

Supplement to “Flat Topological Bands and Eigenstate Criticality in a Quasiperiodic Insulator”

Yixing Fu,¹ Justin H. Wilson,¹ and J. H. Pixley^{1,2,3}

¹*Department of Physics and Astronomy, Center for Materials Theory, Rutgers University, Piscataway, NJ 08854 USA*

²*Center for Computational Quantum Physics, Flatiron Institute, 162 5th Avenue, New York, NY 10010*

³*Physics Department, Princeton University, Princeton, New Jersey 08544, USA*

(Dated: April 11, 2021)

In the following supplemental material we provide details about the phase diagram, the magic-angle transition along the semimetal lines, the machine learning algorithm we have used to support the phase diagram in text, the perturbation theory at second and fourth order in the potential, as well as additional numerical results.

CONTENTS

I. Additional properties of the phase diagram	1
A. Main Phase Diagram	1
B. Dependence on the quasiperiodic wavevector Q	2
C. Chemical Potential dependence of σ_{xy}	2
II. Magic-angle transition	3
III. Machine Learning the localization transition	3
IV. Perturbation theory at second and fourth order	6
V. Finite size effects	7
A. Density of states and conductivity	7
B. Inverse Participation Ratio	8
VI. Extracting the spectral gap	8
A. Critical Exponents at the topological insulator to metal phase transition	10
B. Gapsizes and IPR	10
VII. Dispersion relation, effective mass and Berry curvature	11
A. Effective mass	11
B. Edge states	11
C. Berry curvature of folded bands	11
References	13

I. ADDITIONAL PROPERTIES OF THE PHASE DIAGRAM

A. Main Phase Diagram

Here we augment the phase diagram shown in Fig. 1 of main with several details. Focusing on Fig. S1(left), the magenta line indicates an additional measure of the Anderson localization transition, using a neural net model as elaborated in Section III. Critical eigenstates may appear localized and therefore are not straightforward to diagnose with the IPR and conductivity alone (due to their anomalous scaling with system size). Thus, we use machine learning to provide a conservative measure of the Anderson localized phase, which in certain regimes matches the IPR and conductivity, but in other regimes of the phase diagram extends to larger values of W . In addition, the machine learning approach identifies an area ranging from $M = 4, W = 3$ to $M = 4.5, W = 4$ in the phase diagram (the darker region bounded by pink curve in Fig. S1) as critical, that is not identified by other observables. We leave the details on the CM-AI phase transition for future studies.

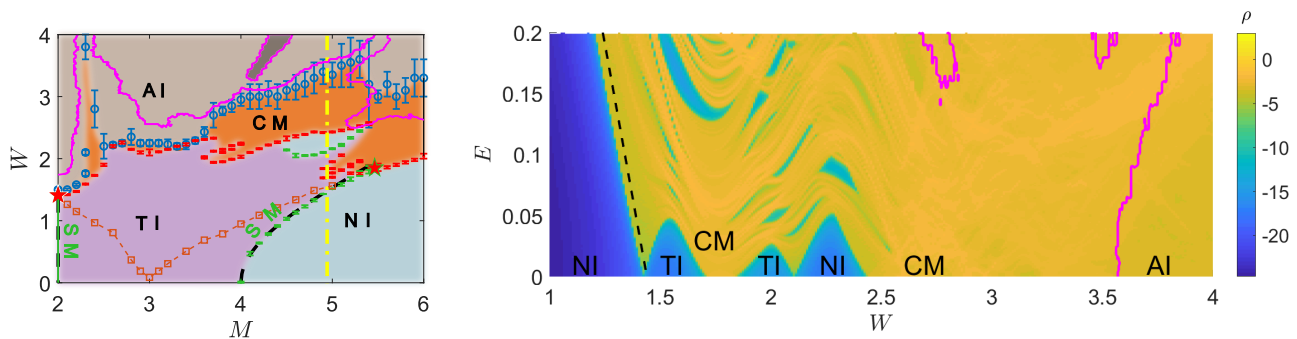


FIG. S1. (a) Full phase diagram with all measures used to diagnose phases and transitions. The magenta line shows the boundary between delocalized or critical phase and localized phase at zero energy, as indicated by the neural net model. The dark region, roughly extending from $M = 4, W = 3$ to $M = 4.5, W = 4$ is indicated as critical phase by the neural net model, but not identified by any other observables. The dashed orange line inside the TI phase shows where the size of the gap centered at $E = 0$ is maximal and thus starts to significantly deviate from perturbation theory. (b) A cut of the phase diagram in energy space represented by the yellow line in (a). Notice the multiple phase transitions, all driven by quasiperiodicity (W) and the higher energy metallic nature. The pink curve represents the boundary to machine-learned, localized eigenstates.

The yellow line starting from $M = 4.9$ is expanded in the parameter space of W, E in Fig. S1(right). The cut shows an example of the very rich sequence of phase transitions, into-and-out-of metallic and topological phases at zero energy, made possible by only increasing W at fixed M . The magenta curve shows the boundary to localized states as determined by the neural net model.

Inside the TI phase with intermediate W , IPR of lowest energy states demonstrate a region of critical scaling well before entering CM phase. Such region of criticality can provide a coarse guide to the parameters (M, W) where flat topological band discussed in the main text is developed. Intriguingly, the onset of such criticality can roughly be traced by the dashed orange line (with \square symbols) originating from $M = 3, W = 0$ in Fig. S1(a) where the size of the gap centered at $E = 0$ no longer follows perturbation theory (see Section IV for perturbation theory and Section VI for numerical calculation of the gaps). The eigenstates at the band edge appear critical, as measured by the IPRs, for a range of W above such line (see Section VI).

B. Dependence on the quasiperiodic wavevector Q

In the main text we have focused on a quasiperiodic wavevector $Q = 2\pi F_{n-2}/F_n$ and linear system size $L = F_n$, where F_n is the n th Fibonacci number. The structure of the phase diagram is strongly dependent on the incommensurate value of Q chosen. Here, we show the zero energy density of states that probes insulating, semimetallic, and metallic parts of the phase diagram (but cannot discern between delocalized and localized wavefunctions) in Fig. S2 for $Q = F_{n-3}/F_n$ (a) and $Q = F_{n-4}/F_n$ (b). With these smaller values of Q , the semimetal-to-metal magic-angle transition along $M = 2$ happens for smaller values of W . Qualitatively, this behavior is captured by the perturbation theory near $W = 0$, where it shows the Dirac cone velocity v is renormalized more strongly for smaller values of Q . However, for even smaller Q , higher order perturbation theory is required to see the velocity renormalizing down to 0. On the other hand, the phase boundary rooted from $M = 4$ and $W = 0$ along the NI-to-TI phase boundary can be predicted well by the perturbation theory (shown as a red line in Fig. S2).

C. Chemical Potential dependence of σ_{xy}

The Chern number is directly determined by the Hall conductivity σ_{xy} . While we use $\rho(0)$ to accurately locate phase boundaries, the numerical calculation of σ_{xy} at large system sizes has more computational complexity. However, σ_{xy} can be used to distinguish trivial and topological states and locate where the Chern bands are in energy. Here, we show an example of a color plot of σ_{xy} at a fixed $M = 4.2$ and varying the Fermi energy E_F and disorder strength W , see Fig. S3. The emergence of a topological phase after the collapse of the (lowest energy) trivial band gap can be seen clearly. In addition, our perturbative result (shown as the red lines) is in excellent agreement with the numerics in locating these phase boundaries for $0 \leq W \lesssim 2$.

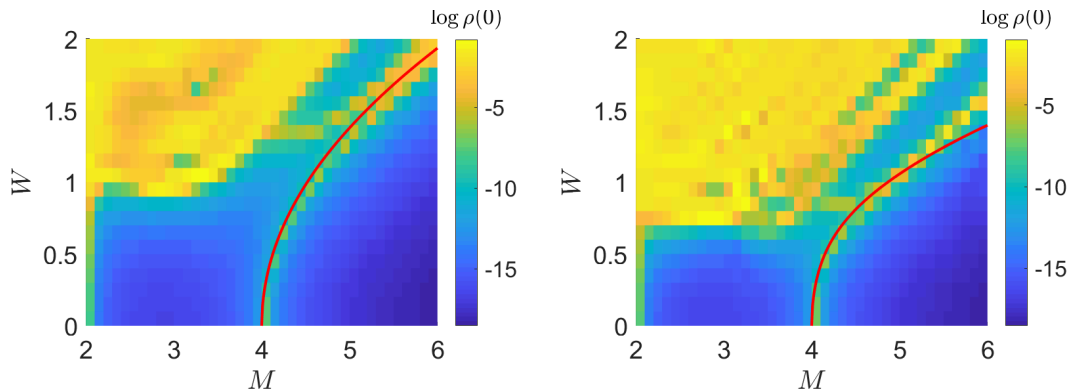


FIG. S2. **Phase diagram determined from density of states at $E_F = 0$ for $Q = F_{n-3}/F_n$ and $Q = F_{n-4}/F_n$.** The red line shown is the result of fourth order perturbation theory used to determine the location of the NI-to-TI transition from the vanishing of the renormalized topological mass \tilde{M} . The perturbative result agrees well with the numerics up to $W \approx 1$.

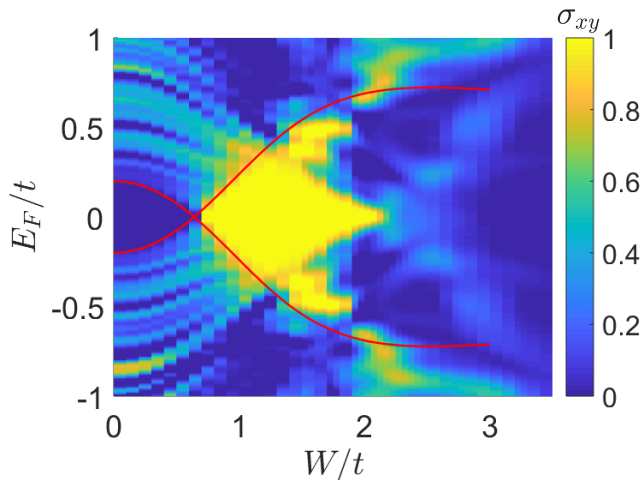


FIG. S3. **Finite energy topological phase diagram.** The Hall conductivity σ_{xy} at various Fermi energies E_F and quasiperiodicity W . The red lines are the perturbation theory prediction of gap size.

II. MAGIC-ANGLE TRANSITION

The semimetal at $M = 2$ is stable with a velocity (calculated from the DOS) that vanishes like $\tilde{v} \sim (W_c(M = 2) - W)^{\beta/2}$ where $W_c(M = 2) = 1.42 \pm 0.02$ and $\beta = 2 \pm 0.3$. This is demonstrated in Fig. S4(a) where \tilde{v} vanishes when $\rho(0)$ rises. Additionally, the wave functions are localized in momentum space when $W < W_c(M = 2)$, and delocalized in momentum space when $W > W_c(M = 2)$ (as indicated in Fig. S4(b) by \mathcal{I}_k being L -independent and $\mathcal{I}_k \sim 1/L^2$, respectively). When the real space IPR is L -independent and the resistivity increases with L and N_c , there is a localization transition with $W_A(M = 2) = 1.50 \pm 0.03$, indicating a small but finite CM phase.

III. MACHINE LEARNING THE LOCALIZATION TRANSITION

In the present model, we found it challenging to pinpoint the Anderson localization transition using conductivity and the inverse participation ratio due to a large number of critical states that can appear localized by some metrics but not others. Therefore, we have supplemented this analysis with a machine learning classification of the single particle wavefunctions.

Machine learning is a class of methods where a non-specialized program can be used to perform a specific task when supplied with an abundance of data. Many machine learning techniques have been applied to various aspects of physics^{S1-S3}. In this work we used Convolutionary Neural Networks (CNN) to distinguish whether a wave function is localized or extended. We train the neural network on a set of wave functions whose localized nature can be easily and

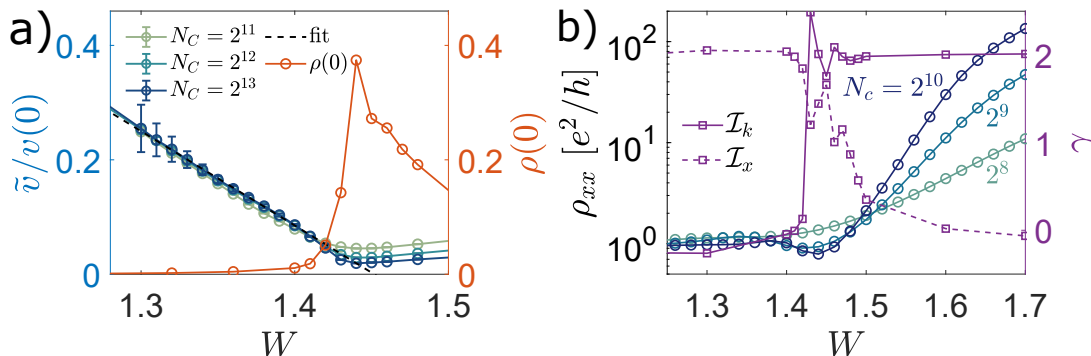


FIG. S4. **The magic-angle transition for the semimetal line $M = 2$.** (a) Renormalized velocity $v/v(0)$ and the resulting finite density of states $\rho(0)$ at the transition, extracted from $\rho(E)$ that is calculated using KPM method with system size $L = 144$, Chebyshev cutoff $N_C = 2^{15}$. See Fig. S7 for further discussion on obtaining \tilde{v} . (b) These plots indicate the appearance of a critical metallic phase $1.4 \lesssim W \lesssim 1.5$ inferred from both the resistivity ρ_{xx} and the scaling of the momentum- and real-space IPRs. ρ_{xx} is calculated using Kubo formula (see main text) with KPM method. The L -dependence of the IPRs is fitted from lowest energy eigenstates obtained using Lanczos method for $L = 89$, $L = 144$, and $L = 233$ to a power law form $\mathcal{I}_\alpha \sim 1/L^{\gamma_\alpha}$, and γ_α is shown as the right vertical axis.

unambiguously determined. Such a neural network automatically applies to all other points in the phase diagram, determining the phase boundaries efficiently and objectively.

A neural network consists of a massive number of nonlinear functions and linear transformations, usually as several “layers,” to replicate any task that distills information from data. Practically, such a combination can be tuned to fit any mapping. Hence, as long as a concrete definition of the task to be executed is available, we can use labelled data as an example to tune the neural network until it replicates the task. Such a process is called “training,” and can be calculated efficiently using modern computers.

In the present context, the problem that we want the neural network to solve is to distinguish localized wavefunctions from extended ones. This task can be thought of as a mapping from the space of all wavefunctions to a binary result of localized or extended. Using a set of wavefunctions labelled in advance, we can train the neural network to capture the relation between wavefunction data and the prediction of a localized phase. Once the training is finished, we can use the neural network model to classify a much larger dataset of wavefunctions, and map out a detailed phase diagram.

One crucial but more technical component of deep learning is the choice of the form and organization (i.e., “architecture”) of the nonlinear functions used to fit^{S4,S5}. In this work, we use a simple version of CNN. The wavefunction classification task is somewhat analogous to figuring out whether the image includes a dog or cat, which is a classical application of CNN. The CNN architecture makes use of a convolution operation prior to applying the nonlinear functions. The convolution operations effectively scrambles but preserves the information at various locations of the input data, and hence makes the model “translational invariant”, i.e. the location of the feature does not affect the output. Such translational invariance allows the neural net model to treat critical and/or localized structures at different locations in the same way.

The neural network methods of machine learning usually suffer from over-fitting that harms the predictive power of the model. Simple and conventional methods against over-fitting including adding regularization terms, use of drop-off layers^{S6} and so on. These methods are practically efficient and sufficient for our purpose.

A summary of the architecture we have used with a convolutionary neural network and drop-off layers is shown in Fig. S5. We experimented with a few different hyper parameters of the CNN architectures, and the model yields similar results. Hence the neural net model we have used is not a result of fine tuning. The robustness across different setups is likely because the localization feature is prominent and less ambiguous as opposed to typical computer vision tasks.

The training set is constructed in two different ways:

1. We look at the images to judge whether each wavefunction is clearly localized or not. The cases in which we are unsure are discarded from the training set. To minimize the affect of systematic bias caused in labelling the training set, we go through wavefunctions at several runs where each set is drawn randomly from the entire collection of wavefunctions and shuffled. Hence, the mislabelling can be considered as a random variation that is independent from the features that do not affect decision boundaries.

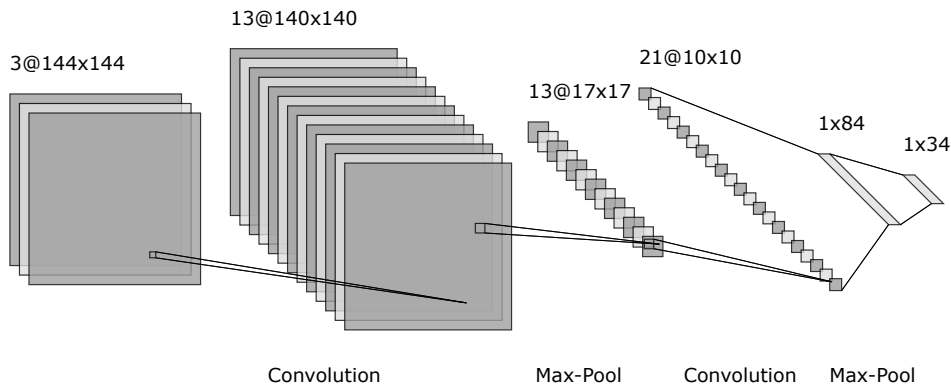


FIG. S5. **Schematic diagram of the neural network structure used for localization detection.** For convolution layers, we apply a convolution operation over a small window to get a data point in the next layer. Max-Pool layer simply takes the maximum of each window to reduce the model size. We also add batch-normalization and dropout layers before and after Max-Pool, but they are not shown here as they do not alter the overall architecture.

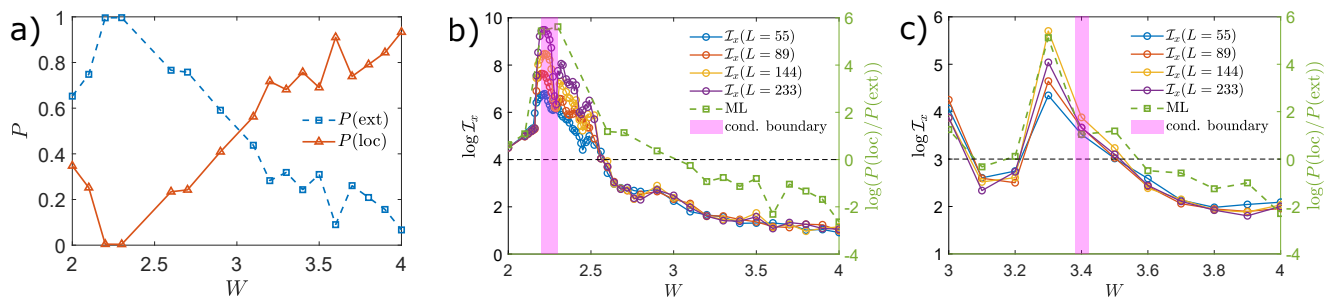


FIG. S6. **Comparing the IPR with the machine learning outcome.** (a) Shows an example of the neural network output for $M = 2.7$, given as the probability of a state being localized [$P(\text{loc})$] or extended [$P(\text{ext})$]. The summarized results are shown for $M = 2.7$ (b) and $M = 4.9$ (c), with comparison against KPM and IPR results. The difference between the two probabilities measures how confidently the model can distinguish localized or extended. Also shown in the figure with the magenta strips is the phase boundary determined by the conductivity, which indicates a transition near $W = 2.25$ for $M = 2.7$, and $W = 3.4$ for $M = 4.9$. Although the three different methods match quite well for $M = 4.9$, for $M = 2.7$ the IPR shows strongly critical behavior up until $W = 2.5$, well after the conductivity appears to vanish. Such critical behavior is detected by the neural net model. For W between 2.3 and 2.5 the IPR shows a strong L dependence and the neural net model predicts an extended phase with high confidence. For a range of W larger than 2.5, the IPR shows a weak L dependence across different system sizes, while in the neural net model $P(\text{loc})$ and $P(\text{ext})$ are quite close to each other.

2. We choose $W > 6$ for localized wavefunction examples, and sample $W = 0$ at various values of M for extended wavefunctions.

The training set of method 2 does not include any of the critical wavefunctions in the CM phase. As a result, the CNN model identifies the critical phase as localized, producing a phase boundary in line with SM/TI to CM transition. This result can also be replicated using the training set from method 1 if we only include extended and fully localized wavefunctions. However, with method 1 we can instead label a dataset such that the non-localized label *includes* critical wavefunctions to provide an interesting complement to the KPM results and is hence included in the main results of Figure 1.

The phase boundary obtained from machine learning between localized and non-localized wavefunctions roughly traces the CM-AI phase boundary provided by the conductivity computed with the KPM for M between 3.8 and 5.4, but it provides a slightly different boundary elsewhere. For $M < 3.8$, the machine learning result labels regions as critical that have a conductivity that looks localized (i.e. σ_{xx} is vanishing with increasing N_c). We further investigate the nature of this region using the inverse participation ratio (IPR) in real and momentum space bases, see Fig. S6. The IPR in this region shows critical behavior that transits into a localized phase at a point that is hard to accurately determine. The machine learning result provides a conservative estimate of where the criticality ends and localization sets in.

In summary, our use of the machine learning method in the present context is to provide an additional measure of

the non-trivial phase boundaries that have a lot of structure. We then use conventional methods (conductivity and the IPR) to validate the physical nature of the phase boundaries.

IV. PERTURBATION THEORY AT SECOND AND FOURTH ORDER

In this section, we provide additional details on the perturbation theory at second and fourth order that correctly captures the renormalization of the topological mass (to describe phase transitions in and out of the TI) and the velocity in the SM phase.

We begin by considering the single-particle Green function

$$\hat{G}_0(\omega) = [\omega - h_0(\omega)]^{-1}, \quad \hat{G}(\omega) = [\omega - h_0(\omega) + V]^{-1} \quad (S1)$$

and use Dyson's equation

$$G(\mathbf{k}, \omega)^{-1} = \omega - h_0(\mathbf{k}) - \Sigma(\mathbf{k}, \omega) \quad (S2)$$

where $\Sigma(\mathbf{k}, \omega)$ is the self-energy at momentum \mathbf{k} including all $G_0(\mathbf{k}, \omega)$ irreducible diagrams. Close to the SM phase near $M = M_1 \equiv 2$ or $M = M_2 \equiv 4$, we express the Hamiltonian in the low-energy limit around the corresponding Dirac node \mathbf{K} as $h_0(\mathbf{K} + \mathbf{q}) = v\mathbf{q} \cdot \boldsymbol{\sigma} + (M - M_i)\sigma_z$ and similarly expand the self-energy to obtain $\Sigma(\mathbf{k} = \mathbf{K} + \mathbf{q}, \omega) = \omega\Sigma_E\sigma_0 + \Sigma_p\mathbf{q} \cdot \boldsymbol{\sigma} + \Sigma_z\sigma_z$ (where σ_0 is the 2-by-2 identity matrix and the $\sigma_{x,y,z}$ are the Pauli matrices). We define the quasiparticle residue Z , the renormalized topological mass \tilde{M} , and renormalized velocity \tilde{v} such that the resulting Green function in the low-energy limit has the form

$$G(\mathbf{k} = \mathbf{K} + \mathbf{q}, \omega) = \frac{Z}{\omega - \tilde{v}\mathbf{q} \cdot \boldsymbol{\sigma} - \tilde{M}\sigma_z}. \quad (S3)$$

Then using Σ_E , Σ_z and Σ_p from $\Sigma(\mathbf{k}, \omega)$ we can express Z , \tilde{M} and \tilde{v} as:

$$Z^{-1} = 1 - \Sigma_E, \quad (S4)$$

$$\tilde{M} - M_i = (M - M_i + \Sigma_z)Z^{-1}, \quad (S5)$$

$$\tilde{v} = v(1 + \Sigma_p/v)Z^{-1}, \quad (S6)$$

To calculate $\Sigma(\mathbf{k}, \omega)$, we treat $V(\mathbf{r})$ perturbatively. In momentum space, V is a delta function connecting \mathbf{k} to $\mathbf{k} \pm Q\hat{x}$ and $\mathbf{k} \pm Q\hat{y}$. Hence, at second order the self energy is

$$\Sigma^{(2)}(\mathbf{k}, \omega) = (W/2)^2 \sum_{\pm, \hat{\mu} = \{\hat{x}, \hat{y}\}} \frac{1}{\omega - h_0(\mathbf{k} \pm Q\hat{\mu})}. \quad (S7)$$

Near $M = 4$, $\mathbf{k} = \mathbf{M} + \mathbf{q}$ with $\mathbf{M} = (\pi, \pi)$, this yields

$$\Sigma_E^{(2)} = -\frac{W^2}{D_2}, \quad (S8)$$

$$\Sigma_p^{(2)} = \frac{W^2}{2} \frac{(4 - M)^2(1 + \cos Q)}{D_2^2} v, \quad (S9)$$

$$\Sigma_z^{(2)} = W^2 \frac{(4 - M) + (\cos Q - 1)}{D_2}, \quad (S10)$$

where $D_2 = (4 - M)^2 + 2(3 - M)(\cos Q - 1)$ is the common denominator that is always positive for $M > 3$. Observe that the numerator of $\Sigma_z^{(2)}$ is also always negative for $M > 4$, and $\Sigma_E^{(2)}$ is always negative. Hence \tilde{M} is renormalized to be smaller as W increases, predicting a critical W where $\tilde{M}(W) = 4$ where TI to CM transition occurs. On the other hand, the direction of velocity renormalization is not obvious from the second-order perturbation theory, and indeed we can only predict the velocity to be renormalized to 0 at fourth-order perturbation theory. This is indicative of scattering off a single Dirac cone, where due to spin selection rules it requires a larger momentum exchange to induce intranode scattering.

The fourth-order perturbation theory includes all of the diagrams that connect the Dirac node to points in the Brillouin zone that are $2Q$ Manhattan distance apart and then back. The fourth order contributions to $\Sigma(\mathbf{k}, \omega)$ are

$$\begin{aligned} \Sigma_E^{(4)} = & \frac{W^4}{4}(-15M^4 + 166M^3 + (-36M^2 + 206M - 295) \cos(3Q) - 732M^2 + \\ & (2M(M(24M - 221) + 697) - 1497) \cos(Q) + (2M(M(13M - 115) + 356) - 770) \cos(2Q) + \\ & 6(M - 3) \cos(4Q) + 1522M - 1260)/D_4 \quad (\text{S11}) \end{aligned}$$

$$\begin{aligned} \Sigma_z^{(4)} = & \frac{W^4}{8}(-10M^5 + 138M^4 - 806M^3 + 2509M^2 + (2M(2M(M(11M - 134) + 622) - 2615) + 4212) \cos(Q) + \\ & 2(M(M(3M(3M - 37) + 538) - 1208) + 1048) \cos(2Q) + 2(451 - 6M(M(3M - 26) + 76)) \cos(3Q) + \\ & 5(M - 3)(3M - 8) \cos(4Q) - 2(M - 3) \cos(5Q) - 4155M + 2904)/D_4 \quad (\text{S12}) \end{aligned}$$

where D_4 is the common denominator

$$\begin{aligned} D_4 = & (-2(M - 3) \cos(Q) + (M - 6)M + 10)^2(-4(M - 2) \cos(Q) + \\ & (M - 4)M + \cos(2Q) + 7)(-2(M - 3) \cos(2Q) + (M - 6)M + 10) \quad (\text{S13}) \end{aligned}$$

With the fourth-order correction, we find that the perturbation theory agrees very well with the numerical results as demonstrated in the main text as well as Figs. S2 and S3. However, fourth order perturbation theory for the velocity renormalization only qualitatively predicts the magic-angle transition where $\tilde{v} = 0$, but at a much larger W than indicated by numerical results. It is natural to expect that this is due to the single node nature of the bandstructure at $\mathbf{M} = (\pi, \pi)$ (all of the scattering is intranode). We anticipate a better prediction of magic angle transition may be achievable only at even higher orders of perturbation theory.

Using exactly the same procedure we can consider the case of M near 2, which is the SM line that divides the two TI regions with opposite sign in the quantum spin Hall effect. From a symmetry point of view it is not surprising that the $M = 2$ SM line is W -independent. This is indeed the case from the perturbation theory, as up to fourth order we have $\Sigma_z^{(2)} = 0$ and $\tilde{M} = M = 2$, hence there is no topological mass renormalization. That means starting from such a SM phase, quasiperiodicity is not driving it out of SM due to curvature in the phase boundary. At second order the velocity only reduces but does not go to zero and the renormalization of \tilde{v} is only due to the quasiparticle residue $\Sigma_E^{(2)} = -\csc(Q/2)^2$, while $\Sigma_p^{(2)} = 0$. This can be understood as follows, at $M = 2$ the two Dirac cones are at $\mathbf{X} = (\pi, 0)$ and $\mathbf{Y} = (0, \pi)$ and being separated by 2π Manhattan distance in momentum space, second order perturbation theory will not be able to induce internode scattering. Whereas, at fourth order the two Dirac cones can be connected by $2Q$ hops in the Brillouin zone. Thus, only fourth order perturbation will be able to predict a vanishing velocity and a magic-angle transition. In line with this reasoning, the renormalized velocity \tilde{v} up to fourth order is

$$\tilde{v} = (v + \Sigma_p^{(2)} + \Sigma_p^{(4)})/Z \quad (\text{S14})$$

Here $\Sigma_p^{(2)}$ vanishes, and the fourth order term

$$\Sigma_p^{(4)} = \frac{W^4}{16^2}(1 + 4 \cos(Q)) \csc(Q/2)^4 v \quad (\text{S15})$$

is negative only when $Q > \cos^{-1}(1/4) \approx 1.82$. Only in this regime does the perturbation theory predict a magic-angle transition. For example, at $Q = 2\pi F_{n-2}/F_n$, it predicts \tilde{v} to vanish at $W_c^{(4)} = 4 \sin(Q/2)(-1 - 4 \cos(Q))^{-1/4} \approx 3.16$. This fourth order perturbative result $W_c^{(4)}$ is an overestimate of the true critical W_c , and thus a more accurate prediction will require higher order perturbation theory. For smaller Q such as $Q = 2\pi F_{n-3}/F_n$, the velocity can never reach 0 at fourth order in perturbation theory. Hence, the magic-angle transition is an even higher order effect than that of $Q = 2\pi F_{n-2}/F_n$. In other words, the reduction of the value of Q requires higher order in perturbation (more Q ‘‘hops’’) to capture internode scattering.

V. FINITE SIZE EFFECTS

A. Density of states and conductivity

In this section we discuss the finite size effects in the KPM calculations of the density of states and the conductivity. The numerical calculation puts the model on a lattice with a linear size L , while the truncation of the polynomial

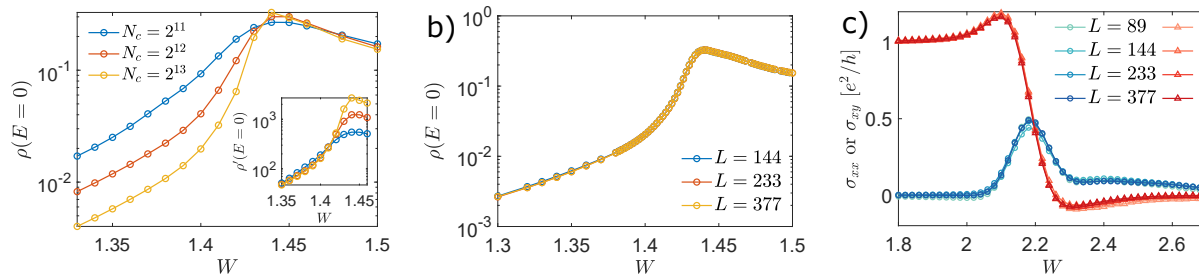


FIG. S7. **Finite size dependence of the density of states and conductivity.** (a) The N_c and (b) L dependence of $\rho(E=0)$ for $M=2$ as a function of W . As the expansion order N_c is increased, the rise of $\rho(E=0, W)$ is much more steep right before the transition. The inset shows $\rho'(E=0)$, computed as $\rho' = N_c\rho$, in the regime this is N_c independent we can extract the renormalized velocity \tilde{v} from this scaling. The L dependence is very weak for $N_c = 2^{13}$. (c) The L dependence of σ_{xy} and σ_{xx} as we fix $N_c = 2^{10}$ for the cut $M = 3.3$. These results demonstrate that we are close to converged in L and the largest finite size effect in the data stems from the finite KPM expansion order.

expansion at expansion order N_c in the KPM controls the energy resolution. As N_c and L increase (attempting to approach the thermodynamic limit), the observables we calculate using the KPM (i.e., the density of state $\rho(E)$, the resistivity ρ_{xx} , the longitudinal and Hall conductivities σ_{xx} and σ_{xy}) all become sharper, allowing us to accurately determine the critical points. In addition, we average over more than 200 samples with random twisted boundary conditions, allowing us to suppress finite size effect rooted from finite L and to avoid emphasizing results from any specific choice of phase in quasiperiodicity.

As an example, we show in detail the N_c and L dependence near the SM-to-CM transition at $M=2$, see Fig. S7(a,b). For fixed $L=233$ and varying N_c , we see that the behavior of the density of states sharpens near the transition, demonstrating that the transition occurs between $W=1.46$ and $W=1.48$. Also we see that at least for density of states at zero energy, taking $L=144$ is adequate for N_c up to 2^{13} , as the $\rho(E=0, W)$ curve does not change at all as we vary L above 144. To determine the renormalized velocity \tilde{v} from the density of states we use the scaling of the SM $\rho(E) \sim \tilde{v}^{-2}|E|$ to obtain the scaling with the KPM expansion order $\rho(0) \sim \tilde{v}^{-2}(1/N_c)$. Here, we have used the fact that the infrared low-energy scale (δE) induced by the finite KPM expansion order is related to N_c via $\delta E = \pi D/N_c$ where D is the total bandwidth of the Hamiltonian. As shown in the inset of Fig. S7(a), we find excellent data collapse of $N_c\rho(0)$ that we use to extract \tilde{v} in the SM phase.

In the main text, we have shown examples of the N_c dependence of the conductivity and the resistivity, where a larger cutoff leads to a steeper change in the $\rho_{xx}(W)$ curve near the critical W , and a longer range of W where σ_{xy} is quantized. In Fig. S7(c) we show the L dependence of the longitudinal and Hall conductivities near W_c as we vary L but fix N_c . In this data we see that for $L=89$ there is a slight variation from the other values of L , but above $L=144$ the $\sigma_{xx}(W)$ and $\sigma_{xy}(W)$ are not changing at all for the different L 's. Again this justifies the use of $L=144$ for N_c up to 2^{10} to calculate conductivity as in the main text.

B. Inverse Participation Ratio

The inverse participation ratios (IPR) reflects the structure of the wavefunction in a particular basis and is commonly used to study Anderson localization transitions. When the IPR is independent from system size, the system is localized; when the IPR scales as $1/L^d$ where L^d is the volume of the d -dimensional lattice, the system is extended. When the IPR vanishes with a power law less than d , the wavefunction is critical. To consistently assess the behavior of the IPR, we calculate the IPR (in both real and momentum space bases) for each combination of parameters at $L=55, 89, 144$ and 233 . Then we fit the log of the IPR $\log(\mathcal{I})$ vs $\log L$, and extract the slope, which is the power law. We demonstrate some examples of such fitting in Fig. S8. The results of the IPR scaling shown in the main text and in Fig. S11 are all computed in this way.

VI. EXTRACTING THE SPECTRAL GAP

The size of the insulating gap corresponds to the topological mass \tilde{M} ; and for the topological insulator to metal phase transition the power-law scaling of the gap size allows us to extract critical exponents at the TI to CM transition as described in the main text. Here we elaborate on the details for extracting spectral gap.

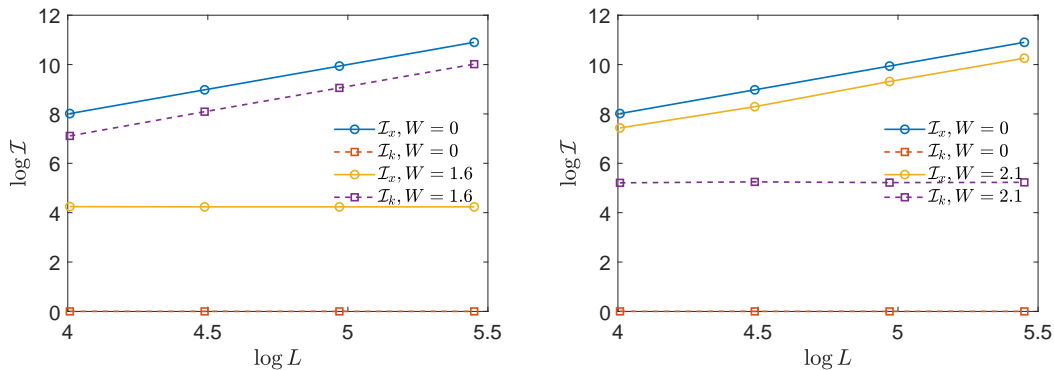


FIG. S8. **L dependence of the IPR** Here we demonstrate two examples of how we determine γ_α for IPR data in the basis $\alpha = x, k$, where $I_\alpha \sim L^{-\gamma_\alpha}$, for $M = 2.7$ (a) and $M = 3.3$ (b). We take a linear fit for $\log I_k$ or $\log I_x$ over $\log L$, then the slope of the fit estimates γ .

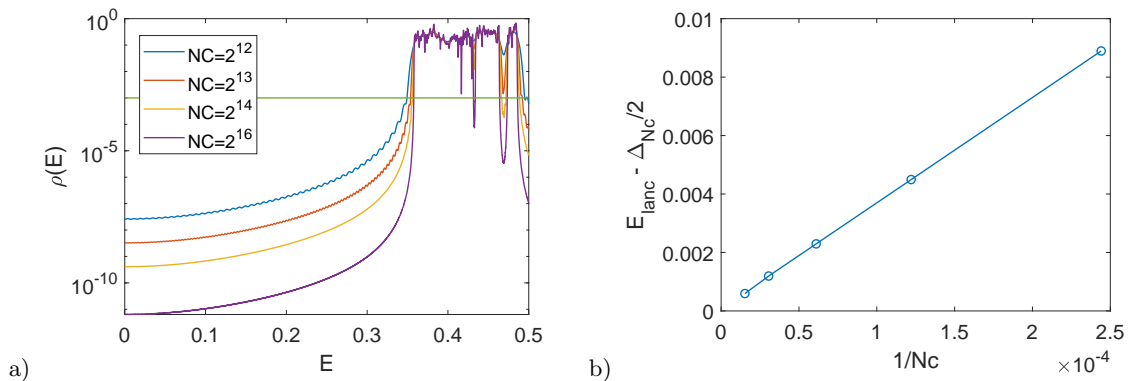


FIG. S9. **Extracting gap from KPM.** Simply choosing a threshold ε makes controlled approximation of gap size in KPM, and matches Lanczos result at large N_c limit. a) Density of state $\rho(E)$ in log scale. The green line marks the threshold $\varepsilon = 0.001$ used for determining Δ using KPM. The crossing point between the green line and $\rho(E)$ curve marks the upper end of the gap; while the lower end is determined similarly. b) The distance between half of the gap size determined by KPM with different cutoff ($\Delta_{N_c}/2$) to the lowest energy state determined by Lanczos (E_{lanc}) as a function of $1/N_c$. At large N_c limit, the difference approaches zero. For $N_c = 8192$ as an example, the KPM calculated Δ is accurate up to $1e-3$.

The size of the insulating gap centered at $E = 0$ is most efficiently calculated by extracting from DOS data in general. We first use the integral of density of state $n(E_1, E_2) = \int_{E_1}^{E_2} \rho(E) dE$ to find a positive energy E_{max} such that $n(0, E_{\text{max}})$ is smaller than $1/2L^2$. Then the range inside $(-E_{\text{max}}, E_{\text{max}})$ with $\rho(E)$ below the chosen threshold ε is the gap. The gap size Δ is the length of the range. Fig. S9 (a) illustrates how the gap is extracted.

In general, KPM involves a broadening of N_c , and hence the choice of threshold would depend on N_c . However, as long as $1 \gg \varepsilon \gg 1/N_c$, the numerical result of Δ from KPM with N_c provide an estimation approximately good up to $1/N_c$. We can also use Lanczos method by setting shift-invert factor σ to be KPM estimated gap edge to obtain the accurate energy of lowest energy state above $E = 0$, which is half of gap size. Fig. S9(b) demonstrate that the difference between KPM estimated gap size and the accurate Lanczos result. As N_c approaches infinity, the difference vanishes.

In fact, for W very close to the TI-CM transition, Lanczos method become more efficient than KPM to accurately determine Δ . Using shift-and-invert Lanczos with shift factor $\sigma = 0$, we can find the minimum of the lowest energy band. Here, we demonstrate a few examples of the gap size extracted using the Lanczos method with various L in supplement to the gap size extracted in the main text using KPM. As shown in Fig. S10(a-c) the result is essentially L independent.

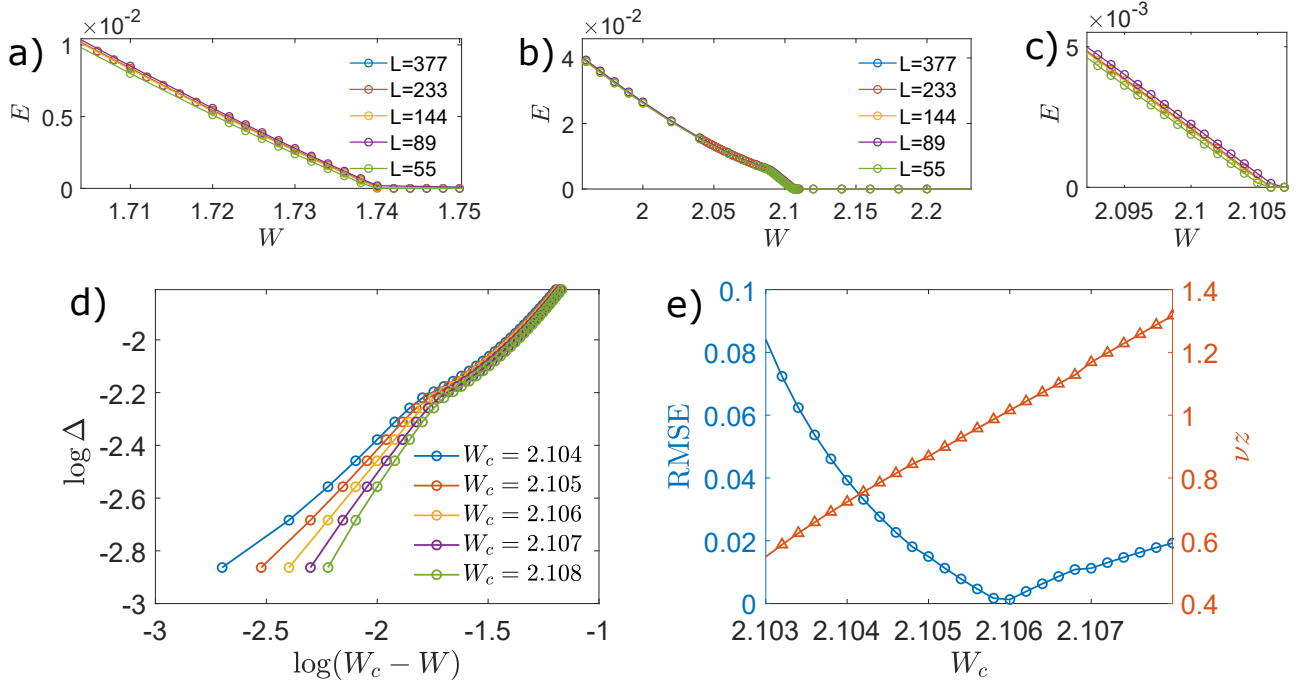


FIG. S10. **Vanishing of the spectral gap.** The gap size as a function of W for $M = 2.4$ (a) and $M = 3.0$ (b) with various system sizes L . (c) is a more zoomed in view of the $M = 3.0$ cut near the transition. In the second row, we show an example of how the combination of critical exponents νz is extracted from the spectral gap data. For a range of choices of W_c , we fit $\log \Delta$ against $\log(W_c - W)$ in the range when $W_c - W$ is under 0.015 and $\Delta > 0.001$ with a straight line. This data is shown in (d). Then we find the point where the root mean square error (RMSE) as shown in (e) of the linear fit is smallest as our best estimation of W_c , where the slope is then νz . For these results, our best estimate of $\nu z = 1.0 \pm 0.1$, with $W_c = 2.106 \pm 0.001$.

A. Critical Exponents at the topological insulator to metal phase transition

The critical exponents at the TI-to-CM phase transition can be extracted from the gap size, as quoted in the main text. Here we demonstrate this process in more detail. The Δ is calculated using Lanczos here, as we focus on a very small range of parameters, making Lanczos more efficient than KPM. We first estimate the critical quasiperiodic strength W_c from the density of states data. Near the estimated W_c , we consider a range of W_c and fit $\log(\Delta)$ over $\log(W_c - W)$, see Fig. S10(d). Then we identify the range of W that $\log \Delta$ is linear to $\log(W_c - W)$ and use least square fit. The best W_c is picked according to the goodness of linear fit, here quantified with root mean square error of the fit, see Fig. S10(e). The slope of the best fit is then νz , where $\Delta \sim (W_c - W)^{\nu z}$. For some cases this critical exponent is difficult to determine accurately because of the very fine phase diagram structure. For a few fixed M cuts shown in the main text, including $M = 2.4$, $M = 3.0$, $M = 3.3$ and $M = 3.8$, we find νz near 1.0. To be precise, for $M = 2.4$ we have $\nu z = 1.06 \pm 0.1$; for $M = 3.0$, $\nu z = 1.00 \pm 0.1$; for $M = 3.3$, $\nu z = 0.95 \pm 0.2$; for $M = 3.8$, $\nu z = 1.13 \pm 0.15$

B. Gapsizes and IPR

Another interesting feature is the boundary where our perturbation theory works to determine the gap size. The perturbation theory expression of the gap size gives a qualitatively correct prediction of the dependency of Δ on W , but the trend has a turning point at some finite W after which it no longer follows the perturbation theory. This is when the lowest band is no longer clearly the topological band, but mixes with other minibands nearby in energy due to the quasiperiodic potential. We show that the IPR of the lowest energy state changes dramatically at the same W where the gap size turns down as shown in Fig. S11.

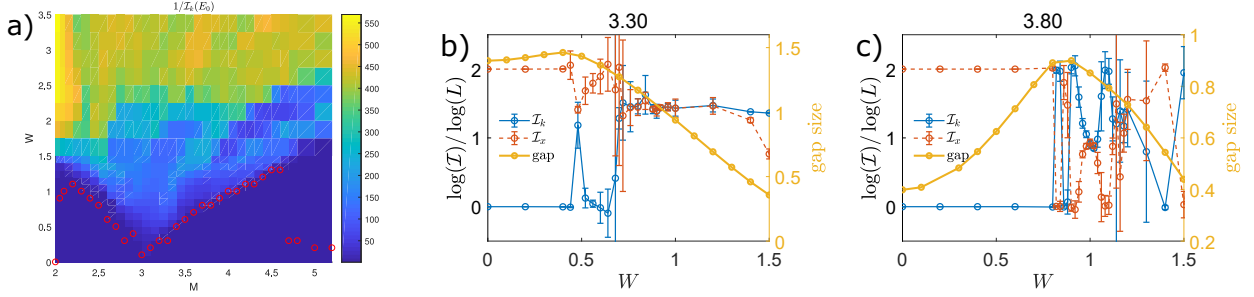


FIG. S11. **Properties of the IPR.** (a) Phase diagram of the momentum space IPR of the lowest eigenstates. The red circles mark where the gap size $\Delta(W)$ changes its trend from increasing to decreasing as determined by the location of the maximum in $\Delta'(W)$. The cuts $M = 3.3$ (b) and $M = 3.8$ (c) show the non-trivial L dependence of the IPR in both real and momentum space start to dramatically change when $\Delta(W)$ begins to turn downward.

VII. DISPERSION RELATION, EFFECTIVE MASS AND BERRY CURVATURE

In this section, we discuss our calculation of eigenstates and wavefunction, and hence dispersion relations and Berry curvature. The twisted boundary condition we implemented, $t_\mu \rightarrow t_\mu e^{i\theta_\mu/L}$, effectively shift momentum \mathbf{k} by $\boldsymbol{\theta}/L$. Without a quasiperiodic potential, this corresponds to a trivial folding of the original energy dispersion $E(\mathbf{k})$ such that $E_n(\boldsymbol{\theta}) = E(2\pi n/L + \boldsymbol{\theta}/L)$ for band n in the Brillouin zone defined by the supercell of size $L \times L$. In other words, $\boldsymbol{\theta}/L$ is the momentum in the mini-Brillouin zone of the superlattice made of $L \times L$ supercells.

To calculate eigenvalues and eigenstates in the interior of the spectrum, we use shift-invert Lanczos method with a shift factor σ . The spectrum is quite dense, so we can only accurately calculate eigenstates near a hard gap in order to have a definite band index for each state as we go to a large system size. Hence, σ is set to be the energy of the desired band edge next to the gap, which can be accurately located using DOS calculated with the KPM. Such a shift factor allows fast convergence to the first n eigenpairs with definite band indices, as long as the $E_i - \sigma$ for all $i \leq n$ is less than the gap size. The computational complexities in both memory and time are dominated by the Cholesky decomposition step in shift-invert Lanczos; whereas the iterations over the Lanczos Ritz vectors to increase the accuracy of eigenvalue and eigenvector calculation are fast. Hence we can always set the error tolerance of Lanczos to be at least 1/100 of the bandwidths of the folded bands.

A. Effective mass

One additional measure of band flatness to consider is the effective mass of the lowest band. We define an effective mass m^* by the expansion of the energy dispersion about its minimum $E(\mathbf{q}) = E(0) + \frac{1}{2m^*} \mathbf{q}^2 + \dots$. Numerically, m^* can be obtained through a quadratic fitting of $E_n(\boldsymbol{\theta})$ near $\boldsymbol{\theta} = \mathbf{0}$, where the n indexes the first, folded band above $E = 0$ (Fig. S12(a), for example). The effective mass can also be obtained, perturbatively, from the pole of the Green function $E_{\text{eff}}(\mathbf{q}) = \pm(\tilde{M} + \mathbf{q}^2/2m^*)$ where $m^* = \tilde{M}/(2\tilde{v}^2)$. At fourth-order, $m^* \sim 10^5(1/t)$ at $W = 3t$, indicating that the QP is flattening the topological bands; our numerics (the blue solid line in Fig. S12(b)) show this effect is even more drastic [Fig. S12(b)].

B. Edge states

For finite size calculation, (twisted) periodic boundary conditions eliminate all edge contribution and keep only the bulk. With open boundary conditions, the edge states can be observed, but we no longer have access to any $\boldsymbol{\theta} \neq \mathbf{0}$. Here, we use twisted boundary conditions only in the x -direction but open boundary conditions along the y -direction, so that we can see the dispersion of the edge states as we vary θ_x , as clearly demonstrated in Fig. S12(c).

C. Berry curvature of folded bands

To clearly characterize the topology of the folded bands, we calculate the Berry curvature directly by replacing momentum with the corresponding twist in the boundary condition. We divided the minizone of size $(2\pi/L)^2$ into a

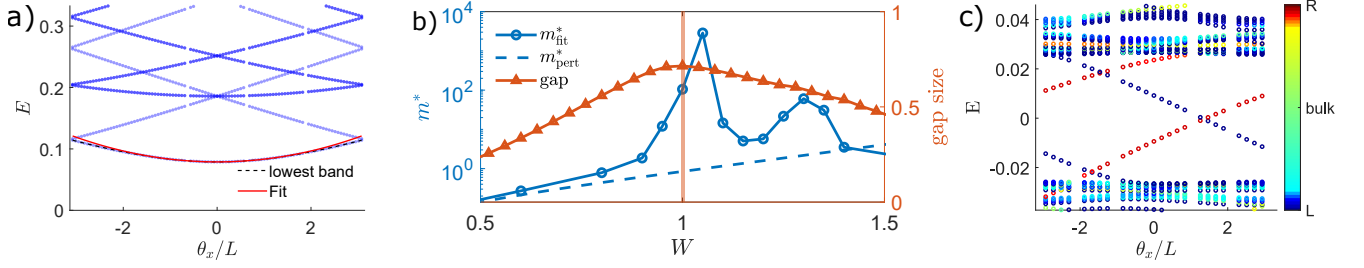


FIG. S12. **Twist Dispersions.** (a) Using the twist dispersion to obtain the effective mass m^* . The red curve is the quadratic fitting result to estimate m^* . The figure shows an example for $M = 4.0$, $W = 0.4$. (b) The effective mass obtained from fitting twist dispersion (m_{fit}^*) and from perturbation theory (m_{pert}^*), compared with gap size. The vertical line marks the W where we scrutinize flat topological bands in the main text. (c) Twist dispersion with open boundary conditions in the y direction and twisted boundary conditions in the x direction. The color corresponds to the location of the eigenstates along the y axis. The red and dark blue states in the bulk gap are the edge states.

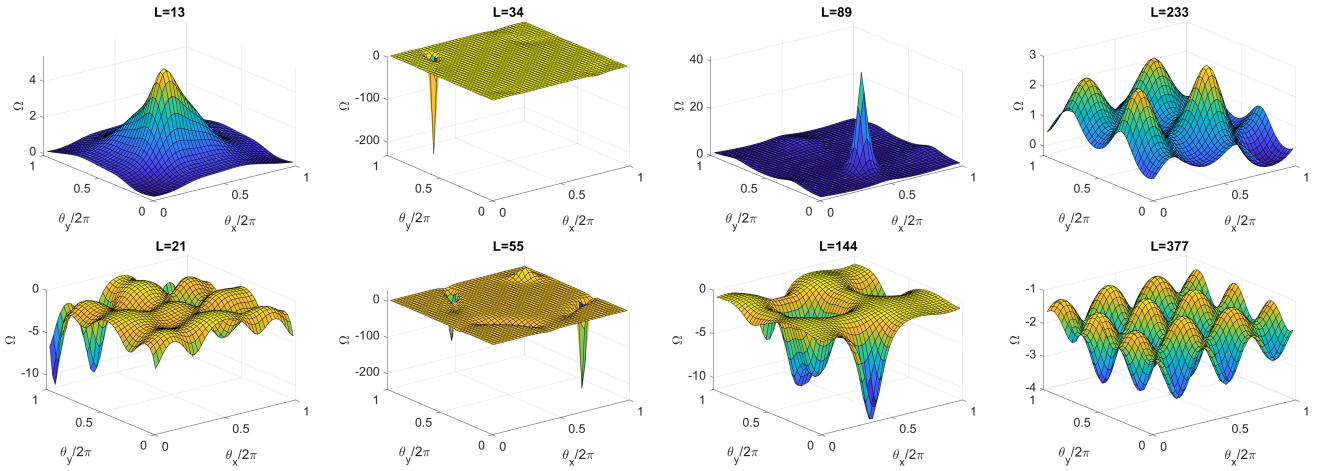


FIG. S13. **Berry curvatures.** The Berry curvature of the first band above the hard gap near $E = -0.5$. The samples shown are at $W = 1.01541$, $M = 4$ and $L = 377$, i.e. at its peak flatness (see Fig. 3(b) of the main text). The first row are system sizes in the sequence of $L = F_n$ with odd n , and the second row for even n . For $L = 55$ and $L = 89$, Berry curvature have clear peaks; while larger L 's see flatter Berry curvature.

grid of size $N_\theta \times N_\theta$ and evaluate $|\psi_n(\boldsymbol{\theta})\rangle$ with $N_\theta = 40$, where n is the band index starting from the band edge near a large gap. We follow Ref.^{S7} to obtain a manifestly gauge invariant calculation that always sum to integer valued Chern number. Note that the method requires the folded band to be gapped, so for example, at small W the calculation will fail. In addition, using Lanczos to compute the eigenstates leaves a phase ambiguity that may change for each calculation. When calculating the Berry curvature on any plaquette, each of the four $|\psi_n(\boldsymbol{\theta})\rangle$ surrounding the plaquette is used twice and it is important to use the same $|\psi_n(\boldsymbol{\theta})\rangle$ to be consistent. At the same time, calculations for different plaquettes are mutually independent, allowing for a straightforwardly parallelized calculation. This can be done either by saving $|\psi_n(\boldsymbol{\theta})\rangle$, or carefully setting the initial vector and the random seed for Lanczos.

Some examples of the calculated Berry curvature are shown in Fig. S13. For each case, the Berry curvature in a given band sums up to the Chern number, which is used for determining the topology of the band structure in main text. To characterize the fluctuations of Berry curvature, the normalized standard deviation of Berry curvature, defined as

$$\frac{\text{s.d.}(\Omega_n)}{\bar{\Omega}_n} \equiv \frac{\sqrt{\sum_{\boldsymbol{\theta}} \Omega_n(\boldsymbol{\theta})^2 - (\sum_{\boldsymbol{\theta}} \Omega_n(\boldsymbol{\theta}))^2}}{\sum_{\boldsymbol{\theta}} \Omega_n(\boldsymbol{\theta})} \quad (\text{S16})$$

is used.

-
- [S1] T. Ohtsuki and T. Ohtsuki, *Journal of the Physical Society of Japan* **85**, 123706 (2016).
 - [S2] J. Carrasquilla and R. G. Melko, *Nature Physics* **13**, 431 (2017).
 - [S3] P. Baldi, P. Sadowski, and D. Whiteson, *Nature communications* **5**, 4308 (2014).
 - [S4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, in *Advances in neural information processing systems* (2012) pp. 1097–1105.
 - [S5] C. M. Bishop, *Pattern recognition and machine learning* (springer, 2006).
 - [S6] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, arXiv preprint arXiv:1207.0580 (2012).
 - [S7] T. Fukui, Y. Hatsugai, and H. Suzuki, *Journal of the Physical Society of Japan* **74**, 1674 (2005).