The deviation from −20 is a finite-size effect and the result will converge to the analytic value as the system size increases.

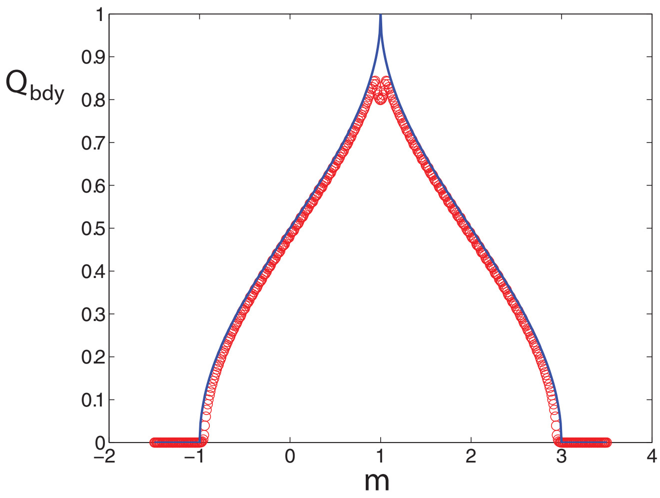

FIG. 8. (Color online) The boundary charge is plotted vs the mass parameter $m$. The solid blue curve represents $\frac{eb_y}{2\pi}$, where $b_y$ is calculated from the solutions to $\cos k_y = -m/t_y$ for a range of $m$ and with $t_y = -1$. The open circles are the numerically calculated boundary charge (per layer) for a system with open boundaries in the $x$ direction ($L_x = 120$) and periodic boundary conditions in the $y$ direction. They match except near $m = 1$, where the cusplike analytic result is cut off in the numerics due to finite-size effects.

know that for the completely gapped WTI phase, each wire contributes $e/2$ boundary charge (modulo $ne$) to an edge normal to the $x$ axis. In comparison, it is clear that for the DSM only the fraction of the wires between the Dirac nodes contribute $e/2$, while the remainder contribute charge $0 (\mathrm{mod}\, e)$.

We can also see that, physically, the bulk polarization manifests as an observable bound charge on the sample edges. In Fig. 7 we show the charge density as a function of position along the open boundary direction for the cylinder geometry mentioned above (see Fig. 6). We have subtracted off the average background charge, and two peaks in the charge density can be seen, one on each end of the sample. The amount of charge localized on each end matches the charge density calculated from Eq. (67) at an interface where the polarization changes from $P_1^x = -\frac{e}{2\pi}\frac{\pi}{3a} = -\frac{e}{6a}$ to zero (we have temporarily restored the lattice constant). More convincingly, in Fig. 8 we show the numerically calculated boundary charge values versus the analytically predicted value of the polarization/boundary-charge over a range of values of $m$ in our square-lattice model. The numerical and analytic results match almost exactly except near $m = 1$, where the analytic result predicts a cusplike shape that is cut off in the numerical calculations from finite-size effects. Interestingly, we see that even though the system is gapless, the charge polarization calculation gives reasonable physical results; e.g., it gives a physically meaningful prediction for the boundary charge. This is unusual, but not unprecedented, as Ref. [61] has shown that one can have a well-defined polarization in a Chern insulator despite the fact it has gapless boundary modes. We comment more about this point later.

Already for just two nodes there are some important subtleties to consider when calculating the polarization. The first subtlety has to do with the direction in which the polarization should point. For example, what determines which boundary has the positive charge in Fig. 7 and which end has a negative charge? The answer to this question is well known: To uniquely specify the polarization, we must apply an inversion-breaking (or $C$-breaking) field that picks the direction of the polarization and then take the limit as the system size goes to infinity before setting the symmetry-breaking perturbation to zero. This is the conventional paradigm for spontaneous symmetry breaking. Thus, in order to uniquely specify the sign of the polarization, and hence effectively the sign of $b_i$, we must turn on a small symmetry-breaking perturbation before we calculate and take the limit in which this perturbation vanishes. This issue arises in Sec. III C when we try to calculate Eq. (67) using field-theoretical methods. To be consistent with the notation in the next section, we call the inversion-symmetry-breaking parameter $m_A$.

The second subtlety is similar in nature and has to do with determining the value of the polarization in a bulk crystalline sample. In fact, in a bulk sample without boundary, since the BZ is periodic and we have no edge states to reference, we cannot determine a unique value for the polarization of a 2D Dirac semimetal. For example, in the simplest case of two nodes, how do we determine the magnitude of the polarization if we do not have a preferred way to take the momentum difference between the Dirac nodes? This is a problem because there are multiple ways to subtract the two momenta in a periodic BZ. For our concrete example, our nodes lie at $\vec{k} = (0, \pm\pi/3)$, and so we could let $\vec{b} = \frac{1}{2}(0, 2\pi/3)$

or, e.g., we could subtract the nodes across the BZ boundary to find $\vec{b}' = \frac{1}{2}(0,4\pi/3)$. The measurable property of the charge polarization is a boundary charge, which is determined by the occupation of the edge states. For two nodes there are two possible cases for how the edge states traverse the edge BZ. If they go through the origin, we should use $\vec{b} = \frac{1}{2}(0,2\pi/3)$, or if they instead go through $k = \pi$, then we should use $\vec{b} = \frac{1}{2}(0,4\pi/3) = (0,\pi) - \frac{1}{2}(0,2\pi/3)$. These two configurations can be interchanged by first adding a weak topological state, whose edge states will traverse the entire edge BZ, and then coupling it to the DSM, which will have the ultimate effect of switching the DSM edge states from one configuration to the other. In Sec. III E we see for the general case that, similar to the case of a polarized Chern insulator, the connection between the bulk value of the polarization and the boundary charge can have a more complicated relationship when more than two nodes are present and there are overlapping regions of edge states in the edge BZ.

### 2. *Symmetry protection of the response of 2D Dirac semimetals*

Before we move on to discuss the EM response due to the time-component $b_0$, we address the important issue of symmetry protection. For the 1D TI and the 2D WTI constructed from stacks of these 1D TI wires, we have only required inversion symmetry to have a well-defined EM response. This symmetry quantizes the 1D polarization to be 0 or $e/2$ on each wire, and as shown in Refs. [8,9], this symmetry is also enough to quantize the polarization (per wire) for the 2D WTI. However, it is well known [81] that for local stability of the Dirac nodes in a DSM, one needs at least the composite $\mathcal{TI}$ symmetry (for $\mathcal{T}^2 = +1$). We would like to understand the importance of this seemingly different requirement for the polarization response and the Dirac-node stability. This issue does not arise, for example, in the 3D Weyl semimetal since the Weyl nodes are locally stable without adding any additional symmetries, and hence, it is important to carefully discuss in the present context.

First, for 1D wires, $\mathcal{TI}$ also quantizes the polarization since $P_1$ is odd under this symmetry. Thus, we could have already constructed a 1D TI and a 2D WTI using this symmetry instead. In fact, the explicit model we have been considering has $\mathcal{TI}$ symmetry as written, and thus we were able to avoid discussing this issue until now.

Importantly, in dimensions greater than 1, $\mathcal{TI}$ symmetry has a crucial effect: It constrains the Berry curvature to satisfy $F(k_x,k_y) = -F(k_x,k_y)$. Since the Berry curvature flux is only defined modulo $2\pi$ on a lattice, this requires that for gapped systems either (i) $F(k_x,k_y) = 0$ or (ii) $F(k_x,k_y) = \pi$, and is constant throughout the BZ (we will only consider the former case [82]). Hence, with this symmetry we expect a vanishing Berry curvature. However, if $F(k_x,k_y)$ is not required to be smooth, we can have singular points in momentum space where $F(k_{xc},k_{yc}) = \pi$; these are exactly the set of Dirac-node locations. Since the Berry flux that passes through a closed manifold, e.g., the BZ, must be a multiple of $2\pi$, this implies that there are an even number of singular points, i.e., fermion doubling. This conclusion immediately implies local stability of the Dirac nodes, because if $\mathcal{TI}$ is preserved and one of the Dirac nodes disappears locally by itself, then there will not be an integer amount of Berry flux in the BZ which is a contradiction.

This constraint, and thus the $\mathcal{TI}$ symmetry itself, is also essential for the 2D charge polarization response of the DSM. Let us illustrate the idea. Suppose we wish to calculate the charge polarization of a crystalline DSM. The physical consequence of a nonvanishing polarization is a boundary charge, so let us specify a particular boundary with a normal vector $\mathbf{G}_N$ in the reciprocal lattice. Let $\mathbf{G}_F$ be the dual vector to $\mathbf{G}_N$, i.e., $G_F^i = \epsilon^{ij} G_N^j$. Then $\mathbf{G}_F$ is the normal vector to a set of lattice lines whose ends terminate on the surface normal to $\mathbf{G}_N$. For example, pick $\mathbf{G}_N = 2\pi\hat{x}$ and $\mathbf{G}_F = 2\pi\hat{y}$. In this case our choice picks out a family of 1D wires parallel to the $x$ direction and stacked in the $y$ direction. Consequently, this gives rise to a family of 1D Bloch Hamiltonians parametrized by the momentum along $\mathbf{G}_F$. In this example we have the family $H_{k_y}(k_x)$ which is parametrized by $k_y$.

To calculate the charge polarization of the DSM with our choice of $\mathbf{G}_N$ (i.e., the polarization parallel to $\mathbf{G}_N$), we can start by asking an important question: How much does the charge polarization of the family of 1D systems $H_{k_y}(k_x)$ vary as $k_y$ is varied? We find

$$\begin{aligned} P_1^x(k_{y2}) - P_1^x(k_{y1}) &= \frac{e}{2\pi}\int_{-\pi}^{\pi} dk_x a_x(k_x,k_{y2}) - \frac{e}{2\pi}\int_{-\pi}^{\pi} dk_x a_x(k_x,k_{y1}) \\ &= \frac{e}{2\pi}\int_{-\pi}^{\pi} dk_x \int_{k_{y1}}^{k_{y2}} dk_y \mathcal{F}(k_x,k_y) \\ &= \frac{e}{2}\sum_{a=1}^{N_{enc}} \chi_a, \end{aligned} \tag{69}$$

where we have used Stokes theorem to replace the line integrals over the Berry connection $\mathbf{a}(\mathbf{k})$ by an area integral over $\mathcal{F}(k_x,k_y) = \partial_{k_x} a_y - \partial_{k_y} a_x$, i.e., the Berry curvature, and we have assumed only one occupied band for simplicity. In the last equality we used the fact that for systems with $\mathcal{TI}$ symmetry the Berry curvature only contains contributions from the singular Dirac points, and the sum runs over all enclosed Dirac nodes. The quantity $\chi_a = \pm 1$, which we call the *helicity* of a Dirac node, indicates whether the flux carried by the node is $\pm\pi$. Thus, two 1D Hamiltonians that are members of the parametrized Hamiltonian family specify cycles in the BZ, and from this result we see that the polarization can only change if the area of the BZ enclosed between those two 1D cycles contains Dirac nodes. This restriction is the key feature of a $\mathcal{TI}$-symmetric system that determines the polarization response. As an aside we note that, since the BZ is a closed manifold, there are two possible ways to choose the region "enclosed" by the closed cycles and this is related to one source of ambiguity in the value of the polarization discussed earlier.

This result in Eq. (69) is generically true given a general family of Bloch Hamiltonians (with $\mathcal{TI}$ symmetry) with some orientation specified by $\mathbf{G}_N$, and parametrized by momentum along $\mathbf{G}_F$. In fact, given two 1D cycles that are members of a parametrized Hamiltonian family in the BZ, then any deformation/rotation of the orientation of the lines, i.e., variation of the choice of the direction vector $\mathbf{G}_N$, will not change the difference in polarization between the two parallel lines unless the lines cross Dirac points during the deformation process. This implies that the *changes* in polarization are always quantized in the presence of $\mathcal{T}\mathcal{I}$ symmetry, which is crucial for being able to determine the polarization from the nodal data.

Since the changes in polarization between different cycles are quantized, we might now ask about the properties of the total polarization. Since each 1D subspace is mapped onto itself by $\mathcal{T}\mathcal{I}$ and the polarization of that 1D system is odd under $\mathcal{T}\mathcal{I}$, we see that the polarization of each of the wires/cycles is quantized to be 0 or $e/2$. The other wires in the family of Hamiltonians either have exactly the same polarization or differ by a quantized amount. For the case with only two nodes, this argument shows that the (fractional part of the) boundary charge, up to an integer per unit cell, is completely determined by the length of $b_i$ that projects onto the edge BZ, which confirms what is predicted in Eq. (67) (more subtleties will arise when we have to consider cases with nodes arising from multiple bands that give rise to overlapping boundary states). The sign of the polarization for two nodes, however, is still ambiguous and can only be determined after a symmetry-breaking parameter is added and after knowing whether we should project the difference between the nodes through the edge BZ origin or the edge BZ boundary. As we discuss more carefully below, for two nodes the overall sign and value of the polarization can also be modified by the addition of nontrivial, occupied weak TI bands to the system. As far as the boundary theory is concerned, this is equivalent to adding an additional flat band of edge states which traverse the entire BZ, and because of the $\mathbb{Z}_2$ nature of the polarization this has to be carefully handled.

### *3. Magnetization response of a 2D Dirac semimetal*

After finishing our discussion of the symmetry protection and the importance of $\mathcal{T}\mathcal{I}$ symmetry for the charge polarization, let us now move on to a discussion of the response due to a nonvanishing $b_0$. We have seen that the spatial part of $b_\mu$ can be interpreted as a charge polarization, and, as shown below, the component $b_0$ represents an orbital magnetization. Before we provide the explicit proof, let us assume that this is the case and support the conjecture with some physical arguments and numerical calculations.

The physical manifestation of a nonvanishing magnetization is a circulating current bound at the edges of the sample. From our conjecture, we should be able to induce such a magnetization by turning on a $b_0$. We can generate a $b_0$ by adding the term $\gamma \sin k_y \mathbb{I}$ to the Dirac semimetal lattice Hamiltonian in Eq. (63). The value of $b_0$ generated would be $b_0 = (\gamma/\hbar)\sin k_{yc}$, where $k_y = k_{yc}$ is the location of the Dirac node (and consequently $-k_{yc}$ is the location of the other node). On topological edges we can immediately see that the addition of this term will cause the flat edge modes to disperse (see Appendix B 2 for a proof). This is also seen the numerical calculations in Fig. 9(a). Thus, the dispersion of the edge modes attached to the Dirac points is exactly what generates the bound current, at least on the edges which actually harbor topological bound states.

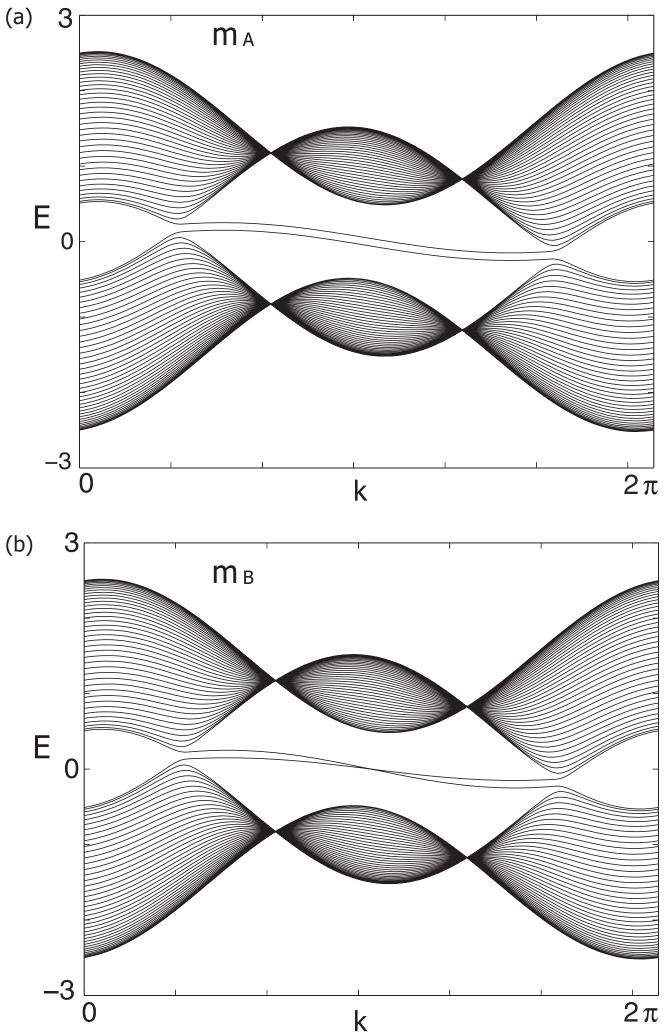

FIG. 9. The Energy spectrum is shown for the DSM with $b_0 \neq 0$ and different masses turned on. (a) With $m_A \neq 0$, we see that the edge modes split and do not cross as they move between the Dirac nodes. (b) With $m_B \neq 0$, it looks like the edge mode dispersion of a Chern insulator and they cross at $k = 0$.

Let us try to confirm this result numerically by calculating the current in the lattice model. Just as for the polarization, to properly calculate the response numerically, there is a subtlety about how to fill the edge states. To do this properly, we again need to choose a small, nonzero inversion-breaking mass before filling the edge modes. In the language of Ref. [61], to properly fill the edge modes in the presence of a nonvanishing $m_A$ we need to use the *adiabatic* filling, not the thermal filling, if we want to calculate the magnetization. One can see the energy spectrum for $b_0 \neq 0$ in Fig. 9(a) with a finite $m_A$ parameter. Adiabatic filling implies filling all of the states, including the edge modes, in the lower half of the spectrum below the energy gap induced by $m_A$. In Fig. 10 we plot the boundary current localized near a single edge vs $b_0$. The bound edge current is exactly $\frac{eb_0}{2\pi}$, which corroborates our conjecture that the magnetization is proportional to $b_0$.

It is interesting to note that in the model in Eq. (63) the $x$ and $y$ directions are very different since we have topological wires oriented along $x$ that are stacked in $y$. This should be

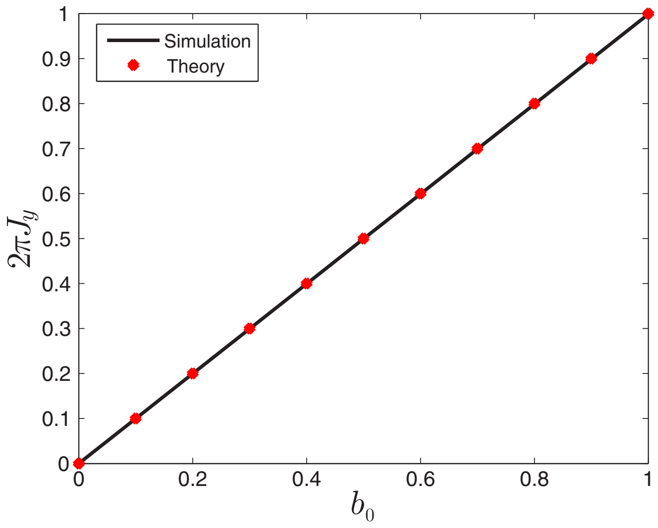

FIG. 10. (Color online) The bound current $J_y$ localized near a single edge vs $b_0$ is plotted for the model in Eq. (63) with $b_y = \frac{\pi}{3}$, $m_A = 10^{-3}$, $L_{x,y} = 120$, and periodic boundary conditions in the $y$ direction. The current matches the field theory prediction.

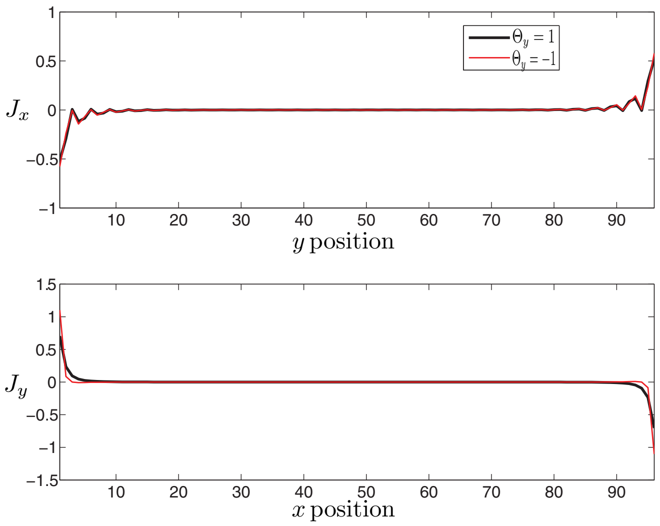

FIG. 11. (Color online) Plots of (a) $J_x$ vs $y$, which is the current on the nontopological edge, and (b) $J_y$ vs $x$, which is the current on the topological edge. This is for the Dirac semimetal considered in previous figures but with a nonzero $b_0$. For this system $b_y = \frac{\pi}{3}$, $\gamma = 0.1$, $m_A = 0.1$, and $L_{x,y} = 96$. There are open boundary conditions in both directions. We note that $j_y$ is exponentially localized, whereas $j_x$ is less-sharply localized and oscillates as it decays into the bulk. The oscillation wavelength coincides with the wave-vector location of the Dirac nodes in $k$ space. With open boundary conditions, we must be careful to properly fill the edge states by using a nonzero inversion-breaking mass term $m_A$. The currents with $\Theta_y = \pm 1$ are plotted in black and red. The total current near the boundaries is identical in both cases and thus the magnetization does not depend on how the edge states traverse on the edge BZ. The slight difference between the current profiles in (b) is due to the fact that the wave functions of the occupied edge modes that determine those boundary currents are different in the two cases with $\Theta_y = \pm 1$; however, the total current is the same.

contrasted with the fact that an orbital magnetization in 2D implies the existence of a bound current on *any* edge (i.e., any interfaces where the magnetization jumps from a finite value to zero). For the topological edges, with normal vectors parallel to the $x$ direction, a nonzero $b_0$ gives the edge modes a nonzero dispersion as shown in Fig. 9(a). The dispersing edge states produce an exponentially localized current $j^y_{\text{bound}}$ that corresponds to the change in magnetization at the edge. However, in the $y$ direction there are no topological edge modes, and it is interesting to consider what happens to $j^x$ on these edges.

We show the result of a numerical calculation in Figs. 11 and 12. In the former, we compare the current profiles of the different edge types in two different cylinder geometries corresponding to the two different edge types. In Fig. 11(a) we show the current on a nontopological edge ($J_x$ on an edge normal to $\hat{y}$), which is still localized on the boundary, but has an oscillatory decay. The wavelength of the oscillation, in fact, matches the wavelength of the Dirac-node wave vectors in momentum space. In Fig. 11(b) we show the current localized on topological edges ($J_y$ on an edge normal to $\hat{x}$) and we can see that each edge carries exponentially localized current with opposite currents on opposite edges.

In Fig. 12 we show the current density on a fully open sample, where we see that all of the current is localized near the edges. The colors are associated with the magnitude of the current parallel to a given edge. Essentially, this is just a different presentation of the data in Figs. 11(a) and 11(b) that shows that on both sets of edges there is a bound current, as expected from the orbital magnetization. Interestingly, on the edges without topological bound states the current oscillates as it decays. However, the magnitude of the current localized near edges of either type is identical, so indeed, even though the model is highly anisotropic, the bulk orbital magnetization generates bound currents on all edges, not just topological ones.

Further, we note that in the case with just two nodes the magnetization has no dependence on whether the edge states go through the origin of the edge BZ or through the boundary of the edge BZ (i.e., at $k_y = \pi$), assuming that the sign of the inversion-breaking parameter and the helicities of the nodes remain the same and only the edge-state locations are flipped. This is explicitly demonstrated in Fig. 11, where the total currents passing through each edge match exactly for these two cases. To generate the second case, where the edge states pass through $\pi$ in the edge BZ, we can choose our square-lattice model with the same parameters as before, except letting $A = -1, t_y = -1$. Below we introduce quantities $\Theta_i$, where $i = x,y$ that track whether the edge states pass through the origin of the edge BZ ($\Theta_i = +1$) or the boundary of the edge BZ ($\Theta_i = -1$) for different directions (e.g., $i = x$ or $y$). As we will see, these signs will enter the expressions for the charge polarizations, but not the magnetization. When there are more than just two nodes the magnetization is affected by the different edge-state configuration possibilities, but not in the same way as the polarization.

Now that we have motivated the EM response of the DSM using some analytic and numeric results on an example model, we now prove these claims using a Dirac semimetal model with two nodes and then go on to generalize to a generic even number of nodes.

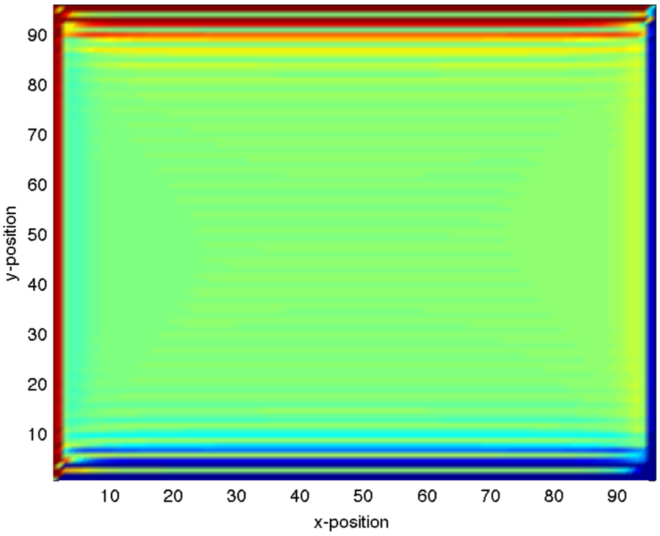

FIG. 12. (Color online) With a similar setup to the previous figure, we use a density plot for the current vs $x,y$ position for the 2D Dirac semimetal with $b_y = \frac{\pi}{3}$, $\gamma = m_A = 0.1$, $L_{x,y} = 96$, and we have open boundary conditions in both directions. We calculated the current density in the $x$ direction and summed it with the current density in the $y$ direction to produce this pseudocolor plot. We see that the currents are spatially localized at the edges, strongly for the one moving along the edges parallel to the $y$ axis and less strongly and oscillatory for the one moving along the edges parallel to the $x$ direction. The total magnitude of the current in the neighborhood of each edge is the same, and the current circulates around the boundaries of the sample.

### C. Derivation of response for continuum Dirac semimetal in 2D

In the previous section we posited a form for the EM response action of the DSM and gave some concrete examples in which the numerical simulations in lattice models matched the response derived from the effective action in Eq. (67). In this section we derive the EM response from a continuum model of the DSM using standard linear response techniques. We derived an example of a continuum Hamiltonian for the DSM in Eq. (66), and we use this as our starting point. After tuning the velocity coefficients to be isotropic, we can write the Hamiltonian for two Dirac cones that exist at the same point in the BZ as

$$H = k_x \mathbb{I} \otimes \sigma^x + k_y \tau^z \otimes \sigma^z, \tag{70}$$

where $\tau^a$ are Pauli matrices representing the two nodes. To this Hamiltonian we add two perturbations, the first of which is a splitting vector $b_\mu = (b_0, b_x, b_y)$ that shifts the two cones apart in momentum (by $2\vec{b}$) and energy (by $2b_0$). With the inclusion of this vector, which, if needed, we allow to be slowly varying in space-time, the Hamiltonian becomes

$$\begin{aligned} H = {} & k_x \mathbb{I} \otimes \sigma^x - b_x \tau^z \otimes \sigma^x + k_y \tau^z \otimes \sigma^z - b_y \mathbb{I} \otimes \sigma^z \\ & + b_0 \tau^z \otimes \mathbb{I}. \end{aligned} \tag{71}$$

The second type of perturbation we allow for is the coupling to external EM fields, which enter the Hamiltonian via minimal coupling $\mathbf{k} \to \mathbf{k} - (e/\hbar)\mathbf{A}$.

To calculate the linear response, we need the current operators that will enter the Kubo-formula calculation. For the EM field the current operators are

$$J_A^x = \frac{\delta H}{\delta A_x} = \frac{e}{\hbar} \mathbb{I} \otimes \sigma^x \equiv \frac{e}{\hbar} \Gamma^x, \tag{72}$$

$$J_A^y = \frac{\delta H}{\delta A_y} = \frac{e}{\hbar} \tau^z \otimes \sigma^z \equiv \frac{e}{\hbar} \Gamma^y, \tag{73}$$

$$J_A^0 = \frac{\delta H}{\delta A_0} = \frac{e}{\hbar} \mathbb{I} \otimes \mathbb{I}. \tag{74}$$

For the splitting vector $b_\mu$, the associated currents are

$$J_B^x = \frac{\delta H}{\delta b_x} = -\tau^z \otimes \sigma^x \equiv \Lambda^x, \tag{75}$$

$$J_B^y = \frac{\delta H}{\delta b_y} = -\mathbb{I} \otimes \sigma^z \equiv \Lambda^y, \tag{76}$$

$$J_B^0 = \frac{\delta H}{\delta b_0} = \tau^z \otimes \mathbb{I} \equiv \Lambda^0. \tag{77}$$

We want to calculate the "topological" response terms for the DSM and, in $(2+1)$ dimensions, we see that such response terms will break either time-reversal or inversion symmetry. It is well known that Dirac fermions in $(2+1)$ dimensions exhibit a parity anomaly that gives rise to a Chern-Simons contribution to the effective action that encodes a nonvanishing Hall conductivity [28,83]. There is a subtlety: To calculate the nonvanishing coefficient, one must introduce a finite, (time-reversal) symmetry-breaking mass parameter that is taken to vanish at the end of the calculation. Since the resulting response coefficient ends up being proportional only to the sign of the symmetry-breaking parameter, it remains nonzero even in the limit where the symmetry breaking is removed. This effect is a manifestation of a quantum breaking of symmetry, i.e., an anomaly. One of the main results of this paper is that we show that the same is true for an inversion-breaking mass term, not just the time-reversal-breaking mass term.

To calculate the responses due to $A_\mu$ or $b_\nu$ perturbations, we need to introduce two different types of symmetry-breaking mass terms,

$$\Sigma_A = \mathbb{I} \otimes \sigma^y, \tag{78}$$

$$\Sigma_B = \tau^z \otimes \sigma^y. \tag{79}$$

These two different mass matrices commute, and thus they are *competing* mass terms. They both separately anticommute with the kinetic part of the Dirac Hamiltonian (including the constant momentum shift $\vec{b}$), and thus the spectrum will be gapped as long as the coefficients $(m_A, m_B)$ of $(\Sigma_A, \Sigma_B)$ are not equal in magnitude. Explicitly, if both mass terms are activated, the energy spectrum is $\pm E_\pm = \pm\sqrt{(k_x - b_x)^2 + (k_y - b_y)^2 + (m_A \pm m_B)^2}$, which is gapped unless $|m_A| = |m_B|$. These mass terms are very familiar in the literature: $\Sigma_A$ is essentially the inversion-breaking Semenoff mass term [84], and $\Sigma_B$ is the continuum version of the time-reversal-breaking Haldane mass term [28].

Generically, in linear response, we find contributions to the effective action of the form

$$S_{\text{eff}}[A_\mu, b_\nu] = \int \frac{d^3 p_1}{(2\pi)^3} \mathcal{A}_\mu^a(p_1) \Pi_{ab}^{\mu\nu}(p_1) \mathcal{A}_\nu^b(-p_1), \tag{80}$$

which has been written in the Fourier-transformed basis, and where $a,b = A,B$, $\mathcal{A}^A_\mu = A_\mu$, and $\mathcal{A}^B_\mu = b_\mu$. The linear response calculation (or equivalently the calculation of the quadratic term in the effective action) amounts to the calculation of the long-wavelength, dc limit of the generalized polarization tensor

$$\Pi^{\mu\nu}_{ab}(\nu,\mathbf{q}) = \frac{\hbar}{2}\int \frac{d\omega d^2p}{(2\pi)^3}\mathrm{tr}\big[J^\mu_a G(\omega+\nu,\mathbf{p}+\mathbf{q})J^\nu_b G(\omega,\mathbf{p})\big], \tag{81}$$

where $\mu,\nu = 0,x,y$, $a,b = A,B$, and $G(\omega,\mathbf{p})$ is the space and time Fourier transform of the single-particle Green's function of the unperturbed ($b_\mu = 0, A_\mu = 0$) Dirac model.

The calculation of $\Pi^{\mu\nu}_{ab}$ is sensitive to the choice of symmetry-breaking masses $m_a$. Since we are only interested in extracting the topological terms in the semimetallic limit, we can consider two cases: (i) $|m_A| > |m_B| = 0$ and (ii) $|m_B| > |m_A| = 0$. We briefly consider the case when both mass terms are nonzero in Sec. III F.

The Fourier transform of the unperturbed Green's function in either of these limits will be

$$\begin{aligned} G(\omega,p) &= \frac{1}{\omega - p_x\Gamma^x - p_y\Gamma^y - m_c\Sigma^c} \\ &= \frac{\omega + p_x\Gamma^x + p_y\Gamma^y + m_c\Sigma^c}{\omega^2 - |\mathbf{p}|^2 - m_c^2}, \end{aligned} \tag{82}$$

where the label $c = A$ or $B$, and is not summed over. The topological terms in the polarization tensor can be calculated by extracting the terms proportional to odd powers of the symmetry-breaking mass:

$$\begin{aligned} \Pi^{\mu\nu}_{ab}(\nu,\mathbf{q}) = {} & \frac{\hbar}{2}\int \frac{d\omega d^2p}{(2\pi)^3} f(\omega+\nu,\mathbf{p}+\mathbf{q}) f(\omega,\mathbf{p}) \\ & \times \mathrm{tr}\big\{J^\mu_a m_c \Sigma_c J^\nu_b(\omega + p_x\Gamma^x + p_y\Gamma^y) \\ & + J^\mu_a[\omega + \nu + (p_x+q_x)\Gamma^x \\ & + (p_y+q_y)\Gamma^y]J^\nu_b m_c\Sigma^c\big\}, \end{aligned} \tag{83}$$

$$f(\omega,\mathbf{p}) = \frac{1}{\omega^2 - |\mathbf{p}|^2 - m_c^2}. \tag{84}$$

Now, to be explicit, let us consider case (i), where $m_A$ is the nonvanishing mass term. We can extract the leading term in the external frequency/momentum, which we find to be

$$\begin{aligned} \Pi^{\mu\nu}_{ab}(\nu,\mathbf{q}) &= 4\frac{e}{2}m_A\epsilon^{\mu\rho\nu}(iq_\rho)\sigma_{ab}\int \frac{d\omega d^2p}{(2\pi)^3}[f(\omega,\mathbf{p})]^2 \\ &= \frac{4\pi^2}{(2\pi)^3}\frac{e}{2}\frac{m_A}{|m_A|}\epsilon^{\mu\rho\nu}(iq_\rho)\sigma_{ab} \\ &= \frac{e}{4\pi}(\mathrm{sgn}\ m_A)\epsilon^{\mu\rho\nu}(iq_\rho)\sigma_{ab}, \end{aligned} \tag{85}$$

where $q_\rho = (\nu,\mathbf{q})$ is the external 3-momentum, $\sigma_{AB} = \sigma_{BA} = 1$, and $\sigma_{AA} = \sigma_{BB} = 0$. This leads to a term in the effective action

$$S^{(A)}_{\mathrm{eff}}[A_\mu,b_\nu] = \frac{e}{2\pi}(\mathrm{sgn} m_A)\int dt d^2x \epsilon^{\mu\nu\rho}A_\mu\partial_\nu b_\rho. \tag{86}$$

This result exactly matches Eq. (67), except for the factor of $\mathrm{sgn} m_A$, which we already motivated as being necessary to pick the sign of the charge polarization.

From this continuum calculation for two nodes we can extract the polarization and magnetization in a nice way as

$$(M,\epsilon^{ij}\Theta_i P_i) = \frac{e}{4\pi}(\mathrm{sgn} m_A)\sum_{a=1}^{2}\chi_a K_{a,\mu}, \tag{87}$$

where $K_{a,\mu}$ are the momentum and energy locations of the nodes, and the $\chi_a$ are the helicities of the nodes. Even for just two nodes the polarization calculated in the continuum approximation is ambiguous since the edge states connecting the nodes could pass through the origin or boundary of the edge BZ. We have corrected for this in Eq. (87) by adding the extra signs $\Theta_i = \pm 1$ $(i = x,y)$, which indicate exactly if the edge states run through the origin on the edge BZ ($\Theta_i = +1$) or through $\pi$ ($\Theta_i = -1$), as motivated before. In general, when not on the square lattice, there is one value of $\Theta$ for each independent spatial direction. These extra signs only enter the formula for the polarization, not the magnetization as mentioned above.

A precise definition of $\Theta_i$ can be determined purely from the bulk properties of the system by calculating the Wilson line of the Berry connection along the 1D Bloch Hamiltonian subspace that projects onto $k = \pi$ in the respective edge BZs (it is analogous to calculating the weak invariant $\nu_i$). This Wilson line can only take two values because of the $\mathcal{TI}$ symmetry, and its trivial (nontrivial) value corresponds to $\Theta_i = +1(-1)$. It is important to note that a knowledge of $\Theta_i$ is not contained in the manifold of band touchings alone and requires some knowledge of the occupied bands. For the two-node case, this implies that the (fractional part of the) polarization can only be determined up to an overall sign if we only have knowledge of the continuum band-touching points and their locations in the BZ. However, the magnetization does not share this particular ambiguity due to $\Theta_i$. This might be attributed to its more isotropic nature.

In this continuum picture we could also imagine having more flavors of fermions with different nodal locations, but with Hamiltonians of essentially the same form as in Eq. (70). Since we want to eventually consider lattice models with multiple pairs of nodes, we know there must be an even number of total nodes, say $2N$, and an equal number $N$ of them have opposite helicity. From our calculation, the polarization/magnetization for such a system might be trivially generalized as

$$(M,\epsilon^{ij}\Theta_i P_i) = \frac{e}{4\pi}\sum_{a=1}^{2N} g_a\chi_a K_{a,\mu},$$

where $g_a$ represents the sign of the symmetry-breaking mass for each Dirac point. Unfortunately, this simple generalization has a few caveats, one example being that it does not take care of the $\mathbb{Z}_2$ nature of the edge states, which becomes important when edge states from different blocks overlap in the edge BZ. We discuss the details of this generalization in Sec. III E.

If we repeat this calculation for case (ii), where $m_B$ is nonvanishing, the result is almost identical, except for the replacement of the matrix $\sigma_{ab}$ by the Kronecker $\delta_{ab}$; i.e., the polarization tensor is

$$\Pi^{\mu\nu}_{ab}(\nu,\mathbf{q}) = \frac{\hbar e_a^2}{4\pi}(\mathrm{sgn} m_B)\epsilon^{\mu\rho\nu}(iq_\rho)\delta_{ab}, \tag{88}$$

where the charge $e_A = e/\hbar$ and $e_B = 1$. Now this gives rise to two terms in the effective action,

$$S^{(B)}_{\mathrm{eff}}[A_\mu,b_\nu] = \frac{e^2}{2h}(\mathrm{sgn}\ m_B)\int dtd^2x\epsilon^{\mu\nu\rho}A_\mu\partial_\nu A_\rho + \frac{\hbar}{4\pi}(\mathrm{sgn}\ m_B)\int dtd^2x\epsilon^{\mu\nu\rho}b_\mu\partial_\nu b_\rho. \tag{89}$$

The first term is the conventional Chern-Simons term which yields a Hall conductivity of $\sigma_{xy} = \frac{e^2}{h}(\mathrm{sgn} m_B)$, which consists of $\frac{e^2}{2h}(\mathrm{sgn} m_B)$ from each of the two Dirac cones. This type of response has been discussed extensively in the literature, so we do not dwell on it here. The second term, which does not yield an EM response since it is independent of $A_\mu$, is discussed in Appendix C.

### D. Physical interpretation of the Dirac semimetal response

The topological EM response of the DSM is more complicated than the 1D band metal because the response density and current depend on derivatives of $b_\mu$, not just the vector itself. When the time-reversal mass term $m_B$ dominates, and there are only two nodes, then we just generate the well-known Chern insulator phase [28] or its associated parity anomaly in the limit $m_B \to 0$ [83]. In this section we consider the less well-known case of when $m_A$ dominates and the resulting inversion-breaking semimetal limit. This will help us solidify an appropriate definition of charge polarization and magnetization for 2D Dirac semimetals, akin to the definition provided for Chern insulators in Ref. [61]. In Sec. III F we revisit the case when $m_B$ is nonvanishing and consider the effects of a finite $m_A$ term in the Chern insulator phase.

#### 1. Response in the inversion-breaking limit ($m_A$ dominating regime)

Let us consider the limit in which the inversion-breaking mass $m_A$ dominates over the time-reversal mass $m_B$ and then send them both to zero (with $m_B \to 0$ first). In that limit the response that we derived is given by

$$S^{(A)}_{\mathrm{eff}}[A_\mu,b_\mu] = \frac{e}{2\pi}(\mathrm{sgn} m_A)\int dtd^2x\epsilon^{\mu\nu\rho}A_\mu\partial_\nu b_\rho.$$

The current from this effective action is given by

$$\begin{aligned} j^\alpha &= \frac{e}{2\pi}(\mathrm{sgn} m_A)\epsilon^{\alpha\mu\nu}\partial_\mu b_\nu \\ &\Rightarrow \rho = \frac{e}{2\pi}(\mathrm{sgn} m_A)(\partial_x b_y - \partial_y b_x), \\ j^i &= \frac{e}{2\pi}(\mathrm{sgn} m_A)\epsilon^{ij}(-\partial_0 b_j + \partial_j b_0). \end{aligned} \tag{90}$$

To simplify, let us assume that $m_A \to 0^+$ so that we can replace $\mathrm{sgn} m_A = +1$.

These equations can be more easily interpreted if we replace $b_i$ via the polarization $P^i_1 = -\frac{e}{2\pi}\epsilon^{ij}b_j$ to generate

$$\rho = -\partial_i P^i_1, \quad j^i = \partial_0 P^i_1 + \frac{e}{2\pi}\epsilon^{ij}\partial_j b_0.$$

We immediately recognize these equations as the contributions to the charge density and current from gradients and time derivatives of the polarization. It is also easy to interpret the term involving $b_0$, as it just represents the contribution to the current from gradients in the magnetization. We can let $M = \frac{e}{2\pi}b_0$ be the out-of-plane magnetization, from which we finally arrive at

$$\rho = -\partial_i P^i_1, \quad j^i = \partial_0 P^i_1 + \epsilon^{ij}\partial_j M, \tag{91}$$

which are the familiar constituent relations for bound charge density and bound charge current in 2D. Thus, we see that, in the limit where $m_A$ dominates over $m_B$ and then tends to zero, the DSM will exhibit an effective polarization and magnetization if $b_i$ and $b_0$ are nonzero, respectively. Bound charge and current manifest at interfaces or boundaries where the bulk values of $b_\mu$ are changing and are the consequence of the topological response.

The relation between $b_\mu$ and the bulk magnetization and polarization makes an important physical connection between generic EM quantities $(P^i_1,M)$ and the quantities $(\epsilon^{ij}b_j,b_0)$ that are determined by the energy and momentum locations of the nodal Dirac points in the electronic spectrum. Accordingly, we can rewrite the effective action as

$$S^{(A)}_{\mathrm{eff}}[A_\mu,b_\mu] = \int dtd^2x\big[M(b_\mu)B + P^i_1(b_\mu)E_i\big], \tag{92}$$

where we have included the dependence on $b_\mu$. Writing the action this way is interesting because it highlights that the DSM can have a well-defined polarization, something that is usually reserved for gapped insulators. From this we see that one possible signature of a clean DSM with nondegenerate nodes would be a semimetal phase with $\mathcal{TI}$ symmetry and a nonvanishing charge polarization/magnetization.

#### 2. Polarization and boundary charge

Using the model for the DSM introduced above, let us revisit the origin of the bound charge and bound current from a more microscopic picture. From the effective action we see that we need $b_y$ to change with $x$ or vice versa to generate a nonzero charge density. To produce a nonzero current, we need $b_0$ to vary with $x$ or $y$. The easiest way to do either of these is to have an interface or boundary. First, suppose we have a boundary where $b_y$ changes with $x$ as $b_y = b_y\Theta(x - x_0)$, where $\Theta(x)$ is a step function. From the response action we should have a bound charge density

$$\rho = (\mathrm{sgn} m_A)\Theta_y\frac{eb_y}{2\pi}\delta(x - x_0), \tag{93}$$

where we recall that the $\Theta_y$ in this formula is needed to capture the correct sign of the boundary charge for lattice systems [cf. Eq. (87)].

The magnitude of the charge density determined by the bulk response action exactly matches the boundary charge we find in the DSM model from the edge modes stretched between the two nodes. The role of the value of $(\mathrm{sgn} m_A)\Theta_y$ is to fix which edge has the occupied states and, subsequently, which edge is unoccupied. Due to the inversion-breaking mass, each boundary state on one edge will be occupied and contribute $e/2$ charge on that boundary for each edge mode. On the other edge, all of the boundary modes will be unoccupied, and each contributes a deficit charge of $-e/2$. The total number of occupied states on the edge is given by the distance spanned by the edge states between the two nodes multiplied by $\frac{L_{\text{edge}}}{2\pi}$, which, in total, is $L_{\text{edge}}\frac{2b_y}{2\pi}$. So, the total charge at the positive edge is given by $L_{\text{edge}}\frac{e}{2}\times\frac{b_y}{\pi}=L_{\text{edge}}\frac{eb_y}{2\pi}$. This implies a polarization of $\frac{eb_y}{2\pi}$, as expected. Thus, we see that while the charge response in the 1D semimetal is controlled by the bulk states, here it manifests as a property of the boundary modes. This is due to the charge density depending on derivatives of $b_i$ instead of $b_i$ itself.

### *3. Orbital magnetization and boundary current*

Next, let us consider the microscopic origin of the magnetization. The bound current that exists on interfaces when $b_0$ is nonvanishing, i.e., when there is a bulk magnetization, is more delicate. For example, the magnetization, as far as the 2D system is concerned, is isotropic and thus should give rise to bound currents on *any* interface, not just an edge with low-energy modes. We already showed in Figs. 11 and 12 that, even though the DSM model we have chosen is inherently anisotropic, there are bound currents on all of the edges. Let us now prove that this boundary current is indeed connected to the bulk orbital magnetization.

First, to generate a nonvanishing $b_0$ in the DSM model, we can add a kinetic energy term $\epsilon(k)=\gamma\sin k_y\mathbb{I}$ to the Hamiltonian $H_{2\text{DWTI}}(k)$ in Eq. (63). If the Dirac nodes are separated in the $k_y$ direction and located at $\vec{k}=(0,\pm k_{yc})$, as for our earlier parameter choice, then this simple kinetic term will generate an energy difference of $2\gamma\sin k_{yc}\equiv 2\hbar b_0$ between the Dirac nodes. Note that this term breaks both $T$ and $\mathcal{I}$ but preserves the composite symmetry $T\mathcal{I}$, which is required for the local stability of the Dirac nodes. Since it breaks $T$, in principle, a magnetization would be allowed by symmetry.

Next we can calculate the magnetization for this model according to the results of Refs. [59,85] using

$$M=\frac{e}{2\hbar}\int\frac{d^2k}{(2\pi)^2}\text{Im}\{\langle\partial_x u_-|[H(k)+E_-(k)]|\partial_y u_-\rangle - \langle\partial_y u_-|[H(k)+E_-(k)]|\partial_x u_-\rangle\}, \tag{94}$$

where $E_-(k),|u_-\rangle$ are the energy and Bloch functions of the lower occupied band, $H(k)=\epsilon(k)+H_{2\text{DWTI}}(k)$, and the derivatives are with respect to momentum. To properly calculate this quantity, we need to turn on a small, but finite, $m_A$ and then set it to zero at the end of the calculation. From symmetry, and from the fact that the extra kinetic term is proportional to the identity matrix, the only terms that contribute to the nonvanishing magnetization are those proportional to $\epsilon(k)$, and we find the simplification [59]

$$M=\frac{e}{2\hbar}\int\frac{d^2k}{(2\pi)^2}2\epsilon(k)F_{xy}(k), \tag{95}$$

where $F_{xy}(k)$ is the Berry curvature.

For small $m_A$ we know that $F_{xy}$ is sharply peaked at each of the two Dirac nodes. For example, when $m_A=0$, then $T\mathcal{I}$ is preserved, and the Berry curvature is a $\delta$-function source at each node. When $m_A\neq 0$ the contributions of the two Dirac points to the Berry curvature have *opposite* signs because of their opposite helicities. Thus, we can see that if $\epsilon(k)$ had the same value for both Dirac nodes, then $M$ would vanish. In the semimetallic limit $m_A\to 0$, which is the limit of physical interest, the magnetization becomes

$$M=(\text{sgn}m_A)\frac{e\Phi_{\text{Dirac}}}{4\pi^2\hbar}\sum_{a=1}^{N_{\text{Dirac}}}\epsilon(\vec{K}_a)\chi_a, \tag{96}$$

where $\vec{K}_a$ is the location of the $a$th Dirac point, $\epsilon(\vec{K}_a)$ is the energy of the $a$th Dirac point, $\chi_a$ is the sign of the Berry phase around the Fermi surface of each Dirac point for an infinitesimally positive chemical potential, and $\Phi_{\text{Dirac}}$ is the constant Berry curvature flux carried by each Dirac point in the gapless limit, i.e., $\Phi_{\text{Dirac}}=\pi$. In terms of $b_0$ for our single pair of Dirac points, we find $M=(\text{sgn}m_A)\frac{e}{2\pi}b_0$, as expected. While this is the general result for the bulk magnetization, even for a lattice model, the connection to a boundary current must be carefully addressed if there are Dirac nodes with overlapping edge modes that can cancel in a $\mathbb{Z}_2$ fashion. We discuss this more in Sec. III E.

Now that we have explicitly determined the relationship between bulk magnetization and the energy locations of the nodal points, let us try to connect the response to the edge-state properties as mentioned earlier. Consider our simple two-node DSM model with $m_A>0$ on a cylinder with periodic boundary conditions in the trivial direction ($y$ direction) and open boundary conditions in the topological direction ($x$ direction). With this choice of orientation the system will exhibit gapless boundary modes. Let us add in the term $\epsilon(k)=\gamma\sin k_y\mathbb{I}$ to generate a nonvanishing $b_0$. The sample thus has $b_0=b_0[\Theta(x)-\Theta(x-L_x)]$, where we have chosen the cylinder to lie between $x=0$ and $x=L_x$. The current density near the left edge ($x=0$) is given from the response action by

$$j_L^y=-\frac{e}{2\pi}b_0\delta(x). \tag{97}$$

The total current traveling within a region near $x=0$ is simply $J_L^y=\int_{-\delta}^{\delta}dx j_L^y=-\frac{eb_0}{2\pi}$. Of course, the *total* current in the $y$ direction will vanish once we take both edges into consideration.

Now we can use this result to compare to the current carried by the edge states. In Fig. 9(a) we show the energy spectrum for the DSM in a cylinder geometry for a nonzero $\gamma$ and a nonzero $m_A>0$. We see that the edge states are attached to the Dirac nodes (slightly gapped by $m_A$), and their dispersion is $\epsilon_{\text{edge}}(k_y)=-\gamma\sin k_y$ (for a derivation, see Appendix B 2). When $m_A$ is identically zero, then at half-filling each edge branch will be occupied up to $E=0$ (which happens at $k_y=\pi$ for our model), and the boundary currents vanish. When $m_A\neq 0$ then the remaining states on the left edge become occupied, which generates a current; the other edge will now have an excess of unoccupied (hole) states which produce a current in the opposite direction. If we take the limit as $m_A\to 0$, then the boundary current will persist since the electrons cannot scatter from one edge to the other as long as translation symmetry is preserved, and the edges remain far enough apart to prevent an interedge hybridization gap. The edge electrons will remain in their "adiabatically" filled state (in the language of Ref. [61]).

Let us now calculate the magnitude of the edge current in these conditions. Explicitly, the current on the left edge when all of the boundary modes are occupied is

$$\begin{aligned} J_L^y &= \frac{e}{2\pi\hbar}\int_{k_{y0}}^{k_{yc}} dk_y \frac{\partial\epsilon_{\text{edge}}(k_y)}{\partial k_y} \\ &= -\frac{e\gamma}{2\pi\hbar}[\sin k_{yc} - \sin k_{y0}] \\ &= -\frac{e}{2\pi}\left[\frac{\gamma}{\hbar}(\sin k_{yc} - \sin k_{y0})\right] = -\frac{eb_0}{2\pi}, \end{aligned} \tag{98}$$

where $k_{y0}$ is the energy up to which the edge state is occupied when $m_A = 0$, and $k_{yc}$ is the point up to which the additional occupied states are filled when the entire edge branch is occupied. Thus, we see that on the sides of the system that have topological edge states, the current is completely accounted for by the boundary modes.

As discussed above, the nonvanishing bulk magnetization also implies that there should be bound currents on edges that do not have low-energy topological boundary modes. Current conservation also indicates that on finite-sized systems, where all boundaries are open, the edge currents from a gapless edge must flow somewhere after hitting a corner. Indeed, this is confirmed in Figs. 11 and 12. Though we do not have a simple argument to derive the magnitude of the edge current on nontopological edges, we found numerically that the magnitudes of the currents localized on each edge are the same.

### E. General formulation of response for 2D DSM

Let us now consider a generic $\mathcal{TI}$-invariant DSM which harbors an even number of Dirac cones. Each Dirac cone $D_a$ $(a = 1,2,\ldots,2N)$ in the semimetal is specified by the data $(\chi_a,\hbar\bar{\mathbf{K}}_a,\epsilon_a,g_a)$, which are the helicity, momentum-space location of the Dirac node, energy of the node, and the sign of an infinitesimal local mass term at the Dirac point, respectively. The helicity indicates whether the winding of the (pseudo)spin around a Fermi surface at a Fermi energy above the node gives rise to a Berry phase of $\pm\pi$ (i.e., $\chi_a = \pm 1$). All of the response coefficients in which we are interested arise from anomalous terms which, even for gapless Dirac nodes, depend on how the gapless point was approached from a gapped phase; this is why we must include the $g_a$. Another way to think about this is that the choice of $g_a$ determines the sign of the symmetry-breaking response for each pair of Dirac nodes.

Let us now consider the generalization of our earlier continuum formula to the case with many flavors. Following Ref. [34], in the ultraclean limit we can associate a conserved current $j^\mu_{(a)}$ to each Dirac cone and a matching gauge field $A_{(a)\mu}$. Each Dirac cone contributes a term to the effective response action of the form

$$S^{(a)}_{\text{eff}}[A_{(a)}] = \chi_a g_a \frac{e^2}{4h}\int d^3x\epsilon^{\mu\nu\rho}A_{(a)\mu}\partial_\nu A_{(a)\rho}. \tag{99}$$

This gauge field contains two pieces: (i) the contribution from the EM gauge potential, and (ii) the energy-momentum shift of each Dirac node. Thus, we have $A_{(a)\mu} = A_\mu + \frac{\hbar}{e}\bar{K}_{(a)\mu}$, where $\bar{K}_{(a)\mu}$ tells us the energy-momentum location of the node such that $\bar{K}_{(a)0} = \epsilon_a/\hbar$, and $A_\mu$ is the true EM vector potential. With this specified, we can rewrite the action in a more transparent manner:

$$\begin{aligned} S[A,K_{(a)}] = \frac{e^2}{4h}\sum_{a=1}^{2N}\chi_a g_a \int d^3x\,\epsilon^{\mu\nu\rho}\left(A_\mu + \frac{\hbar}{e}\bar{K}_{(a)\mu}\right) \\ \times\partial_\nu\left(A_\rho + \frac{\hbar}{e}\bar{K}_{(a)\rho}\right). \end{aligned} \tag{100}$$

Let us now try to extract the important EM contributions to the response. The term containing only powers of $A_\mu$, and none of $K_{(a)\mu}$, is simply

$$S_1[A] = \frac{C_1 e^2}{2h}\int d^3x\,\epsilon^{\mu\nu\rho}A_\mu\partial_\nu A_\rho, \tag{101}$$

where $C_1$ is the total Chern number given by $C_1 = \frac{1}{2}\sum_{a=1}^{2N}\chi_a g_a$. We do not discuss the extra terms in the effective action which are independent of $A_\mu$ here. To understand them better, we can reformulate the response theory using an analog of the $K$ matrix formalism familiar from the Abelian fractional quantum Hall (FQH) states [86]. This discussion lies outside the main scope of the text and we defer it to a brief discussion in Appendix C.

To extract the mixed term that represents the charge polarization and magnetization, we will, for simplicity, restrict ourselves to particular configurations of the $g_a$. When there are more than two nodes, the concept of a single inversion- or time-reversal-breaking mass term is not clearly defined when given a full set of $g_a$. One appropriate generalization of the inversion-breaking mass is to have the sign of the mass fixed to be the same for *all* nodes. That is, $g_a = +1$ or $g_a = -1$ for all values of $a$. For this choice let us call $g_a = g$ for all $a$. This choice has some immediate consequences: (i) The total Chern number vanishes, i.e.,

$$C_1 = \frac{1}{2}\sum_{a=1}^{2N}\chi_a g_a = g\frac{1}{2}\sum_{a=1}^{2N}\chi_a = 0, \tag{102}$$

since there is a generic constraint $\sum_a \chi_a = 0$ coming from the $\mathcal{TI}$ symmetry of the lattice model; and (ii) the Chern number for any pair of opposite-helicity Dirac nodes vanishes. With this constraint we can determine the polarization and magnetization from the band-touching data as discussed below. On the other hand, if we allowed each $g_a$ to have varying signs, the determination of the polarization can be become more complicated since the Chern number of individual pairs of Dirac nodes need not always vanish (cf. Sec. III F) [87]. Henceforth, we fix the all $g_a = g > 0$.

Now, using this choice for the set of $g_a$ we can extract the mixed contribution to the action

$$S_2[A,b] = \frac{e}{2\pi}\int d^3x\epsilon^{\mu\nu\rho}b_\mu\partial_\nu A_\rho, \tag{103}$$

where

$$b_\mu = \frac{g}{2}\sum_{a=1}^{2N}\chi_a\bar{K}_{(a)\mu}. \tag{104}$$

This is the more general formulation of the two-node formulas we had derived previously, and the magnetization and polarization are given as $eb_\mu = 2\pi(M,\epsilon_{ij}P_1^j)$.

Equations (101), (103), and (104) are the general continuum results and are similar to the types of formulas one finds for 3D Weyl semimetals, for example. However, these formulas only provide the correct *bulk* results for a lattice system if there are an even number of edge-state branches (on one edge) that pass through the boundary of the edge BZ. This issue is addressed by simply including an extra sign ($\Theta_j$) in the polarization for each spatial direction as discussed above. Given our choice of an edge, this automatically determines a reciprocal lattice vector normal to the edge $\mathbf{G}_N$ and its dual vector $\mathbf{G}_F$ with components $G_{Fi} = \epsilon_{ij} G_{Nj}$. The $\Theta_{\hat{n}}$, which is essentially a weak index, is determined by the Wilson line integral $\Theta_{\hat{n}} = \exp[i \int_{\vec{G}_N} a_i(\vec{k}) dk^i |_{k_{\text{edge}}=\pi}]$, i.e., the line integral of the adiabatic connection across the BZ along the momentum direction normal to the edge, and evaluated at $k_{\text{edge}} = \pi$ (where $k_{\text{edge}}$ is the momentum tangent to the edge). Physically, the quantity $\Theta_{\hat{n}}$ determines whether the effective 1D wire Hamiltonian at $k_{\text{edge}} = \pi$ has a trivial ($\Theta_{\hat{n}} = +1$) or nontrivial ($\Theta_{\hat{n}} = -1$) polarization. To be explicit, let us consider an orthogonal lattice basis $\vec{a}_1 = \hat{x}$ and $\vec{a}_2 = \hat{y}$ and pick an edge with normal vector $\vec{a}_1$. We have $\vec{G}_N = 2\pi \hat{x}$, $\vec{G}_F = 2\pi \hat{y}$, and $k_{\text{edge}}$ runs over all values of $k_y$. For this choice we have the definitions $\Theta_x = \exp[i \int_{-\pi}^{\pi} dk_x a_{k_x}(k_x,\pi)]$ and $\Theta_y = \exp[i \int_{-\pi}^{\pi} dk_y a_{k_y}(\pi,k_y)]$.

With this correction we arrive at the bulk values of the polarization and magnetization of a DSM, which are valid even in a lattice model:

$$P^i = \frac{e\epsilon^{ij}\Theta_j}{4\pi} \sum_{a=1}^{2N} \chi_a g_a \bar{K}_{a,j} \tag{105}$$

and

$$M = \frac{eg}{4\pi} \sum_{a=1}^{2N} \chi_a \epsilon_a. \tag{106}$$

Since the magnetization is isotropic and can even generate currents on edges without low-energy edge-state branches, we might have anticipated (and we actually confirmed numerically above) that the formula would not be dependent of the values of the weak indices $\Theta_j$.

While these results are what one would find via a bulk calculation of the polarization and magnetization, we should check whether they satisfy the correct physical properties for these quantities. The physical manifestation of the polarization is due to the surface theorem that implies that the boundary charge is proportional to the dot product between the polarization and the normal vector to the edge. Hence, a useful definition of polarization in a DSM should reproduce the correct boundary charge. Similarly, the boundary current should be related to the magnetization. We have already carried out this program for two nodes and found, up to some benign ambiguities, that the polarization and magnetizations determined from these bulk formulas match the expected boundary charge and current. Let us now discuss the complications that arise when there are more than two nodes.

Since we have already shown how this works out for two Dirac nodes, our goal is to determine the connection between the bulk value of the polarization in Eq. (105) and the boundary charge for a generic (even) number of nodes. Unfortunately, as we note below, when we go beyond four nodes the connection between the bulk value of the polarization and the boundary charge can be a bit byzantine. Ultimately, the boundary charge is decided by the arrangement and filling of the low-energy edge states that span between edge-projected Dirac points. When multiple edge branches overlap, a coupling between them, even if it is only localized on the boundary, can dramatically effect the boundary charge. In the most general configuration of nodes, the polarization can be calculated as a sum of the (signed) momentum-space locations of the Dirac nodes projected into the corresponding edge BZ. Unfortunately, the *signs* that enter the linear combination must be determined from the edge-state occupation and do not generically match the bulk result. This is similar to the complication found in Ref. [61], where a precise surface theorem for the bulk polarization in a Chern insulator is only defined when the occupations of the edge branches are included. Here the occupation can change at each Dirac point and thus there can be many possibilities for the boundary charge.

The issue of edge-state overlap is challenging to deal with and can have important effects since the edge states are only stable modulo 2, unlike, for example, the chiral boundary states of a 3D Weyl semimetal, which have an integer classification. If we constrain ourselves to four nodes, then we can determine the correct set of signs that enter the calculation of the boundary charge in the presence of generic couplings between overlapping edge states. We now present an appropriate $\mathbb{Z}_2$ modified construction that captures a well-defined value for the boundary charge as determined from the bulk nodal data. Our result shows that using Eq. (105) is still valid as long as one replaces the set of helicities $\chi_a$ with a modified set $\bar{\chi}_a$ (to be defined below) that takes into account the $\mathbb{Z}_2$ cancellation. The modified set of helicities depends on the particular edge projection of interest and can be easily determined from our construction below.

The construction is as follows. First, given a set of Dirac nodes in the bulk, we choose an arbitrary pairing between the nodes with opposite helicity (the reason they have to be opposite helicity is that the edge states always traverse between nodes with opposite helicity). This is always possible since there are an even number of nodes, and an equal number with positive and negative helicity. The final result does not depend on how this pairing is chosen (modulo the ambiguity of the polarization to adding occupied bands with quantized polarization) as long as the sign of $g_a$ is the same for all nodes (otherwise, we have to worry about pairs of nodes contributing a nonvanishing Chern number instead of a polarization as mentioned in Ref. [87]. Next, depending on the value of $g = +1(-1)$ we draw oriented lines between each nodal pair with the arrow pointing from the negative helicity to the positive helicity (positive helicity to negative helicity). The oriented lines should not cross the boundaries of the BZ that is centered at the $\Gamma$ point (for now). We show two examples of this in Figs. 13(a) and 13(b) for two different choices of nodal pairs.

Now, if we want to calculate the boundary charge, we begin by projecting the energy spectrum onto the associated edge BZ. This will generate the nodal locations in the edge BZ, as well as projections of the oriented lines (as shown in the subfigures in Figs. 13(a)–13(d). It is at this stage that the complications

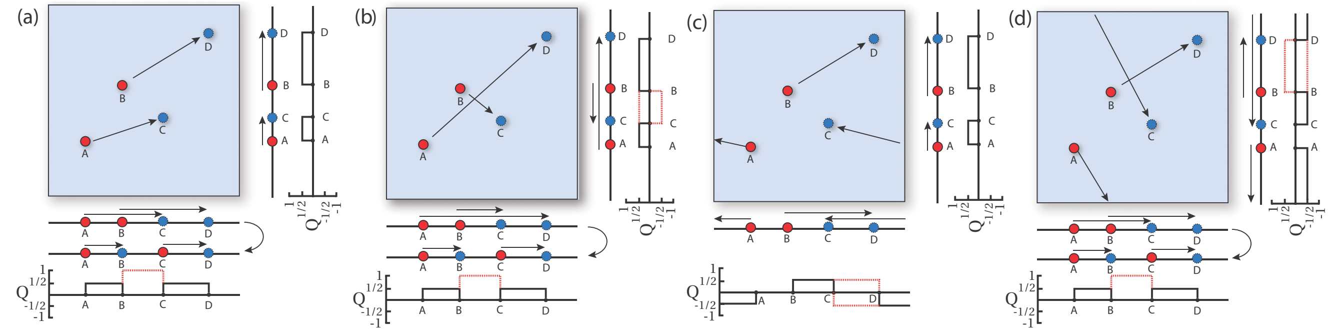

FIG. 13. (Color online) In (a) and (b) we show two arbitrary pairings of the four Dirac nodes with opposite (red and blue) helicity in the 2D square-lattice BZ. In (c) and (d) we show cases when $\Theta_i$ for $i = x,y$ both take their nontrivial values such that the edge states/Dirac-node pairing pass through the BZ boundary. Below and to the right of each 2D BZ we show the projection onto the respective edge BZs. In cases where there are overlapping edge states and a $\mathbb{Z}_2$ cancellation we show the resulting modified effective helicities in a second projected edge BZ subfigure. Finally, we show a diagram for each edge-state projection showing the calculated boundary charge resolved vs $k_{\text{edge}}$ when a uniform half-filled background charge has been subtracted and in units of $e$. The black curves show the results after the cancellation of overlapping edge states and the black + red curves show the result if the overlapping edge-state regions all contribute. In (c) and (d) there are cases where there are red curves below and above the axes. These are contributions coming from each overlapping edge state which cancel when added together.

begin. If the projections of the oriented lines do not overlap at any point in the edge BZ, then one can calculate the boundary charge by (i) multiplying the length of the each of the oriented lines in momentum space by $\frac{e}{2}\frac{1}{2\pi}$, (ii) then multiplying the result by a sign that is +1 if the oriented line points from left to right in the projected edge BZ or −1 if it points from right to left, (iii) and subsequently adding up all of the contributions for all of the pairs of nodes/oriented lines.

However, if there are some overlapping lines, then we have to carefully handle the general $\mathbb{Z}_2$ cancellation, which we can eventually take into account by flipping the helicities of some of the nodes [88]. To determine which helicities should be flipped, we can use the following procedure. For each Dirac node there is one line emanating from itself to its partner. If the number of lines overlapping a Dirac point with the same orientation is even (including its own), then we must flip its helicity; otherwise we leave it unchanged.

In both cases, i.e., whether or not the helicity is flipped, we remove the oriented lines for intervals where there an even number of overlapping lines, which *always* leaves alternating intervals. We show the general rules for flipping helicities in the schematic diagrams in Fig. 14. We can subsequently take these newly determined helicities and plug them into Eq. (105) to generate a value for the polarization that satisfies the surface theorem normal to the chosen edge. We note that this process must be carried out for each choice of edge individually, and the modified helicities for one edge may not work for a different edge. We see an example of this in Fig. 13(a), where when projected onto the $x$ axis the helicities of nodes B and C are changed, while for the projection onto the $y$ axis, none of the nodes have modified helicities. If we now remove the regions over which pairs of overlapping lines exist, as exemplified by the rules in Fig. 14, the boundary charge can be determined geometrically from the remaining oriented lines as in the nonoverlapping case, or simply by plugging into the polarization formula with the modified helicities.

So far this algorithm still misses an important possibility that we have discussed earlier, namely the possibility that the edge states pass through the boundary of the edge BZ $k_{\text{edge}} = \pi$ instead of the origin. We can take these effects into account in our geometric algorithm above by making a simple extension. If $\Theta_{\hat{n}}$ takes its nontrivial (trivial) value then an odd (even) number of oriented lines should pass through the BZ boundaries normal to $\mathbf{G}_N$. This generalizes our discussion above where we have shown zero lines (i.e., an even number) passing through the BZ boundary. We show some examples of nontrivial $\Theta_x$ and $\Theta_y$ in Figs. 13(c) and 13(d). The results are independent of which oriented lines are chosen to pass through the BZ boundary (modulo the ambiguities in the polarization discussed above). If one does not want to include any information about the occupied bands, i.e., does not, or is not able to, calculate $\Theta_{\hat{n}}$, then the overall sign of the polarization is ambiguous, as well as the addition of a quanta of $e/2$ boundary charge per unit cell, which could arise from fully occupied bands carrying a weak invariant. However, this ambiguity essentially exists anyway since one could layer an extra 2D weak TI on top of the 2D DSM and couple the modes of the weak TI to those of the DSM and effectively change between the different values of $\Theta_{\hat{n}}$. Thus, maybe the

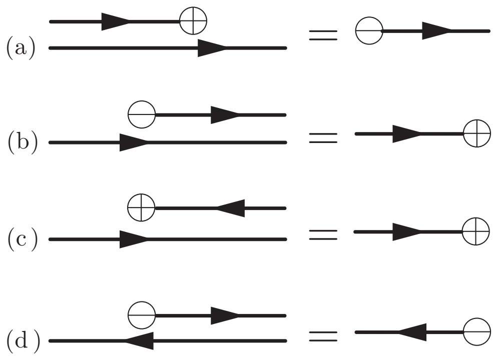

FIG. 14. Rules for helicity modifications and line removal for use with the determination of the boundary charge for Dirac semimetals with four nodes. The signs inside the enclosed circles represent the helicities, and the oriented arrows refer to the nodal pairing in the text.

main advantage to knowing $\Theta_{\hat{n}}$ is to compare with numerical calculations of model systems where a precise model has been specified, and the total weak invariant of the occupied bands is unambiguously specified.

While this algorithm works for four nodes (some extra details are given in the caption of Fig. 13), with six nodes or higher the algorithm is not independent of how the Dirac nodes are chosen to be paired, and it can also give ambiguous results when there are more than two sets of edge states overlapping in a single region. Even if one knows precisely how the edge states connect between the different nodes, e.g., by diagonalizing the system with open boundaries, and thus how the Dirac nodes are "paired," the sign of the polarization due to regions of the edge BZ with more than two overlapping edge states depends, in detail, on the coupling between those edge states. In fact, one can effectively switch the signs of the helicities of the Dirac nodes in pairs (not necessarily the same pairs that are connected via edge states) by modifying boundary terms that couple the different DSM edge-state branches. The final value for the polarization is still a signed sum of the momenta of the bulk Dirac nodes, but the signs that determine the boundary charge have to be determined from the occupation of the edge-state branches, similar in spirit to what is done for the Chern insulator in Ref. [61]. We expect this type of issue to arise whenever the edge states are $\mathbb{Z}_2$ stable as opposed to $\mathbb{Z}$ stable in the chiral case.

Similar complications can arise for the magnetization. The boundary current can depend on the how different pairs of edge states are coupled to each other and on the precise filling of the edge states. Since the edge states are dispersing, even more complicated configurations can arise where the edge states enter the bulk bands arising from other sets of Dirac nodes. It quickly becomes tedious trying to numerically match the boundary current calculation to a bulk result since it depends on both the energies and occupations of the edge states, which can be affected in many ways. In some simple cases where the edge states do not enter the bulk bands, and hence, the occupation of the edge states change at the Dirac nodes themselves, then the bulk result can be recovered. We leave it to future work for an exhaustive treatment of generic edge-state configurations.

### F. General comments about the 2D Dirac semimetal response

*(i) Symmetries of $b_\mu$ in 2D.* Let us discuss the transformation properties of $b_\mu$ under time reversal ($T$), charge conjugation ($C$), and inversion symmetry ($\mathcal{I}$). Since in 2D we know that $b_0$ is proportional to a magnetization, and $b_i$ is proportional to a polarization, we can easily determine their symmetry properties:

$$T : b_0 \to -b_0, \quad C : b_0 \to b_0, \quad \mathcal{I} : b_0 \to -b_0, \tag{107}$$

and

$$T : b_i \to b_i, \quad C : b_i \to b_i, \quad \mathcal{I} : b_i \to b_i. \tag{108}$$

Note that they are both even under $C$, which is due to the fact that our convention for $b_\mu$ in 2D still has the charge factored out. The other thing to note is that $M \sim \mathrm{sgn}(m_A) b_0$ and $P_1^i \sim \mathrm{sgn}(m_A)\epsilon^{ij} b_j$ and $\mathrm{sgn}(m_A)$ is odd under inversion (and parity). When this is taken into account we find that $M$ and $P_1^i$ transform appropriately. In fact, the symmetry properties of $b_\mu$ in 2D match those in 1D.

*(ii) Comments on the electromagnetic response.* The response actions in this section all essentially depend on derivatives of $b_\mu$. Thus, for a homogeneous system there is no charge or current response. This pattern alternates between spatial dimensions. In 1D, 3D, 5D, ..., when the low-energy Fermi surfaces are represented by chiral/Weyl nodes, then the EM response will be a bulk phenomena that does not depend on derivatives of $b_\mu$ whereas in 2D, 4D, 6D, ..., when the Fermi surface arises from Dirac nodes, then the response depends on derivatives of $b_\mu$ which are most commonly generated at interfaces and boundaries.

*(iii) Dependence of the response coefficients on shifts of the origin of the Brillouin zone or the energy reference point.* One might be worried that defining physical quantities in terms of the energy/momentum locations of the nodal points might be problematic since the definitions might depend on arbitrary choices of, e.g., the origin of the BZ or the zero-reference point for energy. Let us consider changing both of these to see what effects they have. In fact, most of the results that follow have been discussed extensively in Refs. [61,85], albeit in a slightly different context, and we go through their arguments here for completeness. For our purpose here we ignore the complication of the $\mathbb{Z}_2$ cancellations.

To illustrate the point, let us take $\bar{K}_{(a)\mu} \to \bar{K}_{(a)\mu} + \Delta k_\mu$. Let us consider the spatial components of $b_\mu$ first, which are related to the polarization $\vec{P}_1$. We can write down the polarization in terms of Bloch wave functions as

$$\vec{P}_{1[\vec{k}_0]} = \frac{e}{(2\pi)^2}\mathrm{Im}\int_{[\vec{k}_0]} d^2k\, \langle u_k|\nabla_k|u_k\rangle, \tag{109}$$

where we have included the dependence of the origin of the BZ by $\vec{k}_0$. Under a change of the origin from $\vec{k}_0 \to \vec{k}_0 + \Delta\vec{k}$, it can be shown generally [61] that the polarization changes by

$$\vec{P}_{1[\vec{k}_0+\Delta\vec{k}]} = \vec{P}_{1[\vec{k}_0]} - \frac{eC_1}{2\pi}\hat{z} \times \Delta\vec{k}, \tag{110}$$

where $C_1$ is the first Chern number. Thus, we see that the polarization itself can seemingly depend on the choice of the origin of the BZ, but only when the Chern number is nonvanishing. When discussing the polarization for the 2D DSM we have been careful to require that $C_1 = 0$ and, hence, we never have this problem. However, even for nonvanishing Chern number it turns out that this issue can be resolved. In fact, there is a discussion in Ref. [61] about a well-defined polarization for Chern insulators. To make sense of this, those authors showed that we need to recall that what is physically meaningful is the *change* in polarization under an adiabatic change of an internal parameter of the system. They show that as long as the same origin in the BZ is used for measuring the initial and final polarization of the system, the results remain consistent.

It is interesting that in our case we find that a shift of $\vec{k}_0$ in Eq. (104) produces exactly the same result as Eq. (110). That is, under $\bar{\mathbf{K}}_{(a)} \to \bar{\mathbf{K}}_{(a)} + \Delta\mathbf{k}$, we see that

$$\Delta P_1^i = \frac{e\epsilon^{ij}}{4\pi}\sum_{a=1}^{2N}\chi_a g_a \Delta k_j = \frac{eC_1\epsilon^{ij}\Delta k_j}{2\pi}, \tag{111}$$

which is the same as Eq. (110), even with a nonzero Chern number. Thus, the effective $\vec{b}$ can change when the origin of the BZ is redefined, but only if the Chern number is nonvanishing.

In this case it is shifted according to the formula derived in Ref. [61] for the charge polarization in a Chern insulator, and any possible ambiguity can be dealt with along those arguments without any issues.

Now we look into what happens with the time component of $b_\mu$. Increasing $b_0$ at a Dirac node is equivalent to reducing the chemical potential at the node or shifting the reference of zero energy for that point. For the purposes of calculations we can interpret a shift in the global reference point in energy as a global change to the chemical potential for the overall system. The magnetization for a Bloch system is defined to be

$$M = \frac{e\epsilon^{ij}}{2\hbar}\int\frac{d^2k}{(2\pi)^2}$$
$$\times \operatorname{Im}\sum_n\int_{\epsilon_{nk}\leqslant\mu}\langle\partial_{k_i}u_{nk}|H_k+\epsilon_{nk}-2\mu|\partial_{k_j}u_{nk}\rangle. \quad (112)$$

Following Ref. [85], we see from this relation that

$$\frac{dM}{d\mu} = -\frac{eC_1}{h} \Rightarrow \Delta M = -\frac{eC_1\Delta\mu}{h}. \quad (113)$$

In fact, this general result exactly matches what we find from our definition of $b_\mu$. Under $\bar{K}_{(a)0} \to \bar{K}_{(a)0} - \frac{\Delta\mu}{\hbar}$, we see that

$$\Delta b_0 = -\frac{1}{2\hbar}\sum_{a=1}^{2N}\chi_a g_a \Delta\mu \Rightarrow \Delta M = -\frac{eC_1\Delta\mu}{h}. \quad (114)$$

Thus, we again see that $b_0$ changes under a redefinition of the origin of energy, but only when the Chern number is nonzero. In this case it changes in the exact same way as a nontrivial Chern insulator.

*(iv) Polarization in a Chern insulator.* Finally, before moving on to the 3D cases, we will discuss a related system with just two Dirac nodes, and nonvanishing time-reversal and inversion-breaking masses. This case, which represents a Chern insulator with broken inversion symmetry, was discussed in Refs. [61,85], where they have defined an electric polarization/magnetization for a Chern insulator. The exact details of the model they considered are a bit different since they use the honeycomb Haldane model [28] with both a nonzero inversion-breaking Semenoff mass and a time-reversal-breaking Haldane mass. In fact, they tune the size of the Haldane mass by changing an adiabatic parameter $\alpha$. What this translates to in the context of our DSM model on the square lattice is that they are working with both an $m_A$ and an $m_B$ turned on. In our language, the analogous Hamiltonian is

$$H = \sin k_x\sigma^x + [-m_A + m_B(\alpha)\sin k_y]\sigma^y$$
$$+ (1 - m - \cos k_x - \cos k_y)\sigma^z, \quad (115)$$

where $m_B = m_A(1+\alpha)$ is tuned as a function of the adiabatic parameter $\alpha$.

Now let us describe the polarization of this system in terms of the edge-state filling. As shown in Appendix B, the energy of the edge states, on edges parallel to the $y$ direction, is given by $E_{L/R}(k_y) = \pm(m_B\sin k_y - m_A)$ where the ± signs are correlated with the left/right edges. In the limit that $m_B \neq 0, m_A = 0$, we have a Chern insulator which is completely inversion symmetric. If we fill all the states with $E \leqslant \mu = 0$ on the edge, then there will be an equal number of filled edge states on both edges, and hence a vanishing polarization. When we turn on an $m_A$, the energy of the edge states is shifted, and filling all the edge states with $E \leqslant \mu$ creates an imbalance between the two edges, depending on how large $m_A$ is. In this case, there are more edge states filled on one edge compared to the other, which leads to a polarization (and possibly magnetization if the nodal energies are shifted). This effect is illustrated in Fig. 15, where we compare the edge spectrum with and without an $m_A$ turned on. The plot has parameters $b_y = \pi/3$, $m_A = 0, 0.1$, and $m_B = 0.2$. The crossing of the edge states moves to the right as we start increasing the $m_A$.

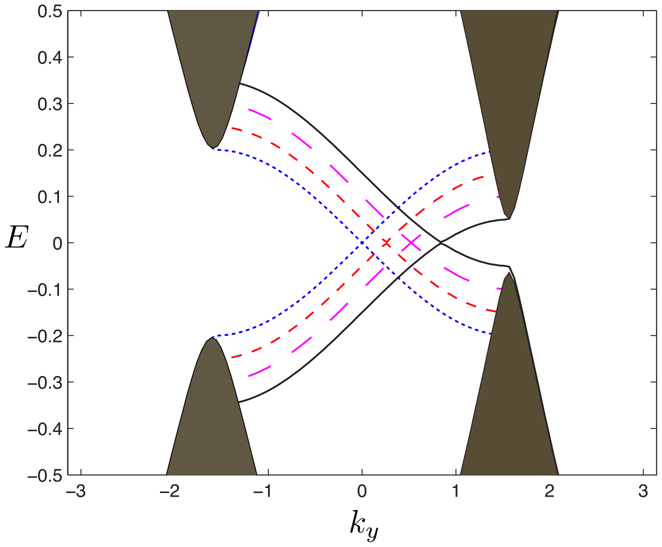

FIG. 15. (Color online) We illustrate the dispersion of the edge states of the model $H = \sin k_x\sigma^x + (m_B\sin k_y - m_A)\sigma^y + (1 - m - \cos k_x - \cos k_y)\sigma^z$ in the limits $m_A = 0, 0.05, 0.10, 0.15$, $m_B = 0.2$ with various dashed lines. The Dirac nodes are located at $\pm\frac{\pi}{2}$. The crossing point has shifted to a nonzero $k_y$ once we turn on an inversion-breaking mass and moves towards one of the Dirac points as $m_A$ is increased further. This leads to a nonzero polarization which is decided by the ratio of $m_A$ and $m_B$, while the sign of the polarization is decided by $m_A$.

In the case when both the masses are finite, we can think of this polarization/magnetization as still arising from an energy and momentum difference, but modified from its original value of $b_\mu$ to new a value we call $\Phi_\mu$. The spatial component is given by $\Phi_i = \sin^{-1}\frac{m_A}{m_B}$. When there is a $T\mathcal{I}$-breaking term $t_p\sin k_y\mathbb{I}$ in our Hamiltonian, we also generate a magnetization dependent on $\Phi_0 = t_p\frac{m_A}{m_B}$. Of course, since the edge states exist between the bulk Dirac nodes this heuristic description only makes sense when $\Phi_\mu$ is lesser than the $b_\mu$ coming from the original gapless bulk Dirac nodes. In the semimetallic limit where $m_A, m_B \to 0$, the polarization/magnetization will actually depend on the ratio of the masses $\Xi = m_A/m_B$ as they are tuned to zero. When the time-reversal-breaking mass is much greater than the inversion-breaking mass, the polarization and magnetization depend on the quantity $\Phi_\mu$ instead of $b_\mu$. However, in the limit that the inversion-breaking mass is much greater than the time-reversal-breaking mass, the polarization/magnetization will depend only on the $b_\mu$ derived from the locations of the Dirac nodes in the gapless limit as we have discussed throughout this section. There is a switch between the two different behaviors when $m_A = m_B \sin k_c$, where $k_c$ is a location of a Dirac node in the gapless limit. In either case, there is no polarization/magnetization without turning on an inversion-breaking mass.

In our general discussions above, if we allowed for the set $g_a$ to take generic values, then to calculate the polarization we would have to use a combination of the results for vanishing Chern number and nonvanishing Chern number on a case-by-case basis for each region of the edge BZ with edge states.

## IV. 3D TOPOLOGICAL SEMIMETALS

There has been a series of recent works that lay out the theory of EM response in Weyl semimetals (WSMs) [34,53–58] and build on the seminal ideas of Nielsen and Ninomiya from three decades ago [31]. We compliment these results in several ways. First, we include lattice-regularized numerical calculations of the response, which show precisely under what conditions the continuum field-theory response calculations can be applied, and most notably when a nonzero current due to the CME can be observed in lattice models. We connect the numerical results with our earlier discussion of the 1D semimetal using a map between the 3D WSM in a uniform magnetic field and many degenerate copies of the 1D semimetal, which can be applied at low energy. We also provide an analytic description of the boundary modes for a lattice model of the WSM, the response behavior of a heterojunction between two different WSMs, and a discussion of the anomaly cancellation which connects the bulk and surface response.

Following this we move on to consider the response of 3D Dirac semimetals in Sec. IV B. Since there is not a similar Nielsen-Ninomiya no-go theorem for lattice Dirac fermions, the Dirac semimetals can come in different varieties. The first type has the Dirac node(s) appearing at the special time-reversal invariant momenta in the BZ. This type is reported to have been realized in $Cd_3As_2$ [47,48]. The other variety is more closely related to the WSM and is essentially a time-reversal and inversion-symmetric version of the WSM where each Weyl node, which exists at generic points in the BZ, is replaced by two copies of the Weyl node, but with opposite chirality; i.e., Weyl nodes are replaced with 3D Dirac nodes at generic points in the BZ. These have recently been confirmed experimentally in $Na_3Bi$ [46].

It is this second type, which was recently dubbed a $Z_2$ nontrivial 3D DSM [49], to which our response theory applies, and unfortunately it is yet to be realized in real material samples. We predict a quasitopological EM response for these materials which is related to the known EM response of the quantum spin Hall insulator [3,62]. In particular, we discuss the response of the 3D DSM when there is a magnetic film in contact with the sample surface. Magnetization domain walls on the surface can generate a line of zero modes along the domain wall and hence give rise to some transport phenomena in these materials including bound charge and currents. We have seen that when discussing Dirac semimetals, we must enforce extra symmetries to provide local stability for the nodes. In 3D, to guarantee local stability of the Dirac nodes, one must require several preserved spatial symmetries, and only certain crystalline space groups support stable nodes [44,49], though we do not focus much more on this in this article.

### A. Response for 3D Weyl semimetal

A simple model for the WSM phase can be formulated with two bands,

$$H_{\mathrm{WSM}} = \gamma \sin k_z \mathbb{I} + \sin k_x \sigma^x + \sin k_y \sigma^y + (2 - m - \cos k_x - \cos k_y - \cos k_z)\sigma^z. \tag{116}$$

This model has two Weyl nodes at $(k_x,k_y,k_z) = [0,0,\pm\cos^{-1}(-m)]$. The identity matrix term generates a difference in energy between the nodes. Around the two nodes, we have linear dispersion $\epsilon_\pm \approx \pm v_F|\mathbf{k}|$, and each of the nodes acts as a monopole of Berry curvature. The Berry curvature flux contained in a Fermi surface surrounding each node can be $\pm 2\pi$ depending on whether the node enclosed is of positive or negative chirality. This property also leads to surface states whose Fermi surfaces consist of open line segments traveling between the projections of the nodes onto the surface BZ [32]. As mentioned before, we follow the convention used in Ref. [54] and define $\vec{b}$ as half the momentum separation in the Weyl nodes and $b_0$ as half the energy difference between them (when there are more than two nodes, this needs to be appropriately generalized). So, in the two-band model we have here, $b_z = \cos^{-1}(-m)$ and $b_0 = (\gamma/\hbar)\sin b_z$.

To calculate the EM response, we can use a continuum description of two Weyl nodes. Following the calculation in Ref. [54], in the continuum approximation we have the following low-energy four-band Hamiltonian:

$$H = \tau^z \vec{\sigma} \cdot \vec{k} + \tau^z b_0 + \vec{\sigma} \cdot \vec{b}. \tag{117}$$

When written as a Lagrangian density coupled to an EM gauge field the four-vector $b_\mu = (b_0,\vec{b})$ appears as an axial gauge field in the action

$$S[b,A] = -\int d^4x\, \overline{\psi}(i\not\partial - e\not A - \not b\gamma^5)\psi, \tag{118}$$

just as it does in the 1D case. We can remove the field $b_\mu$ through a chiral rotation, and hence use the Fujikawa method to evaluate the chiral anomaly which appears due to the noninvariance of the measure under this finite chiral transformation. This is very similar to the derivation we had for the 1D model. This calculation gives us a hint that breaking Lorentz invariance, as we have done in the 1D model, is an essential part of the mechanism to produce a nonzero response. The response action was calculated to be [54]

$$S_{\mathrm{eff}}[A] = -\frac{e^2}{2\pi h}\int d^4x\, \epsilon^{\mu\nu\rho\sigma} b_\mu A_\nu \partial_\rho A_\sigma. \tag{119}$$

We can easily interpret the form of the effective response action since it appears just like an interpolation between the WTI phase generated from a stack of 2D Chern insulators and the normal insulator phase, as was discussed in Sec. I. The current and charge density, assuming $b_\mu$ is homogeneous in space-time, are given by

$$\rho = \frac{e^2}{\pi h}\vec{b}\cdot\vec{B}, \tag{120}$$

$$\vec{j} = \frac{e^2}{\pi h}(\vec{b}\times\vec{E} - b_0\vec{B}). \tag{121}$$

The term in the current involving the electric field is the anomalous QHE of the WSM. The other terms depend on the magnetic field $\vec{B}$ and can be easily interpreted using an analogy to the 1D semimetal, as we now show.

### *1. Understanding the Weyl semimetal response using a quasi-1D description*

To make the mapping to the 1D system we need to apply a uniform magnetic field to the 3D WSM. Consider the two-band model with $b_z \neq 0$. Let us assume that we have a magnetic field turned on in the $z$ direction so that we have $F_{xy} = -B_z$. It is well known, and we reproduce the calculation below, that a Weyl node in a uniform magnetic field has a low-energy zeroth Landau level with dispersion $E_0 = \chi k_z - b_z$ near the Weyl node with chirality $\chi$. It is this level that is responsible for the low-energy EM response in Eq. (120). We see that the zeroth Landau level only disperses along the magnetic field direction and passes through the Weyl node with the direction of the Fermi velocity given by the chirality of the node. Thus, the application of the uniform magnetic field generates a quasi-1D mode at low energy. For a pair of Weyl nodes, as would be found in the simplest WSM, there are two low-energy branches, which, together, effectively form the same low-energy theory as many copies of the 1D (semi-) metal discussed earlier. Thus, the low-energy description is almost identical to the previous 1D semimetal discussion, except that each state has a degeneracy which is set by the total flux of the magnetic field through the $x$-$y$ plane. We denote this degeneracy by $N_\Phi = \frac{B_z L_x L_y}{\Phi_0}$, where $\Phi_0 = \frac{h}{e}$ is the fundamental flux quantum. Thus, in a uniform magnetic field, the low-energy physics of the WSM is equivalent to multiple copies of the 1D semimetal. As seen below, the description is even more apt because, in a lattice-regularized model, the zeroth Landau level modes arising from each Weyl node connect at high energy and form multiple copies of the usual 1D tight-binding band structure.

Let us try to reproduce the charge density predicted in Eq. (120) by using the 1D model. There is a subtlety as to how the states are filled. Of course, if the zeroth Landau level is completely filled or completely empty, then there will be no interesting response. In this case there will be a background charge density of some integer charge per unit cell, but no current will flow in the filled band, and thus there will be no static CME. This was discussed in detail in Ref. [55]. While a filled band can give rise to Lorentz violation because of the inherent lattice structure (e.g., the spatial components of $b_\mu$ can be half a reciprocal lattice vector), the field theory calculations for the semimetal are not sensitive to this. In fact, they can only predict the response from a partially filled band which provides an explicit fractional amount of Lorentz violation (fractional meaning a fraction of a fully filled band). This is similar to the idea of Ref. [71] in which the low-energy structure only determines the fractional part of the response. To match the field-theory calculation we need to assume that the zeroth Landau level is only filled to a chemical potential $\mu = 0$, which implies that the band is partially filled. For example, to calculate the density response we need to count the number of states filled in the zeroth Landau level, which is simply

$$Q = N_\Phi e L_z \int_{-b_z}^{b_z} \frac{dk_z}{2\pi} \tag{122}$$

$$\Rightarrow \rho = \frac{e^2 b_z B_z}{\pi h}, \tag{123}$$

which matches Eq. (120). Before we attempt to understand the properties which lead to a nonzero current, let us look at the zeroth Landau level structure of the WSM in more detail to see how $b_0$ fits into the discussion.

### *2. Zeroth Landau level structure in a Weyl semimetal*

In this section, we proceed to show that $b_0$ can be thought of in a way similar to what we discussed in Sec. II for the 1D model. In the usual case a $b_0$ is produced by shifting the Weyl nodes in energy with respect to each other. We show that when this is the case the zeroth Landau level is shifted in momentum parallel to the magnetic field. So, shifting the nodes in energy acts like an electric field ($k$ is shifted) on the zeroth Landau level. As in 1D, we can also generate a $b_0$ by adding an intrinsic term which generates a velocity difference in the dispersion at the two Weyl nodes; we discuss this case as well. We now show some continuum calculations to justify these statements and then reproduce the same by a simple numerical lattice calculation.

Consider a four-band continuum model for the WSM (a single pair of nodes) where only $b_z \neq 0$ [54]. A possible Hamiltonian is given by

$$H = \tau^z \otimes \sigma^x k_x + \tau^z \otimes \sigma^y k_y + \tau^z \otimes \sigma^z k_z + b_z \mathbb{I} \otimes \sigma^z. \tag{124}$$

To illustrate the effects of a nonvanishing $b_0$ (which is introduced below), we need to include a magnetic field with $k_i \to k_i - eA_i$ and $A_y = B_z x$, where $B_z$ is the uniform magnetic field in the $z$ direction. We note that we have broken translation invariance in the $x$ direction with our choice of Landau gauge, and the eigenvalue equation will be a differential equation in $x$ where we have to replace $k_x \to -i\partial_x$. From now on, this is implicitly assumed. The time-independent Schrödinger equation reads

$$H\psi = E\psi. \tag{125}$$

Following the usual strategy, we can apply $H$ to $\psi$ again to produce $H^2\psi = E^2\psi$. We can evaluate the left-hand side to find

$$\begin{aligned} H^2\psi = \big[&k_x^2 + eB_z \mathbb{I} \otimes \sigma^z + (eB_z)^2(x + k_y/eB_z)^2 \\ &+ k_z^2 + 2b_z k_z \tau^z \otimes \mathbb{I} + b_z^2\big]\psi. \end{aligned} \tag{126}$$

The wave function $\psi$ can be taken to be an eigenstate of $\sigma^z$ for the spin sector, and $\tau^z$ for the orbital sector. Let us denote the eigenvalue of $\sigma^z$ as $\zeta = \pm 1$ and the eigenvalue of $\tau^z$ as $\chi = \pm 1$. Then Eq. (126) is just the harmonic oscillator eigenequation and has the energies

$$E_n(\zeta,\chi,k_z) = \pm\big[2eB_z\big(n + \tfrac{1}{2}\big) + (k_z + \chi b_z)^2 + eB_z\zeta\big]^{1/2}, \tag{127}$$

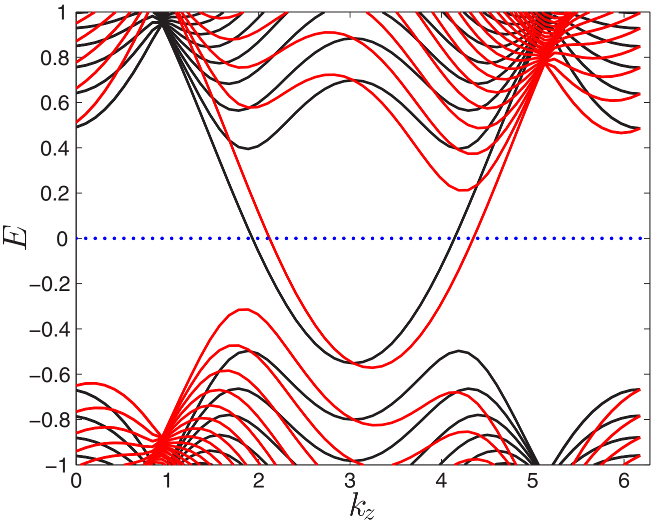

FIG. 16. (Color online) The zeroth Landau level of the WSM in a uniform magnetic field is plotted vs $k_z$ before (in black) and after (in red) switching on a $\gamma$, which gives us $b_0 = (\gamma/\hbar)\sin 2\pi/3 = 0.17$. The blue line is shown to indicate $E = 0$. The model parameters have $b_z = 2\pi/3$, $m = 1/2$, and $L_x = L_y = L_z = 60$ with the magnetic flux per unit cell given by $\phi = 2\pi/60$. $b_0$ was then switched on to plot the curve in red. We see that the Landau level is simply shifted in momentum space and is akin to turning on an external electric field in the 1D model.

with the corresponding wave functions given by

$$\Phi_n(\zeta,\chi,\vec{x}) = N_n\zeta\, e^{-ik_y y - i(k_z+\chi b_z)z} F_n(x + k_y/eB_z) \times \eta, \tag{128}$$

where $N_n$ is a normalization constant, $F_n(x)$ are the Hermite polynomial wave functions, and $\eta = \Lambda(\sigma^z)\otimes\Lambda(\tau^z)$ is a four-component spinor where $\Lambda(\pm 1)$ mean the eigenvectors of $\sigma^z,\tau^z$ given by $\binom{1}{0},\binom{0}{1}$.

To be precise, we need to verify that all of these solutions satisfy Eq. (125). This consistency check eliminates half of the *zero-mode* solutions, and we end up with the result that the zeroth Landau levels have energy,

$$E_0 = \chi k_z - b_z, \tag{129}$$

which depends on the chirality $\chi$ of the Weyl node. This dispersion hits zero energy at $k_z = \pm b_z$, i.e., the location of the Weyl nodes, as expected. These modes also have a degeneracy of $N_\Phi$ for each value of $k_z$ as noted above. In a lattice regularization the zeroth Landau levels of the two Weyl nodes will be connected to each other at high energy (cf. the energy spectrum in Fig. 16).

Now, to turn on a $b_0$ we can add the extra term $\delta H = b_0\tau^z\otimes\mathbb{I}$, which commutes with the initial Hamiltonian. Since it commutes with the original Hamiltonian its primary effect is to shift the energies of the eigenstates. We note that acting on the zeroth Landau level wave functions the energy is shifted by $b_0\chi$, thus leading to the dispersions

$$E_0 = \chi(k_z + b_0) - b_z. \tag{130}$$

This is just a shifted version of the original zeroth Landau level dispersions, and they cross zero energy when $k_z = -b_0 \pm b_z$. So, the conclusion is that $b_0$ shifts the low-energy spectrum of the zeroth Landau level to the right in momentum space, which is the same effect that an external electric field $E_z$ would have. Thus, if the band is partially filled, i.e., when we have explicit Lorentz violation due to the background charge density, this will lead to a nonvanishing current in the presence of an applied B field, but in vanishing applied electric field, which is essentially the CME.

Further pushing the 1D description, let us also show that modifying the relative velocities of the two Weyl points will lead to a similar effect. Consider the Hamiltonian given by

$$\begin{aligned} H = {} & \tau^z\otimes\sigma^x k_x + \tau^z\otimes\sigma^y k_y + \tau^z\otimes\sigma^z k_z \\ & + \mathbb{I}\otimes\sigma^z\alpha k_z + \mathbb{I}\otimes\sigma^z b_z, \end{aligned} \tag{131}$$

where $\alpha \ll 1$. This $\alpha$-dependent term modifies the velocities of propagation in the $z$ direction of the two Weyl nodes. It effectively changes $b_z \to b_z + \alpha k_z$ from our previous analysis. The entire argument for the energies of the zeroth Landau levels from before carries through here, too, and we find a modified zeroth Landau level dispersion of

$$E_0 = \chi k_z - b_z - \alpha k_z. \tag{132}$$

This dispersion crosses zero at $k_z = b_z/(\chi - \alpha) \approx \chi b_z - \alpha b_z + \mathcal{O}(\alpha^2)$. So, near zero energy this term behaves like a momentum shift in the Landau level, and this should give us a nonzero current as we have shown in the 1D model in Sec. II.

To verify these continuum results, we can perform calculations using a simple lattice regularization of the above continuum model. The Hamiltonian is given by

$$\begin{aligned} H = {} & \gamma\sin k_z\mathbb{I} + \sin k_x\sigma^x + \sin k_y\sigma^y \\ & + (2 - m - \cos k_x - \cos k_y - \cos k_z - t_{\mathrm{NNN}}\sin 2k_z)\sigma^z, \end{aligned} \tag{133}$$

where the term proportional to $\gamma$ will cause a shift in energy of the Weyl nodes, and the next-nearest-neighbor term proportional to $t_{\mathrm{NNN}}$ causes a change in the velocity of the zeroth Landau level near the two Weyl nodes.

For $\gamma = 0$ and $t_{\mathrm{NNN}} \neq 0$ the Weyl nodes are given by solving

$$\cos k_z + t_{\mathrm{NNN}}\sin 2k_z = m, \tag{134}$$

which gives us two solutions for $k_z$. Let us try to extract the low-energy Hamiltonians near the nodes in the limit that $t_{\mathrm{NNN}} \ll 1$ by writing the two solutions as $k_z = \pm\kappa_z + \delta k$. We have

$$\cos(\pm\kappa_z + \delta k) + t_{\mathrm{NNN}}\sin(\pm 2\kappa_z + 2\delta k) = m. \tag{135}$$

We can subtract the two equations to find

$$2\sin\kappa_z\sin\delta k - 2t_{\mathrm{NNN}}\cos 2\delta k\sin 2\kappa_z = 0. \tag{136}$$

Using the small angle approximations $\sin\delta k \approx \delta k$, $\cos 2\delta k \approx 1$, we are left with

$$\delta k = 2t_{\mathrm{NNN}}\cos\kappa_z \approx 2t_{\mathrm{NNN}}m. \tag{137}$$

Thus, we see that a nonzero velocity change will lead to a momentum shift of $2t_{\mathrm{NNN}}m$ at the nodal energies. Comparing with the continuum calculation we see that $\alpha b_z = -2t_{\mathrm{NNN}}m$.

We show the numerical results of $\gamma = 0.2$, $t_{\mathrm{NNN}} = 0$ in Fig. 16 and $\gamma = 0$, $t_{\mathrm{NNN}} = 0.2$ in Fig. 17. In both cases we see that near $E = 0$ the zeroth Landau levels are shifted.

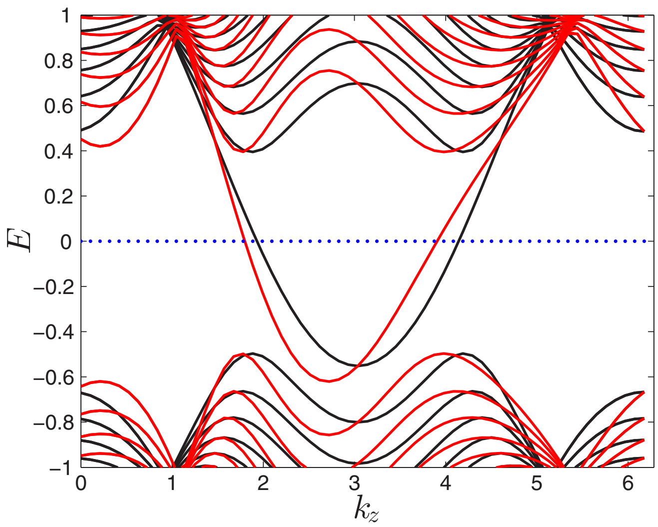

FIG. 17. (Color online) The zeroth Landau level is plotted vs $k_z$ before (in black) and after (in red) switching on a $b_0$ using the NNN velocity term. The blue line is shown to indicate $E = 0$. The model had $b_z = 2\pi/3$, $m = 1/2$, and $L_x = L_y = L_z = 60$, with $\phi = 2\pi/60$. We then switch on a term to change the velocity of the two Weyl nodes with $t_{\text{NNN}} = 0.2$. The shift we expect is then given by $2t_{\text{NNN}}m \approx 0.2$, as seen in the figure. In effect, near $E = 0$ the zeroth Landau level is shifted.

### *3. Response and anomaly cancellation in Weyl semimetals with inhomogeneous $b_\mu$*

So far, all of the response properties that we have considered for the WSM have assumed that $b_\mu$ was constant in space-time. This will not be the case in systems which have boundaries or interfaces across which $b_\mu$ will naturally change. In this section, we closely examine what the bulk action implies for the surface/interface action and how the whole system remains gauge invariant. We recall that the response action is

$$S = -\frac{e^2}{2\pi h}\int d^4x\epsilon^{\mu\nu\rho\sigma}b_\mu A_\nu \partial_\rho A_\sigma. \tag{138}$$

Now, when we take the functional derivative of $S$ with respect to $A_\alpha$ to extract the current, we have to be careful about the behavior of $b_\mu$,

$$j^\alpha = \frac{e^2}{\pi h}\epsilon^{\alpha\mu\rho\sigma}b_\mu \partial_\rho A_\sigma + \frac{e^2}{2\pi h}\epsilon^{\alpha\mu\rho\sigma}A_\sigma \partial_\rho b_\mu. \tag{139}$$

This gives us the usual current we expect for the AHE and CME, along with a term which depends on derivatives of $b_\mu$, but is not manifestly gauge invariant since it depends directly on $A_\mu$. This signals the presence of an anomaly that will arise whenever $b_\mu$ changes.

The Callan-Harvey mechanism provides a straightforward way of understanding this result [89]. To be explicit, let us assume we have an interface in the $x$ direction, located at $x = x_0$, where $b_z$ jumps from a finite value to zero. This is the case in the lattice models we studied in the previous section. Under a gauge transformation ($A_\mu \to A_\mu - \partial_\mu \lambda$) the action transforms as

$$\begin{aligned}\delta_\lambda S &= -\frac{e^2}{2\pi h}\int d^4x\epsilon^{\mu\nu\rho\sigma}b_\mu(-\partial_\nu\lambda)\partial_\rho A_\sigma \\ &= -\frac{e^2}{2\pi h}\int d^4x\epsilon^{\mu\nu\rho\sigma}\partial_\nu b_\mu \partial_\rho A_\sigma \lambda \\ &= \frac{e^2}{2\pi h}\int d^4x\epsilon^{zx\rho\sigma}b_z\delta(x-x_0)\partial_\rho A_\sigma \lambda \\ &= \frac{e^2 L_z b_z}{2\pi h}\int dydt\epsilon^{\rho\sigma}\partial_\rho A_\sigma \lambda \neq 0.\end{aligned} \tag{140}$$

Thus, in order for the system to be gauge invariant there must be localized fermion modes where $b_z$ jumps (except in the case when it jumps in the $z$ direction, since $\delta_\lambda S = 0$ in that case). In fact, for the simple WSM models we have considered, we know that there are such surface/interface states, and they are just straight-line Fermi arcs that stretch between the Weyl nodes projected onto the surface/interface BZ. For a nonzero $b_z$ and a surface with normal vector $\hat{x}$ (just like the interface considered in the previous paragraph), the surface states have a chiral dispersion given by $E(k_y,k_z) = k_y$ at low energy. These chiral modes give rise to the usual chiral anomaly. There is an independent chiral fermion for each value of $k_z$, but the surface states only exist between the Weyl nodes, i.e., only for $-b_z \leqslant k_z \leqslant b_z$. Each 1D chiral mode generates an anomalous contribution to the variation of the boundary/interface action under a gauge transformation [89,90],

$$\delta_\lambda S_{bdry} = -\frac{e^2}{2h}\int dydt\epsilon^{\rho\sigma}\partial_\rho A_\sigma \lambda, \tag{141}$$

where $\rho,\sigma = 0,y$. To calculate the total variation due to all of the modes, we can convert the sum over the independent $k_z$ modes to an integral which generates a factor of $\frac{L_z}{2\pi}2b_z$. We thus find

$$\delta_\lambda S_{bdry}^{(Tot)} = -\frac{e^2 L_z b_z}{2\pi h}\int dydt\epsilon^{\rho\sigma}\partial_\rho A_\sigma \lambda, \tag{142}$$

which exactly cancels the variation coming from the bulk action. Equation (142) is called the consistent anomaly. The consistent anomaly leads to an anomalous Ward identity for current conservation on the edge,

$$\partial_\mu j^\mu_{bdry} = -\frac{e^2 L_z b_z}{2\pi h}\epsilon^{\rho\sigma}\partial_\rho A_\sigma = -\frac{e^2 N_c}{2h}\epsilon^{\rho\sigma}\partial_\rho A_\sigma, \tag{143}$$

where $N_c$ is the total number of modes in the interface/boundary Fermi arc.

Going back to the bulk-current response in Eq. (139), we see that the current naturally splits into two terms: (i) $j^\alpha_{\text{bulk}} = \frac{e^2}{\pi h}\epsilon^{\alpha\mu\rho\sigma}b_\mu \partial_\rho A_\sigma$ and (ii) $\tilde{j}^\alpha_{bdry} = \frac{e^2}{2\pi h}\epsilon^{\alpha\mu\rho\sigma}A_\sigma \partial_\rho b_\mu$. For our interface configuration we find

$$\tilde{j}^\alpha_{bdry} = -\frac{e^2}{2\pi h}\epsilon^{\alpha z x\sigma}A_\sigma b_z \delta(x-x_0). \tag{144}$$

If we integrate this current density over $x$ and $z$, we can combine this current with the current from the consistent anomaly to arrive at the Ward identity for the covariant anomaly (the anomaly that contains all contributions to the boundary current),

$$\partial_\alpha\left(j^\alpha_{bdry}+\tilde{j}^\alpha_{bdry}\right)=-\frac{e^2L_zb_z}{\pi h}\epsilon^{\alpha\sigma}\partial_\alpha A_\sigma. \tag{145}$$

This covariant anomaly precisely matches the bulk-current inflow from $j^x_{\text{bulk}}$ into the boundary/interface. Note that although we have assumed a model which has simple Fermi arcs, the chiral anomaly result is very robust and does not depend on the exact form of the surface-state dispersion, or any other details, only that the states are chiral. Thus, we expect it to hold in any generic model, even in the cases when the Fermi arcs are not straight-line segments, but are curved. This result clearly shows that while the bulk action would predict a gauge-variant response, it is compensated by the surface Fermi arcs states. The same is true when we do not have a physical boundary, but a region in which $b_\mu$ varies in space-time. When $b_\mu$ varies there are two contributions to the boundary current, one arising from the bulk action itself, and the other from the consistent anomalous current required of the boundary states in order to preserve gauge invariance of the bulk and boundary.

#### 4. Numerical results

After these analytic arguments let us explicitly test the predictions with numerical calculations. For illustration, we probe two effects: (i) the CME, which we have tried analyzing using a mapping to the 1D model, and (ii) the charge density response in a system with an inhomogeneous $\vec{b}$. We do this in the context of the two-band WSM lattice model,

$$\begin{aligned}H&=\gamma\sin k_z\mathbb{I}+\sin k_x\sigma^x+\sin k_y\sigma^y\\&\quad+(2-m-\cos k_x-\cos k_y-\cos k_z)\sigma^z,\end{aligned} \tag{146}$$

where $\gamma$ generates a nonzero $b_0$. It is important to note that to perform our numerical calculations we fill the states up to $E=0$; i.e., all states with $E\leqslant 0$ are filled. To illustrate an example of the CME, in Fig. 18 we have plotted the current along the $z$ direction as a function of $b_0$ in the presence of a uniform magnetic field, but no electric field. The predicted current density from the model, assuming a magnetic field in the $z$ direction, is given by

$$j_z=-\frac{eb_0B_z}{\pi h}. \tag{147}$$

The lattice calculation is shown in Fig. 18, and we find exactly this result. For this calculation the magnetic field is implemented using Peierls substitution. We use a Landau gauge to retain translation invariance in one of the directions in the $xy$ plane, and the $z$ direction is also translation invariant. The magnetic field is restricted to have rational flux per unit cell for the spectrum to remain periodic in momentum space.

Another simple effect to test is the density response at an interface where $\vec{b}$ changes. With $B_z\neq 0$, we should have

$$\rho=\frac{eb_zB_z}{\pi h}. \tag{148}$$

So, if we vary $b_z$ in the $x$ direction (with open boundary conditions the $xz$ surfaces host nontrivial surface states), one would expect a varying charge density. In fact, one can see that this is exactly reproduced in numerics and the resultant charge density is plotted in Fig. 19. The bulk charge follows what is predicted by the action in the continuum calculation.

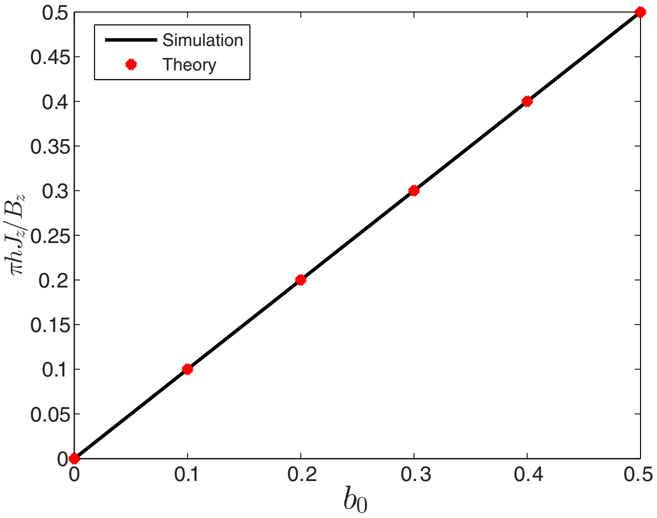

FIG. 18. (Color online) The current is plotted vs $b_0$ for the two-band model of the WSM. The current is linear and the slopes match almost exactly. This plot is generated for $L_x=30$ and the flux per plaquette is $\phi=-2\pi/L_x$. We use $L_y=30$, $L_z=30$, and $b_z=\frac{\pi}{2}$ to generate this plot.

### B. Electromagnetic response of a 3D Dirac semimetal

There has been a lot of recent work predicting and measuring materials candidates for 3D Dirac semimetals [44–48]; however, we are interested in the so-called $Z_2$ nontrivial 3D Dirac semimetals [46,49], where the Dirac nodes appear in

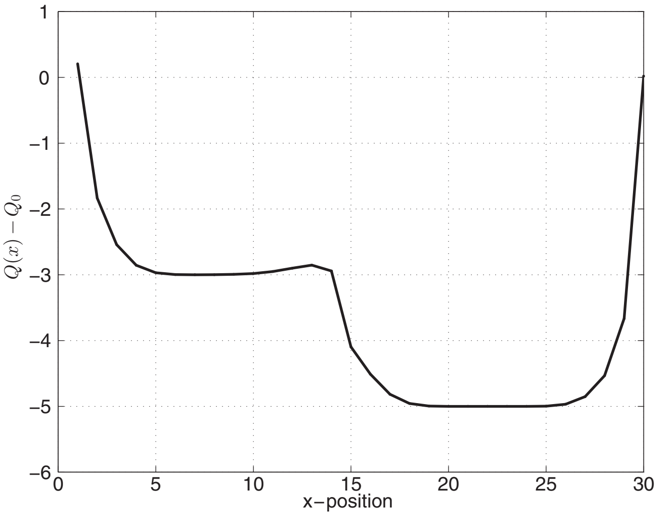

FIG. 19. The charge density is plotted vs position in the $x$ direction with open boundary conditions. The system is composed of a WSM with $b_{z,L}=\pi/5$ for $0<x<L_x/2$ and $b_{z,R}=\pi/3$ for $L_x/2<x<L_x$. The total number of sites in the $x$ direction was $L_x=30$ with magnetic flux per unit cell in the $x$-$y$ plane $\phi=-2\pi/30$. Also, $L_z=30$ and $L_y=30$. The bulk charge density is given by $N_x=-L_zL_yb_zB_z/4\pi^2=-3,-5$, as predicted by the action.

pairs and can exist at generic points in the BZ, as recently measured in $Na_3Bi$. In this section, we discuss an interesting EM probe of this type of 3D DSM and show that it can be derived from the response properties of the 2D time-reversal invariant quantum spin Hall insulator [3,21,22,62,91,92].

In fact, analogous to all of our previous constructions, we can think of the 3D DSM as a layered 2D TI, and in this case it is formed from coupled layers of the quantum spin Hall (QSH) system. The layer construction has aided the discussion and analysis of the other topological semimetals, and we will see that it is very helpful in this case as well. Thus, we begin this section by first examining the response of the QSH insulator itself, since the results can immediately be generalized to stacks of QSH insulators and, hence, the 3D DSM. After reviewing the response of the QSH insulator, we discuss the analogous properties of the DSM and numerically validate our analytical calculations.

The QSH system has an unusual EM response given by [3,62]

$$S[A]=\frac{e}{2\pi}\int d^3x\,\epsilon^{\mu\nu\sigma}A_\mu\partial_\nu\Omega_\sigma, \tag{149}$$

where $\Omega_\mu$ is a gauge field which encodes configurations of inhomogeneous adiabatic perturbations. We clearly define what this means in the following section. Essentially, the configurations of $\Omega_\mu$ are related to possible mass-inducing perturbations of a Dirac-type Hamiltonian. As a consequence of this response term, a magnetic film deposited at the edge of the QSH insulator can generate a localized charge density or adiabatic current if the magnetization is space or time dependent, respectively [3,62]. The edge of the QSH insulator is itself a robust 1D massless Dirac fermion if we preserve time-reversal symmetry. A magnetization on the edge will open a gap, and through the well-known Jackiw-Rebbi mechanism [93] a spatial domain wall in the magnetization will trap a low-energy midgap mode. This mode signals a bound charge of $Q_b=\pm e/2$. Additionally, if the magnetization on one side of the domain wall begins to rotate as a function of time, a quantized adiabatically pumped charge current can flow along the edge through the magnetic junction. Reference [3] showed that both of these phenomena could be derived from Eq. (149). This is the EM signature of the QSH insulator and is closely tied to the response of the 3D DSM.

Now we can construct a stack of QSH insulators. If the layers are weakly coupled, then we will get the conventional WTI state [73,94,95]. If we increase the strength of the interlayer coupling so that we close the bulk gap, we will generate the 3D DSM phase. Just as with the WSM, the edge states of the QSH layers forming the DSM will survive in a certain region of momentum space and will connect the various 3D Dirac nodes with Fermi-surface arcs. We can easily extrapolate the response action of the QSH insulator to the 3D DSM to find

$$S[A]=\frac{e}{2\pi^2}\int d^4x\epsilon^{\mu\nu\rho\sigma}b_\mu A_\nu\partial_\rho\Omega_\sigma. \tag{150}$$

We discuss the consequences of this action below, but first we more carefully recount the analysis for the 2D QSH insulator since its formulation is not as widely known, and we wish for this article to be relatively self-contained.

### *1. Response from the second Chern number*

The discussion in this section closely follows the arguments in Ref. [3], although we only reproduce the necessary ingredients for our discussion of the 3D DSM and leave out some of the details which can be found in the aforementioned reference. In general, the response of the QSH insulator is derived from the second Chern number $C_2$, which is a 4D topological invariant. Since the QSH exists in 2D, the Bloch Hamiltonian is only parametrized by two numbers $k_x,k_y$, which is not enough to generate a nonzero $C_2$. Thus, to probe the EM response properties of the QSH state, we need to couple the system to two additional parameters $\theta(\mathbf{x},t),\phi(\mathbf{x},t)$, which represent adiabatic parameters which vary slowly in space and time so that momentum space is still approximately well defined. The gauge field $\Omega_\mu$ introduced above is a function of space and time, but only through its dependence on $\theta$ and $\phi$.

To be explicit, consider the QSH Hamiltonian given by

$$\begin{aligned}H_{\mathrm{QSH}}(\mathbf{k},\hat{n})&=\sin k_x\Gamma^1+\sin k_y\Gamma^2+(\cos k_x+\cos k_y-2)\Gamma^0\\&\quad+m\sum_{a=0,3,4}\hat{n}_a\Gamma^a,\end{aligned} \tag{151}$$

in which $m>0$, $\Gamma^a$ are the $4\times 4$ Dirac matrices and $\hat{n}=(n_3,n_4,n_0)$ is a 3D unit vector. The $\Gamma^{1,2,3,4}$ are all odd under inversion and time reversal ($T^2=-1$), while $\Gamma^0$ is even under both. The unperturbed QSH insulator will have $n_3=n_4=0$ but $n_0\neq 0$. If we let $\hat{n}$ vary slowly as a function of space-time, we can parametrize it using two adiabatic space-time-dependent parameters via $\hat{n}(\mathbf{x},t)=(\sin\theta(\mathbf{x},t)\cos\phi(\mathbf{x},t),\sin\theta(\mathbf{x},t)\sin\phi(\mathbf{x},t),\cos\theta(\mathbf{x},t))$. The results of Ref. [3] show that in the low-energy continuum limit of $H_{\mathrm{QSH}}$ expanded around the $\Gamma$ point, the gauge curvature of $\Omega$ is directly related to the skyrmion density of the unit vector $\hat{n}$ as

$$\partial_\mu\Omega_\nu-\partial_\nu\Omega_\mu=\tfrac{1}{2}\hat{n}\cdot\partial_\mu\hat{n}\times\partial_\nu\hat{n}. \tag{152}$$

Using Eq. (149) we can write the current in terms of this skyrmion density as

$$j^\mu=\frac{e}{8\pi}\epsilon^{\mu\nu\rho}\hat{n}\cdot\partial_\nu\hat{n}\times\partial_\rho\hat{n}. \tag{153}$$

Now let us consider an important example case. Assume that we have a QSH sheet with a static edge parallel to the $y$ direction and a pair of static magnetic films next to each other on the edge. If the magnetizations of the two films are opposite, this will produce a domain wall on the edge with a magnetization that varies as a function of $y$. In that case we find the parametrization $\theta=\theta(x)$ and $\phi=\phi(y)$. At the location of a $\theta$ domain wall between $\theta=0$ and $\theta=\pi$ there will be an edge. At the location of a $\phi$ domain wall between $\phi=0$ and $\phi=\pi$ there will be a magnetic domain wall. In this geometry we find

$$j^0=\frac{e}{4\pi}\hat{n}\cdot\partial_x\hat{n}\times\partial_y\hat{n}=\frac{e}{4\pi}\sin\theta\times\frac{d\theta}{dx}\frac{d\phi}{dy}. \tag{154}$$

Due to the dependence on the derivatives of $\theta$ and $\phi$, the charge density is localized wherever $\theta(x)$ and $\phi(y)$ are *both* changing. If we have a sharp magnetic domain wall on a sharp edge, then all of the charge density will be localized at the magnetic domain wall, i.e., where the $\theta$ and $\phi$ domain walls intersect. The total charge in the neighborhood of this intersection can be calculated by integrating over $x,y$. The integration is easily performed since the integrand is a total derivative in $x$ and $y$. We just get the integral over the solid angle swept out by $\theta$ and $\phi$, which for this configuration is half the sphere, i.e., $\pm 2\pi$. This yields a bound charge $Q_b = \pm 2\pi \frac{e}{4\pi} = \pm \frac{e}{2}$.

We can similarly find an adiabatic pumping current by having a static edge $[\theta = \theta(x)]$ and sweeping the relative magnetization between the two magnetic films on the edge as a function of time $[\phi = \phi(t)]$ [62]. Everything carries through in exactly same way and we find

$$j^y = \frac{e}{4\pi} \sin\theta \times \frac{d\theta}{dx}\frac{d\phi}{dt}. \tag{155}$$

We can again integrate over $x,t$ to get the total charge transported as the relative magnetization angle sweeps through a full cycle to find, as $\phi : 0 \to 2\pi$, we have $\Delta Q = e$. This current is localized wherever $\theta$ has a sharp change in its value, i.e., on the edge.

We can understand the physics underlying the QSH response from the microscopic behavior of the edge states. In the low-energy limit near the Dirac point, we can write the Hamiltonian for one of the edges of the QSH system (say an edge at $x = 0$) as

$$H_{\text{edge}}(k) = k\sigma^z, \tag{156}$$

where $k$ is the momentum of the coordinate along the edge and we have set the edge velocity to unity. Coupling the magnetic layer to the edge will induce a gap from the proximity exchange (Zeeman) coupling. If the magnetization lies in the plane, then the effective Hamiltonian becomes

$$H_{\text{edge}} + H' = k\sigma^z + m_x\sigma^x + m_y\sigma^y. \tag{157}$$

Let us choose a configuration with $m_x = 0$ and $m_y = m(y)$ is a shifted step function which goes from a negative value to a positive value at $y = 0$. It is well known [93] that this Hamiltonian has an exponentially localized zero mode at the domain wall of $m(y)$ given by

$$\psi = e^{-\int_0^y m(y')dy'} \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \tag{158}$$

when the mass jumps from negative to positive as $y$ increases. On a periodic edge, $m(y)$ will have to have two domain walls to maintain the proper boundary conditions, and the edge will have two zero modes, one at each domain wall. These localized zero modes carry a half charge each. This is the same result found from Eq. (154). To complete the story in the language above, the QSH system itself has a nontrivial value of the $Z_2$ invariant $\theta = \pi$ [3,91]. Thus, its boundary gives a natural place where $\theta$ has a jump from $\pi$ to 0. The spatial dependence of the $\phi$ parameter is due to the magnetization-induced mass.

We can also generate an adiabatically pumped current. To see this we can add a slow, time-dependent perturbation to the edge Hamiltonian,

$$H_{\text{edge}}(k) = k\sigma^z + m\sin\phi(t)\,\sigma^y + m\cos\phi(t)\,\sigma^x, \tag{159}$$

where $\phi(t) = 2\pi t/T$. The mass terms are periodic in time with a period of $T$. From the original work by Thouless [80] we know that as $\phi \to \phi + 2\pi$ an integer amount of charge will be pumped, in this case just a single electron per cycle. This is the same current which is reported in Eq. (155).

### *2. Response of the Dirac semimetal*

Now that we have finished the discussion for a single QSH layer, we are ready to move on to the 3D DSM. We can start from the QSH Hamiltonian, but we need to modify it to include tunneling in the $z$ direction due to the coupled layers. The following model can be used:

$$H_{\text{DSM3}}(\mathbf{k},\hat{n}) = \sin k_x\Gamma^1 + \sin k_y\Gamma^2 + (\cos k_x + \cos k_y + t_z\cos k_z - 3)\Gamma^0 + m\sum_{a=0,3,4}\hat{n}_a\Gamma^a. \tag{160}$$

If the 2D layers are in the QSH phase, then when the tunneling term $t_z$ is weak, the system will be in a WTI phase. As it becomes stronger, eventually the gap will close at one of the time-reversal invariant momenta along the $k_z$ axis and generate a pair of Dirac nodes, hence entering the 3D DSM phase. In a recent work [49] this has been called a $Z_2$ nontrivial Dirac semimetal. From the previous patterns of the EM response, and the known response of the QSH insulator, we can immediately write the response action

$$S_{\text{3D}} = \frac{e}{2\pi^2}\int d^3x dt \epsilon^{\mu\nu\rho\sigma} b_\mu A_\nu \partial_\rho \Omega_\tau \tag{161}$$

for the 3D DSM, where $2b_\mu$ is the energy-momentum separation of the Dirac nodes. We now have a natural family of 2D Bloch Hamiltonians parametrized by $k_z$ $H_{k_z}(k_x,k_y)$. Each of the 2D Hamiltonians, for $k_z$ not at a Dirac node, represents at 2D time-reversal invariant insulator and is classified by the same $Z_2$ invariant as the QSH insulator. As $k_z$ passes through a Dirac node the $Z_2$ invariant jumps from trivial to nontrivial, or vice versa. Thus, one of the regions of $k_z$ between the Dirac nodes will harbor nontrivial topological QSH insulators and thus generate edge states. For each $k_z$ in the topological range we have a contribution of one layer of QSH to the total EM response. This is the meaning of Eq. (161). Reference [49] has shown that this type of semimetal requires a uniaxial rotation symmetry to locally stabilize the Dirac nodes. Our model has such a symmetry ($C_4$ rotation around the $z$ axis) and thus represents a stable $Z_2$ nontrivial DSM. We leave a more general symmetry analysis of the EM response to future work.

Let us look at some examples of the physical phenomena associated with Eq. (161). Just like the case of a single QSH layer, to get a nontrivial response we need to apply a magnetic film to a boundary with nontrivial surface states. As shown in Fig. 20, for Dirac nodes separated in $k_z$ we can coat the $xz$ boundary plane with a magnetic layer. A translationally invariant magnetic domain wall parallel to the $z$ axis in the magnetic layer (see Fig. 20) will create a line of low-energy modes which do not disperse with $k_z$. Thus, for each $k_z$ that contributes a boundary mode we bind a half charge. We numerically calculated the bound charge at a domain wall as a function of $k_z$ and the result is shown in Fig. 21. In this figure we see exactly $e/2$ charge contributed for each value of $k_z$ between the Dirac nodes. For this calculation we used the mass parameter $m = 0.5$ and varied $\phi$ and $\theta$ as functions of $y$ and $x$, respectively, according to Eq. (160).

The bound charge response will also occur in a time-reversal invariant WTI system; however, a new phenomenon which is not available in the WTI is the generation of a current along the domain wall in the direction along which the Dirac nodes are

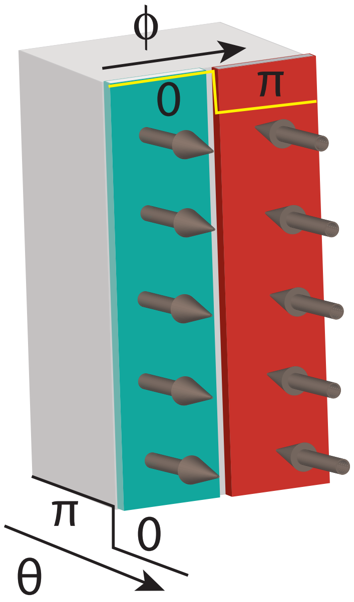

FIG. 20. (Color online) Setup to generate an EM response in a 3D Dirac semimetal. To get a nonzero response, there must be two adiabatic parameters $\theta$ and $\phi$. The parameter $\theta$ represents an interpolation between a 3D Dirac semimetal with, for example, $b_z \neq 0$ to a trivial insulator with $b_z = 0$. The parameter $\phi$ represents a magnetization domain wall on the $xz$ surface plane. There will be a branch of low-energy fermion modes trapped on the domain wall which can bind charge or can carry current if $b_0 \neq 0$.

separated. This can occur if the Dirac nodes are not at the same energy, and it is the 3D DSM analog of the CME in WSMs.

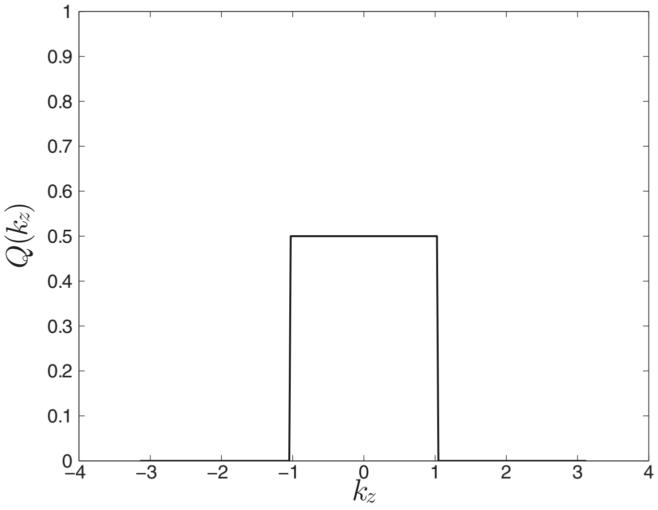

FIG. 21. The localized charge on a magnetic domain wall on the surface of a 3D DSM resolved vs $k_z$, i.e., the direction in which the Dirac nodes are separated in momentum space. We note that there is a half charge bound at the domain wall only for each state satisfying $|k_z| < \cos^{-1} m$. In the plot, we have used $m = 0.5$, which means $b_z = \frac{\pi}{3}$.

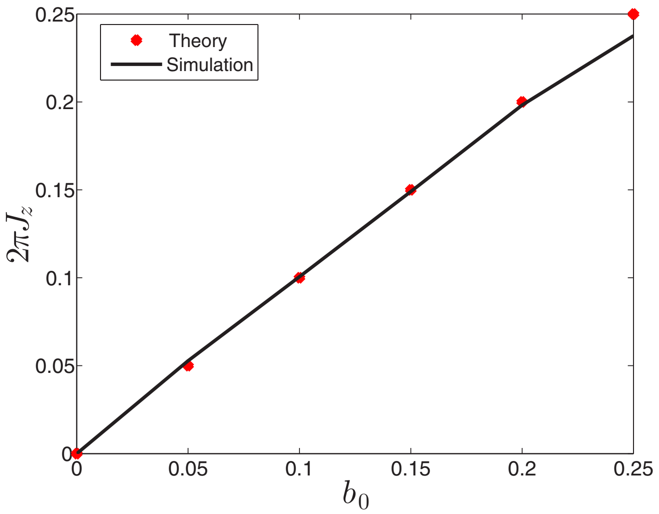

FIG. 22. (Color online) The total current localized at the magnetic domain wall is plotted vs $b_0$ for the 3D DSM. The expected value of the total current localized on the domain wall is $\frac{eb_0}{2\pi}$. The system size is a cube of $L = 30$ lattice sites in every direction with $b_z = \frac{\pi}{2}$. We used open boundary conditions in both the $x, y$ directions and periodic boundary conditions in the $z$ direction. The red dots are the theoretical result and the black line is the numerical result. The deviation arises due to the importance of lattice effects at larger values of $b_0$.

We can generate an energy difference, i.e., $2b_0$, in our Hamiltonian by adding the term $\gamma \sin k_z \mathbb{I}$ to the Hamiltonian in Eq. (160). When we have a magnetic domain wall and a nonzero $b_0$, the localized domain wall states will disperse with energy $E_{dw} = 2\gamma \sin k_z$ and this leads to a nonzero current. We calculated this current numerically as shown in Fig. 22. With a $b_0 \neq 0$, the current is being generated due to the dispersion of the localized edge modes which now have to traverse between the two Dirac nodes in a continuous fashion. The total current localized on the domain wall is given by

$$\begin{aligned} J_z &= \frac{eb_0}{2\pi^2} \int d^2x\, (\partial_x \Omega_y - \partial_y \Omega_x) \\ &= \frac{eb_0}{2\pi^2} \int d\theta d\phi\, \frac{1}{2} \sin\theta = \frac{eb_0}{2\pi}, \end{aligned} \tag{162}$$

which matches the numerical calculation well until $b_0$ is large enough for lattice effects to become important. This mechanism for current generation is reminiscent of the orbital magnetization generation due to currents produced by dispersing edge states in the 2D Dirac semimetal.

## V. DISCUSSION AND CONCLUSIONS

In this paper, we have explored the EM responses of topological semimetals with pointlike Fermi surfaces in various spatial dimensions. We have seen that, generically, the quasitopological contribution to the response depends on a 1-form $b_\mu = (b_0, \vec{b})$, which represents an energy difference $(2b_0)$ and momentum separation $(2b_i)$ of the nodes. To study these systems we first introduced a simple 1D model of a metal, which illustrated some of the general physical principles as well as helped to understand some response properties of the 3D WSM in a uniform magnetic field. This approach works because of the fact that the 1D response is embedded in the

3D WSM response, similar to the 1D TI charge-polarization response being embedded in the 3D axion electrodynamic response [3].

After 1D we then moved on to the case of the 2D Dirac semimetal which was constructed from layered 1D TIs that are stacked and coupled together. The gapless Dirac nodes which occur in this model each have a Chern-Simons response which, when written in terms of the EM gauge field, gives a polarization/magnetization which can be defined for a semimetal and is encoded in the momentum-space positions and energies of the nodes. In this case, an energy difference between the nodes led to an edge current (bulk orbital magnetization) and a momentum separation between the nodes led to a a boundary charge (bulk polarization). The $\mathcal{TI}$ symmetry ensures that the Dirac nodes are locally stable and immediately leads to the viable definition of charge polarization even in this gapless system.

We then moved onto 3D, where we studied the properties of the WSM and tested the predictions of the continuum field theory results with some numerical examples. Furthermore, we showed the precise anomaly cancellation calculation that connects the surface and bulk degrees of freedom. From there, the 3D DSM was then analyzed from the perspective that it is a layered QSH system. The 3D Dirac nodes separate $Z_2$ trivial regions of momentum space from $Z_2$ nontrivial regions, and the resultant response follows from the existence of these nontrivial QSH layers. As such, when a magnetic film is applied to a boundary with nontrivial surface states, we get boundary modes, and bound charge, localized on domain walls of the magnetization. Additionally, a nonzero energy difference in the 3D Dirac nodes produces a localized current which runs along the domain wall.

There are several natural areas to pursue from this point. We have shown that we can understand some topological semimetals, i.e., those with point Fermi surfaces, by stacking topological states in one dimension lower. We only considered the simplest cases in this article, and we have barely scratched the surface of the different 1D and 2D states that could be coupled together to form 2D and 3D semimetal states. Additionally, one could take 1D topological wires and stack them into planes and then subsequently take those planes and stack them into 3D to get a secondary WTI, or, if the interwire coupling is strong enough, a 3D semimetal with line-node Fermi surfaces. In this case the Lorentz violation enters as a 2-form $b_{\mu\nu}$ that will couple to the EM field via $\int d^4x\epsilon^{\mu\nu\rho\sigma}b_{\mu\nu}F_{\rho\sigma}$. In the simplest case, this will give rise to lines of Dirac nodes which will have a polarization and magnetization response controlled by $b_{\mu\nu}$.

Along with determining the EM response, the stacking construction is also useful for discussing the properties of dislocations in WTIs and topological semimetals [17,96]. Additionally, the general pattern of metal/semimetal responses is as follows. For a $D$-dimensional sample, a conventional Fermi surface is a $(D-1)$-dimensional surface in momentum space. The response of this metal is given by a $D$-form $b_{\mu_1\cdots\mu_D}$, which is equivalent to a current via the identification $j^{\alpha}_{(b)} \sim \epsilon^{\alpha\mu_1\cdots\mu_D}b_{\mu_1\cdots\mu_D}$. Generically, when the Fermi surface has dimension $D-q$ (codimension $q$), then the response is controlled by a $D-q+1$ form. These higher forms are sure to yield interesting physical predictions and connections with protected boundary modes. We will explore both of these directions further in Ref. [72].

Another immediate application of our results is to the bulk response action of the 3D topological crystalline insulator protected by mirror symmetry [7,97,98]. It has been shown that alloys of PbSnTe exhibit a mirror-symmetry protected topological phase. If we consider the [001] surface, then there will be four Dirac nodes which all have the same helicity [97], i.e., in our notation for the 2D Dirac semimetal $\chi_a = +1$ for $a = 1,2,3,4$. To define the response, we also need to know the momentum positions of the Dirac nodes and the sign of the local mass terms at the Dirac nodes. Since the four nodes are symmetrically arranged in the surface BZ, let us parametrize their 2D momenta as $\vec{K}_1 = (K,0)$, $\vec{K}_2 = (0,L)$, $\vec{K}_3 = (-K,0)$, and $\vec{K}_4 = (0,-L)$.

The two relevant possibilities for the response coefficients are the Chern number $C_1 = \frac{1}{2}\sum_{a=1}^{4} g_a\chi_a$ and $\vec{b} = \frac{1}{2}\sum_{a=1}^{4} g_a\chi_a\vec{K}_a$. Since the chiralities are all the same, we can replace these with $C_1 = \frac{1}{2}\sum_{a=1}^{4} g_a$ and $\vec{b} = \sum_{a=1}^{4} g_a\vec{K}_a$, where we recall that $g_a$ is the sign of the local mass term at the $a$th Dirac node. Reference [97] showed that there are four possibilities for the $g_a$ due to inversion-breaking perturbations, one particular case being $g_1 = g_4 = -g_2 = -g_3 = 1$. For this set of mass signs $C_1 = 0$ and $\vec{b} = (-K,L)$. If we include the other choices of mass sign, we get the four possibilities $\vec{b} = (\pm K, \pm L)$. This is interesting because if the top surface and bottom surface have different values of $\vec{b}$, then there will be an interfacial region where the polarization changes and there will be bound charge proportional to the difference. Microscopically, this bound charge arises because the domain wall between the two regions of the surface will bind low-energy fermion modes. It would be interesting to explore this further to develop the full response theory, but we will leave this for future work.

## ACKNOWLEDGMENTS

We acknowledge useful conversations with B. A. Bernevig, G. Y. Cho, V. Chua, V. Dwivedi, and especially O. Parrikar. We acknowledge support from ONR Award No. N0014-12-1-0935.

## APPENDIX A: TRANSFORMATION FROM A DIRAC SEMIMETAL ON THE SQUARE LATTICE TO THE HONEYCOMB LATTICE

In this section, we show that graphene can be thought of as an array of $(1+1)$-dimensional TI wires. Let us begin with the 1D TI given by the Bloch Hamiltonian

$$H(\mathbf{k}) = t_x(1 + m - \cos k_x a)\sigma^x + t_x \sin k_x a\,\sigma^y, \tag{A1}$$

where $t_x, m$ are parameters and $a$ is the lattice constant. The system is gapped for all values of $m$ except $m = 0$ or $m = 1$. Let us now induce tunneling in the $y$ direction. In the following, the assumption of $y$ being perpendicular to $x$ is not needed. We could have this tunneling along an oblique direction and orthogonality is not required. In this case the BZ is not a simple square, but it can be a parallelogram. With hopping in the $y$ direction, consider the modified

Hamiltonian

$$\begin{aligned} H(\mathbf{k}) = {}& [t_x + t_x m - t_x \cos k_x a \\ & - t_\theta \cos(k_x a \cos\theta + k_y a \sin\theta)]\sigma^x \\ & + [t_x \sin k_x a + \bar{t}_\theta \sin(k_x a \cos\theta + k_y a \sin\theta)]\sigma^y, \end{aligned} \tag{A2}$$

where we have parametrized the $y$ direction by an angle $\theta$ with respect to the initial $x$ axis.

Let us now look at the graphene Hamiltonian. It is given by

$$\begin{aligned} H_G(\mathbf{k}) = {}& -(t_1 + t_2 \cos\vec{k}\cdot\vec{a}_1 + t_3 \cos\vec{k}\cdot\vec{a}_2)\sigma^x \\ & + (t_2 \sin\vec{k}\cdot\vec{a}_1 + t_3 \sin\vec{k}\cdot\vec{a}_2)\sigma^y, \end{aligned} \tag{A3}$$

where $\vec{a}_{1,2} = \sqrt{3}a(\cos(\pi/6), \pm\sin(\pi/6))$. For an easier comparison, let us rotate this system in the counterclockwise direction in real space by an angle $\pi/6$. The two lattice vectors are now given by $\vec{a}_1 = \sqrt{3}a(\cos(\pi/3), \sin(\pi/3))$ and $\vec{a}_2 = \sqrt{3}a\,(1,0)$. Labeling $\sqrt{3}a = b$, we reduce the Hamiltonian to

$$\begin{aligned} H_G(\mathbf{k}) = {}& -[t_1 + t_2 \cos(k_x b \cos\pi/3 + k_y b \sin\pi/3) \\ & + t_3 \cos k_x b]\sigma^x + [t_3 \sin k_x b \\ & + t_2 \sin(k_x b \cos\pi/3 + k_y b \sin\pi/3)]\sigma^y. \end{aligned} \tag{A4}$$

We note that the Hamiltonians in Eqs. (A4) and (A2) are the same with the following identifications. $t_1 \to -(t_x + t_x m)$, $t_2 \to t_\theta$, $t_3 \to t_x$ with the additional constraint $t_\theta = \bar{t}_\theta$.

Let us now set all parameters in our model (A2) to 1 except for $\bar{t}_\theta$. From our previous statement we know that this will be exactly the same as graphene when $\bar{t}_\theta = t_\theta = 1$. We want to show that the effect of deforming $\bar{t}_\theta$ away from this point is to move the Dirac nodes around in the BZ. Let us look at the gapless points of our model which are the solutions to

$$\sin(k_x a) + \bar{t}_\theta \sin(k_x a \cos\theta + k_y a \sin\theta) = 0, \tag{A5}$$

$$\cos(k_x a) + \cos(k_x a \cos\theta + k_y a \sin\theta) = 1 + m. \tag{A6}$$

In the limit that $\bar{t}_\theta = 1$, we have $(\frac{\pm 1}{a}\cos^{-1}(\frac{1+m}{2}), \frac{\mp(1+\cos\theta)}{a\sin\theta}\cos^{-1}(\frac{1+m}{2}))$ as the gapless points. On the other hand, if $\bar{t}_\theta = 0$, we have $(0, \pm\cos^{-1}(m))$ as the gapless points. As long as $|1+m| < 2$, and $0 \leqslant \bar{t}_\theta \leqslant 1$, we get two gapless points in the spectrum but their location depends generically on the model parameters. In this paper, we always use the model in Eq. (A2) in the limit of $t_x = 1$, $t_\theta = 1$, $\bar{t}_\theta = 0$ for describing Dirac semimetal physics with two bands.

## APPENDIX B: EXACT SOLUTION FOR BOUNDARY STATES IN TOPOLOGICAL SEMIMETAL LATTICE MODELS

In this Appendix we study the edge states of the various topological semimetal lattice models. The solution can be found analytically for the Dirac-type models we have been using following the results of Refs. [23,99]. We begin by solving for the edge states of the two-band lattice Dirac model, i.e., the minimal model for (1 + 1)- and (2 + 1)-dimensional TIs. We then go on to modify these models to form Dirac and WSM states and solve for their boundary modes.

### 1. Exact solution for edge states of the lattice Dirac model

Consider the model given by

$$\begin{aligned} \mathcal{H} &= \epsilon(k) I_{2\times 2} + d_a(k)\sigma^a, \\ d_a(k) &= (A\sin(k_1), d_2(k_2), M(k)), \\ M(k) &= M - 2B[2 - \cos(k_1) - \cos(k_2)], \end{aligned} \tag{B1}$$

where $d_2(k_2)$ is an unspecified, but odd, function of $k_2$ and $A, B, M$ are model parameters. Let us fix the sign of $A > 0$ and $B > 0$. Additionally, we assume that $\epsilon(k) = 0$ for now, but we add it back in later. Note that with $\epsilon(k) = 0$ and $d_2(k_2) = -d_2(-k_2)$ the model is particle-hole symmetric with the symmetry operator $C = \sigma^x$; it is also inversion symmetric with $\mathcal{I} = \sigma^z$. The energy eigenvalues are given by

$$\begin{aligned} E_\pm &= \pm\sqrt{d_a d_a} \\ &= \pm\sqrt{A^2 \sin^2(k_1) + d_2^2(k_2) + M^2(k)}. \end{aligned} \tag{B2}$$

This spectrum is a gapped insulator as long as $\sqrt{d_a d_a} \neq 0$. One gapless critical point of this model occurs when $k_1 = k_2 = M = 0$ and for $M < 0$ ($M > 0$) the model is in a trivial (topological) insulator phase.

When the system is tuned to the nontrivial phase there are gapless edge states which can be shown explicitly in a finite strip geometry or a cylinder geometry. Let us assume that the system has boundaries at $x_1 = 0, L$ and is infinite in the $x_2$ direction. Since we have an inhomogeneous system with open boundaries, we need to Fourier transform the Bloch Hamiltonian back from $k_1$ to $x_1$ via the substitution

$$c_{\vec{k}} = \frac{1}{\sqrt{L}}\sum_j e^{ik_1 x_1(j)} c_{k_2,j}. \tag{B3}$$

This reduces the Hamiltonian to

$$\begin{aligned} \mathcal{H} &= \sum_{k_2,j}\left(\mathcal{M} c^\dagger_{k_2,j} c_{k_2,j} + \mathcal{T} c^\dagger_{k_2,j} c_{k_2,j+1} + \mathcal{T}^\dagger c^\dagger_{k_2,j+1} c_{k_2,j}\right), \\ \mathcal{M} &= A\sin(k_2)\sigma^2 - 2B\left[2 - \frac{M}{2B} - \cos(k_2)\right]\sigma^3, \\ \mathcal{T} &= \frac{iA}{2}\sigma^1 + B\sigma^3. \end{aligned} \tag{B4}$$

Since we are interested in the exponentially localized edge states, we focus on a solution ansatz of the form

$$\psi_\alpha(j) = \lambda^j \phi_\alpha, \tag{B5}$$

where $\lambda$ is a complex number, $j$ is the site index in the $x_1$ direction, and $\phi_\alpha$ is a two-component spinor with $\alpha = 1, 2$. We first look for a solution at $k_2 = 0$, and since the Hamiltonian is particle-hole symmetric, the midgap edge state for this momentum will occur at $E = 0$. Acting with the Hamiltonian at $k_2 = 0$ on our ansatz yields the equation

$$\left[\frac{iA}{2}(\lambda^{-1} - \lambda)\sigma^1 + B(\lambda + \lambda^{-1})\sigma^3 + \mathcal{M}(0)\sigma^3\right]\phi = 0.$$

Multiplying this equation on both sides by $\sigma^3$ gives us

$$\frac{A}{2}(\lambda^{-1} - \lambda)(i\sigma^3\sigma^1)\phi = -[B(\lambda + \lambda^{-1}) + \mathcal{M}(0)]\phi. \tag{B6}$$

The operator $i\sigma^3\sigma^1$ has eigenvalues ±1. First consider $i\sigma^3\sigma^1\phi = -\phi$, under which Eq. (B6) becomes a quadratic equation in $\lambda$, which can be solved to find

$$\lambda_{(1,2)} = \frac{-\mathcal{M}(0) \pm \sqrt{\mathcal{M}^2(0) + (A^2 - 4B^2)}}{A + 2B}. \tag{B7}$$

Thus, from the quadratic equation we have two $\lambda$ solutions for the $-1$ eigenvalue (chirality) of $i\sigma^3\sigma^1$. For every solution $\lambda$ we find that $\lambda^{-1}$ is a solution for $i\sigma^3\sigma^1\phi = +\phi$, and thus for each eigenvalue of $i\sigma^3\sigma^1$ there are two possible values of $\lambda$. Let us label the eigenstates of $i\sigma^3\sigma^1$ as $\phi_\pm$ corresponding to the chiralities. The most general edge state can by written as

$$\psi_j(k_2 = 0) = \left(a\lambda_{(1)}^j + b\lambda_{(2)}^j\right)\phi_+ + \left(c\lambda_{(1)}^{-j} + d\lambda_{(2)}^{-j}\right)\phi_-, \tag{B8}$$

but to satisfy open boundary conditions we must have $a = -b$ and $c = -d$ since $\phi_\pm$ are linearly independent. Additionally, since the mode must be normalizable, we can only keep positive or negative powers of $\lambda$ and thus only one normalizable mode exists (on each edge) as long as the $\lambda$ do not lie on the unit circle, i.e., $|\lambda_{(1,2)}| \neq 1$. If $|\lambda_{(1,2)}| = 1$ an edge-state solution does not exist at all. We also note that solutions with eigenvalues $\lambda$ and $\lambda^{-1}$ are localized on opposite edges of the system based on the form of Eq. (B8).

Now let us generalize this solution for $k_2 \neq 0$. We see that the term $\cos(k_2)$ simply acts as a shift of the parameter $M$ and can be easily accounted for. We also see that $[i\sigma^3\sigma^1,\sigma^2] = 0$ and clearly $[i\sigma^3\sigma^1, I_{2\times2}] = [\sigma^2, I_{2\times2}] = 0$. So the terms $d_2(k_2)\sigma^2$ and $\epsilon(k_2)I_{2\times2}$ can simply be included as $k_2$-dependent shifts of the energy. These terms change the energy dispersion of the edge states, but the eigenstates remain the same. The energy for the edge state for any $k_2$ is given by

$$E_\pm(k_2) = \epsilon(k_2) \mp d_2(k_2). \tag{B9}$$

Importantly, this dispersion does not hold across the entire $k_2$ BZ because there will exist some values of $k_2$ where the values of $\lambda$ coming from a solution of

$$\begin{aligned} \lambda_{(1,2)}(k_2) &= \frac{-m(k_2,M) \pm \sqrt{m(k_2,M)^2 + (A^2 - 4B^2)}}{A + 2B},\\ m(k_2,M) &= -2B[2 - M/2B - \cos(k_2)], \end{aligned} \tag{B10}$$

do not yield normalizable modes. For the edge states to be normalizable, we have to satisfy the condition that $|\lambda_{(1,2)}| \neq 1$, which can be reduced to

$$-2B < m(k_2,M) < 2B \tag{B11}$$

for each $k_2$. The special points in $k_2$ space where the inequalities become equalities are places in the energy spectrum where the edge states merge with the delocalized bulk states. Beyond these special values of $k_2$ the edge states no longer exist. This result, which consists of the dispersion, wave functions, and conditions for normalizability, represents the full analytic solution of the lattice edge states.

### 2. Edge theory for two-dimensional semimetal

Based on the solution for the two-band Dirac model we can immediately adapt it to the case of topological semimetal states with minor modifications. First, let us consider the (2 + 1)-dimensional Dirac semimetal including the possibility of the inversion-breaking ($m_A$) and time-reversal-breaking ($m_B$) mass terms discussed in Sec. III. The Hamiltonian takes the form

$$\begin{aligned} \mathcal{H} &= \epsilon(k)I_{2\times2} + d_a(k)\sigma^a,\\ d_a(k) &= (A\sin k_1, m_A + m_B\sin k_2, M(k)),\\ M(k) &= M - 2B[1 - \cos k_1 - \cos k_2],\\ \epsilon(k) &= \gamma\,\sin(k_2). \end{aligned}$$

Depending on the values of $M$ and $B$, this Hamiltonian can have Dirac nodes at $(0,\pm k_0)$ where $k_0 = \cos^{-1}(-M/2B)$. For a cylinder geometry with open boundary in the $x_1$ direction and periodic boundary conditions in the $x_2$ direction, this model will have edge states when the Dirac nodes exist. The edge states will occur between the Dirac nodes, but depending on the values of $M$ and $B$ they either stretch between the nodes within the BZ or across the BZ boundaries. For a choice such that they stretch within the BZ, the energies of the edge-state branches on the two edges are given by

$$E_\pm = \gamma\,\sin(k_2) \mp |m_A + m_B\sin(k_2)| \quad |k_2| < k_0. \tag{B12}$$

The restriction on the range of $k_2$ arises from a modified condition on normalizability through the relation

$$\begin{aligned} -2B &< m(k_2,M) < 2B,\\ m(k_2,M) &= -2B[1 - M/2B - \cos(k_2)]. \end{aligned} \tag{B13}$$

We can observe several interesting details from Eq. (B12). First we see that if we let $m_A = \gamma = 0$ but $m_B \neq 0$, then the dispersion matches that of the edge states of the (2 + 1)-dimensional Chern insulator [28], as it must since the $m_B$ term is exactly the mass term required to convert a 2D Dirac semimetal into a Chern insulator. If only $m_A$ is nonzero and $m_B = \gamma = 0$, then we get two flat bands, one band on each edge. Finally, if we have $\gamma \neq 0$ and $m_A \neq 0$ but $m_B = 0$, then the two flat bands from the previous case will each disperse, and at half filling there will be bound currents on each edge that, in the limit $m_B \to 0$ give rise to the magnetization discussed in Sec. III. This matches our expectation because if $M$ and $B$ are tuned to values where $k_0 \neq 0$, as we have assumed, then for nonzero $\gamma$ there will be an energy difference between the two Dirac nodes given by $\Delta E = 2|\gamma\sin k_0|$.

### 3. Edge theory in the case of the Weyl semimetal

The WSM also has a Hamiltonian which is given by the form of Eq. (B1) where

$$\begin{aligned} \mathcal{H} &= \epsilon(k_2,k_3)I_{2\times2} + d_a(k)\sigma^a,\\ d_a(k) &= (A\sin k_1, A\sin k_2, M(k)),\\ M(k) &= M - 2B[2 - \cos k_1 - \cos k_2 - \cos k_3], \end{aligned}$$

where we can let $\epsilon(k)$ be a generic function of $k_2,k_3$. This Hamiltonian has two gapless Weyl nodes for $|M/2B| < 1$ at $(k_1,k_2,k_3) = (0,0,\pm k_0)$ where $k_0 = \cos^{-1}(-M/2B)$. Let us assume again that our system has boundaries at $x_1 = 0,L$ and that it is periodic in the other two directions. The main change between this case and the previous ones is that the condition for existence of these edge states at each momentum gets modified because the mass $m(k,M)$ is now parametrized by $k_2$ and $k_3$.

The new normalizability condition that must be satisfied is given by

$$\begin{gathered}-2B < m(k,M) < 2B,\\ m(k,M) = -2B[2 - M/2B - \cos k_2 - \cos k_3].\end{gathered} \tag{B14}$$

The edge-state energies in this case are given by $E_\pm = \epsilon(k_2,k_3) \mp |A\sin k_2|$.

Let us consider a simple case first where $\epsilon(k) \equiv 0$. We want to consider the structure of the boundary modes on a surface with the normal vector in the $x$ direction and the surface BZ in the $(k_2,k_3)$ plane. If we set the chemical potential to zero, we see that there exist Fermi arcs in this plane when $E_\pm = \mp|A\sin k_2| = 0$, which allows for $k_2 = 0,\pi$ and does not explicitly depend on $k_3$. The correct value of $k_2$ depends on the particular choice of $M$ and $B$, so, without loss of generality, let us choose $k_2 = 0$. The boundary-state existence condition of Eq. (B14), which does depend on $k_3$, can be simplified to give us the condition that boundary states are only present when $|k_3| < k_0$. Thus, for this case there exist Fermi arcs that are straight lines which go from $(k_2,k_3) = (0,-k_0)$ to $(k_2,k_3) = (0,k_0)$ in the surface BZ.

To get more nontrivial Fermi-arc shapes, inversion symmetry needs to be broken to lift the degeneracy between the arcs on the two edges. Let us consider the Hamiltonian given by (B14) with $\epsilon(k) = \gamma\sin k_3$. The energy is given by $E_\pm = \gamma\sin k_3 \mp |A\sin k_2|$. With the chemical potential again set at $\mu = 0$ and, for example $\gamma = A/2$, we see that the points in the Fermi arc must satisfy $\sin k_3 = \pm 2\sin k_2$ and Eq. (B14). The solutions to these constraints are complicated functions of $(k_2,k_3)$ and must, in general, be solved numerically.

#### 4. Tunneling between edge states

In this section, we use our model of the boundary states for the topological semimetals to study properties at interfaces between semimetals with different Lorentz-violating parameters, and thus different boundary-state structures. Let us consider the interface between two, semi-infinite 2D DSMs first. Assume that the interface is at $x = 0$, with parameters for $x \leqslant 0$ given by $A,B,M,\gamma$ and for $x > 0$ given by $A',B',M',\gamma'$.

The lattice Hamiltonian for $x \leqslant 0$ is given by

$$\begin{aligned}\mathcal{H} = \left[\sum_{j,k_2=-\infty}^{j=-1} H_j(k_2)\right] + \mathcal{M}c^\dagger_{0,k_2}c_{0,k_2}\\ + \mathcal{T}c^\dagger_{0,k_2}c_{1,k_2} + \mathcal{T}^\dagger c^\dagger_{1,k_2}c_{0,k_2},\end{aligned} \tag{B15}$$

where $H_j$ is the lattice Hamiltonian we have been previously using. To be specific,

$$\begin{aligned}\mathcal{M} &= \gamma\sin k_2\mathbb{I} + (m_A + m_B\sin k_2)\sigma^2\\ &\quad -2B\left[1 - \frac{M}{2B} - \cos k_2\right]\sigma^3, \qquad \text{(B16)}\\ \mathcal{T} &= \frac{iA}{2}\sigma^1 + B\sigma^3.\end{aligned}$$

The Hamiltonian for $x > 0$ has a similar form, just with different parameters. We notice that there is a natural hopping term to connect the two systems. The matrix element for tunneling from site 0 to site 1 is $\mathcal{T}^\dagger$ and the matrix element for tunneling from site 1 to site 0 is $\mathcal{T}$.

Assume that the edge states are of chiralities $c,c'$ which take on the values $+1,-1$. The chirality of the state is simply defined as its eigenvalue under the $i\sigma^3\sigma^1$ matrix discussed in the previous section. The states on the left edge and right edge are given by $\phi_c,\phi_{c'}$, respectively. The Hamiltonian in the edge subspace is given by

$$H = \begin{pmatrix}\langle\phi_c|\mathcal{M}|\phi_c\rangle & \langle\phi_c|\mathcal{T}|\phi_{c'}\rangle\\ \langle\phi_{c'}|\mathcal{T}^\dagger|\phi_c\rangle & \langle\phi_{c'}|\mathcal{M}'|\phi_{c'}\rangle\end{pmatrix}. \tag{B17}$$

We can evaluate the matrix elements in each case by using the fact that $|\phi_\pm\rangle$ are eigenstates of $-\sigma^2$. When the chiralities are opposite, i.e., $cc' < 0$, we have $\langle\phi_\pm|\mathcal{M}|\phi_\pm\rangle = \gamma\sin k_2 \mp (m_A + m_B\sin k_2)$, $\langle\phi_+|\mathcal{T}|\phi_-\rangle = \langle\phi_+|\mathcal{T}|\phi_-\rangle^\dagger = B - A/2$. Off-diagonal terms turn out to be zero if $cc' > 0$, i.e., we have $\langle\phi_+|\mathcal{T}|\phi_+\rangle = \langle\phi_-|\mathcal{T}|\phi_-\rangle = 0$. So, in the case of $cc' > 0$, which is to say we have the same chirality for the edge states, the tunneling Hamiltonian is given by

$$H = \frac{\gamma+\gamma'}{2}\sin k_2\mathbb{I} \pm \left[m_A + \left(m_B + \frac{\gamma-\gamma'}{2}\right)\sin k_2\right]\sigma^3. \tag{B18}$$

We see that the edges do not mix and are only completely gapped when the inversion symmetry is broken (i.e., $m_A$ nonzero). When they are of opposite chiralities, the tunneling Hamiltonian is given by

$$\begin{aligned}H &= \frac{\gamma+\gamma'}{2}\sin k_2\mathbb{I}\\ &\quad \pm \left[m_A + \left(m_B + \frac{\gamma-\gamma'}{2}\right)\sin k_2\right]\sigma^3 + (B - A/2)\sigma^1.\end{aligned} \tag{B19}$$

We see that the term $B - A/2$ when nonzero acts like a mass term and gaps the edge out in this case. In the models we consider, $A = 1, B = -1/2$ and $A - B/2 \neq 0$. In the case when the edge modes have the same chirality the ± signs in Eq. (B18) refer to the chirality itself. In the case when the edge modes have the opposite chirality the ± signs in Eq. (B19) refer to whether the left edge has + or − chirality.

An important thing to notice is that $M$ and $M'$ do not enter the edge Hamiltonians; however, it still has an important effect. The above analysis tells us that the edge modes can gap each other out when they both exist at the same momentum $k_2$. However, it is $M$ and $M'$ that control where the Dirac nodes are and therefore the domain of existence of the edge states in $k_2$. So, those edge states on one edge with a momentum $k_2$ which do not have a counterpart on the other edge will remain gapless regardless. Thus, the edge states will only be removed if the domain of existence overlaps in the two systems.

#### 5. Tunneling in Weyl semimetals

Let us start off with the Hamiltonian which is of the same flavor as before with

$$\begin{aligned}\mathcal{M} &= A\sin k_2\sigma^2 - 2B\left[2 - \frac{M}{2B} - \cos k_2 - \cos k_3\right]\sigma^3,\\ \mathcal{T} &= \frac{iA}{2}\sigma^1 + B\sigma^3.\end{aligned} \tag{B20}$$

Let us assume that again that we have an edge at $x = 0$ and the same setup as the 2D Dirac semimetal. For $x \leqslant 0$ we have parameters $A, B, M$ and for $y > 0$ we have parameters $A', B', M'$. Our analysis from the previous section helps us immensely here. The edge states $|\phi_c\rangle$ are again eigenvectors of $-\sigma^2$. The edge Hamiltonian when we have same chiralities on the two edges is again given by

$$H = \pm A \sin k_2 \mathbb{I}. \tag{B21}$$

On the other hand, when the edge states have opposite chiralities, the edge Hamiltonian is

$$H = \pm A \sin k_2 \sigma^3 + (B - A/2)\sigma^1. \tag{B22}$$

So, yet again, when the edges have opposite chiralities, the term $B - A/2$ acts like a mass term and gaps the modes out. This is, of course, only valid if the edge states exist at the same $k_3$. Edge states with a momentum $k_3$ which do not have a counterpart on the other edge will remain gapless. The $\pm$ signs are related to the same definitions in the previous section. There could be more complications when a term $\epsilon(k_2, k_3)\mathbb{I}$ is added to the Hamiltonian. This modifies the surface Fermi arcs from being straight lines to some other complicated structure. When this happens, only those states on the surface which are degenerate at the same momenta $k_2, k_3$ gap each other out.

## APPENDIX C: *K*-MATRIX FORMALISM

The action in Eq. (99) can be rewritten as

$$S_{\text{eff}} = \frac{e^2}{4h} \int d^3x \, K_{ab} \epsilon^{\mu\nu\rho} A_{(a)\mu} \partial_\nu A_{(b)\rho}, \tag{C1}$$

where $K_{ab} = \chi_a g_a \delta_{ab}$. From these independent currents and gauge fields we can extract the EM response which couples democratically to each Dirac cone via a $2N$-dimensional "charge" vector $t_{\text{EM}} = (e, e, \ldots, e, e)^T$, where $e$ is the electron charge. The Hall conductivity is then $\sigma_{xy} = \frac{1}{2h} t_{\text{EM}}^T K t_{\text{EM}}$. We can also define a valley charge vector $t_V = (\chi_1, \chi_2, \ldots, \chi_{2N})^T$. We can define a valley Hall conductivity via $\sigma_{xy}^V = \frac{1}{h} t_{\text{EM}}^T K t_V$, which determines the valley current in response to an EM field. Finally, we can define a valley-valley Hall conductivity via $\sigma_{xy}^{VV} = \frac{1}{2h} t_V^T K t_V$, which determines the amount of valley current that flows in response to a valley EM field (generated, for example, by strain).

In general, we may have other interesting types of charge vectors $t_S$ if we have more symmetries, e.g., spin-rotation symmetry, or point-group symmetries, that correspond to the quantum numbers carried by the corresponding Dirac cones. We can define charge and valley Hall conductivities of those additional quantum numbers by $\sigma_{xy}^S = \frac{1}{h} t_{\text{EM}}^T K t_S$ and $\sigma_{xy}^{VS} = \frac{1}{h} t_V^T K t_S$. As an example, suppose that we have translation symmetry in space-time, which gives rise to conserved momentum and energy. For translation along the $x$ direction, each Dirac cone has a momentum component $k_{(i)}^x$, leading to a charge vector $t_x = \hbar(k_{(1)}^x, k_{(2)}^x, \ldots, k_{(2N)}^x)^T$. We could use this to define the charge polarization along the $y$ direction as $P_1^y = \frac{1}{2h} t_{\text{EM}}^T K t_x$. This can be written in a more covariant way as $P_1^a = \frac{1}{2h} \epsilon^{ab} t_{\text{EM}}^T K t_b$ and $M = \frac{1}{2h} t_{\text{EM}}^T K t_\epsilon$, where $t_\epsilon = \hbar(\epsilon_{(1)}, \epsilon_{(2)}, \ldots, \epsilon_{(2N)})^T$.

Let us consider a few explicit examples. The simplest case is $N = 1$, where the the Dirac cones are specified, without loss of generality by $(+, \mathbf{P}_{(1)}, \epsilon_1, g_1)$ and $(-, \mathbf{P}_{(2)}, \epsilon_2, g_2)$. Up to global signs, the two possible $K$ matrices are $K_1 = \mathbb{I}$ and $K_2 = \sigma^z$. The $K$ matrix $K_1$ $(K_2)$ corresponds to the case of a time-reversal symmetry (inversion symmetry)-breaking anomalous response. The EM and valley charge vectors for both $K$ matrices are $t_{\text{EM}} = (e, e)^T$ and $t_V = (1, -1)^T$. For $K_1$ we easily find $\sigma_{xy} = e^2/h$, $\sigma_{xy}^V = 0$, and $\sigma_{xy}^{VV} = 1/h$. For $K_2$ we have $\sigma_{xy} = \sigma_{xy}^{VV} = 0$ and $\sigma_{xy}^V = \frac{e}{h}$.

Now let us consider translation invariance so that we can construct a charge vector associated with the energy and momentum of each Dirac point $t_x = (k_{(1)x}, k_{(2)x})$, $t_y = (k_{(1)y}, k_{(2)y})$, $t_\epsilon = (\epsilon_{(1)}, \epsilon_{(2)})$. We can see that the polarization would be $P_1^a = \frac{1}{4\pi} \epsilon^{ab}(k_{(1)b} + k_{(2)b})$ when $K = \mathbb{I}$ and $P_1^a = \frac{e}{4\pi} \epsilon^{ab}(k_{(1)b} - k_{(2)b})$ when $K = \sigma^z$. The magnetization would be given by $M = \frac{e}{4\pi}(\epsilon_{(1)} - \epsilon_{(2)})$ when $K = \sigma^z$ and $M = \frac{e}{4\pi}(\epsilon_{(1)} + \epsilon_{(2)})$ when $K = \mathbb{I}$.

We can also consider a more complicated example with $N = 2$ which will have four Dirac cones. As an explicit example, take $\chi_1 = \chi_2 = 1$ and $\chi_3 = \chi_4 = -1$. The EM and valley charge vectors are $t_{\text{EM}} = (e, e, e, e)^T$ and $t_V = (1, 1, -1, -1)$. We can also define two other useful, linearly independent charge vectors $t_U = (1, -1, -1, 1)$ and $t_W = (1, -1, 1, -1)$. There are $2^4 = 16$ possible $K$ matrices but we only need to consider eight since the other eight differ by an overall sign. These eight are

$$\begin{aligned} &K_1 = \text{diag}[1,1,1,1], \quad K_2 = \text{diag}[1,1,-1,-1], \\ &K_3 = \text{diag}[1,-1,-1,1], \quad K_4 = \text{diag}[-1,1,-1,1], \\ &K_5 = \text{diag}[1,1,-1,1], \quad K_6 = \text{diag}[1,-1,1,1], \\ &K_7 = \text{diag}[-1,1,1,1], \quad K_8 = \text{diag}[1,1,1,-1]. \end{aligned} \tag{C2}$$

We can tabulate their (dimensionless) EM responses via $\frac{1}{2} t_{\text{EM}}^T K t_\alpha$, where $\alpha =$ EM, V, U, and W. We find

$$\left[\begin{array}{ccccc} & \text{EM} & V & U & W \\ K_1 & 2 & 0 & 0 & 0 \\ K_2 & 0 & 2 & 0 & 0 \\ K_3 & 0 & 0 & 2 & 0 \\ K_4 & 0 & 0 & 0 & 2 \\ K_5 & 1 & 1 & 1 & 1 \\ K_6 & 1 & -1 & 1 & -1 \\ K_7 & 1 & -1 & -1 & 1 \\ K_8 & 1 & 1 & -1 & -1 \end{array}\right]. \tag{C3}$$

We note that while all of these results are simple and appealing, we must be careful to handle the cases when the response coefficients have a $Z_2$ nature, i.e., when they are connected to the charge polarization. As shown in the main text, handling the possibility of $Z_2$ cancellation can be taken care of by modifying the product $\chi_a g_a$ for certain nodes. This could, in general, give rise to a modified $K$ matrix, but after that the rest of the formulation would go through. Additionally, since this formalism was derived from independent continuum flavors of Dirac fermions, it may be necessary to modify the sign of certain response coefficients to match the lattice results. Such a sign may be present, for example, for the charge polarization. We called this extra factor $\Theta$ in the main text. We leave a full discussion of these issues to future work.

[1] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).
[2] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).
[3] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, Phys. Rev. B **78**, 195424 (2008).
[4] A. Kitaev, AIP Conf. Proc. **1134**, 22 (2009).
[5] L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).
[6] J. C. Y. Teo, L. Fu, and C. L. Kane, Phys. Rev. B **78**, 045426 (2008).
[7] L. Fu, Phys. Rev. Lett. **106**, 106802 (2011).
[8] T. L. Hughes, E. Prodan, and B. A. Bernevig, Phys. Rev. B **83**, 245132 (2011).
[9] A. M. Turner, Y. Zhang, R. S. K. Mong, and A. Vishwanath, Phys. Rev. B **85**, 165120 (2012).
[10] C. Fang, M. J. Gilbert, and B. A. Bernevig, Phys. Rev. B **86**, 115112 (2012).
[11] J. C. Y. Teo and T. L. Hughes, Phys. Rev. Lett. **111**, 047006 (2013).
[12] R.-J. Slager, A. Mesaros, V. Juričić, and J. Zaanen, Nat. Phys. **9**, 98 (2012).
[13] W. A. Benalcazar, J. C. Teo, and T. L. Hughes, Phys. Rev. B **89**, 224503 (2014).
[14] T. Morimoto and A. Furusaki, Phys. Rev. B **88**, 125129 (2013).
[15] C.-K. Chiu, H. Yao, and S. Ryu, Phys. Rev. B **88**, 075142 (2013).
[16] C. Fang, M. J. Gilbert, and B. A. Bernevig, Phys. Rev. B **87**, 035119 (2013).
[17] T. L. Hughes, H. Yao, and X.-L. Qi, Phys. Rev. B **90**, 235123 (2014).
[18] Y. Ueno, A. Yamakage, Y. Tanaka, and M. Sato, Phys. Rev. Lett. **111**, 087002 (2013).
[19] F. Zhang, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **111**, 056403 (2013).
[20] P. Jadaun, D. Xiao, Q. Niu, and S. K. Banerjee, Phys. Rev. B **88**, 085110 (2013).
[21] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
[22] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).
[23] M. König, H. Buhmann, L. W. Molenkamp, T. Hughes, C.-X. Liu, X.-L. Qi, and S.-C. Zhang, J. Phys. Soc. Jpn. **77**, 031007 (2008).
[24] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **452**, 970 (2008).
[25] J. Moore, Nat. Phys. **5**, 378 (2009).
[26] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. Hor, R. Cava *et al.*, Nat. Phys. **5**, 398 (2009).
[27] T. Zhang, P. Cheng, X. Chen, J.-F. Jia, X. Ma, K. He, L. Wang, H. Zhang, X. Dai, Z. Fang *et al.*, Phys. Rev. Lett. **103**, 266803 (2009).
[28] F. D. M. Haldane, Phys. Rev. Lett. **61**, 2015 (1988).
[29] C.-Z. Chang, J. Zhang, X. Feng, J. Shen, Z. Zhang, M. Guo, K. Li, Y. Ou, P. Wei, L.-L. Wang *et al.*, Science **340**, 167 (2013).
[30] X.-L. Qi, T. L. Hughes, S. Raghu, and S.-C. Zhang, Phys. Rev. Lett. **102**, 187001 (2009).
[31] H. B. Nielsen and M. Ninomiya, Phys. Lett. B **105**, 219 (1981).
[32] X. Wan, A. M. Turner, A. Vishwanath, and S. Y. Savrasov, Phys. Rev. B **83**, 205101 (2011).
[33] A. M. Turner and A. Vishwanath, in *Topological Insulators*, Vol. 6, edited by M. Franz and L. Molenkamp (Elsevier, Amsterdam, 2013), p. 293.
[34] F. Haldane, arXiv:1401.0529.
[35] S. Matsuura, P.-Y. Chang, A. P. Schnyder, and S. Ryu, New J. Phys. **15**, 065001 (2013).
[36] S. Parameswaran, T. Grover, D. Abanin, D. Pesin, and A. Vishwanath, Phys. Rev. X **4**, 031035 (2014).
[37] M. N. Chernodub, A. Cortijo, A. G. Grushin, K. Landsteiner, and M. A. Vozmediano, Phys. Rev. B **89**, 081407(R) (2014).
[38] Z. Jian-Hui, J. Hua, N. Qian, and S. Jun-Ren, Chin. Phys. Lett. **30**, 027101 (2013).
[39] P. Hosur, Phys. Rev. B **86**, 195102 (2012).
[40] C.-X. Liu, P. Ye, and X.-L. Qi, Phys. Rev. B **87**, 235306 (2013).
[41] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).
[42] G. B. Halász and L. Balents, Phys. Rev. B **85**, 035103 (2012).
[43] S. Ganeshan and S. D. Sarma, Phys. Rev. B **91**, 125438 (2015).
[44] S. M. Young, S. Zaheer, J. C. Y. Teo, C. L. Kane, E. J. Mele, and A. M. Rappe, Phys. Rev. Lett. **108**, 140405 (2012).
[45] Z. Wang, Y. Sun, X.-Q. Chen, C. Franchini, G. Xu, H. Weng, X. Dai, and Z. Fang, Phys. Rev. B **85**, 195320 (2012).
[46] Z. Liu, B. Zhou, Z. Wang, H. Weng, D. Prabhakaran, S.-K. Mo, Y. Zhang, Z. Shen, Z. Fang, X. Dai *et al.*, Science **343**, 864 (2014).
[47] M. Neupane, S.-Y. Xu, R. Sankar, N. Alidoust, G. Bian, C. Liu, I. Belopolski, T.-R. Chang, H.-T. Jeng, H. Lin *et al.*, Nat. Commun. **5**, 3786 (2014).
[48] Z. Wang, H. Weng, Q. Wu, X. Dai, and Z. Fang, Phys. Rev. B **88**, 125427 (2013).
[49] B.-J. Yang and N. Nagaosa, Nat. Commun. **5**, 4898 (2014).
[50] T. Meng and L. Balents, Phys. Rev. B **86**, 054504 (2012).
[51] G. Y. Cho, J. H. Bardarson, Y.-M. Lu, and J. E. Moore, Phys. Rev. B **86**, 214514 (2012).
[52] P. Hosur and X. Qi, C. R. Phys. **14**, 857 (2013).
[53] P. Goswami and S. Tewari, Phys. Rev. B **88**, 245107 (2013).
[54] A. A. Zyuzin and A. A. Burkov, Phys. Rev. B **86**, 115133 (2012).
[55] M. M. Vazifeh and M. Franz, Phys. Rev. Lett. **111**, 027201 (2013).
[56] Y. Chen, S. Wu, and A. A. Burkov, Phys. Rev. B **88**, 125105 (2013).
[57] Y. Chen, D. L. Bergman, and A. A. Burkov, Phys. Rev. B **88**, 125110 (2013).
[58] D. Vanderbilt, I. Souza, and F. Haldane, Phys. Rev. B **89**, 117101 (2014).
[59] D. Xiao, W. Yao, and Q. Niu, Phys. Rev. Lett. **99**, 236809 (2007).
[60] D. Xiao, M.-C. Chang, and Q. Niu, Rev. Mod. Phys. **82**, 1959 (2010).
[61] S. Coh and D. Vanderbilt, Phys. Rev. Lett. **102**, 107603 (2009).
[62] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, Nat. Phys. **4**, 273 (2008).
[63] B. I. Halperin, Jpn. J. Appl. Phys. Suppl. **26**, 1913 (1987).
[64] G. Montambaux and M. Kohmoto, Phys. Rev. B **41**, 11417 (1990).
[65] L. Balents and M. P. A. Fisher, Phys. Rev. Lett. **76**, 2782 (1996).
[66] B. A. Bernevig, T. L. Hughes, S. Raghu, and D. P. Arovas, Phys. Rev. Lett. **99**, 146804 (2007).

[67] Generically, the EM response properties of topological semimetals with point nodes is determined by a vector $b_\mu$. If this vector is nonvanishing of this vector, this generally would violate Lorentz invariance since both the time and the space components can be nonvanishing. If only the spatial components were nonzero, as is found, for example, in a WTI, then the system is simply

anisotropic. However, the semimetal response is more general. Throughout this article we often use the term Lorentz violation to imply one of several things. First, any source of anisotropy we term as violating Lorentz invariance, whether it arises from an external field such as an electric field or intrinsically due to the crystal structure or electronic structure of the system in question. The crystal lattice itself provides direct and reciprocal lattice vectors which are signatures of anisotropy, and if there are pointlike Fermi surfaces in the BZ they can also serve as a signal of anisotropy/Lorentz violation. Additionally, the topological semimetals are often described at low-energy by a massless relativistic continuum theory when expanded near the point nodes. Thus, we often encounter three additional sources of Lorentz violation—(i) differences in energy between the nodes; (ii) differences in velocity ("speed of light") of the massless dispersions of the nodes; and (iii) background charge density and/or current density—which give rise to a preferred frame. Thus, when we use the term Lorentz violation we are referring to a general extrinsic or intrinsic source that picks out a "preferred frame" whether it be due to spatial anisotropy or a more general mechanism. These sources are termed as such because they can generate a nonvanishing $b_\mu$.

[68] S. M. Carroll, G. B. Field, and R. Jackiw, Phys. Rev. D **41**, 1231 (1990).

[69] R. Jackiw and S.-Y. Pi, Phys. Rev. D **68**, 104012 (2003).

[70] A. G. Grushin, Phys. Rev. D **86**, 045001 (2012).

[71] F. D. M. Haldane, Phys. Rev. Lett. **93**, 206602 (2004).

[72] S. T. Ramamurthy and T. L. Hughes (unpublished).

[73] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).

[74] K.-Y. Yang, Y.-M. Lu, and Y. Ran, Phys. Rev. B **84**, 075129 (2011).

[75] To be more precise, Chern insulators require a global $U(1)$ charge conservation symmetry for protection, but we do not consider the possibility of superconducting systems.

[76] A. Altland and M. R. Zirnbauer, Phys. Rev. B **55**, 1142 (1997).

[77] J. Zak, Phys. Rev. Lett. **62**, 2747 (1989).

[78] F. Pollmann, E. Berg, A. M. Turner, and M. Oshikawa, Phys. Rev. B **85**, 075125 (2012).

[79] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).

[80] D. J. Thouless, Phys. Rev. B **27**, 6083 (1983).

[81] B. A. Bernevig and T. L. Hughes, *Topological Insulators and Topological Superconductors* (Princeton University Press, Princeton, NJ, 2013).

[82] The authors do not know of any models which realize the latter case. One must also worry about the fact that the total flux must be a multiple of $2\pi$ and, thus, to be well-defined we must have an even number of discretized momentum points. This constraint seems a bit artificial so we do not consider this case further.

[83] A. N. Redlich, Phys. Rev. D **29**, 2366 (1984).

[84] G. W. Semenoff, Phys. Rev. Lett. **53**, 2449 (1984).

[85] D. Ceresoli, T. Thonhauser, D. Vanderbilt, and R. Resta, Phys. Rev. B **74**, 024408 (2006).

[86] X.-G. Wen and A. Zee, Phys. Rev. B **46**, 2290 (1992).

[87] There is always an ambiguity in determining the boundary charge from the bulk Dirac point data, but in cases where the $g_a$ can be different from each other the problem worsens. Even if the total Chern number vanishes there can be cases where, depending on exactly how the edge states connect between nodes, the system could realize a nonzero Chern number in some pairs of Dirac nodes (such that the total sum vanishes) and just a nonzero polarization in other nodes, i.e., effectively a mix of time-reversal and inversion-breaking mass terms for each pair of nodes. The boundary charge in these cases could be determined along the by combining the types analyses in Secs. III E and III F, but the connection between this charge and the bulk calculation for the polarization is more challenging to derive.

[88] Of course, there could be cases with overlapping edge states that do not cancel each other, but to be generic this will require some symmetry. For example, if one edge state carries spin-up and the other carries spin-down, then spin-rotation symmetry could protect from the cancellation. However, this is essentially just converting the edge-state stability from $\mathbb{Z}_2$ to $\mathbb{Z}$. Instead, we consider the generic case without extra symmetries so that the cancellation is taken to always occur for overlapping edge states.

[89] C. G. Callan, Jr. and J. A. Harvey, Nucl. Phys. B **250**, 427 (1985).

[90] T. L. Hughes, R. G. Leigh, and O. Parrikar, Phys. Rev. D **88**, 025040 (2013).

[91] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).

[92] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).

[93] R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976).

[94] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306(R) (2007).

[95] R. Roy, Phys. Rev. B **79**, 195322 (2009).

[96] Y. Ran, Y. Zhang, and A. Vishwanath, Nat. Phys. **5**, 298 (2009).

[97] T. H. Hsieh, H. Lin, J. Liu, W. Duan, A. Bansil, and L. Fu, Nat. Commun. **3**, 982 (2012).

[98] Y. Tanaka, Z. Ren, T. Sato, K. Nakayama, S. Souma, T. Takahashi, K. Segawa, and Y. Ando, Nat. Phys. **8**, 800 (2012).

[99] M. Creutz, Rev. Mod. Phys. **73**, 119 (2001).