

Topological surface-state destruction via trivializing proximity effect: Lattice localization despite continuum criticality

Arthur Niwazuki¹ and Matthew S. Foster^{1,2}

¹*Department of Physics and Astronomy, Rice University, Houston, Texas 77005, USA*

²*Rice Center for Quantum Materials, Rice University, Houston, Texas 77005, USA*

(Dated: November 26, 2025)

In a significant conceptual revision to the tenfold classification scheme for topological insulators and superconductors, it was recently demonstrated that most three-dimensional (3D) classes are simultaneously “localizable” in two distinct, but intricately connected ways: (1) There is no obstruction to Wannier localization of *all* bulk eigenstates, and (2) *Almost all* surface states can be Anderson localized by arbitrarily weak symmetry-preserving quenched disorder. Here we consider the localizable class CI in 3D, and numerically investigate the stability of surface states. We demonstrate that surface states of a bulk class-CI topological lattice model are fragile in that they can be Anderson localized by the combination of weak quenched randomness and hybridization with an additional trivial 2D band (a trivializing proximity effect, TPE). With the TPE, stronger disorder is more destructive to the surface states of the bulk lattice model. By contrast, without additional bands the surface states remain extended, exhibiting robust spectrum-wide quantum criticality. We also investigate the fragility of surface states in effective 2D class-CI continuum Dirac theories, including the chiral limit of the Bistritzer-MacDonald model for twisted bilayer graphene. Although the continuum models exhibit signs of Anderson localization near gap edges for weak disorder, stronger disorder instead appears to *heal* the surface, restoring criticality whilst filling in spectral energy gaps. Our results provide further evidence that effective continuum field theories fail to capture key aspects of surface-state physics in localizable topological phases.

I. INTRODUCTION

A. Localizable topology

One of the main attractions of topological materials [1–4] is the robust conducting edge or surface states that appear at the boundary of an otherwise incompressible bulk. The boundary states are anomalous in that they represent “half” of a normal ($d - 1$)-D conducting system (with the other half transplanted to the opposite side of the d -dimensional bulk). An inferred physical consequence of this anomaly is that edge and surface states are expected to be topologically protected against all manner of weak symmetry-preserving perturbations, including quenched disorder and interactions. This attribute is famously responsible for the precise quantization of the edge-state mediated Hall conductance in the integer quantum Hall effect [5].

This canonical story was however recently upended in Ref. [6], which argued that the standard ten-fold classification scheme for topological insulators and superconductors [7–9] was incomplete. In particular, this work established that topological classes in one, two, and three dimensions (3D) must be further subdivided into two distinct categories: non-localizable and localizable classes. Different from the quantum Hall effect and the usual $\mathbb{Z}_2$ topological insulators, localizable classes exhibit a host of unconventional features: (1) *all* bulk states can be Wannier localized, (2) surface states can be *detached* from the bulk in energy (via a direct or indirect gap) with a suitable symmetry-preserving surface perturbation, and (3) *almost all* surface states can be Anderson localized by arbitrarily weak disorder. *Zero-energy* surface states

remain robustly protected [10, 11].

Non-localizable classes by contrast are guaranteed to possess an obstruction to exponential Wannier localization in the bulk. Edge or surface states are protected from Anderson localization by a spectral flow principle connecting the bulk and boundary spectra. This spectral-flow principle, which does not exist for localizable phases, is responsible for the precise quantization of the quantum Hall conductance involving current flow *through the bulk* in the Corbino geometry [12, 13].

All one-dimensional (two-dimensional) topological phases are localizable (non-localizable), while three of five of the classes in 3D are localizable [14]. The refined classification scheme has since been extended to all dimensions [15–17], with interesting connections established between detachable boundary states and the classification of non-Hermitian topology [16, 17].

In this paper, we consider the localizable class CI, which describes time-reversal invariant, spin SU(2)-symmetric 3D topological superconductors [9]. Employing a bulk three-dimensional lattice model [18], we use kernel-polynomial-method (KPM) numerics [19–21] to demonstrate that surface states away from zero energy can be Anderson localized with weak disorder. We show that localization becomes possible when the surface is coupled to an additional 2D trivial flat band (in such a way that the composite system preserves class-CI symmetries), consistent with the expectations articulated in Ref. [6]. In the latter, it was predicted that coupling trivial degrees of freedom to the surface can open up a spectral gap between bulk and surface states in localizable phases; the detached surface states at nonzero energy are then vulnerable to Anderson localization. We call this

a trivializing proximity effect (TPE) that demonstrates the fragility of surface states in a localizable phase. By contrast, without the TPE we observe robust criticality of the surface states in the presence of disorder, a phenomenon known as “spectrum-wide quantum criticality” [6, 22–25], discussed more below.

B. 2D Dirac surface states and moiré simulators

The focus in this paper is on the two-dimensional (2D) surface states of a bulk 3D class-CI topological superconductor. Typical topological surface states take the form of 2D massless Dirac fermions [1–3]. The latter are by now ubiquitous in both nontopological and topological contexts; examples of the former include monolayer graphene [26–28] and the cuprates [29]. More recently 2D massless Dirac fermions have re-emerged in twisted multilayer (moiré) graphene [30–34].

Although topological boundary Dirac fermions cannot be exactly realized in two-dimensional crystals with local hopping (a consequence of the parity anomaly [35–37]), it is worth emphasizing that they can be *effectively simulated* in such systems when intervalley coupling can be neglected. Hence, a single K valley in graphene (neglecting spin) simulates the surface of a $\mathbb{Z}_2$ topological insulator, and each nodal quasiparticle valley in the d -wave superconducting state of the cuprates is approximately equivalent to the surface of $^3\text{He-B}$ [24]. Bistritzer-MacDonald (BM) models for bilayer and trilayer moiré graphene in the chiral limit simulate class-CI and class-AIII surface states, respectively [21, 38]. Unlike the boundary modes of a topological bulk, none of these simulators are truly robust. For example, simple scalar potential disorder is enough to destroy the analogy between moiré graphene and topological superconductor surface states [44].

In this work we also study the effects of chiral-symmetric (“twist”) disorder [45–53] on the chiral BM model for twisted bilayer graphene (TBLG). This model is effectively topological due to the chiral and particle-hole symmetries needed for class CI, but also because the AB -interlayer tunneling that couples the K valleys between the two layers corresponds to a special realization of this class (discussed in the next subsection). The advent of band gaps at the magic angle (e.g.) already demonstrates the absence of a spectral flow principle for this type of (effective) topological surface theory: in this case, we do not need additional degrees of freedom beyond the moiré potential to detach the “surface” and “bulk” states. Different from our results for the topological lattice model, however, we find that increasing disorder whilst preserving class-CI symmetries *suppresses* localization effects. This hints at a fundamental dichotomy between surface continuum field theories and bulk lattice model descriptions of topological phases, extending beyond the discussion in Ref. [6].

C. Spectrum-wide quantum criticality versus surface-state localizability

The same 10-fold way employed in topological matter also classifies single-particle Hamiltonians with quenched randomness and their Anderson localization physics [54]. This is quite natural, as with problems of either topology or disorder, one stresses robust properties that are independent of model details. A complete classification of 2D Dirac Hamiltonians in the presence of quenched disorder was already known before the discovery of $\mathbb{Z}_2$ topological insulators [55]; here it was noted that classes CI, AIII, and DIII are “doubled,” with special versions of each of these classes admitting only abelian, non-abelian, or gravitational [24] gauge potential disorder. Later it was realized that these special cases are exactly realized at the boundary of 3D topological superconductors [7, 44].

A special feature of the topological superconductor surface-state Dirac models is their exact solvability at zero energy in the presence of disorder [56–61], which establishes the protection of these states from Anderson localization [54]. In the advent of topology, this protection was also argued for on general grounds [10, 11].

The protection of *finite-energy* surface states (i.e., those away from the surface Dirac point) was only investigated recently. In numerical studies that employed effective continuum Dirac models (defined in momentum space to avoid fermion doubling), it was found that finite-energy states also avoid Anderson localization. Instead, finite-energy Dirac surface-states of class CI and AIII topological phases become quantum critical, with universal spatial fluctuations respectively captured by the spin- and integer-quantum Hall plateau transitions [6, 22–25]. This phenomenon was dubbed spectrum-wide quantum criticality (SWQC).

Ref. [6] called into question the ubiquity of SWQC. Focusing on the localizable class AIII, it was demonstrated that a certain symmetry-preserving, homogeneous modification at the surface of a certain bulk topological lattice model (1) allows the surface states to be detached from the bulk (via an indirect or direct energy gap) and (2) enables Anderson localization of all surface states with energy $E \neq 0$. In addition, the perturbing lattice potential is *completely absent* from the effective 2D continuum surface Dirac theory, because it projects to zero in that subspace. In the case of class AIII, the mechanism for surface-state detachment and Anderson localization was shown to be the injection of Berry curvature “from the sky,” i.e. from the surface-bulk spectral interface. The Berry curvature enables Anderson localization of finite-energy surface states, forming a topological Anderson insulator (as in the integer quantum Hall effect or a dirty Chern insulator). At the same time, it is possible to arrange for the surface to possess a certain statistical symmetry so that chiral edge states bounding different random surface domains percolate. This restores surface SWQC, and explains why all finite-energy surface states then exhibit universal integer-quantum-Hall

plateau-transition statistics [6, 23].

The class-AIII 2D continuum Dirac models studied in Ref. [23] correctly capture the fine-tuned SWQC scenario, but fail to encode the localizability of the surface states due to Berry curvature nucleated from the surface-bulk interface. The reason is simple: the *continuum* Dirac spinor contains only half the number of components of the bulk lattice wave functions. In class AIII, the anomalous Berry curvature that enables the destruction of extended surface states manifests only in these neglected components—the microscopic (lattice-scale) surface-state wave functions retain the same orbital content as the bulk.

Our focus here is class CI, which possesses a $T^2 = +1$ time-reversal symmetry [7, 9]. [Class CI describes ordinary time-reversal invariant, spin-SU(2) symmetric s -wave superconductors, which are topologically trivial, in addition to topological variants with nontrivial pairing.] In this case, Berry curvature cannot be the key to disorder-mediated surface-state annihilation. We instead exploit the TPE to demonstrate localizability in a class-CI topological bulk lattice model.

In addition to bulk lattice and 2D continuum BM models, we consider the coupling of the dirty 2D continuum CI Dirac model from Ref. [22] to a continuum flat band. We find that although the coupling to trivial degrees of freedom does allow Anderson localization of some surface states with weak disorder, stronger disorder appears to *heal* the continuum model, restoring SWQC. As with the case of class AIII in Ref. [6], we take this as further evidence that continuum surface field theories (e.g., 2D dirty Dirac equations) are *fundamentally inadequate* for capturing the surface physics of localizable topological phases.

D. Summary sketch of the main results

We preview our main results in Fig. 1. Here we plot a numerically computed parameter θ for several different models, as a function of disorder strength. The parameter θ is (half) the curvature of the so-called *multifractal spectrum* [54], used here to measure the extended or localized nature of 2D surface-state wave functions [see Eqs. (19)–(21), below, for a precise definition]. Larger values of θ indicate a tendency towards Anderson localization. Presented results obtain via fixed realizations of disorder, averaged over large energy windows of surface-state spectra (see Appendix A for details). Two horizontal lines in this figure mark special values of θ . The line denoted “WZW” has $\theta = 1/4$, which is the exact result for the class-CI, quantum-critical zero-energy continuum Dirac surface state in the presence of symmetry-preserving disorder [59, 60]. The line marked “SQHPT” has $\theta = 1/8$, and refers to the spin quantum Hall plateau transition; this is the approximate known result from extensive numerics for the plateau transition in class C [54]. Previous studies of 2D continuum Dirac theories ex-

hibited SWQC, with SQHPT statistics for finite-energy states [22, 25].

Results are exhibited for 5 different models in Fig. 1: (1) the surface states of a bulk lattice model with disorder, (2) the same as (1), but incorporating in addition coupling to a 2D trivial flat band (such that the composite system also resides in class CI), (3) the continuum 2D Dirac model for dirty surface states from Ref. [22], (4) the same as (3), but coupled to a 2D trivial flat band, and (5) the chiral BM model for TBLG, subject to class-CI preserving disorder. Cases (1) and (3) show robust wave function criticality, with the continuum case (3) matching well the SQHPT prediction. Case (2) demonstrates the TPE: finite-energy surface states localize more strongly with increasing disorder when coupled to a trivial 2D

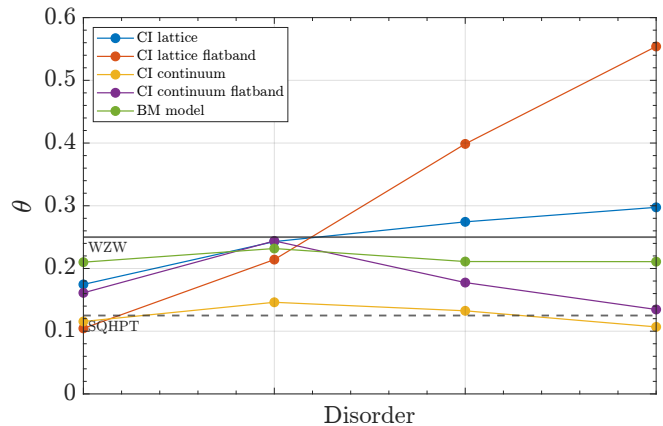


Figure 1. Contrasting the behavior of lattice and continuum models of class-CI topological surface states, subject to disorder and a trivializing proximity effect (TPE). As a proxy for wave function criticality or Anderson localization, we plot the numerically computed curvature parameter θ for the multifractal spectrum (see text), averaged over a certain finite energy window, versus the effective disorder strength. Larger values of θ indicate stronger spatial rarefaction, tending towards Anderson localization. Results are shown for 5 different models: 2D surface states of a 3D class-CI lattice model without and with coupling to a trivial class-CI 2D flat band, the same for states of a 2D continuum Dirac theory of the surface, and the continuum Bistritzer-MacDonald (BM) model for twisted bilayer graphene in the chiral limit, subject to chiral disorder. The BM model is tuned to the first magic angle; the band gaps play a similar role as the coupling to a trivial flat band for the surface states. The lines marked “WZW” and “SQHPT” refer to predictions for zero- and finite-energy *critical* surface states, in the so-called spectrum-wide quantum criticality (SWQC) scenario [6, 22, 25]. The main takeaways are (1) Increasing disorder localizes lattice surface states (away from zero energy) with the TPE, but not significantly without it, and (2) In the 2D continuum theories, increasing disorder has either no effect (states remain critical) or *suppresses* an initial trend towards localization for weak disorder. Disorder strengths are selected relative to each model and can only be qualitatively compared across models; details for these and the selected energy windows are provided in Appendix A.

band of additional states. For cases (4) and (5), weak disorder produces localized states near band edges, as we demonstrate in Sec. III. Somewhat surprisingly, however, cases (4) and (5) show that stronger disorder appears to “heal” the surface, restoring SWQC and *suppressing* the tendency towards localization.

It is worth noting that the lattice-model results effectively describe surface states of an order-of-magnitude smaller surface, compared to the 2D continuum-model results (owing to computational limitations). This is because the surface states occupy a relatively small fraction of the surface Brillouin zone for the bulk lattice model in the slab geometry. Larger relative finite-size corrections are the likely reason why the lattice results without the TPE match better the WZW, rather than SQHPT prediction. Details of the parameters used to produce all data in Fig. 1 are provided in Appendix A.

E. Outline

The rest of this paper is organized as follows. In Sec. II, we introduce the bulk lattice and 2D continuum models employed in this work, and summarize key properties. This includes the effects of hybridizing surface states with trivial 2D bands in the absence of disorder, the incorporation of chiral-symmetric disorder, and comparison to the BM model for TBLG. In Sec. III we present numerical results for the topological robustness of surface states in the presence of disorder and the TPE, using KPM and exact diagonalization for the five models highlighted in Fig. 1. We conclude with a list of key unanswered questions in Sec. IV.

II. MODELS

A. CI bulk lattice model

We consider the 2-orbit-per-site cubic lattice model for a mean-field class-CI topological superconductor introduced in Ref. [18]. The quasiparticle Hamiltonian is

$$H_L^{(3D)} = \int_{\text{BZ}} \frac{d^3\mathbf{k}}{(2\pi)^3} \eta^\dagger(\mathbf{k}) h_L^{(3D)}(\mathbf{k}) \eta(\mathbf{k}),$$

$$\eta(\mathbf{k}) = \left[a_\uparrow(\mathbf{k}), b_\uparrow(\mathbf{k}), a_\downarrow^\dagger(-\mathbf{k}), b_\downarrow^\dagger(-\mathbf{k}) \right]^T, \quad (1)$$

where $\eta(\mathbf{k})$ denotes the Nambu spinor, $\mathbf{k} = \{k_x, k_y, k_z\}$ runs over the 3D cubic Brillouin zone (BZ), and the Bogoliubov-de Gennes Hamiltonian is

$$h_L^{(3D)}(\mathbf{k}) = (-\cos k_x - \cos k_y - \cos k_z - \mu) \tau_3$$

$$+ \Delta_1 (\cos k_x - \cos k_y) \tau_1 \sigma_3$$

$$+ \Delta_2 \sin k_x \sin k_y \tau_1 \sigma_1 - \Delta_2 \sin k_z \tau_1 \sigma_2. \quad (2)$$

Here $\{\tau_i\}$ and $\{\sigma_j\}$ denote Pauli matrices acting respectively on particle-hole and orbital spaces, μ is the chemical potential, Δ_1 encodes the strength of d -wave pairing

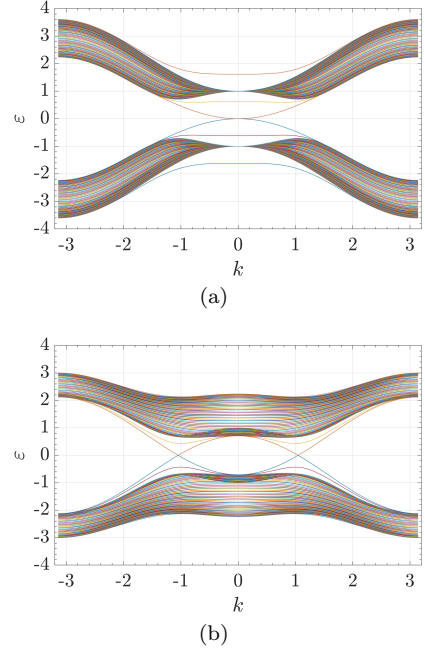


Figure 2. Slab spectrum of the cubic-lattice class-CI topological superconductor model defined by the lattice Hamiltonian in Eq. (2) with winding number $|\nu| = 2$. (a) z -cut, $k_y = 0$ line. The surface states form a quadratic touching, analogous to a single valley of Bernal bilayer graphene. (b) x -cut, $k_z = 0$ line. The surface states appear as two Dirac cones shifted away from the surface Γ point. The parameters for these plots are: $\mu = -2$, $\Delta_1 = 1$, and $\Delta_2 = 1$. In addition, a surface Chern mass $m_c = 1$ is used to gap out one of the two slab surfaces in each plot, so as to energetically isolate the low-energy surface states on the “active” surface. The slab thickness is $N_{z/x} = 60$.

in the $k_{x,y}$ plane, and Δ_2 is an interorbital pairing. We have set the lattice spacing to one and the normal-state hopping amplitude $t = -1$; then this model is topological with winding number $\nu = 2$ when $-3 < \mu < -1$.

1. Surface states

Topological superconductors with $|\nu| = 2$ typically possess two surface Dirac cones (with the Dirac points appearing at *exactly* zero energy, defined relative to the Fermi energy in the paired bulk). For the lattice model in Eq. (2), this is the case for an x -cut surface (parallel to the yz plane). By contrast, for the z -cut surface the model exhibits 2D surface states with a quadratic touching, analogous to a single valley of Bernal-stacked bilayer graphene [26]. The quadratic touching occurs at the Γ point of the surface BZ, and has the low-energy continuum expansion

$$h_z^{(2D)}(k_x, k_y) = \frac{1}{M} \begin{bmatrix} 0 & k^2 \\ \frac{1}{k^2} & 0 \end{bmatrix}, \quad (3)$$

where $k = k_x - i k_y$ and k is measured from the Γ point. The spectra of the z -cut and x -cut surfaces are shown in Fig. 2.

2. Surface states with flat-band coupling

We will demonstrate that the surface states exhibited in Fig. 2(a) become fragile (vulnerable to Anderson localization with *weak* disorder) when these are coupled to an additional 2D flat-band lattice model. Taking the topological bulk to have $z \geq 1$, we can couple it to a flat-band layer at $z = 0$ by adding the Hamiltonian

$$\delta H_{\text{TPE}}^{(3D)} = \frac{\varepsilon_c}{2} \int_{\text{SBZ}} \frac{d^2 \mathbf{k}}{(2\pi)^2} \eta_{z=0}^\dagger(k_x, k_y) \tau_3 \eta_{z=0}(k_x, k_y) + \gamma \left[\sum_{\sigma} \int_{\text{SBZ}} \frac{d^2 \mathbf{k}}{(2\pi)^2} \eta_{z=0}^\dagger(k_x, k_y) \tau_1 \eta_{z=1, \sigma}(k_x, k_y) \right] + \text{H.c.}, \quad (4)$$

Here $\eta_{z=0}$ is a two-component Nambu spinor in particle-hole space of the flat band, ε_c sets the energy of the flat band quasiparticles (at $\pm \varepsilon_c$), and γ denotes the (purely anomalous) coupling strength between the 2D flat band

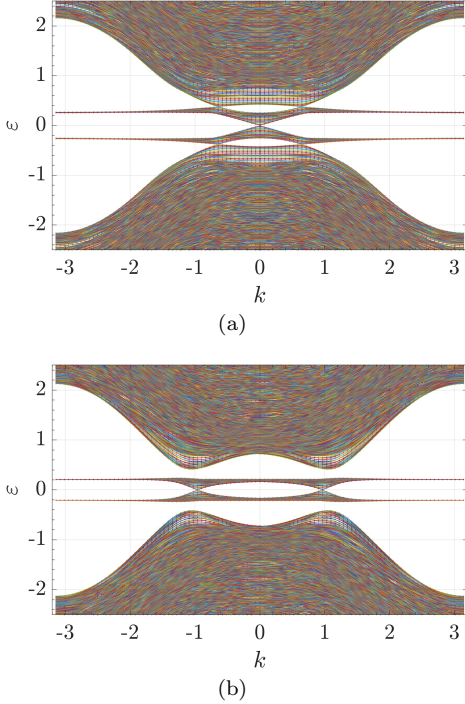


Figure 3. Slab energy spectrum versus k_y of the topological CI lattice model (as in Fig. 2), now coupled to a 2D flat band via Eq. (4). (a) z -cut. (b) x -cut. The model parameters are $\varepsilon_c = 0.2$, $\gamma = 0.35$, $\mu = -2$, $\Delta_1 = 1$, and $\Delta_2 = 1$. Again a Chern mass of size $m_c = 1$ is employed to gap out one of the two slab surfaces. (a) has $N_x = 100$, $N_z = 10$ and (b) is the same with $x \Leftrightarrow z$.

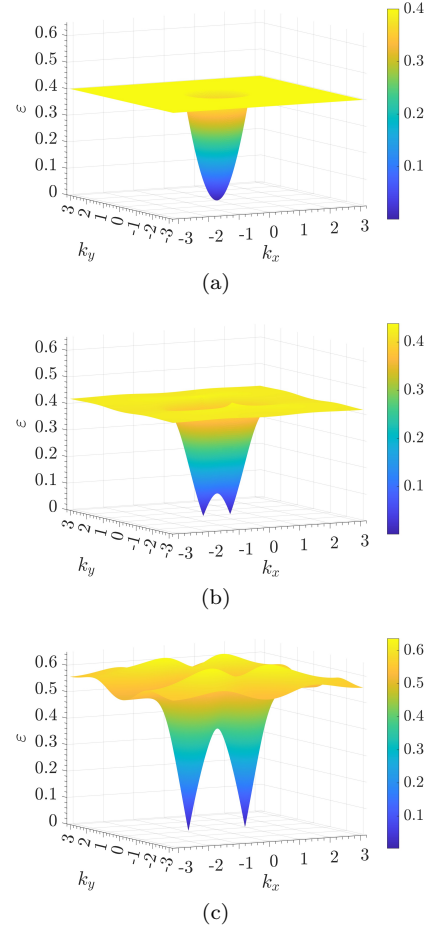


Figure 4. Splitting of the quadratic surface-state touching by proximity coupling. We plot the lowest-energy band for the z -cut bulk CI topological lattice model hybridized with the trivial 2D flat band as in Fig. 3(a), for various hybridization strengths. (a) $\gamma = 0$. (b) $\gamma = 0.2$. (c) $\gamma = 0.6$. Other parameters are $\varepsilon_c = 0.4$, $\mu = -2$, $\Delta_1 = 1$, $\Delta_2 = 1$, $N_z = 20$.

and the top layer of the topological superconductor bulk crystal. The integrals run over the entire surface Brillouin zone (SBZ).

The combined bulk and flat-band system still resides within class CI. It has chiral (physical time-reversal [7]) symmetry

$$\eta_{xyz} \rightarrow M_S (\eta_{xyz}^\dagger)^T, \quad i \rightarrow -i, \quad M_S = \tau_2, \quad (5)$$

and $P^2 = -1$ particle-hole (physical spin π -rotation around $\hat{x}$ [7]) symmetry

$$\eta_{xyz} \rightarrow M_P (\eta_{xyz}^\dagger)^T, \quad M_P = \tau_2. \quad (6)$$

Coupling of the z -surface of the topological bulk to the 2D flat band is shown in Fig. 3(a), whilst an analogous coupling to the x -cut surface is shown in panel (b) of this figure. In the absence of disorder, coupling with the flat band opens up a direct gap between the bulk and surface spectra for coupling to the x surface. For coupling to

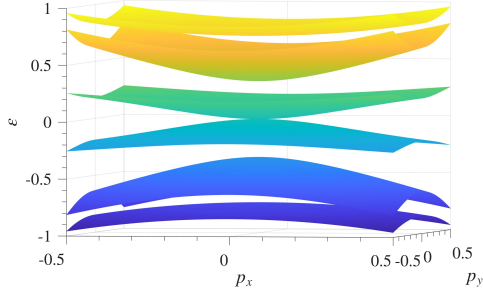


Figure 5. Spectrum of the CI 2D continuum surface-Dirac model hybridized with a class-CI flat band, in the absence of disorder. The parameters are $\varepsilon_c = 0.4$ and $\gamma_0 = 0.35$.

the z -cut surface, there is no gap introduced by the flat band. Nevertheless, we will find that this coupling is sufficient to make the surface states fragile to the effects of disorder.

Similar to the regime of *very small* (rigid) twists in TBLG [32], coupling of the z -cut surface to the flat band splits the quadratic touching of the pristine surface states into two Dirac cones, as shown in Fig. 4. Here we plot the spectrum for the lowest-energy slab band corresponding to the hybridization of the z -cut surface of the topological

bulk with the 2D trivial flat band.

B. CI continuum Dirac surface model

Next we consider the 2D continuum Dirac model for class-CI surface states used to demonstrate SWQC in Ref. [22],

$$h_x^{(2D)} = -i \boldsymbol{\sigma} \cdot \boldsymbol{\nabla} + \mathbf{A}^j(\mathbf{r}) \cdot \boldsymbol{\sigma} \tau_j, \quad (7)$$

where $\boldsymbol{\nabla}$ is the 2D gradient operator, $\{\sigma_i\}$ and $\{\tau_j\}$ are Pauli matrices acting on the spinor and “color” spaces, and the repeated index j is summed over $\{1, 2, 3\}$. Eq. (7) is the effective 2-cone surface Dirac Hamiltonian for the x -cut topological bulk lattice model, Fig. 2(b), after setting the low-energy Fermi velocity of the cones to one. We have in addition incorporated generic class-CI surface disorder, which takes the form of the SU(2) non-abelian vector potential $A_\mu^i(\mathbf{r})$ [7, 44], and which connects the two Dirac colors in exactly the same way as the moiré matrix potential couples together layers in the chiral-limit of the BM model for TBLG [30, 31, 34].

We can further couple this continuum surface theory to a continuum flat band. As in Ref. [22], we formulate the full model in momentum space in order to avoid fermion doubling,

$$H_{\text{TPE}}^{(2D)} = \sum_{\mathbf{p}_1, \mathbf{p}_2} \left[\psi_{\mathbf{p}_1}^\dagger [\mathbf{p}_1 \cdot \boldsymbol{\alpha} \delta_{\mathbf{p}_1, \mathbf{p}_2} + \mathbf{A}^j(\mathbf{p}_1 - \mathbf{p}_2) \cdot \boldsymbol{\alpha} T_j] \psi_{\mathbf{p}_2} + \phi_{\mathbf{p}_1}^\dagger [\varepsilon_c \delta_{\mathbf{p}_1, \mathbf{p}_2} + A_c(\mathbf{p}_1 - \mathbf{p}_2)] \tau_1 \phi_{\mathbf{p}_2} + \sum_{\sigma=1}^2 \{ \phi_{\mathbf{p}_1}^\dagger [\gamma_0 \delta_{\mathbf{p}_1, \mathbf{p}_2} + \gamma_1(\mathbf{p}_1 - \mathbf{p}_2)] \tau_1 \psi_{\sigma, \mathbf{p}_2} + \phi_{\mathbf{p}_1}^\dagger [\gamma_0 \delta_{\mathbf{p}_1, \mathbf{p}_2} + \gamma_2(\mathbf{p}_1 - \mathbf{p}_2)] \tau_2 \psi_{\sigma, \mathbf{p}_2} + \text{H.c.} \} \right], \quad (8)$$

where after a basis change the Dirac kinetic matrices and SU(2) color generators are

$$\{\boldsymbol{\alpha}\} = \{\tau_2 \sigma_2, -\tau_1 \sigma_2\}, \quad \{T_j\} = \{-\tau_3 \sigma_3, \sigma_2, \tau_3 \sigma_1\}. \quad (9)$$

In Eq. (8), $\psi \rightarrow \psi_{\tau, \sigma}$ denotes the 4-component Dirac surface state, while $\phi \rightarrow \phi_\tau$ is a 2-component trivial flat-band Nambu spinor. The flat band has energy ε_c with a random, zero-average component $A_c(\mathbf{r})$. The (purely anomalous) couplings between the Dirac surface and flat-band states are given by the average parameter and random, zero-mean functions γ_0 and $\gamma_{1,2}(\mathbf{r})$, respectively.

The composite system in Eq. (8) is invariant under chiral and $P^2 = -1$ particle-hole transformations required by class CI. These are encoded in the symmetry transformations

$$\psi \rightarrow M_S (\psi^\dagger)^\top, \quad \phi \rightarrow M_S (\phi^\dagger)^\top, \quad i \rightarrow -i, \quad M_S = \tau_3, \quad (10)$$

and

$$\psi \rightarrow M_P (\psi^\dagger)^\top, \quad \phi \rightarrow M_P (\phi^\dagger)^\top, \quad M_P = \tau_2. \quad (11)$$

The particle-hole condition forces γ_0 and $\gamma_{1,2}(\mathbf{r})$ to be real-valued.

In the clean limit, hybridizing the two surface Dirac cones to the flat band opens up (indirect) spectral gaps, similar to the x -cut lattice model, see Fig. 3(b) for the hybridized spectrum of the latter. The hybridized bands of the 2D continuum model are shown in Fig. 5.

1. Disorder

Disorder is incorporated into the continuum model using the random-phase parameterization in momentum space [22, 62],

$$A_{x,y}^i(\mathbf{q}) = \frac{\sqrt{\lambda_a}}{L} \exp \left[i \Theta_{x,y}^i(\mathbf{q}) - \frac{q^2 \xi^2}{4} \right], \quad (12a)$$

$$A_c(\mathbf{q}) = \frac{\sqrt{\lambda_c}}{L} \exp \left[i \Theta_c(\mathbf{q}) - \frac{q^2 \xi^2}{4} \right], \quad (12b)$$

$$\gamma_{1,2}(\mathbf{q}) = \frac{\sqrt{\lambda_g}}{L} \exp \left[i \Theta_{1,2}(\mathbf{q}) - \frac{q^2 \xi^2}{4} \right], \quad (12c)$$

where $\mathbf{q}$ denotes a 2D momentum vector and where each phase angle satisfies $\Theta(\mathbf{q}) = -\Theta(-\mathbf{q})$, but these are otherwise independent, uniformly distributed random variables over $[0, 2\pi)$. The disorder strengths are given by the variances

- λ_a : disorder that scatters between the two colors of the surface theory,
- λ_c : disorder in the flat-band energy,
- λ_g : disorder in the coupling between the surface theory and the 2D flat band.

In Eq. (12), L denotes the linear system size and ξ is a correlation length of order the inverse momentum cutoff [62].

C. Chiral Bistritzer-MacDonald model for TBLG

Finally, we also consider the effects of chiral (“twist” [45–53]) disorder on the chiral limit of the BM model for TBLG. The clean continuum Hamiltonian is [30, 31, 34]

$$h_{\text{BM}}^{(2\text{D})} = -i \boldsymbol{\sigma} \cdot \boldsymbol{\nabla} + \alpha [U_1^*(-\mathbf{r}) \sigma^+ \tau^+ + U_1(\mathbf{r}) \sigma^- \tau^+ + \text{H.c.}], \quad (13)$$

where $U_1(\mathbf{r})$ is the periodic moiré potential

$$U_1(\mathbf{r}) = e^{-i\mathbf{q}_1 \cdot \mathbf{r}} + e^{i\frac{2\pi}{3}} e^{-i\mathbf{q}_2 \cdot \mathbf{r}} + e^{-i\frac{2\pi}{3}} e^{-i\mathbf{q}_3 \cdot \mathbf{r}},$$

$$\mathbf{q}_1 = (0, -1), \quad \mathbf{q}_2 = \left(\frac{\sqrt{3}}{2}, \frac{1}{2} \right), \quad \mathbf{q}_3 = \left(-\frac{\sqrt{3}}{2}, \frac{1}{2} \right). \quad (14)$$

Here $\{\mathbf{q}_{1,2,3}\}$ are the moiré reciprocal lattice vectors (for the symmorphic moiré BZ) and α is the effective dimensionless interlayer tunneling, which gets larger for *smaller* twist angles $\theta \ll \pi/2$ ($\alpha \sim 1/\theta$ [31]). The chiral BM model takes exactly the same form as the continuum 2-color topological surface theory in Eq. (7), with the periodic intercolor coupling involving vector potential components $A_{x,y}^{1,2}$; the components $A_{x,y}^3 = 0$. Therefore the chiral BM model is also in class CI, satisfying

$$-M_S h_{\text{BM}}^{(2\text{D})} M_S = -M_P (h_{\text{BM}}^{(2\text{D})})^\top M_P = h_{\text{BM}}^{(2\text{D})}, \quad (15)$$

with $M_S = \sigma_3$ and $M_P = \sigma_1 \tau_2$.

Perfect flat bands occur at zero energy at the first magic angle $\alpha \simeq 0.586$ [34]. We incorporate uncorrelated disorder into each of the components of the periodic potential,

$$\delta h_{\text{BM}}^{(2\text{D})} = A_1(\mathbf{r}) \sigma_1 \tau_1 + A_2(\mathbf{r}) \sigma_2 \tau_1 + A_3(\mathbf{r}) \sigma_1 \tau_2 + A_4(\mathbf{r}) \sigma_2 \tau_2. \quad (16)$$

These are the same as the $A_{x,y}^{1,2}$ disorder components in the continuum surface model Eq. (12a), and the disorder

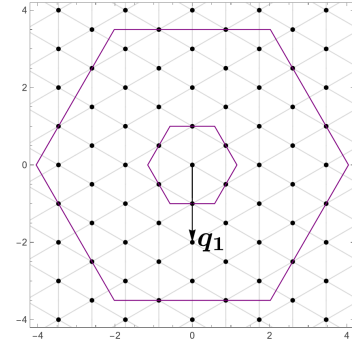


Figure 6. Schematic diagram of the momentum lattice used for the BM model calculation. The inner hexagon is the Brillouin zone of the moiré potential and the outer hexagon is the boundary of the momentum lattice.

is implemented numerically in the same way. Eq. (16) can be regarded as white noise “twist disorder,” because it involves the same interlayer tunneling matrices as the BM model itself. We neglect the additional effects of lattice relaxation, which would enter as an additional abelian vector potential [26, 47] (and change the symmetry class to AIII).

In our subsequent numerical calculations for the BM model, we employ a triangular momentum-space lattice, which is shown in Fig. 6. The boundary hexagon of the lattice has edge length $\frac{1}{\sqrt{3}} (\frac{5}{2} + M_1)$. The lattice constant is $1/(2M_2)$. The system size is L^2 with $L = 5M_2 + 2M_1M_2$.

III. FRAGILITY TO LOCALIZATION

A. Multifractal measures

To characterize the degree of Anderson localization for surface-state wave functions, we employ multifractal analysis [54]. By summing moments of the position-space probability for eigenstates over the surface, we obtain scaling behavior with respect to the surface linear system size L . Anderson localized states with localization length smaller than the system size give moments that do not scale with L , while extended states scale as a simple power of $1/L$. Critical wave functions (such as the topologically protected zero-energy states [7, 10, 11, 59, 60] as well as finite-energy surface states in the SWQC scenario [6, 21–25]) exhibit nontrivial power-law scaling with $1/L$, encoded in the so-called multifractal spectrum of scaling exponents. We introduce these in the context of surface states obtained for the bulk 3D lattice model.

For the lattice model with the z -cut [Fig. 2(a)], we define the probability at site position (x, y) on the surface for a surface eigenstate wave function $\psi_\varepsilon(x, y, z)$ by summing the probability through a depth of n_1 layers (and

renormalizing),

$$p(x, y; \varepsilon) \equiv \frac{\sum_{z=1}^{n_l} |\psi_\varepsilon(x, y, z)|^2}{\sum_{z'=1}^{n_l} \sum_{x', y'} |\psi_\varepsilon(x', y', z')|^2}. \quad (17)$$

We then compute the box probability

$$\mu_b(x, y; \varepsilon) = \sum_{(x', y') \in b(x, y)} p(x', y'; \varepsilon), \quad (18)$$

where $b(x, y)$ denotes a square of linear size b centered at (x, y) . This coarse-graining step allows us to extract multifractal dimensions from scaling with box sizes in a fixed system size L . These are determined from the inverse participation ratio

$$P_b(q; \varepsilon) = \sum_{x, y} [\mu_b(x, y)]^q \propto \left(\frac{b}{L}\right)^{\tau(q)}, \quad (19)$$

where $\tau(q)$ is the multifractal exponent. Unlike $P_b(q; \varepsilon)$, for weak multifractality the exponent $\tau(q)$ is self-averaging in the infinite- L limit. At a certain q , we can fit $\ln P_b(q; \varepsilon)$ averaged over N_s energy-adjacent eigenstates with $\ln(b/L)$ under several different box sizes $\{b\}$ to extract the anomalous part of the spectrum,

$$\Delta(q) = \tau(q) - 2(q - 1). \quad (20)$$

For weak multifractality and small q , this can be approximated by [62]

$$\Delta(q) = -\theta(\varepsilon) q(q - 1). \quad (21)$$

From the fitted $\tau(q)$, we can extract $\theta(\varepsilon)$ near energy ε [and determine the validity of the parabolic approximation to the spectrum $\tau(q)$]. When the states are extended and ergodic, $\theta = 0$. When the states are critical, $\theta > 0$ is finite and (approximately) q independent. Anderson localized states have $\tau(q > 0) = 0$, which formally corresponds to $\theta = \theta_q \rightarrow 2/q$.

B. Surface-state scenarios

As discussed in the Introduction, localizable topological phases such as class CI in 3D [6] have fragile surface states that can be destroyed (e.g., Anderson localized by *weak* disorder). The only exceptions are the protected zero-energy surface states, but these form a set of measure zero in the surface-state spectrum. The localization of all but the zero-energy surface states has important implications for experimental observables; for example, such an almost-localized surface cannot produce a finite surface current at finite temperature (in the absence of interaction-mediated inelastic scattering) [63].

In fact, previous numerical studies of class-CI continuum Dirac surface states instead demonstrated robust spectrum-wide quantum criticality (SWQC) [22, 25], reviewed in Secs. IC and ID, see also Fig. 1. In particular,

it was observed that finite-energy surface states of the Dirac model in Eq. (7) exhibit *universal* fractality, with $\theta \simeq 1/8$, consistent with the known value of the class-C *spin* quantum Hall plateau transition (SQHPT). This is not to be confused with the quantum spin Hall effect and 2D $\mathbb{Z}_2$ topological insulators [1]; the spin quantum Hall effect is an analog of the ordinary integer effect predicted to occur in a topological 2D $d+id$ superconductor [3, 54]. The plateau transition then involves bulk quasiparticle states that carry spin, instead of charge (the latter being ill-defined for quasiparticles in a superconducting state). The observation of class-C states is in and of itself a persistent mystery, at odds with standard arguments that finite-energy states of 10-fold way Hamiltonians always reside in one of the three Wigner-Dyson classes A, AI, or AII [54]. A symmetry argument in favor of class C for finite-energy *Andreev bound* surface states of a class-CI topological phase, based upon replica-sigma-model target manifolds, was advanced in Ref. [22]. We will not dwell on this point here, but return to it as an open question in the conclusion Sec. IV.

Continuum Dirac surface states of the localizable class AIII were likewise found to exhibit ordinary integer quantum Hall criticality [23]. In Ref. [6], it was shown however that this requires statistical fine-tuning in a microscopic bulk lattice model of a class AIII topological phase. In particular, a certain “fragmenting potential” was identified in Ref. [6] with unexpected properties. The fragmenting potential can be turned on as a uniform perturbation at the surface of the lattice model that preserves the AIII symmetry, and induces the following effects: (1) it can open a gap between the surface and bulk spectra, (2) it pushes Berry curvature into the lattice-model surface states, (3) it enables Anderson localization of almost all surface states (by chiral symmetric disorder). And yet, this mechanism is completely absent in the continuum Dirac models for AIII surfaces, because the Berry curvature induced by the fragmenting potential goes into the lattice “microstructure” of the surface states, neglected in continuum treatments [6].

We focus here on an alternative way to destroy surface states proposed in Ref. [6], which is coupling the surface to trivial degrees of freedom in a way that preserves the defining symmetry of the class (trivializing proximity effect, TPE). This mechanism can be employed in both bulk lattice and 2D continuum Dirac models, and enables the opening of a gap in the surface-state spectrum of both. Our goal is to determine under what circumstances (if any) the SWQC scenario persists in the presence of TPE.

C. CI cubic lattice model, TPE

We examine the surface states of the lattice model in Eq. (2), without and with coupling to the trivial flat band encoded in Eq. (4). We calculate the surface states of an $N_x \times N_y \times N_z$ thin slab, with $N_x = N_y \gg N_z$. In order

to focus on the low-energy surface states of the top z -cut interface, we gap out the bottom one with a surface Chern mass $m_c \tau_2$. This breaks the CI symmetry down to C on the bottom layer, but has negligible effect on the bulk or top surface for the thicknesses N_z employed here.

Since we study the z -cut surface, the clean system without the flat-band coupling has the quadratically dispersing surface states (ala Bernal-stacked bilayer graphene) shown in Fig. 2(a). Unlike class AIII, we were unable to identify a uniform surface modification that does not involve additional degrees of freedom, preserves the class-CI symmetry, and which opens up a gap between surface and bulk spectra. In other words there seems to be no homogeneous analog of the fragmenting potential employed in Ref. [6]. An analogous term might

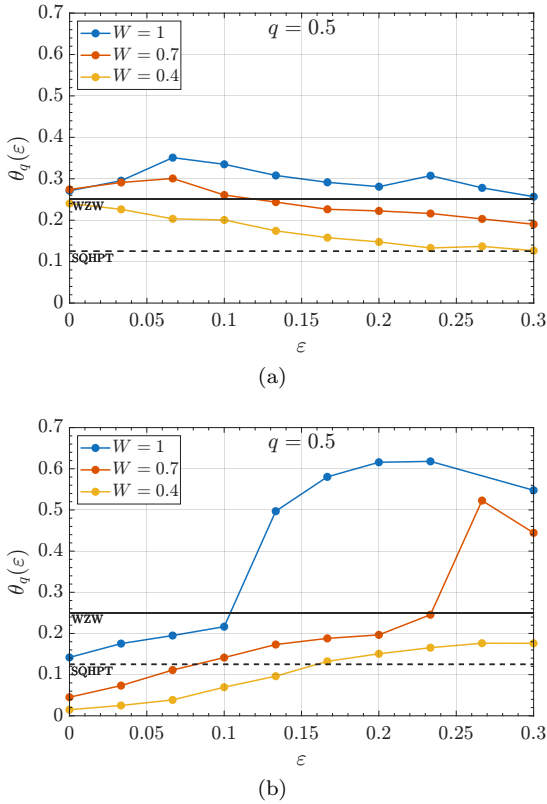


Figure 7. Effects of disorder on the surface states of the CI lattice model [Eqs. (2) and (4)] and the topological proximity effect (TPE). We plot $\theta_q(\epsilon)$ [Eq. (21)] without and with coupling to the flat band for varying disorder strengths. (a) Surface states without coupling to the flat band. The weak effect of increasing the disorder strength on $\theta_q(\epsilon)$ as well as its approximate energy independence are qualitatively consistent with the SWQC scenario [22, 25]. (b) Demonstration of the TPE: lattice surface states hybridized with the trivial flat band. The trend of increasing $\theta_q(\epsilon)$ with increasing disorder for the finite-energy states indicates a drift towards Anderson localization for these states. The overall disorder strength is W (see text) for each curve. Other parameters are $N_y = 72$, $N_z = 10$, $\mu = -2$, $\Delta_1 = 1$, $\Delta_2 = 1$, $m_c = 1$, and for panel (b) $\epsilon_c = 0.4$ and $\gamma = 0.5$.

exist for the x -cut surfaces, but likely requires spatial modulation at the wave vector connecting the two surface Dirac cones shown in Fig. 2(b); again we have not found a homogeneous potential perturbation that can “detach” the surface states.

We calculate the multifractal curvature parameter θ [Eqs. (19)–(21)] for surface states subject to disorder, with and without the flat band. We present results for varying disorder strengths and systems sizes. Disorder with uniform distribution $[-W, W]$ is added to each of the parameters μ , Δ_1 , Δ_2 , ϵ_c , and γ in Eq. (2) and Eq. (4). The ratio between the disorder strengths is kept constant while the overall disorder magnitude is varied. Results are exhibited in Figs. 7 and 8. Here the depth used to define the surface states is $n_l = 1$ without the flat band, and $n_l = 3$ with it (including the flat band layer) [Eq. (17)]. Each $\theta_q(\epsilon)$ at an energy point ϵ is an average from 30 states immediately around that energy. The wave functions are computed using kernel polynomials and the filter and shake method [19–21].

Figs. 7(a) and 8(a) demonstrate that the surface states of the lattice model in the absence of the coupling to the flat band are stable with increasing disorder or system size, with small nonzero values of θ indicating wave func-

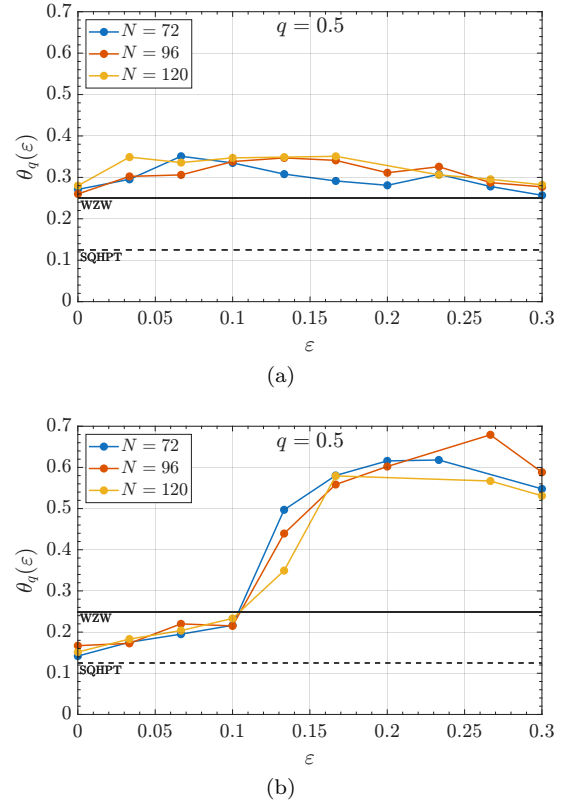


Figure 8. The same as Fig. 7, but for fixed disorder and three different system sizes. (a) Surface states in the absence of the flat band. (b) Surface states hybridized with the flat band. The parameters are $W = 1$, $N_z = 10$, $\mu = -2$, $\Delta_1 = 1$, $\Delta_2 = 1$, $m_c = 1$, and for panel (b) $\epsilon_c = 0.4$, $\gamma = 0.5$.

tion criticality. This is qualitatively consistent with the SWQC scenario, although the obtained $\theta(\varepsilon)$ are generally closer to the zero-energy analytical result $\theta = 1/4$, instead of the class-C SQHPT value $\theta \sim 1/8$. Calculations for the 2D continuum Dirac model shown in Fig. 10 instead converge to the class-C value, consistent with previous work [22]. Finite-size effects are a possible explanation for this lattice versus continuum quantitative discrepancy (without the flat band), given that the continuum calculations effectively simulate a much larger system for the same numerical surface grid size.

By contrast, Figs. 7(b) and 8(b) demonstrate the TPE: the coupling to the flat band enables disorder to push finite-energy surface eigenstates towards Anderson localization. Fig. 7(b) shows the effect increases strongly with increasing disorder. This is one of the main results of this paper, and is consistent with the arguments advanced in Ref. [6]. Note that the lowest-energy surface states remain critically delocalized, as expected [7, 10, 11]. Fig. 8(b) on the other hand indicates only minimal changes to the observed TPE with increasing system size.

For class AIII, it was shown in Ref. [6] that coupling the surface states to a trivial flat band pushes Berry curvature into those states, in a similar fashion as the fragmenting potential. Surface domains with different signs of the fragmenting potential can induce 1D chiral edge modes “in the sky,” i.e. in the induced gap between bulk and surface states. By contrast, we have not found the analog of a fragmenting potential, and chiral edge states cannot span the gap with the class-CI $T^2 = +1$ time-reversal symmetry intact. Despite the tacit identification of SWQC wave functions in the continuum model of class CI surface states [22] with the spin quantum Hall plateau transition, there does not appear to be a mechanism for inducing chiral edge modes on the CI surface without changing the symmetry class. Thus the exact mechanism for the TPE in this case must be different from class AIII.

D. CI continuum model, no TPE

Next we examine the 2D continuum Dirac theory defined by Eq. (7), previously studied in Ref. [22]. We use exact diagonalization to solve for the eigenstates in momentum space (avoiding fermion doubling of the continuum Dirac equation). We compare results for the surface without and with the flat band, Eq. (8). Disorder is incorporated via Eq. (12).

Results without the flat band are exhibited in Figs. 9 and 10. Fig. 9(a) compares the level statistics [64] to the Gaussian orthogonal (GOE) and Gaussian unitary (GUE) ensembles; the results are consistent with the former. This is expected from standard arguments, whereby finite-energy states of class-CI Hamiltonians reside in the orthogonal class AI [22, 54]. The global density of states is shown in Fig. 9(b), and exhibits the low-energy cusp

expected for class CI, see Ref. [22] for a quantitative comparison to the analytical prediction [56]. Despite the GOE statistics, Fig. 10 demonstrates that *most* surface states converge to $\theta \sim 1/8$ with increasing disorder (except those far above the energy cutoff, in the high-energy Lifshitz tails), consistent with class-C spin quantum Hall plateau transition (SQHPT). This is the SWQC scenario [22, 25]. The orthogonal class is instead expected to Anderson localize in 2D.

Fig. 11 shows the effects of disorder on the hybridized continuum 2D surface-state-flat-band system. Weak disorder preserves the spectral gaps opened by hybridization, and localizes states close to the gap edges. This suggests the onset of the TPE. Different from the lattice results shown in Figs. 7(b) and 8(b), however, increasing disorder for the continuum system appears to *heal* the surface. I.e., spectral gaps are washed out, and the multifractal spectrum re-acquires the SWQC demonstrated in Fig. 10 without the flat band hybridization.

The key takeaway is that the *continuum and lattice models behave oppositely with increasing disorder*. The lattice results agree with the picture articulated in Ref. [6], which correlated surface-state fragility with the

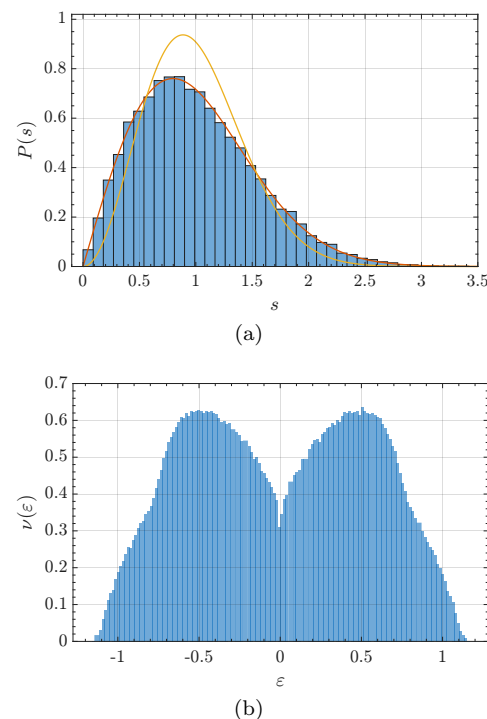


Figure 9. Benchmark results for the CI continuum (2D) Dirac model with disorder [Eq. (7)], without coupling to the flat band. (a) Level statistics compared to the GOE (red) and GUE (orange) predictions. The results are consistent with the GOE. (b) Global density of states $\nu(\varepsilon)$. The cusp near zero energy is expected [56], and was previously numerically analyzed and found to agree well with the analytical prediction [22]. The linear system size is $L = 97$, while the disorder strength and correlation length [Eq. (12a)] are set to $\lambda_a = 2\pi$ and $\xi = \pi$, respectively

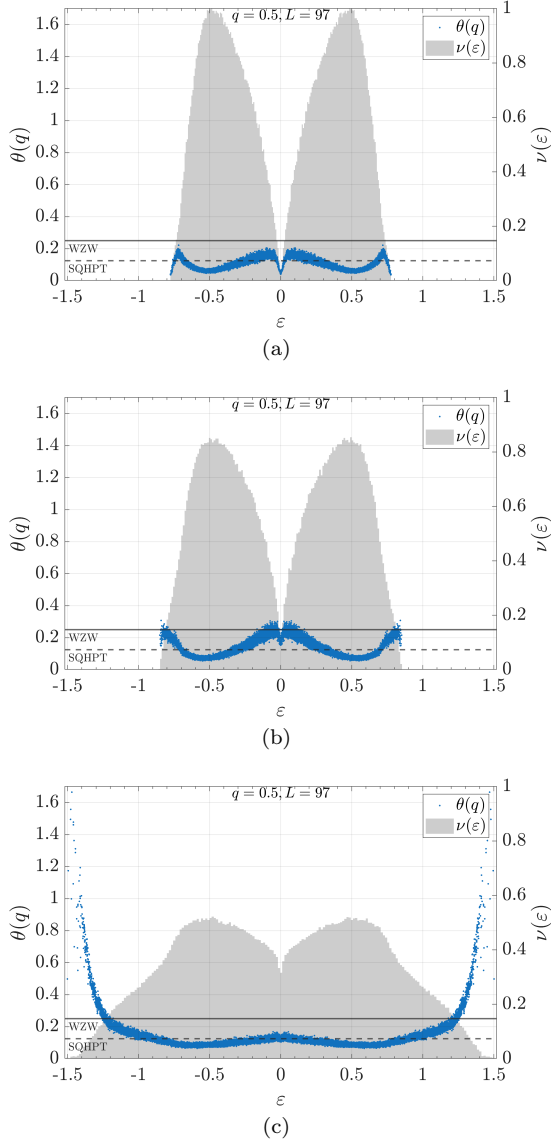


Figure 10. Robust surface states of the CI continuum Dirac model [Eq. (7)] without the flat band, demonstrating SWQC [22, 25]. We plot the multifractal curvature $\theta_q(\varepsilon)$ [Eq. (21)] and the corresponding global density of states $\nu(\varepsilon)$ versus eigenstate energy. Results are shown for three different disorder strengths [Eq. (12a)] in (a–c): $\lambda_a = 0.5, 1, 2\pi$. Increasing the disorder strength creates more states with apparent class-C SQHPT critical statistics, as found previously in Ref. [22]. The other parameters are $L = 97$, $\xi = \pi$, and the boxes used for multifractal analysis are $b = (4, 6, 8)$.

hybridization to trivial degrees of freedom; the combined system is expected to exhibit the TPE. By contrast, the continuum results do *not* show a robust TPE. Unlike the lattice-scale (“fragmenting potential”) device used in Ref. [6] to localize class-AIII surface states, it was argued in this reference that the TPE would destabilize *both* bulk lattice and 2D continuum-model surface states. The surprising finding that only the lattice model shows the expected behavior is a second main result of this paper.

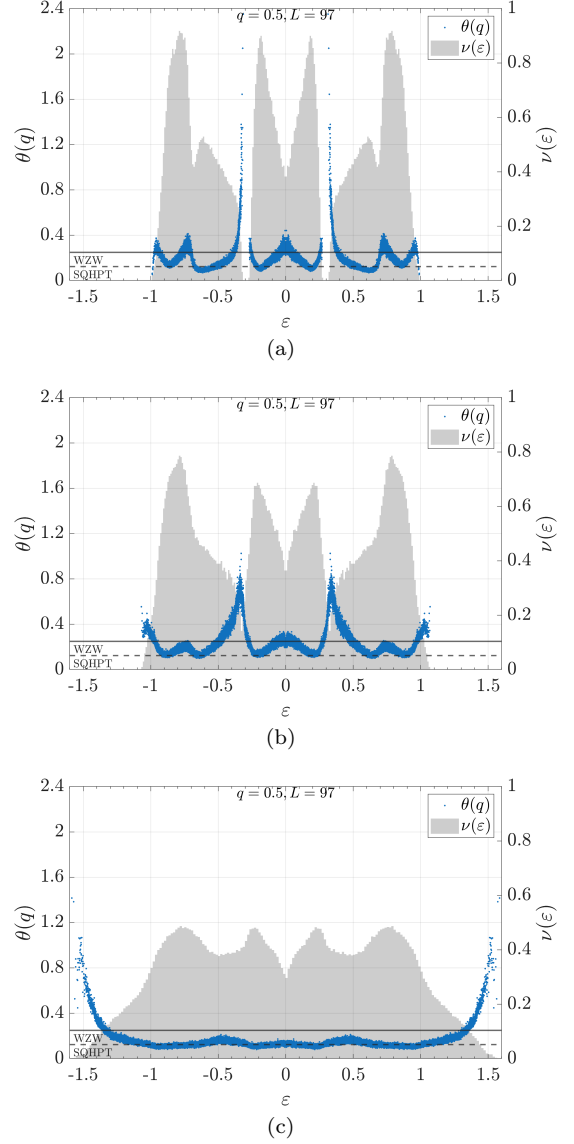


Figure 11. Testing whether the continuum model exhibits the same TPE found for lattice surface states, see Figs. 7 and 8. The class-CI 2D Dirac model is hybridized with the continuum flat band, Eq. (8). The coupling to the flat band opens up gaps in the clean surface spectrum, as shown in Fig. 5. Here we incorporate disorder and examine the effects upon wave function fractality and the spectral gaps. For weak disorder $\lambda_a = 0.2$ (a), the spectral gaps persist and states near the gap edges are strongly localized with $\theta_q(\varepsilon) > 1$. With increasing disorder $\lambda_a = 1$ (b) and $\lambda_a = 2\pi$ (c), however, the spectral gaps wash out and SWQC appears to be restored—compare to Fig. 10(c). These results indicate that the TPE is *not* robust in the continuum model. The other parameters are $\varepsilon_c = 0.4$, $\lambda_c = 0.2$, $\gamma_0 = 0.35$, $\lambda_g = 0.2$, $L = 97$, $\xi = \pi$, $b = (4, 6, 8)$.

Examining Fig. 11, one could argue that the results are not unexpected. The clean flat band, surface-state hybridized system exhibits energy gaps that disappear when the disorder becomes sufficiently strong so as to

“smear out” the position of the flat band. Nevertheless, Fig. 11(c) shows that the outcome for strong disorder is essentially identical to the isolated surface [Fig. 10(c)].

At this point it is worth reiterating that the results for the isolated surface [22] are anomalous and *not* expected. According to the standard logic, all finite-energy states should reside in the orthogonal class AI. For disorder not too strong and in finite volume, 2D wave functions will appear critical with an approximately parabolic multifractal spectrum. The parameter $\theta(\varepsilon)$ is determined by dimensionless conductance $G(\varepsilon)$ of the states [65],

$$\theta(\varepsilon) = \frac{1}{2\pi^2 G(\varepsilon)}. \quad (22)$$

Ignoring slow logarithmic renormalization effects (weak localization), the dimensionless finite-energy $G(\varepsilon)$ should vary inversely with the disorder strength λ_a , so that

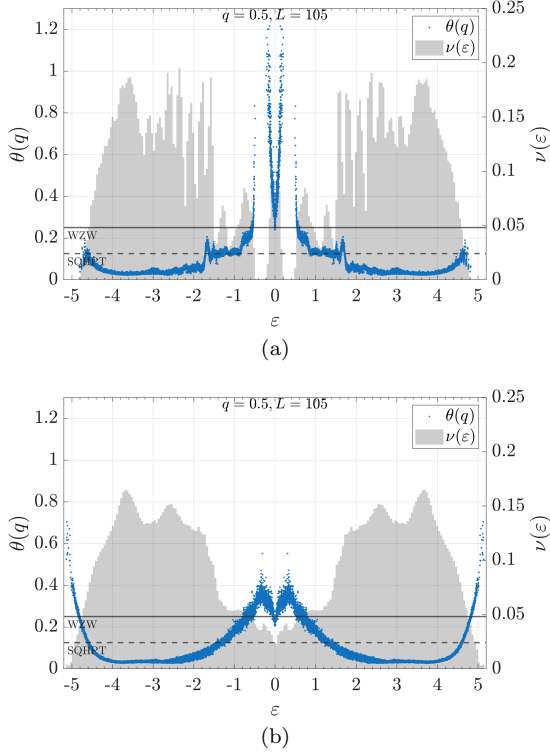


Figure 12. Effects of chiral “twist” disorder [Eq. (16)] on the chiral Bistritzer-MacDonald model for TBLG [Eq. (13)]. This is a special case of the 2D continuum model without the flat band, whereby spectral gaps arise from the (periodic, non-random) moiré potential. We incorporate weak uncorrelated randomness in each component of the latter. The model is tuned to the first magic angle $\alpha \simeq 0.586$. As in Figs. 10 and 11, we plot the multifractal curvature $\theta_q(\varepsilon)$ and density of states $\nu(\varepsilon)$ versus energy ε , for different disorder strengths (a) $\lambda = 0.5$ and (b) $\lambda = 2\pi$. As for the case of the continuum coupled to the flat band Fig. 11, the initial localizing effects of weak disorder appear to be *suppressed* by stronger disorder, “healing” the surface. The moiré parameters (see Fig. 6 and associated discussion in text) are $L = 105$ and $\xi = \pi$.

$\theta(\varepsilon) \sim \lambda_a$. Instead, both Figs. 10 and 11 show convergence to the universal SQHPT value $\theta(\varepsilon) \sim 1/8$ with increasing disorder. The conclusion is that even the flat-band hybridized surface theory seems to show robust SWQC [22, 25], at odds with the lattice-model results obtained above.

E. Bistritzer-MacDonald model

As a final point of comparison, we consider the chiral BM model for TBLG [Eq. (13)], subject to the chiral disorder in Eq. (16). As explained in Sec. II C, the chiral BM model is a special case of the 2D class-CI Dirac continuum model defined by Eq. (7), with four of the six vector-potential components encoding the interlayer (pure AB) tunneling due to the moiré potential, Eq. (II C). At the first magic angle, the chiral BM model has spectral gaps between the zero-energy flat bands and the rest of the spectrum [31, 34]. This already indicates the absence of a spectral flow principle for class-CI surface states.

In this case, we consider the incorporation of “twist” disorder [45–53] in the form of random white-noise additions to the moiré components, Eq. (16). Results are shown in Figs. 12 and 13. As in the case of the continuum model without the moiré potential, but with hybridization to the continuum flat band (Fig. 11), the numerical

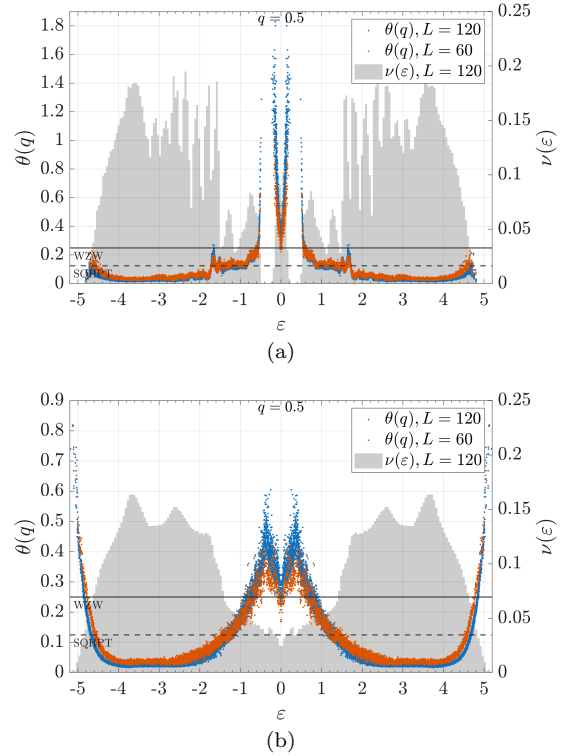


Figure 13. Same as Fig. 12, but with multifractal spectra shown for two different system sizes $L = 90$, box sizes $b = (3, 5, 6)$ and $L = 120$, $b = (4, 6, 8, 10)$. (a) $\lambda_a = 0.5$. (b) $\lambda_a = 2\pi$. The correlation length is $\xi = \pi$.

results for the dirty chiral BM model show that the energy gaps are not stable to disorder. Strongly localized states that appear near the gap edges for weak disorder are “washed out” by stronger disorder. The overall trend of the states, averaged over energy, is a *suppression* of the extent of Anderson localization with increasing disorder, see Fig. 1.

These results obtained here for “twist” dirt in the chiral BM model are very similar to the continuum case without the moiré potential: localization is suppressed by increasing disorder, which also fills in the spectral gaps. For the disorder strengths studied in this case, we do not observe the class-C SWQC for the BM model. This is because the model remains dominated by the periodic BM potential at larger energies for the system sizes that we can access here. We note that although the disorder studied here is chiral, it is different from the fine-tuned disorder discussed in Ref. [51] that preserves the flatness of the zero-energy bands at the first magic angle.

IV. OPEN QUESTIONS

As we have already summarized the main problems explored and resulting findings in the Introduction, here we close with a set of important open questions.

1. What is the physical picture behind the trivialization of surface states in class-CI bulk topological lattice models? For class AIII, a Berry curvature mechanism was uncovered in Ref. [6], reviewed in Sec. IC. This cannot apply to class CI due to the $T^2 = +1$ time-reversal symmetry in the latter. One possibility is quantum geometry [66–69], although for the case studied here the surface-trivial band hybridization does *not* open up a spectral gap, see Fig. 3(a). Moreover, the absence of a robust TPE for the continuum model studied here (with a spectral hybridization gap in the clean limit) suggests that one must confront the quantum geometry of bands that *interpolate* between 2D to 3D in the slab geometry.
2. A second key result of this work is the divergence between lattice and continuum surface models regarding the fragility of class-CI surface states. This points to a fundamental incompleteness of the 2D continuum Dirac equation for describing topological surface states. In particular, the TPE (introduced in Ref. [6] but explored for the first time numerically here) gives conflicting results for bulk lattice versus continuum surface calculations. Are there important implications for twisted moiré materials such as TBLG? This is an especially interesting, but potentially fraught question, given the *lattice-scale quasiperiodicity* (i.e., non-translational invariance) of twisted multilayers.
3. Finally, another persistent mystery is the *delocalization mechanism* for class-CI surface states at fi-

nite energy, i.e. the spectrum-wide quantum criticality scenario. For class AIII, the observation of critical, integer-quantum-Hall-plateau states at all nonzero surface-state energies is understood as the result of a statistical symmetry, tuning between surface-state topological Anderson insulators that are nucleated by Berry curvature from the surface-bulk interface or TPE [6]. From a symmetry-class perspective, this is natural because finite-energy states of any class-AIII system are expected to reside in the unitary class A, which also hosts the quantum Hall effect [54].

By contrast, finite-energy states of a class-CI system (topological or trivial, bulk or surface, lattice or continuum) are expected to reside in the orthogonal class AI [22], a class that is believed to *always localize* in two spatial dimensions [54]. In Ref. [22], it was argued that class-C might alternatively be realized for finite-energy surface states of a class-CI bulk, based in part on a symmetry analysis of replicated nonlinear sigma model target manifolds. The numerical results for SWQC obtained here and in [22, 25] are consistent with the spin quantum Hall plateau transition in class C, but no physical picture has yet emerged to convincingly explain this observation. An alternative scenario could be an (at present unknown) delocalization mechanism for class AI that happens to give a very similar multifractal spectrum as the class-C plateau transition.

ACKNOWLEDGMENTS

We thank L. Trifunovic, P. Brouwer, and R. Queiroz for helpful discussions. This work was supported by the Welch Foundation Grant No. C-1809. This work was supported in part by the Big-Data Private-Cloud Research Cyberinfrastructure MRI-award funded by NSF under grant CNS-1338099 and by Rice University’s Center for Research Computing (CRC).

Appendix A: Parameters for Fig. 1

The data points presented in Fig. 1 of the Introduction are obtained for the five different models studied in this work. The results in this summary figure involve averaging multifractal curvature data over finite energy windows in each model for different disorder strengths. The precise specification of all parameters used to obtain Fig. 1 are as follows,

1. CI lattice model. $W_1 = 0.4, 0.7, 0.85, 1$. Parameters: $N_y = 72$, $N_z = 10$, $\mu = -2$, $\Delta_1 = 1$, $\Delta_2 = 1$, $m_c = 1$. Energy window $(0.1, \infty)$.
2. CI lattice flat band. $W_1 = 0.4, 0.7, 1, 1.2$. Parameters: $\varepsilon_c = 0.4$, $\gamma = 0.5$, $N_y = 72$, $N_z = 10$, $\mu = -2$, $\Delta_1 = 1$, $\Delta_2 = 1$, $m_c = 1$. Energy window $(0.1, \infty)$.

3. CI continuum model. $W = 0.5, 1, \pi, 2\pi$. Parameters: $L = 97$, $\xi = \pi$, $b = (4, 6, 8)$. Energy window $(0.1, 0.4)$.

4. CI continuum model with flat band. $W =$

$0.2, 1, \pi, 2\pi$. Parameters: $\varepsilon_c = 0.4$, $\lambda_c = 0.2$, $\gamma_0 = 0.35$, $\lambda_g = 0.2$, $L = 97$, $\xi = \pi$, $b = (4, 6, 8)$. Energy window $(0.1, 0.6)$.

5. BM model. $W = 0.5, \pi, 2\pi, 7$. Parameters: $M_1 = 5$, $M_2 = 7$, $L = 105$. Energy window $(0.1, 1.5)$.

-
- [1] B. A. Bernevig and T. L. Hughes, *Topological Insulators and Topological Superconductors* (Princeton University Press, 2013).
- [2] M. Z. Hasan and C. L. Kane, Colloquium: Topological Insulators, *Rev. Mod. Phys.* **82**, 3045 (2010).
- [3] X.-L. Qi and S.-C. Zhang, Topological insulators and superconductors, *Rev. Mod. Phys.* **83**, 1057 (2011).
- [4] Y. Ando and L. Fu, Topological crystalline insulators and topological superconductors: From concepts to materials, *Ann. Rev. Condens. Matter Phys.* **6**, 361 (2015).
- [5] For a review, see e.g. A. Altland and B. D. Simons, *Condensed Matter Field Theory*, 3rd edition (Cambridge University Press, Cambridge, England, 2023).
- [6] A. Altland, P. W. Brouwer, J. Dieplinger, M. S. Foster, M. Moreno-Gonzalez, and L. Trifunovic, Fragility of Surface States in Non-Wigner-Dyson Topological Insulators, *Phys. Rev. X* **14**, 011057 (2024).
- [7] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Classification of topological insulators and superconductors in three spatial dimensions, *Phys. Rev. B* **78**, 195125 (2008).
- [8] A. Kitaev, Periodic table for topological insulators and superconductors, *AIP Conf. Proc.* **1134**, 22 (2009).
- [9] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. W. Ludwig, Topological insulators and superconductors: tenfold way and dimensional hierarchy, *New J. Phys.* **12**, 065010 (2010).
- [10] A. M. Essin and V. Gurarie, Delocalization of boundary states in disordered topological insulators, *J. Phys. A* **48**, 11FT01 (2015).
- [11] H. Schulz-Baldes and T. Stoiber, *Harmonic Analysis in Operator Algebras and its Applications to Index Theory and Topological Solid State Systems* (Springer, Cham, Switzerland, 2022).
- [12] R. B. Laughlin, Quantized Hall conductivity in two dimensions, *Phys. Rev. B* **23**, 5632 (1981).
- [13] B. I. Halperin, Quantized Hall conductance, current-carrying edge states, and the existence of extended states in a two-dimensional disordered potential, *Phys. Rev. B* **25**, 2185 (1982).
- [14] In addition to localizable classes CI, CII, and AIII, class DIII, which includes $^3\text{He-B}$ [7], is non-localizable (localizable) for odd (even) winding numbers in three dimensions [6].
- [15] B. Lapierre, L. Trifunovic, T. Neupert, and P. W. Brouwer, Topology of ultra-localized insulators and superconductors, *arXiv:2407.07957*.
- [16] D. Nakamura, K. Shiozaki, K. Shimomura, M. Sato, and K. Kawabata, Non-Hermitian Origin of Detachable Boundary States in Topological Insulators, *Phys. Rev. Lett.* **135**, 096601 (2025).
- [17] K. Shiozaki, D. Nakamura, K. Shimomura, M. Sato, and K. Kawabata, K -theory classification of Wannier localizability and detachable topological boundary states, *Phys. Rev. B* **112**, 075152 (2025).
- [18] A. P. Schnyder, P. M. R. Brydon, D. Manske, and C. Timm, Andreev Spectroscopy and Surface Density of States for a Three-Dimensional Time-Reversal-Invariant Topological Superconductor, *Phys. Rev. B* **82**, 184508 (2010).
- [19] A. Weiße, G. Wellein, A. Alvermann, and H. Fehske, The Kernel Polynomial Method, *Rev. Mod. Phys.* **78**, 275–306 (2006).
- [20] H. Guan and W. Zhang, Dual Applications of Chebyshev Polynomials Method: Efficiently Finding Thousands of Central Eigenvalues for Many-Spin Systems, *SciPost* **11**, 103 (2021).
- [21] X. Zhang, J. H. Wilson, and M. S. Foster, Critical Filaments and Superconductivity in Quasiperiodic Twisted Bilayer Graphene,” *Phys. Rev. B* **111**, 024207 (2025).
- [22] S. A. A. Ghorashi, Y. Liao, and M. S. Foster, Critical Percolation without Fine-Tuning on the Surface of a Topological Superconductor, *Phys. Rev. Lett.* **121**, 016802 (2018).
- [23] B. Sbierski, J. F. Karcher, and M. S. Foster, Spectrum-Wide Quantum Criticality at the Surface of Class AIII Topological Phases: An “Energy Stack” of Integer Quantum Hall Plateau Transitions, *Phys. Rev. X* **10**, 021025 (2020).
- [24] S. A. A. Ghorashi, J. F. Karcher, S. M. Davis, and M. S. Foster, Criticality across the Energy Spectrum from Random Artificial Gravitational Lensing in Two-Dimensional Dirac Superconductors, *Phys. Rev. B* **101**, 214521 (2020).
- [25] J. F. Karcher and M. S. Foster, How spectrum-wide quantum criticality protects surface states of topological superconductors from Anderson localization: Quantum Hall plateau transitions (almost) all the way down, *Ann. Phys.* **435**, 168439 (2021).
- [26] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, The electronic properties of graphene, *Rev. Mod. Phys.* **81**, 109 (2009).
- [27] N. M. R. Peres, Colloquium: The transport properties of graphene: An introduction, *Rev. Mod. Phys.* **82**, 2673 (2010).
- [28] S. Das Sarma, S. Adam, E. H. Hwang, and E. Rossi, Electronic transport in two-dimensional graphene, *Rev. Mod. Phys.* **83**, 407 (2011).
- [29] A. Altland, B. Simons, and M. Zirnbauer, Theories of low-energy quasi-particle states in disordered d -wave superconductors, *Phys. Rep.* **359**, 283 (2002).
- [30] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, Graphene Bilayer with a Twist: Electronic Structure, *Phys. Rev. Lett.* **99**, 256802 (2007).

- [31] R. Bistritzer and A. H. MacDonald, Moiré bands in twisted double-layer graphene, *Proc. Natl. Acad. Sci. (USA)* **108**, 12233 (2011).
- [32] R. de Gail, M. O. Goerbig, F. Guinea, G. Montambaux, and A. H. Castro Neto, Topologically protected zero modes in twisted bilayer graphene, *Phys. Rev. B* **84**, 045436 (2011).
- [33] For a review, see e.g. A. H. MacDonald, Bilayer Graphene’s Wicked, Twisted Road, *Physics* **12**, 12 (2019).
- [34] G. Tarnopolsky, A. J. Kruchov, and A. Vishwanath, Origin of Magic Angles in Twisted Bilayer Graphene, *Phys. Rev. Lett.* **122**, 106405 (2019).
- [35] A. N. Redlich, Gauge Noninvariance and Parity Nonconservation of Three-Dimensional Fermions, *Phys. Rev. Lett.* **52**, 18 (1984).
- [36] G. W. Semenoff, Condensed-Matter Simulation of a Three-Dimensional Anomaly, *Phys. Rev. Lett.* **53**, 2449 (1984).
- [37] F. D. M. Haldane, Model for a Quantum Hall Effect without Landau Levels: Condensed-Matter Realization of the “Parity Anomaly,” *Phys. Rev. Lett.* **61**, 2015 (1988).
- [38] In the case of twisted bilayer, the chiral limit is equivalent to a class-CI surface theory. Incorporating AA interlayer hopping but preserving approximate particle-hole symmetry (the model with “stable topology” [39, 40]) can be viewed as class-CI surface states with time-reversal symmetry breaking (class C) that does not open a gap. Breaking both chiral and particle-hole symmetries in twisted bilayer (“fragile topology” [41–43]) gives a Dirac Hamiltonian in class A that is equivalent to a non-topological spinless graphene monolayer with intervalley coupling.
- [39] Z. Song, Z. Wang, W. Shi, G. Li, C. Fang, and B. A. Bernevig, All Magic Angles in Twisted Bilayer Graphene are Topological, *Phys. Rev. Lett.* **123**, 036401 (2019).
- [40] Z. Song, B. Lian, N. Regnault, and B. A. Bernevig, Twisted bilayer graphene. II. Stable symmetry anomaly, *Phys. Rev. B* **103**, 205412 (2021).
- [41] H. C. Po, L. Zou, A. Vishwanath, and T. Senthil, Origin of Mott Insulating Behavior and Superconductivity in Twisted Bilayer Graphene, *Phys. Rev. X* **8**, 031089 (2018).
- [42] H. C. Po, L. Zou, T. Senthil, and A. Vishwanath, Faithful tight-binding models and fragile topology of magic-angle bilayer graphene, *Phys. Rev. B* **99**, 195455 (2019).
- [43] J. Ahn, S. Park, and B.-J. Yang, Failure of Nielsen-Ninomiya Theorem and Fragile Topology in Two-Dimensional Systems with Space-Time Inversion Symmetry: Application to Twisted Bilayer Graphene at Magic Angle, *Phys. Rev. X* **9**, 021013 (2019).
- [44] Ordinary electric potential disorder enters into the 2D surface theory of a 3D topological superconductor *in an anomalous way*, i.e. as a type of chiral-symmetric vector potential perturbation [7]. This is due to the “fractionalized” version of chiral symmetry that is realized at the boundary of the topological bulk.
- [45] J. H. Wilson, Y. Fu, S. Das Sarma, and J. H. Pixley, Disorder in Twisted Bilayer Graphene, *Phys. Rev. Res.* **2**, 023325 (2020).
- [46] B. Padhi, A. Tiwari, T. Neupert, and S. Ryu, Transport across twist angle domains in moiré graphene, *Phys. Rev. Res.* **2**, 033458 (2020).
- [47] G. Shavit, K. Kolář, C. Mora, F. von Oppen, and Y. Oreg, Strain Disorder and Gapless Intervalley Coherent Phase in Twisted Bilayer Graphene, *Phys. Rev. B* **107**, L081403 (2023).
- [48] N. Nakatsuji and M. Koshino, Moire Disorder Effect in Twisted Bilayer Graphene, *Phys. Rev. B* **105**, 245408 (2022).
- [49] P. A. Guerrero, V.-H. Nguyen, J. M. Romeral, A. W. Cummings, J.-H. Garcia, J.-C. Charlier, and S. Roche, Disorder-Induced Delocalization in Magic-Angle Twisted Bilayer Graphene, *Phys. Rev. Lett.* **134**, 126301 (2025).
- [50] A. Sanjuan Ciepielewski, J. Tworzydło, T. Hyart, and A. Lau, Transport Effects of Twist-Angle Disorder in Mesoscopic Twisted Bilayer Graphene, *Nanotechnology* **36**, 065401 (2025).
- [51] V. Crépel, P. Ding, N. Verma, N. Regnault, and R. Queiroz, Topologically Protected Flatness in Chiral Moiré Heterostructures, *Phys. Rev. X* **15**, 021056 (2025).
- [52] A. Uri, S. Grover, Y. Cao, J. A. Crosse, K. Bagani, D. Rodan-Legrain, Y. Myasoedov, K. Watanabe, T. Taniguchi, P. Moon, M. Koshino, P. Jarillo-Herrero, and E. Zeldov, Mapping the Twist-Angle Disorder and Landau Levels in Magic-Angle Graphene, *Nature* **581**, 47–52 (2020).
- [53] M. Kapfer, B. S. Jessen, M. E. Eisele, M. Fu, D. R. Danielsen, T. P. Darlington, S. L. Moore, N. R. Finney, A. Marchese, V. Hsieh, P. Majchrzak, Z. Jiang, D. Biswas, P. Dudin, J. Avila, K. Watanabe, T. Taniguchi, S. Ulstrup, P. Bøggild, P. J. Schuck, D. N. Basov, J. Hone, and C. R. Dean, Programming Twist Angle and Strain Profiles in 2D Materials, *Science* **381**, 677–681 (2023).
- [54] For a review, see e.g. F. Evers and A. D. Mirlin, Anderson transitions, *Rev. Mod. Phys.* **80**, 1355 (2008).
- [55] D. Bernard and A. LeClair, A classification of 2D random Dirac fermions, *J. Phys. A* **35**, 2555 (2002).
- [56] A.A. Nersisyan, A.M. Tsvelik, F. Wenger, Disorder effects in two-dimensional d-wave superconductors, *Phys. Rev. Lett.* **72**, 2628 (1994).
- [57] A.W.W. Ludwig, M.P.A. Fisher, R. Shankar, G. Grinstein, Integer quantum Hall transition: An alternative approach and exact results, *Phys. Rev. B* **50**, 7526 (1994).
- [58] A.M. Tsvelik, Exactly solvable model of fermions with disorder, *Phys. Rev. B* **51**, 9449 (1995).
- [59] C. Mudry, C. Chamon, X.-G. Wen, Two-dimensional conformal field theory for disordered systems at criticality, *Nucl. Phys. B* **466**, 383 (1996).
- [60] J.-S. Caux, I.I. Kogan, A.M. Tsvelik, Logarithmic operators and hidden continuous symmetry in critical disordered models, *Nucl. Phys. B* **466**, 444 (1996).
- [61] M.J. Bhaseen, J.-S. Caux, I.I. Kogan, A.M. Tsvelik, Disordered Dirac fermions: the marriage of three different approaches, *Nucl. Phys. B* **618**, 465 (1996).
- [62] Y. Z. Chou and M. S. Foster, Chalker scaling, level repulsion, and conformal invariance in critically delocalized quantum matter: Disordered topological superconductors and artificial graphene, *Phys. Rev. B* **89**, 165136 (2014).
- [63] Z. Wang, M. P. A. Fisher, S.M. Girvin, and J. T. Chalker, Short-range interactions and scaling near integer quantum Hall transitions, *Phys. Rev. B* **61**, 8326 (2000).
- [64] G. Livan, M. Novaes, and P. Vivo, *Introduction to Random Matrices: Theory and Practice* (Springer, Cham, 2018).
- [65] F. Wegner, Inverse participation ration in $2 + \epsilon$ dimen-

- sions, [Z. Phys. B **36**, 209 \(1980\)](#).
- [66] J. P. Provost and G. Vallee, Riemannian Structure on Manifolds of Quantum States, *Comm. Math. Phys.* **76**, 289-301 (1980).
- [67] P. Törmä, Essay: Where Can Quantum Geometry Lead Us?, [Phys. Rev. Lett. **131**, 240001 \(2023\)](#).
- [68] J. Yu, B. A. Bernevig, R. Queiroz, E. Rossi, P. Törmä, and B.-J. Yang, Quantum Geometry in Quantum Materials, [arXiv:2501.00098](#).
- [69] N. Verma, P. J. W. Moll, T. Holder, and R. Queiroz, Quantum Geometry: Revisiting electronic scales in quantum matter, [arXiv:2504.07173](#).