



# Consequences of average time-reversal symmetry in disordered antiferromagnetic topological insulators

Yihao Lin <sup>1</sup> and Ji Feng <sup>1,2,\*</sup>

<sup>1</sup>International Center for Quantum Materials, School of Physics, Peking University, Beijing 100871, China

<sup>2</sup>Hefei National Laboratory, Hefei 230088, China



(Received 28 October 2022; accepted 6 March 2023; published 20 March 2023)

Average symmetry protects the topological surface states of topological (crystalline) insulators with time-reversal symmetry from disorder-induced localization. However, the nature of such average symmetry for magnetic topological insulators and, in particular, its connection to surface transport await inspection. Here, we investigate the impact of imperfect magnetic order on an antiferromagnetic topological insulator, which is a solid with a bulk axion field  $\theta = \pi$ . We find that the disordered topological surfaces of an antiferromagnetic topological insulator can be generally gapped and localized. The behavior of topological surfaces is now controlled by a *mesoscopic* average time-reversal symmetry that requires a magnetically imperfect system to be divisible into finite and magnetically neutral slabs. In the presence of this mesoscopic average time-reversal symmetry, the topological surface states will be gapless in the thermodynamic limit and tend to delocalize at a single energy similar to the delocalization transition in the chiral universality class. The notion of average-symmetry-induced delocalization is thus extended to account for magnetic topological insulators, and the spectroscopic and transport signatures clarified herein are relevant to future experimental investigations.

DOI: [10.1103/PhysRevB.107.115138](https://doi.org/10.1103/PhysRevB.107.115138)

## I. INTRODUCTION

A topological crystalline insulator (TCI) is protected by a discrete crystalline symmetry [1,2] whose Hamiltonian cannot be continuously deformed to that of an atomic insulator while maintaining that symmetry and harbors gapless states on symmetry-preserving boundaries. Examining such bulk-boundary correspondence in a broader context beyond the single-particle, clean limit is a topic of current interest [3–8]. Disorder usually breaks the topology-protecting crystalline symmetry and is thus expected to be detrimental to the bulk-boundary correspondence of TCIs [9]. Nonetheless, previous studies indicate that the topological surface states of a TCI cannot be localized by disorder thanks to the existence of *average symmetry* [2,10,11]. In the SnTe-family mirror Chern insulators with nonmagnetic disorder, for example, a topological surface is divided into two kinds of insulating domains with Dirac masses of different signs related by the mirror symmetry, which results in one-dimensional (1D) helical states bound to the domain walls and protected by time-reversal symmetry from backscattering. If two kinds of domains appear with equal likelihood when the disordered system preserves average mirror symmetry, the topological surface evades localization due to the percolation of domain wall states [9]. The idea is clearly very important in clarifying the nature of topological surfaces in real materials with disorder and imperfection that breaks the symmetry underlying the bulk topology.

When the concept of TCI is extended to magnetic systems [12], a representative example is the antiferromagnetic (AFM) topological insulator (TI), which is protected by the symmetry  $S = \Theta T_{1/2}$  ( $\Theta$  is time reversal and  $T_{1/2}$  is a half lattice translation) and characterized by a nontrivial  $\mathbb{Z}_2$  topology [13] and a bulk axion field  $\theta = \pi$  [14]. Unlike the TCIs considered previously, the magnetic order in an AFM TI breaks time-reversal symmetry. The relevant question in this case is then how magnetic disorder affects the spectral and transport properties on topological surfaces, as well as the nature and role of average symmetry, which clearly requires investigation.

Therefore it is the purpose of the present study to examine the impact of magnetic imperfections on the topological surface states of AFM TIs. In finite-supercell models with AFM domain walls, spectral gaps are generally observed on topological surfaces, whose size is a function of a single variable, dubbed the *Ising moment*, characterizing the imperfect AFM orders. It is further discovered that the topological surface states of AFM TIs tend to be gapless in the thermodynamic limit when the system can be divided into finite slabs with zero net magnetization. This is thought to be the *average time-reversal symmetry* preserved on a *mesoscopic scale*, which further affects the transport of topological surfaces. Subsequently, employing the transfer matrix and the stochastic Dirac equation, it is uncovered that though topological surface states are usually localized, delocalization occurs at a single energy when the mesoscopic average time-reversal symmetry is present. Though this delocalization transition is intimately related to that in the chiral universality class, it displays distinct critical behaviors. These results are then argued to be generalizable to disordered AFM TIs beyond the domain wall

\*jffeng11@pku.edu.cn

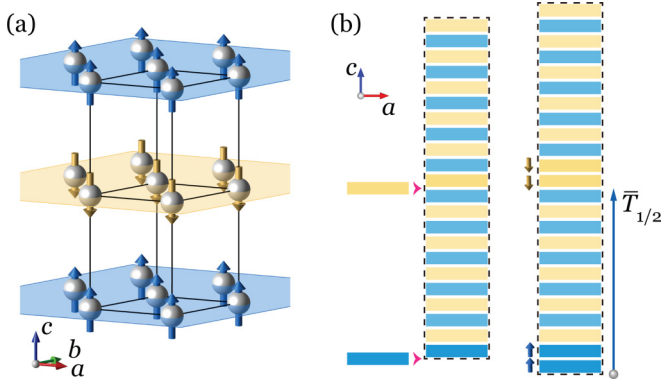


FIG. 1. The effective tight-binding model of  $\text{MnBi}_2\text{Te}_4$ . (a) A lattice model on a hexagonal lattice with layer-dependent on-site exchange field (blue and yellow arrows) along  $c = c\hat{z}$ ,  $\mathbf{a} = (1, 0, 0)$  and  $\mathbf{b} = \frac{1}{2}(1, \sqrt{3}, 0)$ . (b) A ferromagnetic bilayer as a kink in the AFM order. A configuration with periodically arranged kinks can preserve a  $\bar{S} = \Theta\bar{T}_{1/2}$  symmetry.

models studied herein. These findings provide insights into the effects of average symmetry on a disordered magnetic TCI, as well as some of the spectroscopic and transport signatures relevant to experimental measurements on AFM TIs, such as the  $\text{MnBi}_2\text{Te}_4$  family [14–17].

## II. SURFACE SPECTRUM AND ISING MOMENT

We begin with an examination of the topological surface states using a tight-binding (TB) model for an AFM TI,  $\text{MnBi}_2\text{Te}_4$ , introduced in Ref. [18], which is derived from the layered topological insulator  $\text{Bi}_2\text{Se}_3$  [19] with a layer-dependent exchange field (see Appendix A). This model indeed produces an AFM TI in the case of perfect AFM ordering,  $m_\ell = (-1)^\ell$  [see Fig. 1(a)], where  $m_\ell$  is the sign of the exchange field (magnetization) of the  $\ell$ th layer. The particular type of magnetic imperfection to be investigated is the AFM domain wall, which shows up as a ferromagnetic bilayer [or a kink; see Fig. 1(b)] and occurs in the Ising limit (strong anisotropy with an easy  $z$  axis). It is noteworthy that a superlattice with equidistant domain walls with alternating magnetization can possess a composite symmetry  $\bar{S} = \Theta\bar{T}_{1/2}$ . Such a superlattice will be subject to a  $\mathbb{Z}_2$  topological classification protected by  $\bar{S}$  and is necessarily an AFM TI in the weak-field limit because it can continuously deform into a TI with in-gap surface states. However, as the exchange field increases, a topological transition into a trivial insulator ( $\mathbb{Z}_2 = 0$ ) can happen.

This can be intuitively understood: Without interlayer coupling, every monolayer undergoes a Chern-insulator transition in the strong-field limit and develops a chiral edge state (see Appendix A). Adjacent layers with opposite magnetization have opposite Chern numbers, and the counterpropagating chiral edge states are coupled when interlayer hopping is turned on and gapped except at the kinks. The remaining chiral edge states at separated opposite (magnetized) kinks are also gapped by coupling mediated by the topological surface states (of the TI), resulting in an overall spectral gap across the topological surface (i.e.,  $\mathbb{Z}_2 = 0$ ).

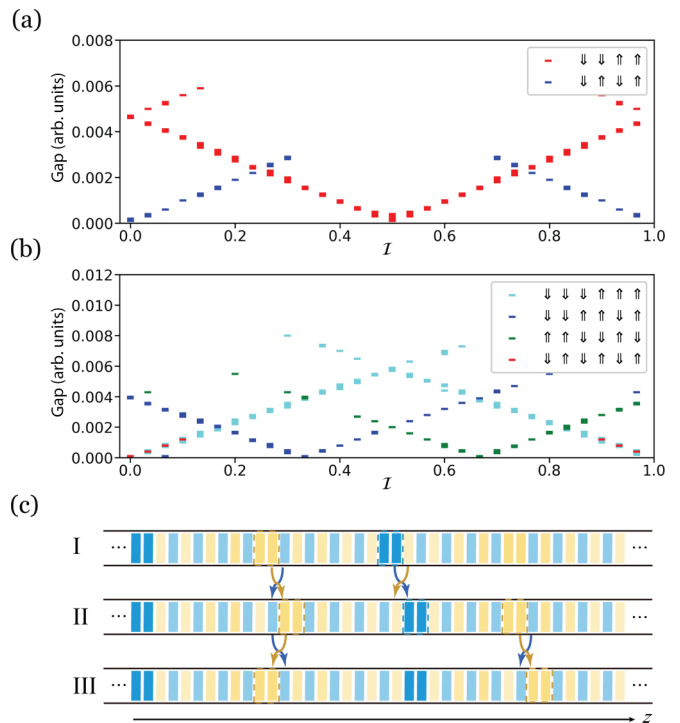


FIG. 2. (010) surface gaps at  $\Gamma$  of 60-layer supercells with (a) four kinks and (b) six kinks, plotted against the Ising moment  $\mathcal{I}$ . (c) An elementary kink migration is accomplished by a transposition of an adjacent pair of oppositely magnetized layers. Processes I  $\rightarrow$  II and II  $\rightarrow$  III demonstrate how the change in  $\mathcal{I}$  due to migration of one kink can be canceled by migration of another.

Hereinafter, we stick to a moderate exchange field in which a superlattice with  $\bar{S}$  symmetry is topologically nontrivial,  $\mathbb{Z}_2 = 1$ . When AFM domain walls are introduced such that the system possesses a net magnetization, the exchange field at the mean-field level will shift the Dirac point. Taking the (010) surface (normal to the  $y$  axis) as an example, the surface spectrum corresponding to a surface Hamiltonian  $\mathcal{H} = v_x k_x \sigma_z + v_z k_z \sigma_x$  is shifted along the  $k_x$  axis when particle-hole symmetry is broken (and the surface electrons possess a finite  $z$  component of real spin) [18]. The degeneracy at a Dirac point shifted away from  $\Gamma$  is accidental and can be easily lifted by higher-order terms such as the intrinsic  $C_3$  warping term in our Hamiltonian [18,20]. In the following, we focus on configurations with zero net magnetization, which requires an equal number of  $\uparrow$  and  $\downarrow$  kinks ( $n_\uparrow = n_\downarrow$ ). By enforcing the average time-reversal symmetry across the whole sample, the (gapped) Dirac point of the surface states is pinned near  $\Gamma$  of the surface Brillouin zone (BZ).

We calculate the band gaps of (010) surface states [21] at  $\Gamma$  of plenty of AFM-ordered supercells made of  $N = 60$  layers with  $n$  randomly placed kinks. Remarkably, the computed surface gap is seen to be a function of a single variable, which we call the *Ising moment*

$$\mathcal{I} = \overline{\ell m_\ell} \bmod 1, \quad (1)$$

where the overbar stands for averaging over all magnetic layers. Evidently,  $\mathcal{I}$  is invariant under supercell multiplications or origin shifts. As exemplified in Figs. 2(a) and 2(b) for  $n = 4$

and 6, respectively, the surface gap has a few branches characteristic of different permutations of kink magnetization. Given a kink permutation, the surface gap is singularly determined by the value of  $\mathcal{I}$ , irrespective of the detailed placements of the kinks. Significantly, the surface states turn out to be gapless at a single special  $\mathcal{I}$ , whose value  $\mathcal{I}^*$  can be one of the following:

$$\mathcal{I}^* = p/n, \quad p = 0, 2, \dots, n - 2. \quad (2)$$

Each  $\mathcal{I}^*$  corresponds to multiple configurations, which can be inferred from the fact that the change in  $\mathcal{I}$  due to the movement of one kink can always be canceled by the movement of another kink, as illustrated in Fig. 2(c).

A pair of  $\uparrow$  and  $\downarrow$  kinks can bump into each other and annihilate. All kinks can be removed by consecutive pair annihilations, ending in the perfect AFM order. Keeping track of the changes in  $\mathcal{I}$  in this process of approaching the perfect AFM order ( $\mathcal{I} = 0.5$ ) provides us with a formula for  $\mathcal{I}$ , up to modulo 1,

$$\mathcal{I} = 0.5 - \frac{1}{N} \sum_{i=1}^n \alpha_i p_i \stackrel{p_i \rightarrow \bar{p}}{=} \mathcal{I}^*, \quad (3)$$

where we define the number of monolayers between the  $i$ th and  $(i + 1)$ th kinks to be  $p_i - 2$ .  $\alpha_i$  is the excess magnetization  $n_{\uparrow} - n_{\downarrow}$  in the first  $i$  kinks and measures the total magnetization transferred from the region below the  $i$ th kink (inclusive) to the region above in order to annihilate all kinks. We observe that  $\mathcal{I}^*$  is attained at equal separations  $\bar{p} = \frac{1}{n} \sum_i p_i = \frac{N}{n}$ .

This result has significant consequences: Assuming that  $|\alpha_i|$ 's possess an upper bound,  $|\alpha_i| < \alpha_{\max}$ , and that there is no correlation between  $p_i$ 's, the deviation  $\mathcal{I} - \mathcal{I}^* \sim |\alpha_{\max}|/\sqrt{n}$  will tend to zero in the thermodynamic limit  $n \rightarrow \infty$  and  $N/n \rightarrow \text{const}$  due to cancellation of statistical fluctuation in  $p_i$ 's. The existence of  $|\alpha_{\max}|$  indicates the presence of the mesoscopic average time-reversal symmetry: If the configuration can be divided into slabs with zero net magnetization, each slab with at most  $2n_{\max}$  kinks, then  $|\alpha_{\max}| \leq n_{\max}$ . The mesoscopic average time-reversal symmetry makes any slab in the configuration have a net magnetization that is bounded instead of increasingly scaled with the thickness of the slab, and results in *gapless topological surface states* in the thermodynamic limit.

### III. SURFACE TRANSPORT IN THE THERMODYNAMIC LIMIT

The transport of topological surface states of the AFM TI with kinks in the thermodynamic limit can be investigated by establishing a transfer matrix description. We will begin with the (010) topological surface and will then generalize the results to any topological surface in the Discussion. The (010) surface is chosen because it preserves the nonmagnetic  $x$ -mirror symmetry,  $\mathcal{M}_x$ , in the  $D_{3d}$  crystal point group of  $\text{MnBi}_2\text{Te}_4$ . We can exploit this mirror operation as a local gauge transformation that conducts a change of bases locally within each site in  $\downarrow$  layers. Since  $\mathcal{M}_x$  flips the  $z$  component of spin, the transformed Hamiltonian will possess a uniform  $+\hat{z}$  exchange field and now describes a nominal perfect ferromagnetic phase. The model at  $k_x = 0$  after this local gauge transformation has uniform on-site terms and uniform (but

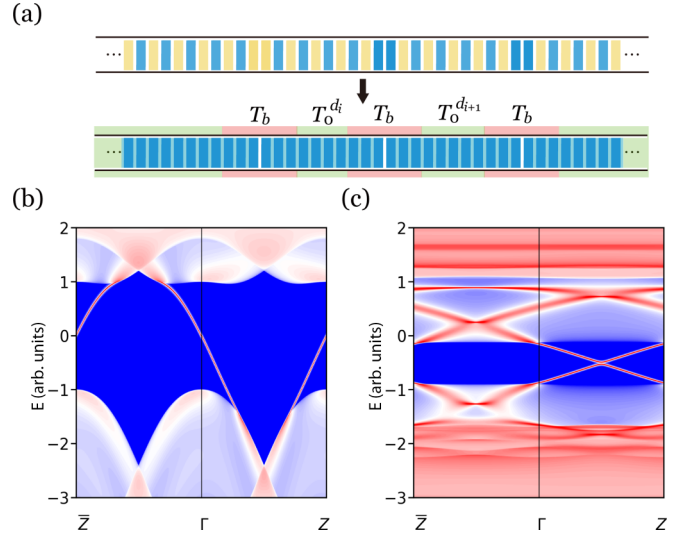


FIG. 3. (a) A local gauge transformation changes the AFM kink model (top) into the FM bond defect model (bottom). The bond defect zones are highlighted as pink blocks, and FM regions are highlighted as green blocks. (b) and (c) (010) surface spectra of a homogeneous FM bulk without bond defects (b) and a superlattice of defect zones (c). When the unit cell is doubled, both spectra will be folded, and a Dirac point will appear at  $\Gamma$  of the folded BZ protected by the  $S$  or  $\bar{S}$  symmetry. Thus the momenta of two surface states at the same energy should be  $k_z$  and  $\pi - k_z$ , respectively.

modified) interlayer hoppings for the original perfect AFM phase. With a bilayer domain wall, the interlayer hopping between these two layers remains unchanged after the local gauge transformation (see Appendix B). In the case of multiple kinks, this gauge transformation leads to a random “bond defect” model [see Fig. 3(a)], which is conducive to a transfer matrix treatment.

A bond defect configuration is composed of homogeneous FM blocks without defects and bond defects gluing adjacent blocks. The homogeneous FM bulk supports a pair of gapless surface modes, as shown in Fig. 3(b), since it corresponds to an  $S$ -symmetric AFM TI before the local gauge transformation. We define a “defect zone” as being composed of  $2L$  layers centered at a bond defect, as depicted in Fig. 3(a).  $L$  is large enough so that surface states at the boundaries of a defect zone can be adequately expanded in terms of the homogeneous FM surface modes [ $L = 3$  is enough for the (010) surface]. Apparently, the superlattice of defect zones also supports gapless surface modes, as shown in Fig. 3(c), as it corresponds to an  $\bar{S}$ -symmetric AFM TI before the local gauge transformation. The existence of gapless surface modes enables us to read the eigenvalues of their transfer matrices across the low-energy regime of interest from their surface spectra. We write down their transfer matrices:

$$T_0(k_z) = \text{diag}[e^{ik_z}, e^{i(\pi - k_z)}] \equiv \sigma_z T(k_z) \quad (4)$$

for a monolayer and

$$T_b(\theta, u) = W(u)T_0(\theta)W(u)^{-1} \quad (5)$$

for a defect zone, where  $T(\xi) = \text{diag}[e^{i\xi}, e^{-i\xi}]$  and  $k_z(\theta)$  is the phase acquired by the branch of the surface mode

propagating in the  $-\hat{z}$  direction across the monolayer (defect zone).  $W(u) = [[1, u]^T, [u^*, 1]^T]$  accounts for the scattering of the homogeneous FM surface modes by the bond defect and encodes the formal constraint of current conservation. The construction of the transfer matrix given here is valid irrespective of any adiabatic  $S$ -invariant deformation on the AFM TI's Hamiltonian.

With these elemental transfer matrices, we can compute the total transfer matrix of a bond defect configuration. One can verify that the total transfer matrix  $T_n$  of a configuration that corresponds to an AFM-ordered supercell with  $n$  kinks that preserves the average time-reversal symmetry satisfies (see Appendix G)

$$\text{tr } T_n = 2 + 16N^2 k_z^2 |u'|^2 (\mathcal{I} - \mathcal{I}^*)^2 + O(|u'|^4), \quad (6)$$

at the energy where  $\theta' = 0$ , with  $\theta'$  and  $u'$  defined by the equation  $T_b(\theta', u') = T(\bar{\phi})T_b(\theta, u)$ .  $\bar{\phi} = k_z(\bar{p} - 2L)$  is the average phase acquired by traveling from one defect zone to the next.  $\theta' \approx \bar{\phi} + \theta$  in the small- $u$  limit, i.e., the sum of phases from traveling between defect zones and from crossing a defect zone. Thus the condition  $\theta' = 0$  indicates a complete phase cancellation, and a traveling mode acquires no net phase *on average* across the entire surface. Therefore Eq. (6) describes the trace of  $T_n$  at  $\Gamma$  in the surface Brillouin zone of the supercell. This is consistent with the eigenvalues of  $T_n$ ,  $\lambda_1 = \lambda_2 = 1$  at  $\mathcal{I} = \mathcal{I}^*$  for  $\theta' = 0$  ( $\text{tr } T_n = 2$ ,  $\det T_n = 1$ ). When  $\mathcal{I} \neq \mathcal{I}^*$ ,  $\text{tr } T_n > 2$  and there are no traveling modes at the energy where  $\theta' = 0$ , which corresponds to the band-gap opening at  $\Gamma$  for  $\mathcal{I} \neq \mathcal{I}^*$  observed in the previous section.

Equation (6) demonstrates that the band-gap closure at  $\mathcal{I} = \mathcal{I}^*$  given by Eq. (3) is generally valid for supercells with arbitrary numbers of pairs of opposite kinks, and a (gapped) Dirac point would show up at  $\Gamma$ . Equation (6) results solely from the generic forms of  $T_0$  and  $T_b$ , which in turn, as we shall see in the Discussion, are dictated solely by  $S$ - ( $\bar{S}$ -)symmetry-protected nontrivial topology. Consequently, the empirical observation regarding the existence and value of  $\mathcal{I}^*$  when the surface gap closes applies to any topological surfaces. Since Eq. (6) holds irrespective of the permutation of kink magnetization, the universal condition  $\theta' = 0$  indicates that different configurations with the same kink density will open a surface gap at the same energy on the same topological surface.

In the thermodynamic limit  $n \rightarrow \infty$  and  $N/n \rightarrow \text{const}$ , the localization of the topological surface is characterized by the Lyapunov exponent [22,23] of  $T_n$ :

$$\gamma = \lim_{n \rightarrow \infty} \frac{\ln \|T_n\|}{n}, \quad (7)$$

which is related to the dimensionless conductance through  $g \sim \text{sech}^2 n\gamma$  [24].  $\gamma > 0$  corresponds to localization, whereas  $\gamma = 0$  in a delocalized case. Assuming that  $p_i$ 's independently follow an identical exponential distribution, we calculate the Lyapunov exponent  $\gamma$  of  $T_n$  [25], as shown in Figs. 4(a) and 4(b). Concerning the distribution of kink magnetizations, we have examined two disordered ensembles: Each sample in ensemble I only preserves average time-reversal symmetry across the whole sample, while each sample in ensemble II preserves the mesoscopic average time-reversal symmetry. A configuration drawn from ensemble II is

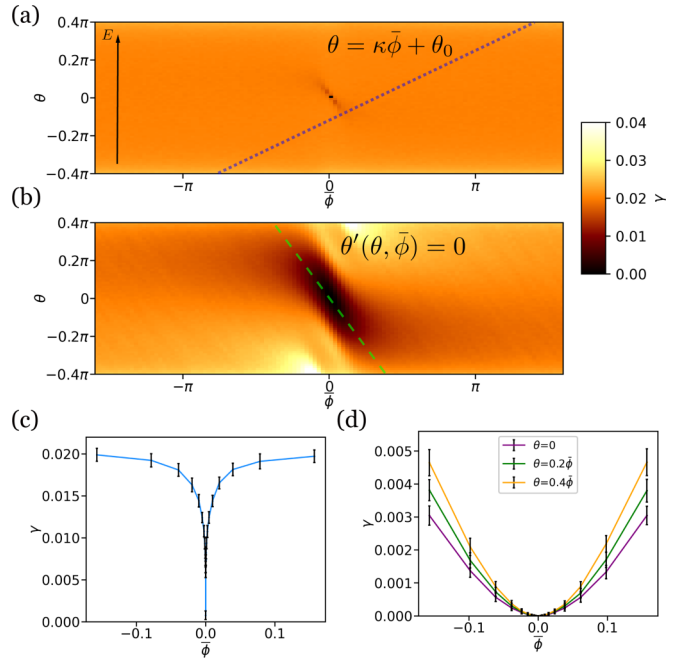


FIG. 4. Computed Lyapunov exponent  $\gamma$  for (a) ensemble I and (b) ensemble II.  $\gamma$  are averaged over 100 samples with  $2^{15}$  kinks for each point in the  $\bar{\phi}$ - $\theta$  plane, with  $u = 0.1e^{i\pi/8}/\cos\theta$ . Dashed lines indicate a typical  $\theta = \kappa\bar{\phi} + \theta_0$  (blue) and  $\theta' = 0$  (green). Critical scaling of  $\gamma$  is shown for (c) ensemble I,  $\gamma \propto 1/\ln|\bar{\phi}|$ , and (d) ensemble II,  $\gamma \propto |\bar{\phi}|^\nu$  with  $\nu \simeq 1.76$ . Vertical bars represent the standard deviations. Note that  $\bar{\phi} = k_z(\bar{p} - 2L)$  and note the linear relation between energy  $E$  and  $k_z$ ,  $\theta$ :  $E = v_0 k_z + \omega_0 = v_d \theta + \omega_d$ , so  $\bar{\phi} \propto E$ .

made up of finite slabs of zero net magnetization, where each slab has at most six kinks.

For a specific set of tight-binding model parameters including the strength of the field,  $\gamma$  should be a function of two independent parameters, energy  $E$  and the kink density. However, we compute it as a function of another set of two independent parameters,  $\theta$  and  $\bar{\phi}$ , to achieve the universality of calculated results across models. Due to the linear dispersion of surface states in the low-energy regime,  $k_z$  and  $\theta$  bear approximate linear relations (fixed by TB model parameters) to  $E$ , and  $\bar{\phi} = k_z(\bar{p} - 2L)$  is also proportional to the average spacing between defect zones. Thus a straight line  $\theta = \kappa\bar{\phi} + \theta_0$  in the parametric plane will correspond to a specific set of TB model parameters with a given kink density.

For ensemble I,  $\gamma$  is typically finite over the energy range examined as displayed in Fig. 4(a), suggesting localized surface states in the presence of imperfect magnetic order. The configurations in ensemble I generally correspond to  $\mathcal{I} \neq \mathcal{I}^*$  because  $|\alpha_i| = |n_{\uparrow} - n_{\downarrow}|$  in the first  $i$  kinks is now approximately a sum of  $i$  independent random integers  $\pm 1$  sampled with equal probability and typically of  $O(\sqrt{i})$ , so the deviation  $\mathcal{I} - \mathcal{I}^*$  will be of  $O(1)$ . Delocalization appears only at  $\theta = \bar{\phi} = 0$ , when the homogeneous bulk and the defect zone superlattice have surface modes of equal energy at  $\Gamma$ , and  $\theta' = 0$  at the very same energy. According to Eq. (6),  $\text{tr } T_n = 2$  due to  $k_z = 0$  and  $T_n$  exceptionally supports traveling eigenmodes at  $\theta' = 0$ , which display a propensity for



delocalization. The tendency toward delocalization at  $\theta' = 0$  becomes more pronounced in ensemble II, where a tilted oval-shaped region with vanishing  $\gamma$  appears near the origin and is more likely to be observed. Beyond the oval region, the Lyapunov exponent stays finite indicating localization. Now we see the second consequence of the mesoscopic average time-reversal symmetry: The generally localized topological surface states tend to *delocalize at a single energy*.

#### IV. DISCUSSION

We have seen that the mesoscopic average time-reversal symmetry has important impacts on the bulk-boundary correspondence of AFM TIs: leaving topological surface states gapless and driving a delocalization transition at a single energy in the thermodynamic limit. We can view the  $S$  symmetry as the clean limit of the mesoscopic average time-reversal symmetry as it enforces neutralization of magnetization within any adjacent bilayers, and in this sense, the latter is a kind of ‘‘average  $S$  symmetry.’’ The mesoscopic average time-reversal symmetry discussed is physically relevant for two considerations. First, a mediated AFM interaction between neighboring kinks will favor the cancellation of the net magnetization. Second, the magnetic dipolar interaction that leads to magnetic domains in ferromagnetic materials tends to suppress magnetization on macroscopic scales [26]. Here, we put our theory in contact with systems described by stochastic Dirac equations (SDEs) to understand the critical behaviors of the delocalization transition near zero energy. The map to SDEs also allows us to generalize our conclusions to AFM TIs with various types of interlayer imperfections beyond AFM domain walls.

The computed delocalization transition is closely related to that in disordered systems belonging to the chiral universality class [27–29]. First, we note that both  $T_0$  and  $T_b$  satisfy  $\sigma_1 T_{0/b} \sigma_1 = -T_{0/b}^*$ ; so the total transfer matrix  $T_n$  of an arbitrary bond defect configuration can be decomposed into the product of a sequence of transfer matrices ( $M_i$ ) that satisfy an effective time-reversal symmetry  $\sigma_1 M_i \sigma_1 = M_i^*$ ,  $\forall i$  [30].  $T_n = M_n \cdots M_1$  can be seen as the evolution operator of a 1D SDE (see Appendix F) [31].

$$[-i\sigma_3 \partial_x + V(x, E) + m(x, E)\sigma_1]\Psi = 0, \quad (8)$$

which is invariant under  $\sigma_1 \mathcal{K}$  ( $\mathcal{K}$  for complex conjugation). This is similar to the one-dimensional random hopping model [27,28], whose Schrödinger equation in the continuum limit near the band center takes the form

$$h\Psi = [-i\sigma_3 \partial_x + m(x)\sigma_2]\Psi = E\Psi, \quad (9)$$

where  $h$  satisfies a chiral symmetry  $\sigma_1 h \sigma_1 = -h$ . Eigenstates of Eq. (9) show delocalization at  $E = 0$  if the average mass  $\langle m(x) \rangle = 0$  where the critical scaling of the Lyapunov exponent near  $E = 0$  is  $\gamma \sim 1/\ln|E|$ . In fact, the delocalization criticality in ensemble I near  $\theta = \bar{\phi} = 0$  shows a similar scaling  $\gamma \sim 1/\ln|\bar{\phi}|$  see Fig. 4(c) and note that  $E \propto \bar{\phi}$ . Furthermore,  $\gamma$  in ensemble I possesses a finite fluctuation at delocalization as in the chiral universality class, which means that a typical sample at  $\theta = \bar{\phi} = 0$  is generally expected to be a bad conductor.

The Dirac equation in Eq. (8) also helps explain the delocalization behaviors in ensemble II within the oval-shaped parametric region tilted along  $\theta' = 0$ .  $\theta' = 0$  leads to the average potential  $\langle V \rangle = 0$  since both are proportional to the average phase accumulated by traveling modes across the sample. The topological surface states of a sample drawn from ensemble II are gapless at  $\theta' = 0$ . On the other hand, a finite  $\langle m \rangle$  definitely opens a gap. Therefore it can be inferred that  $\langle m \rangle = 0$  for the gapless ensemble II. Thus samples in ensemble II at  $\theta' = 0$  are indeed close to the delocalization criticality. However, the delocalization transition in ensemble II is computed to exhibit an algebraic scaling

$$\gamma \sim \bar{\phi}^\nu \quad (10)$$

with  $\nu \sim 1.76$ , which results in a broadened region with vanishing  $\gamma$ . Furthermore, the fluctuation of  $\gamma$  is suppressed at delocalization, and a topological surface of a sample with magnetic disorder akin to ensemble II may actually correspond to a good conductor. We believe that this peculiarity of ensemble II is closely related to the mesoscopic average time-reversal symmetry: Samples in ensemble II by design can be divided into slabs with zero net magnetization, and as proved in Appendix F, the mass terms of each slab can effectively cancel out when  $\langle V \rangle = 0$ , leading to a delocalized state.

It is important to put our results in the context of recent studies on disordered  $\theta = \pi$  topological systems, which also point to disorder-induced surface localization [32–34]. The surface localization in these studies is a result of the breaking of average time-reversal symmetry necessary to access the axion insulator phase. In contrast, the surface localizes in our systems with average time-reversal symmetry, which is usually expected to result in surface delocalization [10]. The departure of our results from the conventional wisdom can be attributed to the surface anisotropy of time-reversal breaking in the presence of background magnetic order. The surface of a disordered  $\theta = \pi$  bulk phase can be described by a 2D random-mass Dirac Hamiltonian where surface delocalization is governed by transport properties of 1D chiral modes bound to the walls between insulating domains of positive and negative masses. It is usually assumed that the random Dirac Hamiltonian can be mapped to the Chalker-Coddington network model [35,36] that takes into account the quantum tunneling between chiral modes at nodes and effectively describes a 2D percolation picture of surface delocalization. With the background layered magnetic order, however, the orientation of chiral edge modes bound to almost parallel domain walls is highly anisotropic. Especially, in the presence of translation invariance along edges of layers as in our model, the disorder problem can be effectively reduced to 1D with a good quantum number  $k_x$ , where the 2D percolation picture apparently does not apply.

We are then in a position to clarify the connection between our results and the nontrivial topology of an AFM TI. To this end, let us consider a more generic configuration that loses the  $S$  symmetry due to any of the various imperfections, e.g., fluctuation in the exchange field strength, on-site potential, hopping strength, and magnetic domain wall or kinks, etc., although within each layer homogeneity stays intact. We study an arbitrary imperfect slab from the disordered configuration,

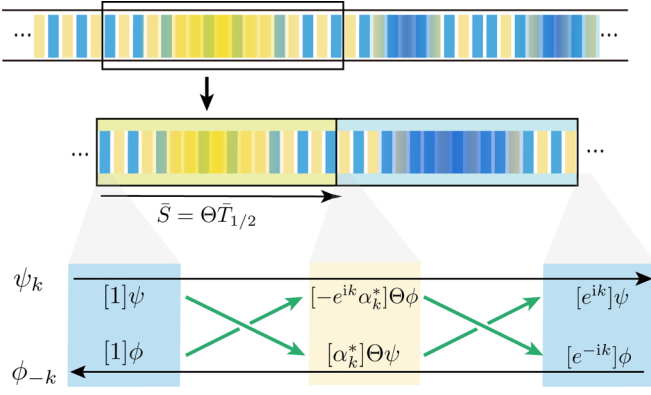


FIG. 5. Any imperfect slab in a disordered configuration can be made into an  $\bar{S}$ -symmetric superlattice with its time-reversal images.  $\psi_k$  and  $\phi_{-k}$  are two branches of superlattice surface modes along the  $k_z$  axis, and  $\bar{S}$  (green arrows) specifies the form of the transfer matrix of the imperfect slab as described in the text.

whose cleaving surfaces on both ends are separately located in a perfect AFM block. An  $\bar{S}$ -symmetric superlattice can be built from a supercell composed of the imperfect slab and its time-reversal image, as depicted in Fig. 5. The superlattice can be an AFM TI [37] and thus accommodates two branches of gapless surface states along the  $k_z$  axis ensured by the nontrivial topology. The wave functions of these surface states near both ends of the slabs (within perfect AFM layers) will be linear superpositions of the surface modes of the perfect AFM TI at the same energy, which enables a transfer matrix description of the transport property of the slab in terms of how coefficients in the linear superpositions evolve across the slab.

At a given energy  $E$ , if the wave function at a  $\uparrow$  layer (within a perfect AFM block) can be expanded in terms of two components  $\psi$  and  $\phi$ , then the wave function at a  $\downarrow$  layer (within a perfect AFM block), by  $S$  symmetry in the perfect AFM TI, can be expanded in terms of  $\Theta\phi$  and  $\Theta\psi$ . We consider evolution of expansion coefficients using  $\psi$  and  $\phi$  ( $\Theta\phi$  and  $\Theta\psi$ ) as bases in  $\uparrow$  ( $\downarrow$ ) layers and choose  $\psi$  and  $\phi$  to be the wave functions of two branches of surface modes  $\psi_{0k}$  and  $\phi_{0-k}$  of the perfect AFM TI at a  $\uparrow$  layer, dubbed  $\psi_0$  and  $\phi_0$ .  $S$  symmetry relates  $\psi_{0k}$  and  $\phi_{0-k}$ :

$$S\psi_{0k} = \alpha_k \phi_{0-k}, \quad S\phi_{0-k} = \beta_{-k} \psi_{0k}, \quad (11)$$

where  $\alpha_k$  and  $\beta_k$  are phase factors that satisfy  $\alpha_k^* \beta_{-k} = -e^{-ik}$  since  $S^2 = -T_1$  ( $T_1$  is a lattice translation). With these relations we can write down the transfer matrix  $T_{\uparrow\downarrow}$  going from a  $\uparrow$  layer to its adjacent  $\downarrow$  layer, and  $T_{\downarrow\uparrow}$  from a  $\downarrow$  layer to the next  $\uparrow$  layer,

$$T_{\uparrow\downarrow} = \alpha_k^* \begin{bmatrix} e^{-ik} & 0 \\ 0 & 1 \end{bmatrix}, \quad T_{\downarrow\uparrow} = \alpha_k \begin{bmatrix} -1 & 0 \\ 0 & e^{-ik} \end{bmatrix}. \quad (12)$$

Upon fixing the  $U(1)$  gauge degree of freedom to our advantage by assigning  $\alpha_k = -e^{ik/2}$ , we have  $T_{\uparrow\downarrow} = T_{\downarrow\uparrow}$  taking the form identical to  $T_0$  [38].

Similar arguments apply to the  $\bar{S}$ -symmetric superlattice (see Fig. 5). The transfer matrix  $M_1$  going from the slab to its time-reversal image as well as  $M_2$  from the time-reversal image to the next supercell will take the form identical to  $T_0$  but

in a different set of bases. Transformed back to the bases  $\psi_0$  and  $\phi_0$ , the transfer matrix of the imperfect slab must take the form  $T_b$  of a bond defect in Eq. (5) as required by the current conservation. Therefore, without knowing the details of the imperfection associated with the superlattice, the  $\bar{S}$  symmetry already dictates the form of the surface transfer matrices. Consequently, the generic imperfect configuration considered can be described by the same SDE as Eq. (8). Similar behaviors to those of ensemble II are then expected when a (properly generalized) mesoscopic average time-reversal symmetry is preserved.

## ACKNOWLEDGMENTS

We are grateful for stimulating discussions with X. C. Xie and H. Jiang. We acknowledge the financial support from the National Natural Science Foundation of China (Grants No. 12274003, No. 11725415, and No. 11934001), the National Key R&D Program of China (Grants No. 2018YFA0305601 and No. 2021YFA1400100), and the Innovation Program for Quantum Science and Technology (Grant No. 2021ZD0302600).

## APPENDIX A: THE EFFECTIVE TIGHT-BINDING MODEL OF $\text{MnBi}_2\text{Te}_4$

Here we describe the lattice model of an AFM TI used in the main text. We follow Zhang *et al.* [18] for the effective tight-binding model of  $\text{MnBi}_2\text{Te}_4$ , by regularizing the four-band effective  $\mathbf{k} \cdot \mathbf{p}$  model of  $\text{Bi}_2\text{Se}_3$  [19],

$$\begin{aligned} \mathcal{H}_0(\mathbf{k}) = & \varepsilon_0(\mathbf{k}) \mathbb{1}_{4 \times 4} + M(\mathbf{k}) s_0 \sigma_3 + v(k_x s_x + k_y s_y) \sigma_1 \\ & + v_z k_z s_z \sigma_1 + w(k_+^3 + k_-^3) s_0 \sigma_2, \end{aligned} \quad (A1)$$

on a 3D hexagonal lattice. In Eq. (A1),  $k_{\pm} = k_x \pm ik_y$ ,  $\varepsilon_0(\mathbf{k}) = C_0 + C_1 k_z^2 + C_2(k_x^2 + k_y^2)$ , and  $M(\mathbf{k}) = M_0 + M_1 k_z^2 + M_2(k_x^2 + k_y^2)$ .  $s_i$  and  $\sigma_i$  are Pauli matrices ( $i = 1, 2, 3$ ) or the unit matrix ( $i = 0$ ) for spin and orbital degrees of freedom, respectively. The hexagonal lattice is defined by the lattice vectors  $\mathbf{a}_1 = (1, 0, 0)$ ,  $\mathbf{a}_2 = \frac{1}{2}(1, \sqrt{3}, 0)$ , and  $\mathbf{c} = (0, 0, 1)$ . Regularization on this lattice results in a non-magnetic tight-binding model with only nearest-neighbor hoppings

$$H_0 = \sum_{\ell \mathbf{q}} \psi_{\ell \mathbf{q}}^\dagger D_0(\mathbf{q}) \psi_{\ell \mathbf{q}} + \psi_{\ell \mathbf{q}}^\dagger J_{\mathbf{q}} \psi_{\ell+1 \mathbf{q}} + \psi_{\ell+1 \mathbf{q}}^\dagger J_{\mathbf{q}}^\dagger \psi_{\ell \mathbf{q}}, \quad (A2)$$

where  $\ell$  is the layer index,  $\mathbf{q} = (q_x, q_y)$  is the in-plane crystal momentum, and  $\psi_{\ell \mathbf{q}}$  ( $\psi_{\ell \mathbf{q}}^\dagger$ ) is a four-component vector of annihilation (creation) operators.  $D_0(\mathbf{q})$  and  $J_{\mathbf{q}}$  are given by

$$\begin{aligned} D_0(\mathbf{q}) = & [\varepsilon_0(\mathbf{q}) \mathbb{1}_{4 \times 4} + M(\mathbf{q}) s_0 \sigma_3 \\ & + v(q_x s_x + q_y s_y) \sigma_1 + w(q_+^3 + q_-^3) s_0 \sigma_2 \\ & + 2C_1 \mathbb{1}_{4 \times 4} + 2M_1 s_0 \sigma_3]_{\text{regularized}}, \\ J_{\mathbf{q}} = & -C_1 \mathbb{1}_{4 \times 4} - M_1 s_0 \sigma_3 + \frac{1}{2i} v_z s_z \sigma_1, \end{aligned} \quad (A3)$$

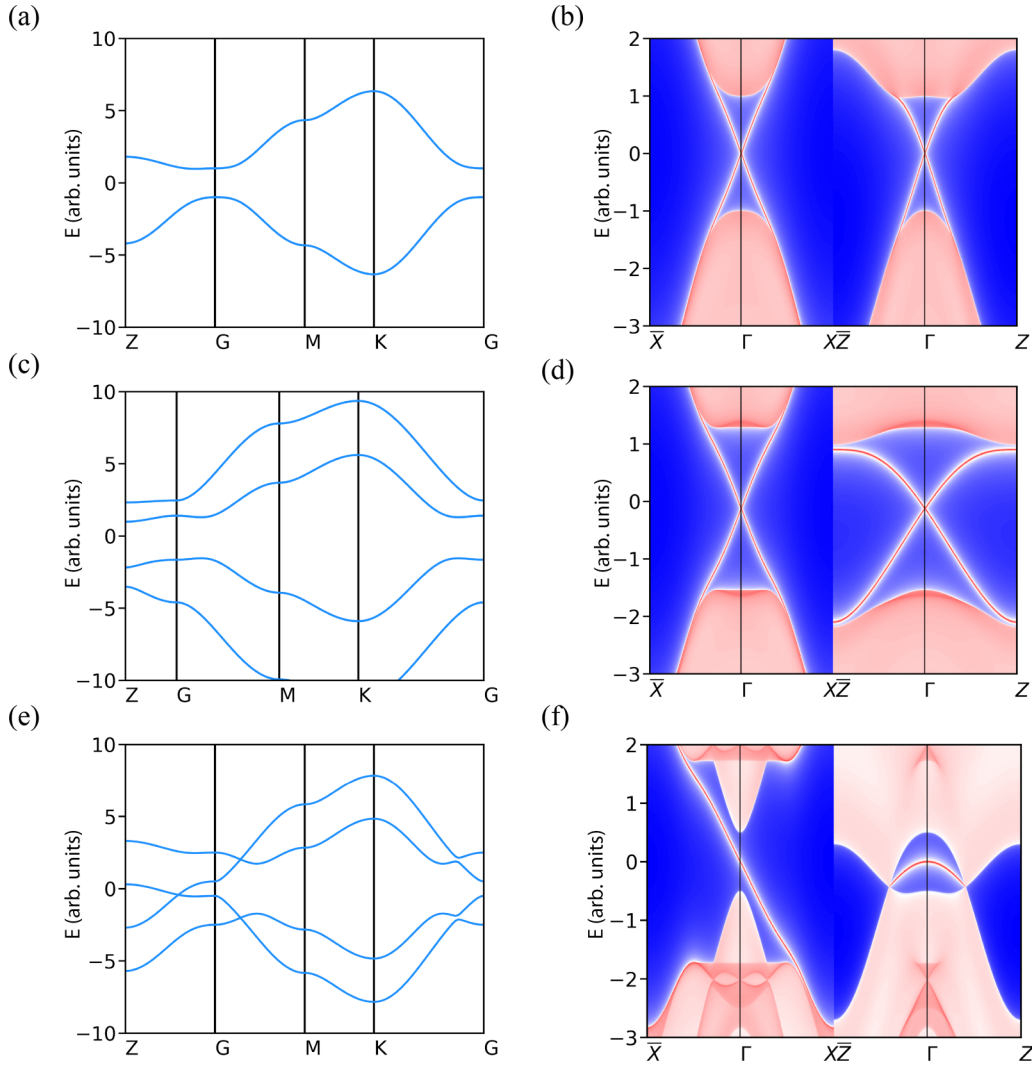


FIG. 6. Bulk and surface spectra of effect lattice model of  $\text{MnBi}_2\text{Te}_4$  in (a) and (b) the paramagnetic phase ( $m = 0$ ), (c) and (d) the AFM phase [ $m_l = (-1)^l$ ], and (e) and (f) the FM phase ( $m_l = 1$ ).

where lattice regularization within each monolayer is accomplished by the following substitutions:

$$\begin{aligned}
 q_x &\rightarrow \frac{1}{3}[2\sin(q_1) + \sin(q_2) + \sin(q_3)], \\
 q_y &\rightarrow \frac{1}{\sqrt{3}}[\sin(q_2) - \sin(q_3)], \\
 q_x^2 + q_y^2 &\rightarrow 4 - \frac{4}{3}[\cos(q_1) + \cos(q_2) + \cos(q_3)], \\
 q_x^2 - q_y^2 &\rightarrow \frac{4}{3}[-2\cos(q_1) + \cos(q_2) + \cos(q_3)], \\
 q_x q_y &\rightarrow \frac{2}{\sqrt{3}}[-\cos(q_2) + \cos(q_3)], \\
 q_+^3 + q_-^3 &\rightarrow 16[-\sin(q_1) + \sin(q_2) + \sin(q_3)].
 \end{aligned} \tag{A4}$$

Here,  $q_1$ ,  $q_2$ , and  $q_3$  are defined as

$$q_1 = \mathbf{q} \cdot \mathbf{a}_1, \quad q_2 = \mathbf{q} \cdot \mathbf{a}_2, \quad q_3 = q_1 - q_2. \tag{A5}$$

We then add a layer-dependent Ising field modulated by the order parameter  $m_\ell = \pm 1$  to each layer

$$D_q(m_\ell) = D_0(\mathbf{q}) + m_\ell \cdot m_s \sigma_0, \tag{A6}$$

which finally leads to the Hamiltonian:

$$H = \sum_{\ell q} \psi_{\ell q}^\dagger D_q(m_\ell) \psi_{\ell q} + \psi_{\ell q}^\dagger J_q \psi_{\ell+1q} + \psi_{\ell+1q}^\dagger J_q^\dagger \psi_{\ell q}. \tag{A7}$$

We have used  $M_0 = -1$ ,  $M_1 = 1$ ,  $M_2 = 1$ ,  $C_0 = 0$ ,  $C_1 = -0.3$ ,  $C_2 = 0$ ,  $v_z = 1.5$ ,  $v = 1.5$ ,  $w = 1.5/16$ , and  $m = 1.5$  (default for magnetic phases) as our choice of parameters throughout this paper. With this choice, we can reproduce the bulk topology of  $\text{MnBi}_2\text{Te}_4$  in different magnetic phases [16], namely, a TI in the paramagnetic phase, an antiferromagnetic (AFM) TI in the AFM phase, and a Weyl semimetal in the ferromagnetic (FM) phase. The bulk and surface spectra in the nonmagnetic, AFM, and FM regimes are shown in Fig. 6.

Now we describe the electronic structure of an AFM-ordered supercell with kinks, noting especially a topological phase transition driven by varying exchange strength  $m$ .

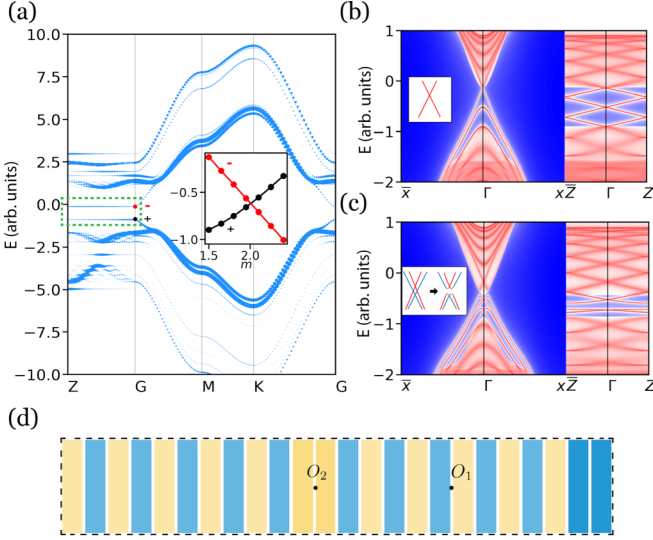


FIG. 7. Topological transition of an  $\bar{S}$ -symmetric supercell. (a) Unfolded bulk band structure of the  $\bar{S}$ -symmetric supercell shown in (d) at  $m = 1.5$ . Kink modes are highlighted (in the green box). The inset shows the energy of kink modes as a function of  $m$ , which suggests a band inversion. (b) (010) surface spectra at  $m = 1.5$  (before band inversion). (c) (010) surface spectra at  $m = 2.25$  (after band inversion); the inset shows how the surface states are hybridized. (d) Structure of the  $\bar{S}$ -symmetric supercell.  $O_1$  and  $O_2$  are the centers of inversions  $P_1$  and  $P_2$ .

Shown in Fig. 7(a) is the bulk band structure of the  $\bar{S}$ -symmetric superlattice in Fig. 7(d), unfolded [39] back to the Brillouin zone of the nonmagnetic primitive cell. The bands are at least doubly degenerate due to the symmetry  $P_1\Theta$ , where  $P_1$  is an inversion centered at  $O_1$ . The system also preserves an inversion symmetry  $P_2$  centered at  $O_2$ . Remarkably, there are two groups of quasi-2D bands dispersionless in  $k_z$  across the band gap, which are localized on the kinks (see Fig. 8). The upper and lower degenerate pair of kink modes possess opposite parity at  $\Gamma$ . Increasing the exchange field drives a band inversion between them, leading to a topological transition from the nontrivial  $\mathbb{Z}_2 = 1$  phase to the trivial  $\mathbb{Z}_2 = 0$  phase.

As described in the main text, this topological transition is accomplished through each magnetic monolayer undergoing a Chern-insulator transition. This is evident by noting that without interlayer coupling, the effective Hamiltonian of each monolayer resembles that of  $\text{Hg}_{1-y}\text{Mn}_y\text{Te}$  quantum wells [40], which transforms into a Chern insulator in the strong-exchange-field limit. The chiral edge state of a monolayer is coupled by interlayer hopping with its counterpropagating (two-half) partner(s) from adjacent layers with opposite Chern numbers, and becomes gapped except at kinks. Chiral edge states at opposite kinks merge into helical states and hybridize with the Dirac surface states of (AFM) TI, leading to an overall surface gap as shown in Figs. 8(b) and 8(h). This picture is further supported by the surface spectrum calculation of supercells shown in Figs. 8(c) and 8(g), where two kinks are of the same magnetization. Their surface spectra possess two chiral states in the strong-field limit as shown in Figs. 8(d) and 8(h).

## APPENDIX B: THE LOCAL GAUGE TRANSFORMATION

The Hamiltonian Eq. (A1) has the point group symmetry  $D_{3d}$ , which includes an  $x$ -mirror symmetry  $\mathcal{M}_x = -i s_x \sigma_3$  (according to Sec. III of Ref. [19],  $\mathcal{M}_x = \mathcal{R}_x(\pi)\mathcal{P} = i s_x \sigma_0$  before a gauge transformation  $U_1 = \text{diag}(1, -i, 1, i)$ , differing from our convention by an insignificant minus sign). As  $\mathcal{M}_x$  flips the exchange field along the  $z$  axis, we have the following symmetry relations:

$$\begin{aligned} \mathcal{M}_x D_{\mathbf{q}}(m_\ell) \mathcal{M}_x^{-1} &= D_{\bar{\mathbf{q}}}(-m_\ell), \\ \mathcal{M}_x J_{\mathbf{q}} \mathcal{M}_x^{-1} &= J_{\bar{\mathbf{q}}}, \end{aligned} \quad (\text{B1})$$

with  $\bar{\mathbf{q}} = (-q_x, q_y)$ .

We consider a local gauge transformation to the Hamiltonian Eq. (A7) that applies  $i\mathcal{M}_x$  to every  $\downarrow$  layer, and the resultant Hamiltonian is

$$H' = \sum_{\ell q} \psi_{\ell q}^\dagger D_{\ell q} \psi_{\ell q} + \psi_{\ell q}^\dagger J_{\ell q} \psi_{\ell+1q} + \psi_{\ell+1q}^\dagger J_{\ell q}^\dagger \psi_{\ell q}, \quad (\text{B2})$$

where

$$D_{\ell q} = \begin{cases} D_{\mathbf{q}}(+1), & m_\ell = +1 \\ D_{\bar{\mathbf{q}}}(+1), & m_\ell = -1 \end{cases}$$

and the interlayer hopping  $J_{\ell q}$  depends on the magnetization of two layers it connects in the kink model:

$$J_{\ell q} = \begin{cases} J_{\mathbf{q}}, & \uparrow\uparrow \\ i\mathcal{M}_x J_{\mathbf{q}} i\mathcal{M}_x = J_{\bar{\mathbf{q}}}, & \downarrow\downarrow \\ J_{\mathbf{q}} i\mathcal{M}_x = i\mathcal{M}_x J_{\bar{\mathbf{q}}}, & \uparrow\downarrow \\ i\mathcal{M}_x J_{\mathbf{q}}, & \downarrow\uparrow. \end{cases}$$

The model in Eq. (B2) describes a bond defect model at  $k_x = 0$  ( $\mathbf{q} = \bar{\mathbf{q}}$ ) that possesses a uniform on-site term  $D_{\ell q} = D_{\ell k_y}$  and uniform interlayer hoppings  $J_{\ell q} = i\mathcal{M}_x J_{k_y}$  except at the center of kinks  $J_{\ell q} = J_{k_y}$ .

## APPENDIX C: CALCULATION DETAILS OF FIGURES 2 AND 4

We have generated a series of AFM-ordered supercells with kinks, whose surface gaps are calculated using the iterative Green's function method [21] and plotted against the Ising moment in the main text. The supercells are generated by first shuffling 30  $\uparrow$  and 30  $\downarrow$  layers and performing simulated annealing on the magnetic moments assuming an AFM interlayer coupling. Supercells with a net moment or with domain wall separation of  $< 3$  layers are discarded. In the end, we obtain, from 4000 generated samples, 934 supercells with two domain walls, 2025 supercells with four domain walls, and 892 supercells with six domain walls. Surface gaps of supercells with two domain walls are also found as a function of Ising moments but are not displayed because supercells with equal  $\mathcal{I}$  are equivalent up to translation, and the confinement effect cannot be excluded as a confounding factor.

The total transfer matrix  $T_n$  of a bond defect configuration discussed in the main text is

$$T_n = T_0^{d_n}(k_z) T_b(\theta, u) \cdots T_0^{d_1}(k_z) T_b(\theta, u), \quad (\text{C1})$$

where  $d_i = p_i - 2L$  is the number of monolayers between the  $i$ th and  $(i+1)$ th defect zones. We calculate the Lyapunov



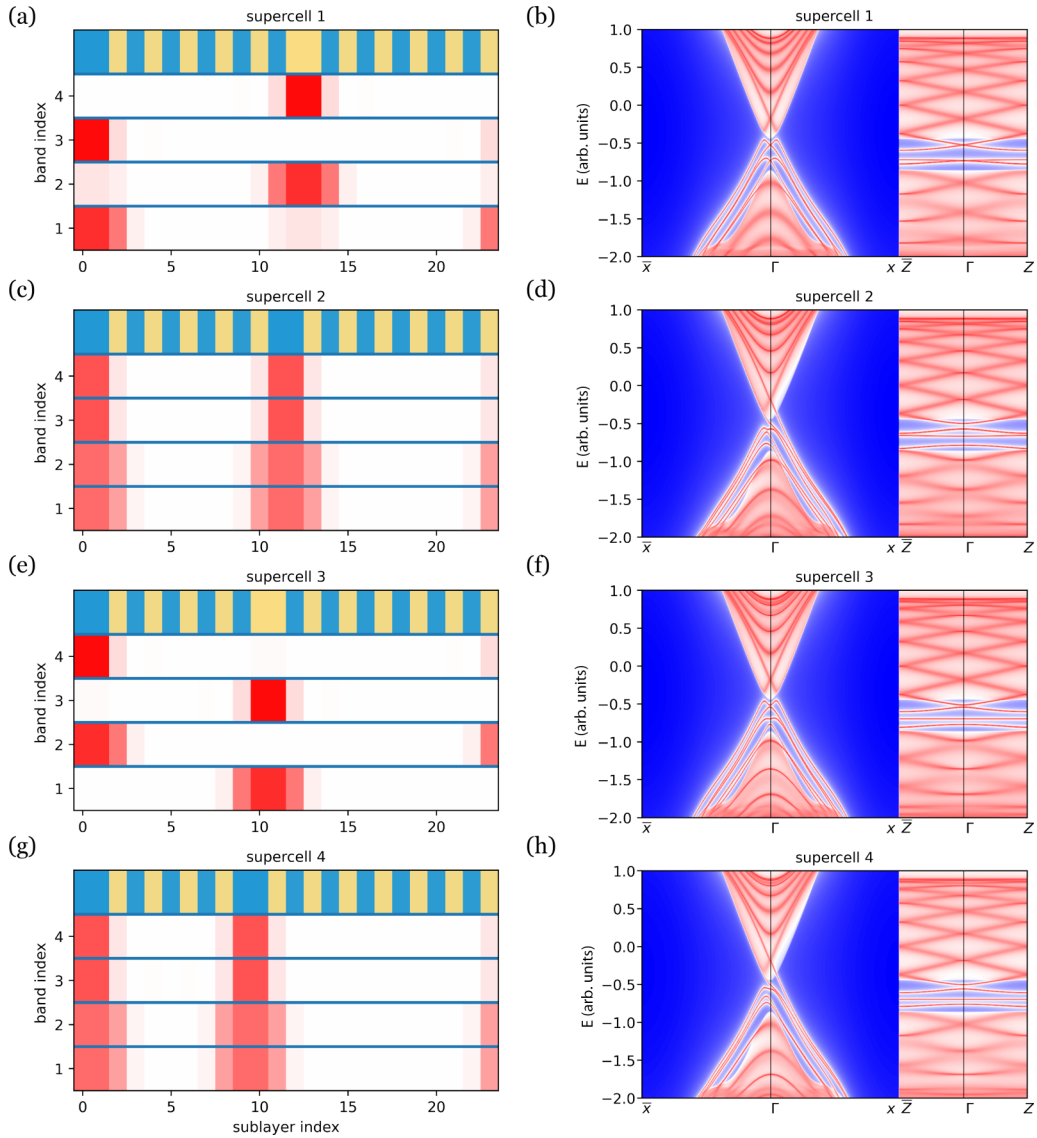


FIG. 8. (a), (c), (e), and (g) Local density of kink modes at  $\Gamma$  in different supercells at  $m = 2.25$  (after band inversion). (b), (d), (f), and (h) (010) surface spectra of the same supercells.

exponent of  $T_n$  by QR decomposition [25] with  $T_0^{d_i}(k_z)$  calculated as  $\sigma_z^{d_i} T(\phi_i)$ , where  $\phi_i = k_z d_i$ . On the premise that separations of the neighboring kinks independently follow an identical exponential distribution,  $\phi_i$  is sampled as a continuous variable via  $\phi_i = \bar{\phi}x$  with  $x$  drawn according to the probability density  $p(x) = \exp(-x)$ ,  $x \in [0, \infty)$ . As only whether  $d_i$  is even or odd enters into  $\sigma_z^{d_i}$ , they are sampled as a binary sequence. In ensemble I,  $d_i$ 's are sampled as a sequence of random integers  $\pm 1$ , fulfilling the condition  $n_\uparrow = n_\downarrow$ . In ensemble II,  $d_i$ 's of each slab are randomly chosen from all possible sequences of  $d_i$ 's of maximum length 6 satisfying  $n_\uparrow = n_\downarrow$ .

#### APPENDIX D: DERIVATION OF EQUATION (3)

The definition of the Ising moment

$$\mathcal{I} = \frac{1}{N} \sum_{\ell=1}^N \ell m_\ell \pmod{1} \quad (\text{D1})$$

is simple for numerical calculation but not intuitive. As mentioned in this paper, keeping track of the change in  $\mathcal{I}$  in restoring the perfect AFM ordering ( $\mathcal{I} = 0.5$ ) by pair annihilation of opposite kinks  $\uparrow$  and  $\downarrow$  leads to an expression of  $\mathcal{I}$  in terms of separations between kinks

$$\mathcal{I} = 0.5 - \frac{1}{N} \sum_{i=1}^n \alpha_i p_i, \quad (\text{D2})$$

where  $p_i - 2$  is the number of monolayers between the  $i$ th and  $(i + 1)$ th kinks and  $\alpha_i$  is  $n_\downarrow - n_\uparrow$  calculated from the  $(i + 1)$ th to the  $n$ th kinks.

An elementary kink migration is accomplished by a transposition of an adjacent pair of oppositely magnetized layers, which can be either  $\uparrow\downarrow \mapsto \downarrow\uparrow$  (transposition A) or  $\downarrow\uparrow \mapsto \uparrow\downarrow$  (transposition B) and results in a kink moving by  $\pm 2$  layers. Transposition A (B) changes  $\mathcal{I}$  by  $2/N$  ( $-2/N$ ). Neighboring kinks with identical (opposite) magnetization are separated by an odd (even) number of layers. To annihilate a neighboring

pair of opposite kinks, say, the  $i$ th kink  $\uparrow$  and the  $(i+1)$ th kink  $\downarrow$ , we have to apply transposition A to the  $\uparrow$  kink  $\frac{p_i-2}{2}$  times to bring two kinks in touch and one more transposition A to annihilate them. This process causes a total change of the Ising moment  $\Delta\mathcal{I} = p_i$ . There will be net  $\alpha_i \uparrow$  kinks propagating upward across this  $p_i$  slab to restore zero net magnetization on both sides, and thus we arrive at Eq. (D2).

We set up a symbolic representation of AFM-ordered supercells with kinks. As usual,  $\uparrow$  and  $\downarrow$  stand for kinks with upward and downward magnetization, respectively. We further use  $\boxed{p'_i}$  to denote a perfectly AFM-ordered separation between the  $i$ th and  $(i+1)$ th kinks made of  $p'_i = p_i - 2$  layers. We list here the results of the Ising moments for supercells with four and six kinks with  $\mathcal{I}^*$  read from the surface gap calculation labeled as follows.

(i) Supercell “ $\downarrow \boxed{p'_1} \uparrow \boxed{p'_2} \downarrow \boxed{p'_3} \uparrow \boxed{p'_4}$ ,”  $\mathcal{I}^* = 0$

$$\mathcal{I} = 0.5 + \frac{p_1 + p_3}{N} \pmod{1}. \quad (\text{D3})$$

(ii) Supercell “ $\downarrow \boxed{p'_1} \downarrow \boxed{p'_2} \uparrow \boxed{p'_3} \uparrow \boxed{p'_4}$ ,”  $\mathcal{I}^* = 0.5$

$$\mathcal{I} = 0.5 + \frac{p_1 + 2p_2 + p_3}{N} \pmod{1}. \quad (\text{D4})$$

(iii) Supercell “ $\downarrow \boxed{p'_1} \downarrow \boxed{p'_2} \downarrow \boxed{p'_3} \uparrow \boxed{p'_4} \uparrow \boxed{p'_5} \uparrow \boxed{p'_6}$ ,”  $\mathcal{I}^* = 0$

$$\mathcal{I} = 0.5 + \frac{p_1 + 2p_2 + 3p_3 + 2p_4 + p_5}{N} \pmod{1}. \quad (\text{D5})$$

(iv) Supercell “ $\downarrow \boxed{p'_1} \downarrow \boxed{p'_2} \uparrow \boxed{p'_3} \uparrow \boxed{p'_4} \downarrow \boxed{p'_5} \uparrow \boxed{p'_6}$ ,”  $\mathcal{I}^* = 1/3$

$$\mathcal{I} = 0.5 + \frac{p_1 + 2p_2 + p_3 + p_5}{N} \pmod{1}. \quad (\text{D6})$$

(v) Supercell “ $\uparrow \boxed{p'_1} \uparrow \boxed{p'_2} \downarrow \boxed{p'_3} \downarrow \boxed{p'_4} \uparrow \boxed{p'_5} \downarrow \boxed{p'_6}$ ,”  $\mathcal{I}^* = 2/3$

$$\mathcal{I} = 0.5 - \frac{p_1 + 2p_2 + p_3 + p_5}{N} \pmod{1}. \quad (\text{D7})$$

(vi) Supercell “ $\downarrow \boxed{p'_1} \uparrow \boxed{p'_2} \downarrow \boxed{p'_3} \uparrow \boxed{p'_4} \downarrow \boxed{p'_5} \uparrow \boxed{p'_6}$ ,”  $\mathcal{I}^* = 0$

$$\mathcal{I} = 0.5 + \frac{p_1 + p_3 + p_5}{N} \pmod{1}. \quad (\text{D8})$$

We notice that  $\mathcal{I}^*$  is achieved when separations are equal  $p_i = N/n$ . We further mention that  $\sum_i p_i = N$  can be used to cancel a specific  $p_i$  in the above expressions, which amounts to choosing a different supercell with  $\boxed{p'_i}$  at the end and will not change the value of the Ising moment  $\mathcal{I}$ .

#### APPENDIX E: SURFACE TRANSFER MATRIX OF A DEFECT ZONE

We have defined a defect zone composed of  $2L$  layers centered at a bond defect.  $L$  should be large enough to write

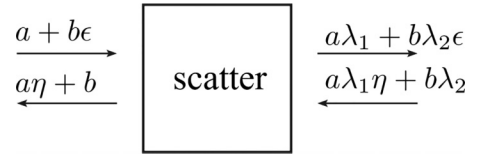


FIG. 9. A generic scattering state of  $T_b$ .

down its surface transfer matrix  $T_b$  in the channel basis of the homogeneous FM bulk.

The generic eigendecomposition of  $T_b$  is

$$T_b = \begin{pmatrix} 1 & \epsilon \\ \eta & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} \begin{pmatrix} 1 & \epsilon \\ \eta & 1 \end{pmatrix}^{-1}, \quad (\text{E1})$$

where  $\lambda_1 = e^{i\theta}$  and  $\lambda_2 = -e^{i\theta}$  supports traveling modes. By considering a scattering state made of the linear combination of eigenvectors of  $T_b$  (see Fig. 9), we are led to the following constraints due to flux conservation:

$$\begin{aligned} (1 - |\lambda_1|^2)(1 - |\eta|^2) &= 0, \\ (1 - |\lambda_2|^2)(1 - |\epsilon|^2) &= 0, \\ (\epsilon - \eta^*)(1 - \lambda_1^* \lambda_2) &= 0. \end{aligned} \quad (\text{E2})$$

From the third line in Eq. (E2), we find  $\epsilon = \eta^*$  if  $\lambda_1^* \lambda_2 \neq 1$ , which is exactly our case.

#### APPENDIX F: MAPPING TO 1D DIRAC EQUATION

The surface transfer matrices of a single FM layer  $T_0$  and a defect zone  $T_b$  both satisfy  $\sigma_1 T_{0/b} \sigma_1 = -T_{0/b}^*$ . Thus any composite transfer matrix  $M$  composed of an even number of  $T_{0/b}$  will satisfy the effective time-reversal constraint  $\sigma_1 M \sigma_1 = M^*$  characteristic of Hamiltonians of orthogonal class [30,41]. It follows that  $M$  has the following polar decomposition:

$$M = \exp(i\rho\sigma_3) \exp(\eta\sigma_2) \exp(i\zeta\sigma_3). \quad (\text{F1})$$

Clearly, configurations in ensemble I and ensemble II can be easily divided into such slabs whose transfer matrices  $M_i$  satisfy the time-reversal constraint and equivalently describe the solution of a 1D SDE:

$$[-i\partial_x + V(x, E)\sigma_3 - im(x, E)\sigma_2]\Psi(x) = 0, \quad (\text{F2})$$

where  $V(x, E) = -\sum_i (\zeta_i(E)\delta(x - x_{i1}) + \rho_i(E)\delta(x - x_{i3}))$  and  $m(x, E) = -\sum_i \eta_i(E)\delta(x - x_{i2})$  describe an energy-dependent potential and mass term due to point scatters ( $x_{i1} = x_i - 0^+$ ,  $x_{i2} = x_i$ ,  $x_{i3} = x_i + 0^+$ ,  $x_i < x_{i+1} \forall i$ ). The solution of the Dirac Eq. (F2) can be expressed as

$$\begin{aligned} \Psi(x, E) &= \mathcal{P} \exp \left[ -i \int_0^x dx' (V(x', E)\sigma_3 - im(x', E)\sigma_2) \right] \\ &\times \Psi(0, E), \end{aligned} \quad (\text{F3})$$

where the path-ordering  $\mathcal{P}$  ensures separation of the transfer matrix into the product of exponentials as in Eq. (F1). Equation (F2) has been transformed into

$$[-i\sigma_3\partial_x + V(x, E) + m(x, E)\sigma_1]\Psi(x) = 0 \quad (\text{F4})$$

in the main text, a more popular form found in the literature [27,29,42].

A configuration in ensemble II by construction has a natural division in which each slab has zero magnetization. The

transfer matrix of one such slab is related to the Ising moments at the energy when its own  $\theta' = 0$  as discussed in this paper. This connection seems to suggest the following relations:

$$\rho_j(E) + \zeta_j(E) \sim \sum_{i \in S_j} (k_z d_i + \theta), \quad (\text{F5a})$$

$$\eta_j(E) \sim 4u' \sum_{i \in S_j} \alpha_i (k_z d_i + \theta), \quad (\text{F5b})$$

where  $S_j$  stands for the  $j$ th slab and the summations run over all the kinks within this slab. At  $\theta' = 0$ , we have

$$-\int dx V(x) = \sum_j \rho_j(E) + \zeta_j(E) \sim \sum_i (k_z d_i + \theta) = 0, \quad (\text{F6})$$

which means  $\theta = -k_z \bar{d}$ . The sum of  $m(x)$  becomes

$$\begin{aligned} -\int dx m(x) &= \sum_j \eta_j(E) \sim 4u' \sum_i \alpha_i k_z (d_i - \bar{d}) \\ &= 4u' N k_z (\mathcal{I} - \mathcal{I}^*). \end{aligned} \quad (\text{F7})$$

This ‘‘explains’’ how nonzero  $\mathcal{I} - \mathcal{I}^*$  typically corresponds to a nonzero average mass that opens the surface gap. The relation Eq. (F7) is supported by Eq. (6) in the main text,  $\text{tr} M_i = 2 + \eta_i^2 + O(\eta^4)$  when its own  $\theta' = \rho_i + \zeta_i = 0$ .

It remains to establish that the mass of each slab can be meaningfully added, as we merge slabs into one. This can be proved by noting the following identity:

$$\begin{aligned} \exp(\eta_1 \sigma_2) \exp(i\xi \sigma_3) \exp(\eta_2 \sigma_2) \\ \simeq \exp(i\xi_1 \sigma_3) \exp(\eta \sigma_2) \exp(i\xi_2 \sigma_3) + o(\text{2nd order terms}) \end{aligned} \quad (\text{F8})$$

with  $\eta = \eta_1 + \eta_2$ ,  $\xi_1 = \xi \eta_2 / \eta$ , and  $\xi_2 = \xi \eta_1 / \eta$  when  $\eta \neq 0$ . Thus the composition  $M(\rho_1, \eta_1, \xi_1) M(\rho_2, \eta_2, \xi_2)$  will be a transfer matrix  $M(\rho, \eta, \xi)$  with  $\eta = \eta_1 + \eta_2$ . We will provide a proof of the conjecture Eqs. (F5a) and (F5b) next.

## APPENDIX G: DERIVATION OF EQUATION (6)

The transfer matrix  $T_b$  of a defect zone has the following decomposition:

$$T_b = W(u) T_0(k_z) W(u)^{-1} = \exp(i\rho \sigma_3) \exp(\eta \sigma_1) \exp(i\xi \sigma_3) \sigma_3.$$

In the limit of small  $u$ ,  $\rho + \zeta = \theta$ , and  $\eta \sim O(|u|)$ . The total transfer matrix  $T_n$  of a bond defect configuration is given by

$$\begin{aligned} T_n &= T_0^{d_n}(k_z) T_b(\theta, u) \cdots T_0^{d_1}(k_z) T_b(\theta, u) \\ &= \exp(i\phi_n \sigma_3) \sigma_3^{d_n} T_b \cdots \exp(i\phi_1 \sigma_3) \sigma_3^{d_1} T_b(\theta, u) \\ &= \exp(-i\xi \sigma_3) [\exp(i\Psi_n \sigma_3) \sigma_3^{d_n} \exp(\eta \sigma_1) \sigma_3] \\ &\quad \cdots [\exp(i\Psi_1 \sigma_3) \sigma_3^{d_1} \exp(\eta \sigma_1) \sigma_3] \exp(i\xi \sigma_3), \end{aligned} \quad (\text{G1})$$

with  $\Psi_i = \rho + \zeta + \phi_i$ . Using  $\sigma_3 \exp(\eta \sigma_1) \sigma_3 = \exp(-\eta \sigma_1)$ , we will eliminate all  $\sigma_3^{d_i}$  terms in Eq. (G1), which leads to

$$\begin{aligned} T_n &= \exp(-i\xi \sigma_3) \exp(i\Psi_n \sigma_3) \exp(m_n \eta \sigma_1) \\ &\quad \cdots \exp(i\Psi_1 \sigma_3) \exp(m_1 \eta \sigma_1) \exp(i\xi \sigma_3), \end{aligned} \quad (\text{G2})$$

where  $m_i = \pm 1$  and  $m_1 = -1$ . It can be shown that  $m_i$  and  $m_{i+1}$  will have opposite (identical) signs when  $d_i$  is even (odd). This is similar to the situation of kink magnetization, i.e., an even (odd) number of monolayers separate opposite (identical) kinks. Thus  $m_i$ 's just indicate the sign of the  $z$  component of the  $i$ th kink magnetization (up to an overall minus sign). This one-to-one correspondence between  $m_i$ 's and the sequence of kink magnetization leads us to consider the counterpart of pair annihilation,

$$\begin{aligned} \cdots \exp(\pm \eta \sigma_1) \exp(i\Psi_i \sigma_3) \exp(\mp \eta \sigma_1) \cdots \\ \simeq \cdots \exp\left(i \frac{\Psi_i}{2} \sigma_3\right) \exp(\pm 2\eta \Psi_i \sigma_2) \exp i \frac{\Psi_i}{2} \sigma_3 \cdots \quad (\text{G3}) \\ \simeq \cdots \exp(i\Psi_i \sigma_3) \cdots \exp(\pm 2\eta \Psi_i \sigma_2) + o(\text{2nd order terms}), \end{aligned} \quad (\text{G4})$$

where we have used the fact that  $\exp(\pm 2\eta \Psi_i \sigma_2)$  commutes with other  $\exp(i\Psi_i \sigma_3)$  or  $\exp(\pm \eta \sigma_1)$  terms up to second order in  $\Psi_i$  and  $\eta$ . After annihilating all pairs of opposite kinks, the coefficient before  $\sigma_2$  will bear a similar mathematical structure to that of the Ising moment. We find

$$\begin{aligned} T_n \simeq \exp\left(i \sum_i \Psi_i \sigma_3\right) \exp\left(\pm 2\eta \sum_i \alpha_i \Psi_i \sigma_2\right) \\ + o(\text{2nd order terms}). \end{aligned} \quad (\text{G5})$$

where the  $\pm$  sign should be  $+$  ( $-$ ) when the first kink is  $\Downarrow$  ( $\Uparrow$ ). After identification  $\Psi_i = k_z d_i + \theta$  and  $\eta = 2u'$ , Eq. (G5) confirms the conjecture Eqs. (F5a) and (F5b) in Appendix F and therefore validates all the deductions from it, including Eq. (6) in the main text.

- [1] L. Fu, Topological Crystalline Insulators, *Phys. Rev. Lett.* **106**, 106802 (2011).
- [2] T. H. Hsieh, H. Lin, J. Liu, W. Duan, A. Bansil, and L. Fu, Topological crystalline insulators in the SnTe material class, *Nat. Commun.* **3**, 982 (2012).
- [3] K. Nomura, M. Koshino, and S. Ryu, Topological Delocalization of Two-Dimensional Massless Dirac Fermions, *Phys. Rev. Lett.* **99**, 146806 (2007).
- [4] A. Vishwanath and T. Senthil, Physics of Three-Dimensional Bosonic Topological Insulators: Surface-Deconfined Criticality

and Quantized Magnetoelectric Effect, *Phys. Rev. X* **3**, 011016 (2013).

- [5] H. Isobe and L. Fu, Theory of interacting topological crystalline insulators, *Phys. Rev. B* **92**, 081304(R) (2015).
- [6] D. F. Mross, A. Essin, and J. Alicea, Composite Dirac Liquids: Parent States for Symmetric Surface Topological Order, *Phys. Rev. X* **5**, 011011 (2015).
- [7] M. A. Metlitski and A. Vishwanath, Particle-vortex duality of two-dimensional Dirac fermion from electric-magnetic duality

- of three-dimensional topological insulators, *Phys. Rev. B* **93**, 245151 (2016).
- [8] H. Song, S.-J. Huang, L. Fu, and M. Hermele, Topological Phases Protected by Point Group Symmetry, *Phys. Rev. X* **7**, 011020 (2017).
- [9] Y. Ando and L. Fu, Topological crystalline insulators and topological superconductors: From concepts to materials, *Annu. Rev. Condens. Matter Phys.* **6**, 361 (2015).
- [10] L. Fu and C. L. Kane, Topology, Delocalization via Average Symmetry and the Symplectic Anderson Transition, *Phys. Rev. Lett.* **109**, 246605 (2012).
- [11] I. C. Fulga, B. van Heck, J. M. Edge, and A. R. Akhmerov, Statistical topological insulators, *Phys. Rev. B* **89**, 155424 (2014).
- [12] R.-X. Zhang and C.-X. Liu, Topological magnetic crystalline insulators and corepresentation theory, *Phys. Rev. B* **91**, 115317 (2015).
- [13] R. S. K. Mong, A. M. Essin, and J. E. Moore, Antiferromagnetic topological insulators, *Phys. Rev. B* **81**, 245209 (2010).
- [14] D. Zhang, M. Shi, T. Zhu, D. Xing, H. Zhang, and J. Wang, Topological Axion States in the Magnetic Insulator  $\text{MnBi}_2\text{Te}_4$  with the Quantized Magnetoelectric Effect, *Phys. Rev. Lett.* **122**, 206401 (2019).
- [15] M. M. Otrokov, I. I. Klimovskikh, H. Bentmann, D. Estyunin, A. Zeugner, Z. S. Aliev, S. Gaß, A. U. B. Wolter, A. V. Koroleva, A. M. Shikin, M. Blanco-Rey, M. Hoffmann, I. P. Rusinov, A. Yu. Vyazovskaya, S. V. Eremeev, Yu. M. Koroteev, V. M. Kuznetsov, F. Freyse, J. Sánchez-Barriga, I. R. Amiraslanov *et al.*, Prediction and observation of an antiferromagnetic topological insulator, *Nature (London)* **576**, 416 (2019).
- [16] J. Li, Y. Li, S. Du, Z. Wang, B.-L. Gu, S.-C. Zhang, K. He, W. Duan, and Y. Xu, Intrinsic magnetic topological insulators in van der Waals layered  $\text{MnBi}_2\text{Te}_4$ -family materials, *Sci. Adv.* **5**, eaaw5685 (2019).
- [17] Y. Gong, J. Guo, J. Li, K. Zhu, M. Liao, X. Liu, Q. Zhang, L. Gu, L. Tang, X. Feng, D. Zhang, W. Li, C. Song, L. Wang, P. Yu, X. Chen, Y. Wang, H. Yao, W. Duan, Y. Xu *et al.*, Experimental realization of an intrinsic magnetic topological insulator, *Chin. Phys. Lett.* **36**, 076801 (2019).
- [18] R.-X. Zhang, F. Wu, and S. Das Sarma, Möbius Insulator and Higher-Order Topology in  $\text{MnBi}_{2n}\text{Te}_{3n+1}$ , *Phys. Rev. Lett.* **124**, 136407 (2020).
- [19] C.-X. Liu, X.-L. Qi, H. Zhang, X. Dai, Z. Fang, and S.-C. Zhang, Model Hamiltonian for topological insulators, *Phys. Rev. B* **82**, 045122 (2010).
- [20] L. Fu, Hexagonal Warping Effects in the Surface States of the Topological Insulator  $\text{Bi}_2\text{Te}_3$ , *Phys. Rev. Lett.* **103**, 266801 (2009).
- [21] M. P. L. Sancho, J. M. L. Sancho, J. M. L. Sancho, and J. Rubio, Highly convergent schemes for the calculation of bulk and surface Green functions, *J. Phys. F: Met. Phys.* **15**, 851 (1985).
- [22] I. Y. Gol'dsheid and G. A. Margulis, Lyapunov indices of a product of random matrices, *Russ. Math. Surv.* **44**, 11 (1989).
- [23] B. Kramer and A. MacKinnon, Localization: theory and experiment, *Rep. Prog. Phys.* **56**, 1469 (1993).
- [24] J. L. Pichard and G. André, Many-channel transmission: Large volume limit of the distribution of localization lengths and one-parameter scaling, *Europhys. Lett.* **2**, 477 (1986).
- [25] K. Geist, U. Parlitz, and W. Lauterborn, Comparison of different methods for computing Lyapunov exponents, *Prog. Theor. Phys.* **83**, 875 (1990).
- [26] N. Ashcroft and N. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976).
- [27] L. Balents and M. P. A. Fisher, Delocalization transition via supersymmetry in one dimension, *Phys. Rev. B* **56**, 12970 (1997).
- [28] M. Steiner, Y. Chen, M. Fabrizio, and A. O. Gogolin, Statistical properties of a localization-delocalization transition in one dimension, *Phys. Rev. B* **59**, 14848 (1999).
- [29] F. Evers and A. D. Mirlin, Anderson transitions, *Rev. Mod. Phys.* **80**, 1355 (2008).
- [30] P. A. Mello and J.-L. Pichard, Symmetries and parametrization of the transfer matrix in electronic quantum transport theory, *J. Phys. I* **1**, 493 (1991).
- [31] A. Comtet, C. Texier, and Y. Tourigny, Products of random matrices and generalised quantum point scatterers, *J. Stat. Phys.* **140**, 427 (2010).
- [32] K. Nomura and N. Nagaosa, Surface-Quantized Anomalous Hall Current and the Magnetoelectric Effect in Magnetically Disordered Topological Insulators, *Phys. Rev. Lett.* **106**, 166802 (2011).
- [33] H. Li, H. Jiang, C.-Z. Chen, and X. C. Xie, Critical Behavior and Universal Signature of an Axion Insulator State, *Phys. Rev. Lett.* **126**, 156601 (2021).
- [34] Z.-D. Song, B. Lian, R. Queiroz, R. Ijan, B. A. Bernevig, and A. Stern, Delocalization Transition of a Disordered Axion Insulator, *Phys. Rev. Lett.* **127**, 016602 (2021).
- [35] J. T. Chalker and P. D. Coddington, Percolation, quantum tunnelling and the integer Hall effect, *J. Phys. C: Solid State Phys.* **21**, 2665 (1988).
- [36] C.-M. Ho and J. T. Chalker, Models for the integer quantum Hall effect: The network model, the Dirac equation, and a tight-binding Hamiltonian, *Phys. Rev. B* **54**, 8708 (1996).
- [37] This is provided that the exchange field is not too strong or, more generally, if the  $\bar{S}$  superlattice can adiabatically deform into the AFM TI in the clean limit without breaking the  $\bar{S}$  symmetry. Adiabatic connection to the 3D TI is convenient in the construction of such deformation but is not necessary.
- [38] This means that transfer matrices in Eqs. (4) and (5) can be obtained without local gauge transformation that depends on the mirror symmetry, though doing so was convenient.
- [39] W. Ku, T. Berlijn, and C.-C. Lee, Unfolding First-Principles Band Structures, *Phys. Rev. Lett.* **104**, 216401 (2010).
- [40] C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Quantum Anomalous Hall Effect in  $\text{Hg}_{1-y}\text{Mn}_y\text{Te}$  Quantum Wells, *Phys. Rev. Lett.* **101**, 146802 (2008).
- [41] M. Caselle and U. Magnea, Random matrix theory and symmetric spaces, *Phys. Rep.* **394**, 41 (2004).
- [42] H. Mathur, Feynman's propagator applied to network models of localization, *Phys. Rev. B* **56**, 15794 (1997).