

Origin of Mott Insulating Behavior and Superconductivity in Twisted Bilayer Graphene

Hoi Chun Po,¹ Liujun Zou,^{1,2} Ashvin Vishwanath,¹ and T. Senthil²

¹*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

²*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*



(Received 4 June 2018; revised manuscript received 13 August 2018; published 28 September 2018)

A remarkable recent experiment has observed Mott insulator and proximate superconductor phases in twisted bilayer graphene when electrons partly fill a nearly flat miniband that arises a “magic” twist angle. However, the nature of the Mott insulator, the origin of superconductivity, and an effective low-energy model remain to be determined. We propose a Mott insulator with intervalley coherence that spontaneously breaks $U(1)$ valley symmetry and describe a mechanism that selects this order over the competing magnetically ordered states favored by the Hund’s coupling. We also identify symmetry-related features of the nearly flat band that are key to understanding the strong correlation physics and constrain any tight-binding description. First, although the charge density is concentrated on the triangular-lattice sites of the moiré pattern, the Wannier states of the tight-binding model must be centered on different sites which form a honeycomb lattice. Next, spatially localizing electrons derived from the nearly flat band necessarily breaks valley and other symmetries within any mean-field treatment, which is suggestive of a valley-ordered Mott state, and also dictates that additional symmetry breaking is present to remove symmetry-enforced band contacts. Tight-binding models describing the nearly flat miniband are derived, which highlight the importance of further neighbor hopping and interactions. We discuss consequences of this picture for superconducting states obtained on doping the valley-ordered Mott insulator. We show how important features of the experimental phenomenology may be explained and suggest a number of further experiments for the future. We also describe a model for correlated states in trilayer graphene heterostructures and contrast it with the bilayer case.

DOI: [10.1103/PhysRevX.8.031089](https://doi.org/10.1103/PhysRevX.8.031089)

Subject Areas: Condensed Matter Physics

I. INTRODUCTION

Superconductivity occurs proximate to a Mott insulator in a few materials. The most famous are the cuprate high- T_c materials [1]; others include layered organic materials [2], certain fullerene superconductors [3], and some iron-based superconductors [4]. In these systems, there is a complex and often poorly understood relationship between the Mott insulator and the superconductor, which has spurred tremendous research activity in condensed matter physics in the past 30 years. Very recently, in some remarkable developments, both Mott insulating behavior and proximate superconductivity have been observed in a very different platform: two layers of graphene that are rotated by a small angle relative to each other [5,6].

Twisted bilayer graphene (TBG) structures have been studied intensely in the past few years [7–18]. The charge density is concentrated on a moiré pattern which forms

(at least approximately) a triangular lattice [8,9,11,14,15]. The electronic states near each valley of each graphene monolayer hybridize with the corresponding states from the other monolayer. When the twisting angle is close to certain discrete values known as the magic angles, theoretical calculations show that there are two nearly flat bands (per valley per spin) that form in the middle of the full spectrum that are separated from other bands [12]. When the carrier density is such that the chemical potential lies within these nearly flat bands, interaction effects are expected to be enhanced. At a filling of $1/4$ or $3/4$ (denoted $\nu = -2$ and $+2$, respectively, with full band filling denoted $\nu = +4$) of these nearly flat bands, Ref. [5] reports insulating behavior at very low temperatures. At such fillings, band insulation is forbidden, which leads naturally to the expectation that these are correlation-driven (Mott) insulators. Doping the Mott insulator at $1/4$ band filling—with either electrons or holes—reveals superconductivity at low T [6].

A number of other striking observations are made in Refs. [5,6] about both the Mott insulator and the superconductor from transport studies in a magnetic field. The Mott insulation is suppressed through the Zeeman coupling of the magnetic field at a low scale of approximately $5T$ —roughly the same scale as the activation gap inferred from

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

zero field resistivity. Quantum oscillations are seen in the hole-doped state with a frequency set (in the hole-doped side) by the density deviation from the Mott insulator. The degeneracy of the corresponding Landau levels is half of what might be expected from the spin and valley degrees of freedom that characterize electrons in graphene. The superconductivity occurs at temperatures that are high given the low density of charge carriers. Just like in other doped Mott insulators, there is a dome of superconductivity with T_c reaching an “optimal” value at a finite doping. The superconductivity is readily suppressed in accessible magnetic fields—both perpendicular and parallel to the plane.

The observation of these classic phenomena in graphene gives new hope for theoretical progress in addressing old questions on Mott physics and its relationship to superconductivity. They also raise a number of questions. What is the nature of the insulators seen at these fractional fillings? How are they related to the observed superconductivity? On the theoretical side, what is an appropriate model that captures the essential physics of this system?

In this paper, we make a start on addressing these questions. The two nearly flat bands for each valley found in the band structure have Dirac crossings at the moiré K points (but not Γ). We argue that these Dirac crossings are protected by symmetries of the TBG system. We show that this protection precludes finding a real-space representation of the nearly flat bands in terms of Wannier orbitals localized at the triangular moiré sites, in contrast to natural expectations. Thus, a suitable real-space lattice model is necessarily different from a correlated triangular-lattice model with two orbitals (corresponding to the two valleys) per site. We instead show that a representation that is faithful to the Dirac crossings is possible on a honeycomb lattice with two orbitals per site, but even this representation has some subtleties. First, one cannot implement a natural representation of all the important symmetries in the problem, which include spatial symmetries, time reversal, and a separate conservation of electrons of each valley [which we dub $U_v(1)$]. Second, since the charge density is concentrated at the moiré triangular sites (which appear as the centers of the honeycomb plaquettes), the dominant interaction is not an on-site Coulomb repulsion on the honeycomb sites. Rather, it is a “cluster-charging energy” that favors having a fixed number of electrons in each honeycomb plaquette, which makes this model potentially rather different from more standard Hubbard models with on-site interactions.

Armed with this understanding of the microscopics, we can begin to address the experimental phenomenology. We propose that this system spontaneously breaks the valley $U_v(1)$ symmetry—we call the resulting order “intervalley coherent” (IVC). We discuss microscopic mechanisms that stabilize IVC symmetry breaking. We point out that, even when the IVC is fully polarized, it cannot, by itself, lead to

a fully insulating state but rather leads to a Dirac semimetal. The development of a true insulator needs a further symmetry breaking (or some more exotic mechanism) to gap out the Dirac points. We show that, once the valley symmetry is spontaneously broken, the physics at lower energy scales can be straightforwardly formulated in terms of a real-space honeycomb lattice tight-binding model with a dominant cluster-charging interaction and other weaker interactions. We outline a number of different possible routes in which a true insulator [19] can be obtained in such an IVC-ordered system. A concrete example is a state that further breaks C_3 rotational symmetry. We show how doping this specific IVC insulator can explain the phenomenology of the experiments. We present a possible pairing mechanism due to an attractive interaction mediated by Goldstone fluctuations of the IVC phase. We describe and contrast features of other distinct routes by which the IVC state can become a true insulator at $\nu = \pm 2$. We propose a number of future experiments that can distinguish between the different routes through which an IVC can become a true insulator.

In addition, very recently, a heterostructure of ABC -stacked trilayer graphene and boron nitride (TLG/hBN), which also forms a triangular moiré superlattice even at a zero twist angle, was studied [20]. This system also features nearly flat bands that are separated from the rest of the spectrum. Correlated Mott insulating states are seen at fractional fillings of the nearly flat band. Unlike the TBG, here the nearly flat band has no Dirac crossing, which makes the details of the two systems potentially rather different. In particular, the nearly flat band of the TLG/hBN can be modeled in real space as a triangular-lattice model with two orbitals per site, supplemented with interactions. However, the hopping matrix elements are, in general, complex (but subjected to some symmetry restrictions). We describe some properties of this model and suggest that this system offers a good possibility to realize novel kinds of quantum spin-orbital liquid states.

II. ELECTRONIC STRUCTURE OF TWISTED BILAYER GRAPHENE: GENERAL CONSIDERATIONS

A. Setup

First, to establish the notation, let us consider a graphene monolayer, with lattice vectors \mathbf{A}_1 and \mathbf{A}_2 (see Appendix A for details). The honeycomb lattice sites are located at $\mathbf{r}_{1,2} = \frac{1}{2}(\mathbf{A}_1 + \mathbf{A}_2) \mp \frac{1}{6}(\mathbf{A}_1 - \mathbf{A}_2)$, where the $-$ and $+$ signs are, respectively, for the sites labeled by 1 and 2.

Now consider the moiré pattern generated in the twisted bilayer problem. For concreteness, imagine we begin with a pair of perfectly aligned graphene sheets, and consider twisting the top layer by an angle θ relative to the bottom one. Now we have two pairs of reciprocal lattice vectors, the original ones \mathbf{B}_a and $\mathbf{B}'_a = R_\theta \mathbf{B}_a$. Like Refs. [7,12], we

approximate the moiré superlattice by the relative wave vectors, leading to a periodic structure with reciprocal lattice vectors $\mathbf{b}_a = \mathbf{B}'_a - \mathbf{B}_a = (R_\theta - I)\mathbf{B}_a$. For small θ , we can approximate this by $\mathbf{b}_a = \theta \hat{z} \times \mathbf{B}_a$. Thus, the moiré pattern also has triangular-lattice symmetry, but it is rotated by 90° and has a much larger lattice constant. Note that, in this dominant harmonic approximation, questions of commensuration or incommensuration are avoided, since no comparison is made between the other sets of harmonics of the moiré superlattice that are not commensurate to the dominant one.

Let us now briefly review the low-energy electronic structure of monolayer graphene to set the notation. Parameterizing $\mathbf{k} = \sum_{j=1,2} g_j \mathbf{B}_j$ for $g_j \in (-1/2, 1/2]$, the Bloch Hamiltonian for the nearest-neighbor hopping model is

$$H(\mathbf{g}) = t(e^{-i(2\pi/3)(g_1-g_2)} + e^{-i(2\pi/3)(g_1+2g_2)} + e^{i(2\pi/3)(2g_1+g_2)})\sigma^- + \text{H.c.} \quad (1)$$

Note that, as a general property of our present choice of Fourier transform, the Bloch Hamiltonian is not manifestly periodic in the Brillouin zone (BZ). Rather, for any reciprocal lattice vector \mathbf{B} , we have $H(\mathbf{k} + \mathbf{B}) = \eta_{\mathbf{B}} H(\mathbf{k}) \eta_{\mathbf{B}}^\dagger$, where $\eta_{\mathbf{B}} = \text{diag}(e^{-i\mathbf{B} \cdot \mathbf{r}_a})$. One can now pass to a continuum limit near each Dirac point $\mathbf{K} = (2\mathbf{B}_1 - \mathbf{B}_2)/3$ and $-\mathbf{K}$. We then have the linearized Hamiltonian

$$\begin{aligned} H(\mathbf{K} + \mathbf{k}) &= -\hbar v_F \mathbf{k} \cdot \boldsymbol{\sigma}; \\ H(-\mathbf{K} + \mathbf{k}) &= \hbar v_F \mathbf{k} \cdot \boldsymbol{\sigma}^*, \end{aligned} \quad (2)$$

where $\hbar v_F = \sqrt{3}ta/2$ in our simple nearest-neighbor model. Since $H(\mathbf{q})$ is not periodic in the BZ, expanding about the other equivalent Dirac points leads to a slightly modified form of the Hamiltonian (due to conjugation by some η). In second quantized notation, we can write the continuum Hamiltonian:

$$\begin{aligned} \hat{h}_+ &= -\hbar v_F \int d^{-2}\mathbf{k} \hat{\psi}_{+\mathbf{k}}^\dagger (\mathbf{k} \cdot \boldsymbol{\sigma}) \hat{\psi}_{+\mathbf{k}}; \\ \hat{h}_- &= +\hbar v_F \int d^{-2}\mathbf{k}' \hat{\psi}_{-\mathbf{k}'}^\dagger (\mathbf{k}' \cdot \boldsymbol{\sigma}^*) \hat{\psi}_{-\mathbf{k}'}, \end{aligned} \quad (3)$$

where the momentum integration is understood to be implemented near the Dirac point momentum by introducing a cutoff $|\mathbf{k}| \leq \Lambda$ and $d^{-2}\mathbf{k} = [(dk_x dk_y)/(2\pi)^2]$. The symmetry implementation on the continuum fields is tabulated in Appendix A. For example, C_3 -rotation symmetry is represented as $\hat{C}_3 \hat{\psi}_{\pm\mu\mathbf{k}} \hat{C}_3^{-1} = e^{\mp i(2\pi/3)\sigma_3} \hat{\psi}_{\pm\mu; C_3\mathbf{k}}$, where $\mu = t, b$ is the layer index.

Next, we couple the degrees of freedom in the two layers of graphene and arrive at a continuum theory for the twisted bilayer graphene system [7,12]. First, we note that the

rotated Bloch Hamiltonian of a monolayer can then be identified as $H_\varphi(\mathbf{k}) = H(R_{-\varphi}\mathbf{k})$. Linearizing about the rotated K point $R_\varphi\mathbf{K}$ obtains $\hat{h}_\pm(\varphi)$, with $\hat{h}_\pm(\varphi)$ defined by replacing $\boldsymbol{\sigma} \mapsto \boldsymbol{\sigma}_\varphi$ in \hat{h}_\pm , where

$$\boldsymbol{\sigma}_\varphi \equiv e^{-i\varphi\sigma_3/2} \boldsymbol{\sigma} e^{i\varphi\sigma_3/2}. \quad (4)$$

Focusing on a single valley, say, \mathbf{K} , the continuum theory [7,12] of the twisted bilayer graphene system is described by the Hamiltonian $\hat{H}_{\text{cont}} = \hat{H}_{\text{Dirac}} + \hat{H}_T$, where

$$\begin{aligned} \hat{H}_{\text{Dirac}} &= \hat{h}_+(\varphi_t) + \hat{h}_+(\varphi_b); \\ \hat{H}_T &= \int_0^\Lambda d^2\mathbf{k} \hat{\psi}_{+b;\mathbf{k}}^\dagger T_{q_1} \hat{\psi}_{+t;\mathbf{k}+\mathbf{q}_1} + \text{H.c.} \\ &+ \text{symmetry-related terms,} \end{aligned} \quad (5)$$

t and b , respectively, denote the top and bottom layers, and we set $\varphi_t = \theta/2$ and $\varphi_b = -\theta/2$. Here, we introduce $\mathbf{q}_1 \equiv R_{-\theta/2}\mathbf{K} - R_{\theta/2}\mathbf{K}$, which characterizes the momentum transfer between the electronic degrees of freedom of the two layers [12]. Assuming T_{q_1} is real, the symmetries of the system, which we discuss in the following subsection, constrain T_{q_1} [21] to take the form $T_{q_1} = w_0 - w_1\sigma_1$, where $w_{0,1}$ are real parameters [22]. Similarly, one can generate the omitted symmetry-related terms by applying symmetries on T_{q_1} .

B. Symmetries of the continuum theory

Let us discuss how the symmetries of the graphene monolayer are modified in the twisted bilayer problem within the dominant harmonic approximation. We see that, in addition to the moiré translation symmetry, we have C_6 rotation, time reversal, and a mirror symmetry. Furthermore, a $U(1)$ valley symmetry that allows us to assign valley charge to the electrons emerges in the low-energy limit. The generator of C_6 rotation and time reversal flips the valley charge, while reflection leaves it invariant.

Microscopically, the stacking pattern of the two layers can be specified as follows [12,21,23]: First, we align the two layers perfectly in a site-on-site manner, corresponding to the ‘‘AA-stacking’’ pattern, and then rotate the top and bottom layers about a hexagon center by angles $\theta/2$ and $-\theta/2$ clockwise, respectively; second, we shift the top layer by a vector \mathbf{d} parallel to the plane. For generic values of θ and \mathbf{d} , one expects that almost all of the spatial symmetries are broken.

However, within the dominant harmonic approximation, it is found that, on top of possessing moiré lattice translation symmetries, the effective theory is also insensitive to \mathbf{d} [12]. This finding implies that, given θ , the effective theory will at least possess all the exact symmetries for any choice of \mathbf{d} . A particularly convenient choice is when we take $\mathbf{d} = \mathbf{0}$. In this case, we can infer all the point-group

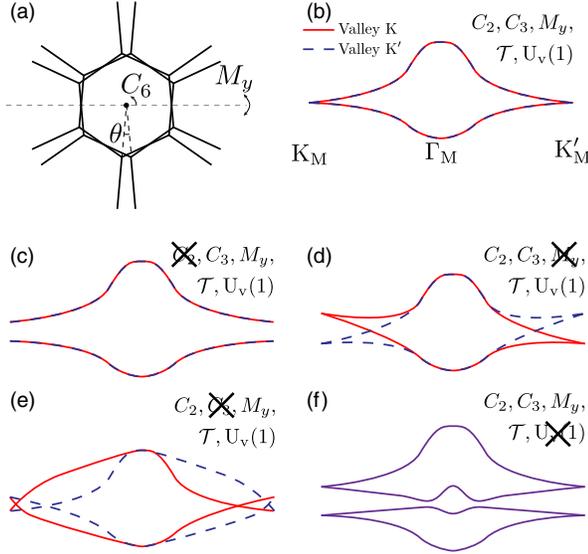


FIG. 1. Effective symmetries and constraints on band structures. (a) The effective symmetries of the twisted bilayer graphene system can be inferred by inspecting the point-group symmetry of a hexagon center in real space, taken to be the rotation axis of the layers. (b) Schematic band structure along a high-symmetry path in the moiré Brillouin zone. (c)–(f) Effect of symmetry breaking. (c) Breaking the C_2 rotation gaps out the Dirac points. (d) An external perpendicular electric field breaks the mirror M_y symmetry, which only modifies the energetics but cannot open a band gap at charge neutrality [7]. (e) When C_3 rotation is broken, but the combined symmetry of twofold rotation C_2 and time reversal \mathcal{T} is preserved, the Dirac points remain protected, although unpinned from K_M and K'_M . (f) When valley conservation $U_v(1)$ symmetry is broken, one can no longer label the bands using their valley index. The gaplessness at charge neutrality is no longer symmetry required, although, depending on detailed energetics, there can still be remnant Dirac points. In contrast, at the quarter filling relevant for the observed Mott physics, there are necessarily Dirac points present in this case. The other symmetry-breaking patterns listed above also do not open band gaps at quarter filling.

symmetries of the system by focusing on the center of the hexagons [Fig. 1(a)]. Aside from the rotational symmetries generated by the sixfold rotation C_6 , we see that there is an additional mirror plane M_y , which, in fact, combines a mirror perpendicular to the 2D plane together with an in-plane mirror which flips the top and bottom layers. Strictly speaking, this process leads to a twofold rotation in 3D space, but when restricting our attention to a 2D system it acts as a mirror.

In summary, the effective theory will at least have the following spatial symmetries: lattice translations, a sixfold rotation, and a mirror, which allows one to uniquely identify its wallpaper group (i.e., 2D space group) as $p6mm$ (numbered 17 in Ref. [24]). Having identified the symmetries of the system, one can derive the model following a phenomenological approach by systematically incorporating all symmetry-allowed terms with some cutoff

[21]. We tabulate the explicit symmetry transformation of the electron operators in Appendix A.

In the effective theory, the degrees of freedom arising from the microscopic K and K' points are also essentially decoupled [7,12], which is because, for a small twist angle satisfying $|\sin\theta| \ll 1$, we have $|\mathbf{b}_a| \ll |\mathbf{K}|$, and therefore the coupling between the K and K' points is a very high-order process. Hence, on top of the usual electron-charge conservation, the effective theory has an additional, emergent $U_v(1)$ conservation corresponding to the independent conservation of charge in the two valleys K and K' . Henceforth, we refer to this as “valley conservation.” The valley-charge operator is given by

$$\hat{I}_z = \int \mathrm{d}^2\mathbf{k} (\hat{\psi}_{+\mathbf{k}}^\dagger \hat{\psi}_{+\mathbf{k}} - \hat{\psi}_{-\mathbf{k}}^\dagger \hat{\psi}_{-\mathbf{k}}). \quad (6)$$

Note that, as time reversal \mathcal{T} interchanges the K and K' valleys, it is not a symmetry of a single valley. Similarly, C_6 also interchanges the two valleys; thus,

$$\hat{T} \hat{I}_z \hat{T}^{-1} = -\hat{I}_z, \quad (7)$$

$$\hat{C}_6 \hat{I}_z \hat{C}_6^{-1} = -\hat{I}_z, \quad (8)$$

$$\hat{M}_y \hat{I}_z \hat{M}_y^{-1} = +\hat{I}_z. \quad (9)$$

We then see that their combined symmetry $C_6\mathcal{T}$ is a symmetry in the single-valley problem. In fact, one can check that the symmetry of the single-valley problem is described by the magnetic space group 183.188 (BNS notation; Ref. [25]). We tabulate the generating symmetries in Table I.

TABLE I. Summary of key effective symmetries. From top to bottom, the listed symmetries are time reversal, moiré lattice translation, a perpendicular 2D mirror, threefold rotation, combined symmetry of twofold rotation and time reversal, and valley $U_v(1)$ conservation. For any symmetry g , it either commutes ($\eta_g = +1$) or anticommutes ($\eta_g = -1$) with the valley-charge operator \hat{I}_z .

Symmetry	η_g	Remarks
\mathcal{T}	-1	Broken by valley polarization $\langle I_z \rangle \neq 0$
t_a	+1	...
M_y	+1	Broken by perpendicular electric field
C_3	+1	Pins Dirac points to K_M and K'_M
$C_2\mathcal{T}$	+1	Protects the local stability of the Dirac points
$\exp(-i\theta\hat{I}_z)$	+1	...

III. LOW-ENERGY THEORY: TWO-BAND PROJECTION

Formally, the continuum effective theory [7,12] we describe corresponds to an infinite-band problem for each valley. However, near charge neutrality it is found that, for some range of angles, the moiré potential can induce additional band gaps at a certain commensurate filling of the moiré unit cell. The “nearly flat bands” identified near the magic angle correspond to two bands per valley, separated from all other bands by band gaps, that form Dirac points at the K_M and K'_M points in the moiré BZ. These bands correspond to the relevant degrees of freedom for the correlated states observed in Refs. [5,6], and, in the following, we focus our attention to the properties of these bands. In this section, we always focus on a single valley, say, that corresponding to the K point in the microscopic description.

A. Symmetry-enforced band contacts

A salient feature of the effective theory is the presence of Dirac points at charge neutrality, whose velocity is strongly renormalized and approaches zero near the magic angle [7,12]. The stability of the Dirac points can be understood from symmetries: For a single valley, K_M is symmetric under the (magnetic) point group generated by $C_6\mathcal{T}$. In particular, $(C_6\mathcal{T})^2 = C_3$, and therefore we can label each band at K_M by its C_3 eigenvalue, which takes a value in $\{1, \omega = e^{-i(2\pi/3)}, \omega^*\}$. In particular, a band with C_3 eigenvalue ω is necessarily degenerate with another with eigenvalue ω^* , as $(C_6\mathcal{T})^2 \neq 1$ on these bands and enforces a Kramers-like degeneracy. The observed Dirac points at charge neutrality correspond precisely to this two-dimensional representation [Fig. 1(b)].

While we allude to the presence of $C_6\mathcal{T}$ symmetry in explaining the stability of the Dirac points, these band contacts are actually locally stable so long as the symmetry $(C_6\mathcal{T})^3 = C_2\mathcal{T}$ is kept. This stability can be reasoned by noting that $C_2\mathcal{T}$ quantizes the Berry phase along any closed loop to $0, \pi \bmod 2\pi$, and a Dirac point corresponds precisely to the case of a nontrivial π Berry phase [26].

Let us now consider the effect of breaking the various symmetries (spontaneously or explicitly) in the system. First, as $C_2\mathcal{T}$ is crucial in protecting the local stability of the Dirac points, once it is broken, the Dirac points can be immediately gapped out [Fig. 1(c)]. However, as long as $C_2\mathcal{T}$ symmetry is preserved, a small breaking of any other point-group symmetries will not lead to a gapped band structure at charge neutrality. For instance, the mirror M_y maps K_M to K'_M , and its presence ensures only that the two inequivalent Dirac points are at the same energy. Therefore, even when a perpendicular electric field is externally applied such that M_y is broken, as in the setup in Refs. [5,6], it can only induce an energy difference between the two Dirac points [7] [Fig. 1(d)]. This result should be

contrasted with the case of Bernal-stacked bilayer graphene, whose quadratic band touching at charge neutrality can be gapped by an external electric field [27]. Alternatively, if C_3 symmetry is broken, the Dirac points are unpinned from K_M and K'_M [Fig. 1(e)]. As such, for a sufficiently strong C_3 breaking, a band gap might open at charge neutrality if the Dirac points could meet their oppositely charged partners and annihilate. (Though, as we argue later, this situation is impossible without further symmetry breaking [28].)

Now consider the case when valley conservation is spontaneously broken by an IVC, i.e., the valley charge \hat{I}_z is no longer conserved. In this case, we should first consider the full four-band problem consisting of both valleys. At, say, K_m , the combined symmetry of $M_y\mathcal{T}$ ensures that the Dirac points from the two valleys are degenerate. While such degeneracy is lifted in the presence of an IVC, as long as the remaining symmetries are all intact, we can only split the degeneracy according to $4 = 2 \oplus 2$ [Fig. 1(f)]. This remaining twofold degeneracy rules out an interpretation of the experimentally observed Mott insulator as a Slater insulator with a spatial-symmetry-respecting (ferro) IVC incorporated at the Hartree-Fock level. Instead, one must either introduce additional symmetry breaking, say, that of C_3 or lattice translations, or consider an IVC which also breaks some additional spatial symmetries. We elaborate on these points in Sec. VI. We also note that an essentially identical argument holds for the case of spontaneously ferromagnetic order leading to fully spin-polarized bands of I_z ordering. In this way, it connects to the quarter-filled Mott insulator we are interested in.

B. Triangular versus honeycomb lattice

A conventional route for understanding the correlated states observed in Refs. [5,6] is to first build a real-space tight-binding model for the relevant bands and then incorporate short-range interactions to arrive at, say, a Fermi-Hubbard model. Typically, the orbital degrees of freedom involved in the tight-binding model can be identified from either applying chemistry insight or, more systematically, studying the projected density of states for the relevant bands, both of which are inapplicable to the current moiré potential problem; furthermore, an understanding of the structure of the wave functions is required. Indeed, as is noted in Refs. [9,23,29], the local density states for the flat bands are well localized to the AA regions of the moiré pattern, which form a triangular lattice. This theoretical prediction is also confirmed experimentally [8,11,14]. Based on this observation, it is natural to consider a real-space model starting from effective orbitals centered at the AA sites, which corresponds to a tight-binding model defined on the triangular lattice [5,6]. In addition, by treating the two valleys separately, one envisions a model with two orbitals localized to each of the triangular sites (i.e., AA regions of the moiré pattern).

From symmetry representations, however, we can immediately rule out such a model. This observation can be readily inferred from the computed band structure [7,12,21,23] [Fig. 4(j)]: While the two bands are non-degenerate at Γ , as we have explained, they form symmetry-protected Dirac points at K_M and K'_M . Using such a pattern of degeneracies, one can infer the possible symmetry representations at these high-symmetry points, and from a real-space analysis [30–32] one finds that a triangular-lattice model always leads to the same symmetry representation at all three of the high-symmetry points; i.e., they are either all nondegenerate or are all Dirac points. This result is inconsistent with the observed pattern of degeneracies, which rules out all triangular-lattice models.

In fact, the degeneracy pattern described is familiar—it corresponds exactly to the monolayer graphene problem. One can further check that this is the only possible solution using the methods described in Refs. [30,32]. Symmetrywise, this correspondence implies that any tight-binding model must correspond to orbitals forming a honeycomb lattice. To reconcile with the predicted and observed local density of states [8,9,11,14,23], however, these orbitals must have nontrivial shapes: Although each orbital is centered at a honeycomb site, which corresponds to the AB/BA region of the moiré pattern, the weight of the orbitals is mainly localized to the AA sites. Therefore, we expect the shape of the orbitals to resemble a (three-lobed) fidget spinner [Figs. 3(a) and 3(b)].

C. Obstructions to symmetric Wannier states

Our symmetry analysis suggests that one should model the system by orbitals centered at the AB/BA regions of the moiré potential, which form a honeycomb lattice. A minimal tight-binding model of a single valley would then be

$$\hat{H}_{\text{minimal}} = \sum_{\rho_i} t_{\rho_i} e^{i\phi_{\rho_i}} \hat{c}_r^\dagger \hat{c}_{r+\rho_i} + \text{H.c.}, \quad (10)$$

where \hat{c}_r^\dagger is an electron-creation operator centered at a honeycomb site (for a single valley) and ρ_i connects two i th nearest-neighbor sites. Given that this equation describes a single valley which breaks time-reversal symmetry, the hoppings are, in general, complex unless constrained by a space-group symmetry.

A pedestrian approach involves optimizing the parameters $\{t_{\rho_i}, \phi_{\rho_i}\}$ to reproduce the energy eigenvalues obtained from the continuum description. Would this be a good starting point for building up a real-space effective model upon which we can incorporate interaction terms? Contrary to usual expectations, we argue that such an approach has a serious flaw in capturing certain essential properties. Specifically, we show that, while the energy eigenvalues may be well approximated, the topology of the resulting Bloch wave functions will necessarily be incorrect. This result has important dynamical consequences,

relating to the stability of band contacts under different symmetry assumptions, which, in turn, dictate whether an insulator will result at particular fillings. In particular, we find two symmetry obstructions to deriving a single-valley tight-binding model. The first concerns the symmetry representations of M_y : We find that the two bands have opposite M_y eigenvalues of ± 1 , whereas, from a real-space analysis [30–32], one can show that the two bands in a tight-binding model must have the same M_y eigenvalue.

There is a second, more serious, obstruction: Aside from a quantized Berry phase of π for any closed loop encircling a single Dirac point, one can further define a \mathbb{Z} -valued winding number [28]. In contrast to the conventional case of graphene, the two inequivalent Dirac points in the single-valley model are known to have the *same* winding number [5,28,33]. As the net winding number of the Dirac points arising in any two-band tight-binding model would necessarily be zero, we can then conclude that there is an obstruction for a symmetric real-space description; i.e., there is an obstruction for constructing localized Wannier functions that reproduces just the two bands of interest, represents $C_2\mathcal{T}$ naturally, and preserves valley quantum numbers. A more detailed description of this obstruction, by relating it to the anomalous surface state of a three-dimensional topological phase, is contained in Appendix B. Essentially, this argument invokes three key ingredients: (i) a two-band model, (ii) $C_2\mathcal{T}$ symmetry, and (iii) net winding of the Dirac points in the Brillouin zone.

We return to the question of tight-binding models in Sec. VIII, but, for the discussion below, we work directly in the momentum space in the manifold of states spanned by the nearly flat bands.

IV. INTERVALLEY COHERENT ORDER: PHENOMENOLOGICAL MOTIVATION

We first describe some important clues from experiments [5,6] on the nature of both the Mott state and the superconductor. We begin with the observation that—at optimal doping—an in-plane magnetic field suppresses the superconductivity when the Zeeman energy scale is of the order of the zero field T_c . This observation shows that the superconductor has spin-singlet pairing. Upon hole doping the $\nu = -2$ insulator, quantum oscillations are seen with a frequency set by the density of doped holes in perpendicular B fields exceeding approximately $1T$. This result tells us that the “normal” metallic state and the superconductor that emerges from it should be regarded as doped Mott insulators: The charge carriers that are available to form the normal state Fermi surface or the superconducting condensate are the doped holes. Thus, the hole-doped superconductor retains information about the Mott insulator. In contrast, electron doping this Mott insulator leads very quickly to quantum oscillations with a high frequency that is set by the deviation of the charge density from the charge neutrality point ($\nu = 0$). This result

may indicate a first-order transition between a metal and Mott insulator on the electron-doped side. It will be important to search for signs of hysteresis in transport experiments as the gate voltage is tuned. As the superconductor is better developed and characterized on the hole-doped side, we restrict our attention to hole doping from now on.

A further important clue from the quantum oscillation data is that the Landau levels (per flux quantum) are twofold degenerate, whereas one would expect fourfold degeneracy coming from the spin and valley degeneracy. The doped holes have thus lost either their spin or valley quantum numbers (or some combination thereof). Losing spin makes it hard to reconcile with spin-singlet pairing that can be suppressed with a Zeeman field. Thus, we propose instead that the valley quantum number is lost. The simplest option [34] then is that the valley quantum number is frozen due to symmetry breaking, i.e., $\langle \mathbf{I} \rangle \neq 0$. Here, we may define \mathbf{I} using the electron operators $\hat{c}(\mathbf{k})$ for the nearly flat-band states:

$$\mathbf{I} = \sum_{a,b,n,\alpha,\mathbf{k}} \hat{c}_{a\alpha}(\mathbf{k})^\dagger \boldsymbol{\tau}_{ab} \hat{c}_{b\alpha}(\mathbf{k}), \quad (11)$$

where $a, b = \pm$ correspond to the valley index, α is the spin index, n labels the two bands for each valley, and $\boldsymbol{\tau}$ denotes the standard Pauli matrices.

A nonzero expectation value for I^z breaks time-reversal symmetry, which leads to a sharp finite-temperature phase transition in $2d$ and would likely have been detected in the experiments. Given the absence of any evidence of a sharp finite-temperature transition, we propose that the ordering is in the pseudospin xy plane. These phenomenological considerations therefore lead us to an IVC-ordered state.

We note that, for IVC ordering to be useful to explain the quantum oscillations, it has to occur at a scale that is large compared to the scales set by the magnetic field. Specifically, the band splitting due to IVC ordering must be bigger than the Landau level spacing approximately 15–30 K at the biggest fields used (of the order of $5T$), which means that the IVC order is much more robust than the superconductivity and occurs at a higher temperature scale. We further need the IVC order to be present already in the Mott insulator, so that upon doping it can impact the quantum oscillations.

Thus, our view is that the first thing that happens as the sample is cooled from a high temperature is IVC ordering. This order then sets the stage for other phenomena to occur at lower temperatures (the Mott insulation or the superconductivity).

V. SIMPLE THEORY OF THE IVC-ORDERED STATE

We now describe a mechanism that stabilizes IVC ordering and describe the properties of the resultant state.

Interestingly, to treat this stage of the problem, it is sufficient to work within a momentum-space formulation, which enables us to sidestep the difficulties elaborated in Sec. III C with a real-space tight-binding formulation.

Consider the nearly flat bands in the limit of strong Coulomb repulsion. Note that the dominant part of the interaction is fully $SU(4)$ invariant. We expect that the Coulomb interaction prefers an $SU(4)$ ferromagnetic state—similar to the $SU(4)$ ferromagnetism favored by Coulomb interaction in the zeroth Landau level in monolayer graphene [35–37] or in the extensive literature on flat-band ferromagnetism [38]. Indeed, the difficulties with Wannier localization of the nearly flat bands also suggest that, when Coulomb interactions dominate, an $SU(4)$ ferromagnetic ground state will be favored. The band dispersion, however, is not $SU(4)$ symmetric, and hence there will be a selection of a particular direction of polarization in the $SU(4)$ space. To address this, we consider the energies of different orientations of the $SU(4)$ ferromagnet within a simple Hartree-Fock theory. Specifically, we compare a spin-polarized state, a pseudospin I^z -polarized state, and the IVC state with I^x polarization.

Assume a Hamiltonian

$$H = H_0 + V \quad (12)$$

with

$$H_0 = \sum_{an\mathbf{k}} \epsilon_{an}(\mathbf{k}) c_{an\alpha}^\dagger(\mathbf{k}) c_{an\alpha}(\mathbf{k}). \quad (13)$$

Similarly to before, a is the valley index, α is the spin index, and n labels the two bands for each valley. The dispersion $\epsilon_{an}(\mathbf{k})$ is independent of the spin, and, due to time reversal, $\epsilon_{an}(\mathbf{k}) = \epsilon_{-an}(-\mathbf{k})$. We assume a simple form of interaction:

$$V = \frac{g}{N} \sum_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}} c_{a\alpha}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{a\alpha}(\mathbf{k}_1) \cdot c_{a'\alpha'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{a'\alpha'}(\mathbf{k}_2), \quad (14)$$

where N is the number of \mathbf{k} points in the moiré Brillouin zone. Repeated indices are summed over here. This interaction actually has an $SU(8)$ symmetry, but it is strongly broken down to $SU(4)$ by the difference in dispersion between the two bands and eventually down to $U(2) \times U(2)$ by the asymmetry of the dispersion under $\mathbf{k} \rightarrow -\mathbf{k}$. Each $U(2)$ factor corresponds simply to independent $U(1)$ charge and $SU(2)$ spin-conservation symmetries of the two valleys.

We also remark that Eq. (14) is overly simplified, for it does not incorporate form factors arising from the modulation of the Bloch wave functions over the BZ when projecting onto the nearly flat bands. With such form factors included, the interaction projected onto the nearly flat bands should be written as

$$V = \frac{g}{N} \sum_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}} \Lambda_{nn'}^a(\mathbf{k}_1 + \mathbf{q}, \mathbf{k}_1) \Lambda_{mm'}^{a'}(\mathbf{k}_2 - \mathbf{q}, \mathbf{k}_2) \cdot c_{ana}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{an'a}(\mathbf{k}_1) \cdot c_{a'ma'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{a'm'a'}(\mathbf{k}_2) \quad (15)$$

with the form factors given by the Bloch wave functions of the states in the nearly flat bands via

$$\Lambda_{nn'}^a(\mathbf{k}_1, \mathbf{k}_2) = \langle u_{an}(\mathbf{k}_1) | u_{an'}(\mathbf{k}_2) \rangle, \quad (16)$$

where $|u_{an}(\mathbf{k})\rangle$ is the Bloch wave function of a state in the nearly flat bands labeled by valley index a , band index n , and momentum \mathbf{k} (it has no dependence on the spin indices). These form factors are potentially important for the present problem due to the nontrivial band topology present in the valley-resolved band structure. Our preliminary analysis suggests that the results of the Hartree-Fock calculation are modified in the ultra-flat-band limit, i.e., when the interaction term overwhelms the kinetic energy, whereas the key conclusions below are stable within a range of intermediate interaction strengths. In view of this analysis, in the following, we first pursue the simplified Hartree-Fock theory and leave the task of settling down the real ground state for future (numerical) studies; it is an interesting question answering whether the actual experimental systems demand a more sophisticated treatment.

Details of the Hartree-Fock calculation are presented in Appendix E. In summary, we find that the IVC state has a lower energy than both spin- and I^z -polarized states. The physical reason is that, for both the spin- and I^z -polarized states, the order parameter is conserved, and, hence, there is a linear shift of the band when the order parameter is nonzero. In contrast, due to the $k \rightarrow -k$ dispersion anisotropy, the IVC order parameter does not commute with the Hamiltonian. IVC order thus does not simply shift the band but modifies it more significantly. Assuming a near full polarization in the Hartree-Fock Hamiltonian, the noncommutativity leads to an extra energy gain at second order in the IVC state compared to the spin-polarized or I^z -polarized states.

Note that, in the presence of $U(2) \times U(2)$ symmetry, the spin-singlet IVC state is degenerate with states that have spin-triplet IVC ordering with an order parameter $I^x S$. The selection between the singlet and triplet IVC order has to occur due to other terms in the Hamiltonian that have been ignored so far. We do not attempt to pin down the details of this selection in this paper and simply assume, as suggested by the phenomenology, that the spin-singlet IVC is preferred and discuss its consequences.

Next, we turn to a description of the properties of the IVC state. We assume that the order parameter is large and first study its effects on the band structure. In the absence of valley ordering, at the two Dirac points, there is a fourfold band degeneracy. As explained in Sec. III A, the valley ordering splits this fourfold degeneracy into two sets of twofold degenerate Dirac points. When the order parameter

is large, the four nearly flat bands split into two sets of two bands [Fig. 1(f)]. At quarter filling, we fill the bottommost band, which, however, results in not a Mott insulator but a Dirac semimetal. Thus, the IVC state by itself does not lead to a Mott insulator, and a further mechanism is needed. We discuss this in the next section. We note that the semimetals obtained from planar valley order versus I^z order are rather different, the latter being similar to spin-ordered states. Furthermore, while additionally breaking C_3 symmetry alone can eventually gap the Dirac points of the IVC semimetal, the same is not true of spin- or I^z -ordered semimetals, which need further symmetry breaking due to their Dirac points carrying the same chirality.

Going beyond the mean field, the universal properties of the IVC-ordered state are determined by its symmetry breaking. It has a Goldstone mode with linear dispersion at the longest scales. Furthermore, it has a finite-temperature Berezinskii-Kosterlitz-Thouless transition, which has weak signatures in standard experimental probes [39].

VI. INTERVALLEY COHERENT MOTT INSULATORS: GENERALITIES AND A CONCRETE EXAMPLE

We see that IVC ordering by itself gives us only a Dirac semimetal and not a Mott insulator. We now consider the physics below the IVC-ordering scale. First, we note that that, once $U_v(1)$ symmetry is broken, there is no difficulty with writing down a real-space tight-binding model for the two lowest bands. This model lives on the honeycomb lattice and must be supplemented with interactions. Naively, we might imagine that the dominant interaction is an on-site Hubbard repulsion. However, we know that the orbital shapes are such that the actual charge density is concentrated at the original triangular sites, i.e., at the center of the hexagons of the honeycomb lattice. Now, if there is an electron at a honeycomb site r , its wave function is spread equally between the three hexagonal plaquettes that the site r is a part of. The integral of the modulus square of the wave function in any one such plaquette is $1/3$. The total charge that is localized at the center R of each hexagon (i.e., the triangular moiré sites) is therefore

$$Q_R = \sum_{r \in \text{hexagon}} \sum_{r\alpha} \frac{n_\alpha(r)}{3}. \quad (17)$$

Now, let us make the reasonable approximation that the primary Coulomb interaction is on site on the triangular-lattice sites R . Then, in terms of the honeycomb model, the appropriate Coulomb interaction is a ‘‘cluster Hubbard’’ term:

$$H_U = U \sum_R (Q_R - 2)^2. \quad (18)$$

We also specialize to $\nu = -2$ when this honeycomb lattice is half filled. This interaction penalizes charge fluctuations

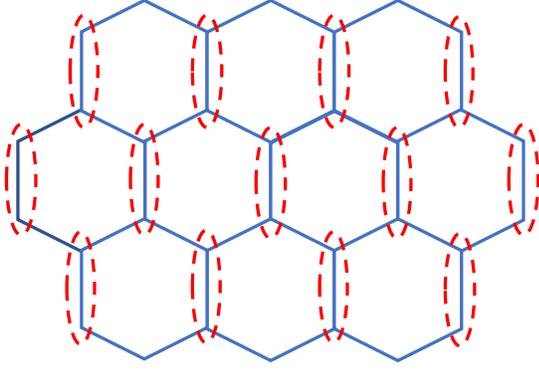


FIG. 2. A C_3 symmetry-breaking order which preserves other symmetries.

on each honeycomb plaquette. Thus, a suitable model Hamiltonian at scales much smaller than the intervalley coherence scale takes the form

$$H = H_t + H_U, \quad (19)$$

$$H_t = -\sum_{r,r'} \sum_{\alpha} t_{rr'} c_{r\alpha}^{\dagger} c_{r'\alpha} + \text{H.c.} \quad (20)$$

For the usual Hubbard model with a strong on-site repulsion, the Mott insulating state has the usual two-sublattice Néel order. However, when the cluster-charging energy is dominant, this is not obviously the case. We therefore allow ourselves to consider a few different possibilities for the Mott insulator. Naturally, in all these options, the charge gap of the insulator is much lower than the scale of IVC ordering. In the experiments, the charge gap is estimated to be about 5 K. The IVC ordering should then occur at a much higher scale, consistent with what we already concluded based on the phenomenology. In this section, to be concrete, we focus on a particular Mott insulator where the C_3 rotation symmetry is spontaneously broken while preserving other symmetries (Fig. 2).

In passing, we note that insulators driven by cluster charging have been considered in a number of different contexts before. For instance, Refs. [40,41] study models with extended Coulomb interactions as a route to access Wigner-Mott insulators. Cluster-charging insulators have also been long studied [42–44] as a platform for various fractionalized insulating phases of matter. In contrast to these earlier works, where a dominant cluster-charging interaction is simply postulated and the resulting physics explored, here we identify a natural mechanism for such an interaction.

A. C_3 -broken insulator

Breaking the C_3 symmetry allows gapping out the Dirac points and leads to an insulator. As the C_3 breaking order parameter increases, the two Dirac points move towards

each other [Fig. 1(e)] and eventually annihilate to produce a fully gapped insulator. This annihilation (and, correspondingly, the gap minimum just into the insulator) occurs at either the Γ or M point depending on the details of the dispersion. Note that, within this picture, the C_3 breaking also occurs at a scale bigger than the approximately 5 K charge gap of the Mott insulator. Clearly, the excitations above the charge gap are ordinary electrons, and their gap can be readily closed by a Zeeman field.

Upon doping this insulator, charge enters as ordinary holes and forms a small Fermi pocket. This pocket is centered at either Γ or M depending on the location of the minimum insulating gap. In either choice, due to the absence of C_3 symmetry, there is just a single such Fermi pocket which will accommodate the full density of doped holes. Because of the intervalley ordering, these holes are valley polarized in the I^x direction. Naturally, this explains the quantum oscillation experiments—the frequency is set by the density of doped holes, and the Landau level degeneracy (per flux quantum) is only twofold (from the spin).

A natural pairing mechanism emerges from the coupling of the holes to Goldstone fluctuations of the intervalley order, as we now elaborate. In the presence of an intervalley condensate, an appropriate effective action is

$$S = S_0[\psi] + S_1[\psi, \theta], \quad (21)$$

$$S_0 = \int d\tau \left(\int d^2x \bar{\psi} (\partial_{\tau} + \mu) \psi + \int d^2k \bar{\psi}_k h_k \psi_k \right), \quad (22)$$

$$S_1 = \int d\tau d^2x \Phi_0 (e^{-i\theta} \bar{\psi}_+ \psi_- + \text{c.c.}). \quad (23)$$

Here, ψ is a continuum electron field that represents the electrons in the low-energy nearly flat bands, θ is the phase of the intervalley condensate, and Φ_0 is its amplitude. Note that $h_k = \epsilon_s(k) + \epsilon_a(k)\tau^z$ is a 2×2 matrix for each k point [45]. We allow for slow Goldstone fluctuations of the phase and obtain a convenient form of the electron-electron interaction induced by these fluctuations. To that end, we first define new fermion variables χ through

$$\psi = e^{i\theta\tau_z/2} \chi. \quad (24)$$

This redefinition removes the θ dependence from S_1 , but S_0 now takes the form

$$S_0[\psi] = S_0[\chi] + S'_0[\chi, \theta], \quad (25)$$

$$S'_0[\chi, \theta] = \int_{x,\tau} \frac{i}{2} \partial_i \theta \bar{\chi} \tau^z \chi + \frac{1}{2} \partial_i \theta J_i^v(x). \quad (26)$$

Here, J_v is the contribution to the $U_v(1)$ current from the fermions. It is conveniently written down in momentum space as

$$J_i^v(q) = \int d^2k \bar{\chi}_{k+q} \frac{\partial h_k}{\partial k_i} \chi_k. \quad (27)$$

Now we assume that Φ_0 is near maximum polarization and diagonalize the χ Hamiltonian obtained from $S_0[\chi] + S_1[\chi]$. As discussed earlier, there are two sets of bands per spin (corresponding to $I_x = \pm 1$) that are well separated from each other. The low-energy electrons are those that have valley polarization $I_x \approx 1$. We wish to obtain the coupling of these electrons to the θ fluctuations. For the bands with $I_x \approx 1$, we write

$$\chi_{+\alpha} = \chi_{-\alpha} \equiv d_\alpha. \quad (28)$$

It follows that $\bar{\chi} \tau^z \chi \approx 0$ and similarly $\bar{\chi} \{[\partial \epsilon_s(k)]/\partial k_i\} \tau^z \chi_k \approx 0$. The only nonvanishing coupling, therefore, is to the contribution from $\epsilon_a(k)$. We get

$$J_i^v(q) \approx \int d^2k \bar{d}_{k+q} \frac{\partial \epsilon_a(k)}{\partial k_i} d_k \equiv \int d^2k v_i^a(k) \bar{d}_{k+q} d_k. \quad (29)$$

Now we assume we have integrated out the fermions everywhere except in the close vicinity of the Fermi surface, which gives a long-wavelength, low-frequency effective action for the θ fluctuations of the form

$$S_{\text{eff}}[\theta] = \int_{q,\omega} K \left(\frac{\omega^2}{v^2} + q^2 \right) |\theta(q, \omega)|^2. \quad (30)$$

Here, K is the phase stiffness of the θ field, and v is the velocity of the linear dispersing θ fluctuations. We now integrate out θ to get an effective interaction between the electrons:

$$S_{\text{int}} = - \int_{q,\omega} \frac{q^2}{32K(\frac{\omega^2}{v^2} + q^2)} |J_i^v(q, \omega)|^2, \quad (31)$$

which is an attractive interaction. Anticipating that the important regime for pairing is $|\omega| \ll vq$ for an approximate treatment, we set $\omega = 0$ in the prefactor to get a simplified effective interaction:

$$S_{\text{int}} = - \frac{1}{32K} \int_{q,\omega} |J_i^v(q, \omega)|^2. \quad (32)$$

We emphasize again that—within our approximate treatment—the only contribution to J_i^v comes from the *antisymmetric* part of the “normal” state dispersion. This attractive interaction can now be treated within a BCS mean field and leads to a superconducting state.

Note that, in real space, since the large repulsion will be on the hexagon center and not on the honeycomb site, there is no particular reason to disfavor on-site s -wave pairing. Though we do not give a detailed description of the pairing symmetry, the route to superconductivity sketched

above naturally leads to a spin-singlet superconductor. Furthermore, it forms out of a “normal” metal of ordinary holes through a BCS-like pairing mechanism. We expect then that Zeeman fields of the order of T_c efficiently suppress the superconductivity except possibly at very low doping (where eventually phase fluctuations kill T_c). At low doping, and when one is near a high-symmetry point of the Brillouin zone (which is consistent with the fact that there is no additional degeneracy seen in quantum oscillations), the antisymmetric part of the dispersion is expected to be constrained by symmetry to be small. For example, near the Γ point, it vanishes as the cube of the crystal momentum, which would lead to a small valley current (the derivative of the antisymmetric dispersion with respect to momentum) and, hence, a weakening of the coupling to valley Goldstone modes, as the doping is reduced. However, if C_3 rotation symmetry is broken, the antisymmetric dispersion can include a term that is linear in momentum, leading to a nonvanishing valley current at small doping.

It is also important to ask if a conventional pairing mechanism due to coupling to phonons might be operative. We note that the bandwidths of the nearly flat bands (approximately 10 meV) are much smaller than the typical phonon energies in graphene. Thus, the magic-angle twisted bilayer graphene system is far from a regime in which phonon effects can be treated within the usual adiabatic approximation. Furthermore, the observation of superconductivity only in the vicinity of the correlated insulator appears unnatural within a phonon-based theory. Nevertheless, it is possible that phonon effects play a role in various aspects of our physics and contribute to the effective interactions between the electrons. We leave for the future a proper treatment of the electron-phonon coupling in these systems and focus instead here on electron-electron interaction effects, which clearly must play an important role given the proximity of the superconductor to a correlation-driven Mott insulator.

VII. OTHER POSSIBLE MOTT INSULATING STATES

The C_3 -broken insulator is a concrete example of how an intervalley condensate of the twisted bilayer system can eventually become a Mott insulator. However, given the current experimental information, it is not clear that this is uniquely dictated. Therefore, we sketch a few different Mott insulating states and present some of their phenomenological consequences.

- (1) *Translation-broken insulator*.—Broken moiré translations—for instance, Kekule ordering on the effective honeycomb lattice—can also gap out the Dirac points. The properties of this state and its evolution into the doped superconductor are similar to the C_3 -broken insulator discussed above.

- (2) *Antiferromagnetic insulator*.—This state is the familiar Mott insulator of the usual honeycomb Hubbard model. Upon doping, it is expected to evolve into a spin-singlet superconductor as seen in numerical studies of the $t - J$ honeycomb model [46]. The pairing symmetry appears to be $d + id$. It will be interesting to look for signatures of broken time-reversal symmetry if this scenario is realized. Furthermore, this state is known to have quantized spin and thermal Hall effects and associated gapless edge states [47,48]. Other properties related to this state are also discussed in the literature [49,50].
- (3) *Featureless Mott insulators*.—Given that the honeycomb lattice features two sites in the unit cell, it evades the Lieb-Shultz-Mattis theorem and allows for a featureless ground state (i.e., a gapped insulator with neither topological order nor symmetry breaking) at half filling [51–56]. Pictorially, this state is viewed as a spin-singlet Cooper pair of electrons being localized on orbitals composed of equal-weight superpositions of the hexagons of the honeycomb lattice. While model wave functions of this phase have been constructed, the interactions that can drive a system into this phase remain to be understood.
- (4) *Quantum spin liquids*.—The simplest possibility is a fully gapped quantum spin liquid. In this case, there are neutral spin-1/2 excitations (spinons) in the insulator. Upon doping, a natural possibility is that the charge goes in as bosonic holons (spinless charge- e quasiparticles) whose condensation leads to superconductivity. This process is the classic resonating valence bond (RVB) mechanism [1] for superconductivity in a doped Mott insulator. However, in this scenario, at low doping the superconducting T_c will not have anything to do with the spin gap (measured by the Zeeman scale needed to suppress pairing).

We do not attempt to decide between these different options in this paper. However, we outline experiments that can distinguish between them in Sec. X.

VIII. TIGHT-BINDING MODELS

A. For TBG

As we argue, there is an obstruction for writing down any tight-binding model for the single-valley problem. A natural way out, therefore, is to instead consider the four-band problem consisting of both valleys. Our earlier argument requiring the orbitals to be centered on the sites of the honeycomb lattice still applies, but now with two orbitals associated with every site. In addition, to resolve the mirror-eigenvalue obstruction we describe, these two orbitals should transform oppositely under the mirror symmetry. These orbitals, however, cannot have definite valley charge, for otherwise the problem is reduced back to

the earlier case with each valley considered separately. Instead, it is natural to demand each of the two orbitals to be a time-reversal singlet, which leads to a standard representation for the symmetry group of $p6mm$ together with time reversal. However, due to the aforementioned anomaly, the representation of valley $U_v(1)$ is necessarily complicated, and we address that later.

Forgetting about valley conservation for the time being, the construction of Wannier functions becomes a rather standard problem, and well-established protocols apply. In particular, we construct well-localized Wannier functions using the projection method [57], starting from a set of well-localized trial wave functions as the “seed” of the Wannier functions (Appendix C). Specifically, we start with the continuum theory of Ref. [12], as described around Eq. (5), with the parameters $w_0 = 110$ meV and $w_1 = 120$ meV. The success of the construction hinges crucially on having a nonsingular projection everywhere in the BZ, which can be monitored by ensuring that the overlap between the seed and the actual Bloch wave functions neither vanish nor diverge anywhere in the BZ [57]. Using this approach, we construct Wannier functions for a particular choice of parameters, detailed in Appendix C, for the four nearly flat bands near charge neutrality (spin ignored). Our trial wave functions attain a minimum and maximum overlap of 0.38 and 3.80, respectively, indicating a satisfactory construction. Indeed, the Wannier functions we obtain are quite well localized [Figs. 3(a) and 3(b)], with approximately 90% of their weight contained within one lattice constant from the Wannier center. In addition, the Wannier functions we construct are exponentially localized [Figs. 3(c) and 3(d)], as anticipated from the nonsingular trial wave-function projection.

Having constructed the Wannier functions, one can readily extract an effective tight-binding model \hat{H}^{WF} by first projecting the full Bloch Hamiltonian into the Wannier basis and then performing Fourier transform. Because of the exponential tail, however, the resulting tight-binding model has infinite-range hopping despite an exponentially suppressed amplitude. To capture the salient behavior of the model, it is typically sufficient to keep only the bonds with strength larger than some cutoff t_c . In other words, t_c serves as a control parameter, and one recovers the exact band structure in the limit of $t_c \rightarrow 0$, albeit at the cost of admitting infinite-range hoppings.

The obtained band structures for different value of t_c are plotted in Figs. 4(a)–4(d). We find that a fairly long-range model (see Appendix D for parameters), keeping terms that connect sites up to two lattice constants apart, is needed to capture all the salient features of the energetics. It should be noted that the range of the approximate models generally depends on the localization of the Wannier function, and in this work we have not optimized the Wannier functions. It is therefore possible that, by further optimization, one may capture the energetics more faithfully using only shorter-range terms.

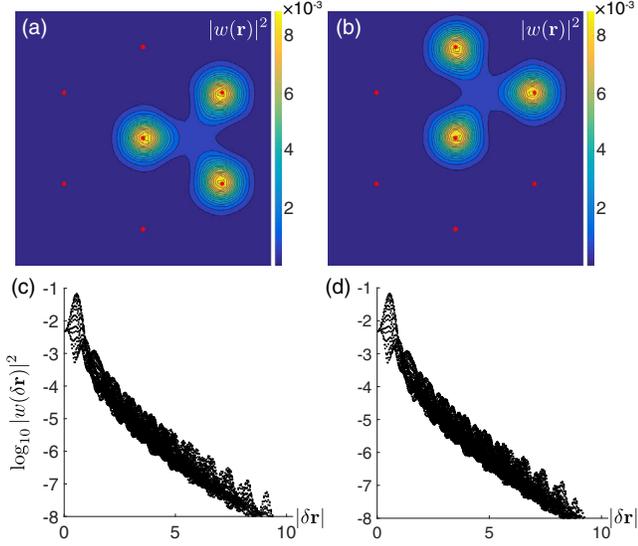


FIG. 3. Constructed Wannier functions. As both valleys are taken into account, there are four nearly flat bands (spin ignored), giving rise to four Wannier functions per unit cell. Results shown are for a pair related by C_6 symmetry; results for the other pair, which look nearly identical, are presented in Appendix C. (a),(b) Amplitude of the Wannier function $w(\mathbf{r})$. Red stars indicate the AA regions in the moiré potential. (c),(d) The amplitude decays exponentially when the distance from the Wannier center, $\delta\mathbf{r}$ (measured in units of the moiré lattice constant), is $\delta\mathbf{r} \gtrsim 3$.

Although spatial and time-reversal symmetries are respected in the tight-binding model, valley conservation is explicitly broken, which is because our Wannier functions cannot be chosen to represent valley conservation naturally, similar to the case of topological insulators [58], and, therefore, any truncation of the transformed Hamiltonian generically introduces explicit valley-conservation symmetry breaking. Furthermore, one can ask how the operator I_z is represented. In particular, we want to construct the projection operators for the single-particle problem, $\mathcal{P}_\pm = \frac{1}{2}(1 \pm I_z)$, which project into the valleys. It would be most desirable if one can formulate I_z , and hence \mathcal{P}_\pm , directly in real space. Given that I_z is also a free-electron Hermitian operator, this formulation amounts to finding a symmetric Wannier representation of I_z . However, we find that there are again obstructions, which mirror exactly the obstructions we face when attempting to construct Wannier functions for the single-valley two-band model of the nearly flat bands, i.e., a mismatch in the mirror eigenvalues, as well as a nonzero net charge of the Dirac points. Such an inheritance of the obstructions is presumably a manifestation of the underlying anomaly of the single-valley description.

Towards recovering valley conservation.—It is desirable to restore valley conservation even approximately, for our truncated model, and we describe a method below. Recall that we denote the Hermitian valley-charge operator in the continuum effective theory by \hat{I}_z [Eq. (6)]. Projecting \hat{I}_z into the four-band subspace of the nearly flat bands and

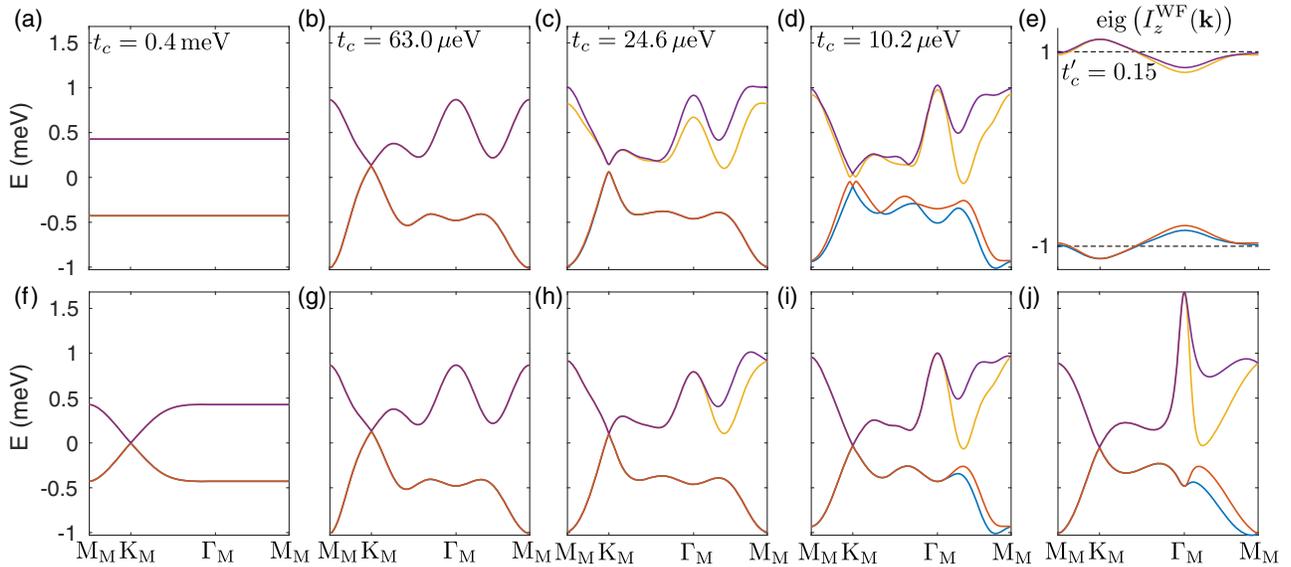


FIG. 4. Effective tight-binding model for the nearly flat bands. (a)–(d) Effective tight-binding models for the nearly flat bands are derived by projecting the Hamiltonian in the continuum theory into the Wannier basis. Bonds with strength $< t_c$ are truncated from the model, and the resulting band structures for three choices of t_c are shown. (e) One can also derive the effective valley-charge operator \hat{I}_z^{WF} , using the same procedure with cutoff t'_c . Similar to the Hamiltonian, the truncation of terms in the effective operator leads to an error, and so the eigenvalues of $I_z^{\text{WF}}(\mathbf{k})$ have a small deviation from the exact values of ± 1 . (f)–(i) Valley conservation can be reenforced on the effective model through projection. (f)–(i) correspond, respectively, to the effective models shown in (a)–(d). As t_c is reduced, the effective model more faithfully reproduces the salient energetics features of the continuum theory, the latter of which is plotted in (j).

rotating into the Wannier function basis, we arrive at another Hermitian operator defined on the honeycomb sites, which we can simply interpret as yet another Hamiltonian-like object in our problem. Similar to the earlier discussion for the Hamiltonian, the effective \hat{I}_z operator \hat{I}_z^{WF} has infinite-range hopping with an exponentially suppressed amplitude, and it is natural to approximate it by truncation, keeping again only terms with a strength larger than some t'_c . Such a truncation, however, introduces a deviation in the eigenvalues of \hat{I}_z from the physical values of ± 1 [Fig. 4(e)]. To fix this deviation, therefore, we can further perform spectral flattening of the corresponding bands to ± 1 in momentum space. This procedure is well defined so long as a band gap is sustained between the second and third bands of \hat{I}_z^{WF} , which is generically true as long as t'_c is chosen to be reasonably small. We denote this flattened version by \hat{I}_z^{WF} . Physically, this corresponds to an approximation of the actual representation of the valley-charge operator in our tight-binding model, and again our approximation can be made exact in the limit of $t'_c \rightarrow 0$.

We can now restore valley conservation in our effective Hamiltonian. To this end, for $n \in \mathbb{Z}$ define the projection operator

$$\hat{P}_n = \sum_{\alpha} |n, \alpha\rangle \langle n, \alpha|, \quad (33)$$

which projects into the sector with \hat{I}_z^{WF} eigenvalue n (in the many-body Hilbert space) and satisfies $\hat{P}_n \hat{P}_m = \hat{P}_n \delta_{n,m}$. We can now define

$$\hat{H}^{\text{WF}} \equiv \sum_{n \in \mathbb{Z}} \hat{P}_n \hat{H}^{\text{WF}} \hat{P}_n, \quad (34)$$

for which valley conservation is restored. In essence, through this procedure we introduce a pair of Hermitian operators $(\hat{H}^{\text{WF}}, \hat{I}_z^{\text{WF}})$, corresponding, respectively, to the Hamiltonian and the valley charge, that converge to the exact operators in the limit $t_c, t'_c \rightarrow 0$.

The valley projection procedure we describe applies equally well to an interacting Hamiltonian \hat{H}_U^{WF} obtained by projecting the microscopic interaction terms to our Wannier basis, which again would not be automatically valley conserving due to the truncation errors. For the free part of the Hamiltonian, however, the projection procedure can be greatly simplified, because the Bloch states of \hat{I}_z^{WF} , which equal those of \hat{I}_z by definition, are known, and using which we can decompose $H^{\text{WF}}(\mathbf{k})$ into the valley-conserving and valley-breaking parts. The projection then proceeds simply by retaining only the valley-conserving part. More concretely, write the Bloch ‘‘Hamiltonian’’ of the valley-charge operator as $\tilde{I}_z(\mathbf{k}) = \Psi_{+\mathbf{k}} \Psi_{+\mathbf{k}}^\dagger - \Psi_{-\mathbf{k}} \Psi_{-\mathbf{k}}^\dagger$, where $\Psi_{\pm\mathbf{k}}$ are 4×2 matrices. Note that the columns of $\Psi_{\pm\mathbf{k}}$ are simply the \pm -valley-charge eigenstates. We can then perform the projection by

$$\tilde{H}^{\text{WF}}(\mathbf{k}) = \Psi_{+\mathbf{k}}^\dagger H^{\text{WF}}(\mathbf{k}) \Psi_{+\mathbf{k}} + \Psi_{-\mathbf{k}}^\dagger H^{\text{WF}}(\mathbf{k}) \Psi_{-\mathbf{k}}, \quad (35)$$

giving an easy way to perform the described projection.

As is shown in Figs. 4(f)–4(i), the projected effective tight-binding model reexhibits all the symmetry features of the bands from the continuum theory [Fig. 4(j)], for any choice of truncation parameter t_c . In particular, Figs. 4(a) and 4(f) represent the simplest model which demonstrates the utility of our approach, with the valley projection alone converting an otherwise hopping-free Hamiltonian into one exhibiting the charge-neutrality Dirac points. We further remark that, although generically H^{WF} does not respect $U_v(1)$, the effective Hamiltonian corresponding to Fig. 4(b) comes very close to being $U_v(1)$ symmetric in terms of its energetics along the high-symmetry line. We provide an explicit tabulation of the bands in this H^{WF} and that of I_z^{WF} in Appendix D.

B. Nearly flat bands in trilayer graphene-boron nitride moiré superlattices

Recently, Mott insulating phases (but not superconductivity, at the time of writing) were observed [20] in a heterostructure of *ABC* trilayer graphene encapsulated in boron nitride (TLG/hBN), where a moiré superlattice is present even at zero twisting between the graphene layers. Four minibands are observed close to neutrality, whose bandwidth and separation can be tuned by a vertical electric field. Half filling one of the nearly flat bands results in a Mott insulator.

We remark that the symmetry setup for this trilayer heterostructure bears more resemblance to the Bernal-stacked bilayer graphene than the TBG system we discuss above. In particular, the absence of Dirac points among the minibands suggests that no C_2 symmetry is present, and the system is potentially described by a wallpaper group for which the two sublattices of the honeycomb lattice are no longer symmetry related (say, the wallpaper group $p3m1$, No. 14 [24]). If that is indeed the case, one expects the valley-resolved band structure to admit a tight-binding model defined on the triangular lattice, although it remains to be checked whether or not the charge density profile exhibits any nontrivial features, akin to that found for the TBG system [Fig. 3(c)]. It will be of great interest to derive a concrete real-space effective model for the TLG/hBN, but we leave this derivation as a future work.

IX. MODEL FOR CORRELATED STATES IN TRILAYER GRAPHENE HETEROSTRUCTURE

In this section, we briefly consider the case of triangular moiré superlattices in a trilayer graphene heterostructure. Correlated insulating states were observed very recently in this system [20]. Just like in the twisted bilayer, here, too, there are nearly flat bands that are separated from the rest of the spectrum. However, unlike the TBG, here there are no

Dirac crossings in this nearly flat band, and the low-energy degrees of freedom are in the trivial representation of C_3 . In addition, it is known that a vertical electric field can induce a gap for ABC -stacked trilayer graphene, and, depending on the direction of the electric field, the band structure can have a zero Chern for one direction of the electric field or a nonzero Chern number for the other direction of the electric field [59]. It is thus reasonable that, in the case where the direction of the electric field is such that the nearly flat bands possess no net Chern number, the nearly flat band can be modeled in real space by a triangular-lattice model with two orbitals (corresponding to the two valleys) per site supplemented with interactions. However, some care is still necessary. Time reversal and C_2 both act by flipping the valley index. Thus, the band dispersion $\epsilon_{\pm}(k)$ within a single valley is not symmetric under $k \rightarrow -k$:

$$\epsilon_{\pm}(k) \neq \epsilon_{\pm}(-k), \quad (36)$$

though

$$\epsilon_{+}(k) = \epsilon_{-}(-k) \quad (37)$$

is satisfied. A real-space tight-binding description on the triangular lattice therefore takes the form

$$H'_{\text{trilayer}} = -\sum_{RR'} \sum_{a=\pm} \sum_{\alpha} t_{RR'} c_{R\alpha a}^{\dagger} (e^{i\phi_{rr'}\tau_z})_{aa'} c_{R'a'\alpha} + \text{H.c.}, \quad (38)$$

with $t_{rr'}$ real and positive. The phases $\phi_{rr'}$ are, in general, nonzero. The phase ϕ even on nearest-neighbor bonds cannot be removed, as, in general, the symmetries permit a nonzero flux Φ for any single valley through an elementary triangle (and the opposite flux for the other valley).

When the Coulomb interaction dominates, the $SU(4)$ ferromagnet with a further selection of IVC order is once again a possibility. From a real-space point of view, the projection of the Coulomb interaction to the Wannier basis used to formulate the tight-binding model leads to an appropriate interaction Hamiltonian. If the Wannier orbitals are not tightly confined to each triangular site, then there will be a significant *intersite* ferromagnetic Hund's exchange which promotes $SU(4)$ ferromagnetism with a further selection of IVC ordering.

It is interesting to consider the limit where the Wannier functions are sufficiently tightly localized that such a ferromagnetic intersite exchange is weak and can be ignored. In that limit, to obtain a *minimal* model for this system we restrict the hopping to just be nearest neighbor and include an on-site repulsion. The minimal model then takes the form

$$\begin{aligned} H_{\text{trilayer}} &= H_0 + V, \\ H_0 &= -\sum_{(RR')} \sum_{a=\pm} \sum_{\alpha} t_{RR'} c_{R\alpha a}^{\dagger} (e^{i\phi_{rr'}\tau_z})_{aa'} c_{R'a'\alpha} + \text{H.c.}, \\ V &= \frac{U}{2} \sum_R (N_R - N_0)^2 \end{aligned} \quad (39)$$

with $\phi_{12} + \phi_{23} + \phi_{31} = \pm\Phi$ with a $+$ sign for up-facing triangles and a $-$ sign for down-facing ones. Here, sites 1, 2, and 3 are assumed to be arranged counterclockwise on each triangle. N_R is the total electron number at site R , and N_0 controls the filling factor. As in previous sections, this Hamiltonian has a $U(2) \times U(2)$ symmetry corresponding to independent $U(2)$ rotations of each valley in addition to the discrete symmetries described above. The model thus needs to be supplemented with further weaker interactions that break the continuous symmetry down to $U(2)$, though we do not specify them here. Note that if the flux $\Phi = 0$, then the model actually has an even higher $SU(4)$ symmetry. Also, this model has an extra C_2 rotational symmetry that flips the valleys, which should be viewed as an emergent symmetry of the model defined above, which should be broken by other terms. In particular, it should be differentiated from a microscopic C_2 symmetry; if such a microscopic symmetry was present, it would combine with time-reversal symmetry and protect an odd number of Dirac points in the single-valley trilayer graphene band structure, which suffers from a parity anomaly [60,61]. Our effective model does not suffer from this parity anomaly, because this microscopic C_2 symmetry is absent.

The minimal model above allows for a discussion of the Mott insulator in the strongly correlated regime of large U at integer N_0 . In the experiments, Mott insulators at fillings $N_0 = 1, 2$ have been reported. In the large- U limit, the effective model takes the form of a “spin-orbital” Hamiltonian on a triangular lattice that has four states per site: two spins and two valleys. A systematic t/U expansion is readily performed to yield this spin-orbital Hamiltonian. At $O(t^2/U)$, the “superexchange” is not sensitive to the flux Φ , and we end up with an $SU(4)$ quantum antiferromagnet on the triangular lattice (This is not actually correct. As shown in Ref. [62], there is a second order process which depends on the phase of t . This leads to a term that breaks $SU(4)$ symmetry. However we anticipate that the more interesting dependence on flux comes from the third order term identified below.). For $N_0 = 1$, the $SU(4)$ spins are in the fundamental representation, while for $N_0 = 2$ they are in the six-dimensional representation.

Antiferromagnetic models of $SU(4)$ spins have been studied on a variety of lattices with different motivations (for some representative recent papers, see Refs. [63–65]). It seems likely that they go into “paramagnetic” states that preserve $SU(4)$ symmetry. However, a new feature in the present problem is the presence of the flux Φ in the underlying Hubbard model which breaks $SU(4)$ to $U(2) \times U(2)$. This

feature modifies the spin-orbital model at $O(t^3/U^2)$. In the experiments, the ratio of Coulomb interactions to the bandwidth of the nearly flat bands may be controlled by a perpendicular electric field, and it may thus be possible to tune the strength of these third-order terms relative to the second-order ones. In Appendix F, we derive the spin-orbital Hamiltonian to third order, showing how the flux Φ leads to new terms not present in an $SU(4)$ -invariant model. We, however, leave for the future a detailed study of these interesting spin-orbital models.

At any rate, we emphasize that this trilayer system is thus *qualitatively* different from the twisted bilayer graphene where we argue that a real-space triangular-lattice description is not possible due to Dirac crossings within the nearly flat bands.

X. PROPOSED FUTURE EXPERIMENTS

As discussed in previous sections, the ideas presented in this paper suggest a number of experiments which will be extremely useful in revealing the physics. Here, we reiterate and elaborate on some of these suggestions.

A crucial clue from the existing experiments is that an in-plane field suppresses the superconductivity—at optimal doping—when the Zeeman energy is of the order of the zero-field T_c . This suppression indicates spin-singlet pairing and that T_c at optimal doping is associated with the loss of pairing. It will be extremely useful to study this systematically as a function of doping. For the doped C_3 -broken insulator, the superconductivity may be driven by the pairing of a small Fermi surface of electrons. Then (except perhaps at very small doping), T_c and the critical Zeeman scale continue to track each other as the doping is decreased. In contrast, if the pairing (in the form of singlet valence-bond formation) already happens in the Mott insulator—as in the usual RVB theory, or with the featureless Mott insulator—then, with decreasing doping, T_c and the critical Zeeman field should part ways significantly.

A second crucial clue from the experiments is the 2, 4, 6, 8, ... degeneracy pattern of the Landau fan emanating from the Mott insulator. We propose that this pattern is due to the freezing of the valley degree of freedom, which can be distinguished from the alternate possibility that there is spin freezing by studying the quantum oscillations in a tilted field. Zeeman splitting, if it exists, should show up in a characteristic way as a function of the tilt angle.

Our proposal is the intervalley phase coherence at a scale higher than both the superconducting $T_c \approx 1.5$ K and the Mott insulating scale approximately 5 K. The valley symmetry is, as usual, related to translational symmetry of the microscopic graphene lattice. In the twisted bilayer, there is an approximate translation symmetry that holds at some short scale associated with translation by one unit cell of the microscopic graphene lattices. Under this approximate translation operation, electrons at the different valleys get different phases, which is a $U_v(1)$ rotation. Therefore, intervalley

ordering strongly breaks this approximate short translation symmetry. Within each moiré site, the density of states is uniform at the lattice scale when there is no intervalley ordering but oscillates once this order sets in. This difference may be detectable through scanning tunneling microscopy (though, if the bilayer graphene is fully encapsulated by boron nitride, it may be challenging to see the graphene layer).

Assuming there is intervalley ordering, if the undoped Mott insulator develops antiferromagnetic order, it appears likely that the doped superconductor is a spin-singlet $d_{x^2-y^2} + id_{xy}$ superconductor, which spontaneously breaks time-reversal symmetry. In contrast, for a doped C_3 -broken state, either s -wave or $d + id$ spin-singlet superconductivity seem possible. It is also useful to directly search for broken C_3 or moiré translational symmetry in the experiments. Finally, the very different behavior in quantum oscillations between electron and hole doping away from the Mott insulator suggests that there may be a first-order transition into the Mott state as it is approached from the charge neutrality point, which will lead to a hysteretic response as the gate voltage is tuned towards charge neutrality from the Mott insulator.

XI. CONCLUSION

In this paper, we addressed some of the theoretical challenges posed by the remarkable observations of Mott insulating states and proximate superconductivity in twisted bilayer graphene.

We proposed that both the Mott insulator and the superconductor develop out of a state with spontaneous intervalley coherence that breaks independent conservation of electrons at the two valleys. We described a mechanism for the selection of this order over other spin- or valley-polarized states owing to the peculiarities of the symmetry realization in the band structure. We showed that intervalley ordering by itself does not lead to a Mott insulator and described possible routes through which a Mott insulator can develop at a low temperature. A specific concrete example is a C_3 -broken insulator. We showed how doping such an insulator leads to an understanding of the quantum oscillation data and presented a possible pairing mechanism for the development of superconductivity. We described potentially useful experiments to distinguish the various possible routes to a Mott insulator from an intervalley coherent state.

Our work was rooted in a microscopic understanding of the twisted graphene bilayer. We showed that the momentum-space structure of the nearly flat bands places strong constraints on real space descriptions. In particular, contrary to natural expectations, we showed that a real-space lattice model is necessarily different from a correlated triangular-lattice model with two orbitals (corresponding to the two valleys) per site. This difference is due to a symmetry-enforced obstruction to constructing Wannier functions centered at the triangular sites that can capture the Dirac crossings of the nearly flat bands. We showed that a honeycomb lattice

representation may be possible but requires a nonlocal implementation of valley $U_v(1)$ symmetry. In our description of the intervalley ordered state and its subsequent low-temperature evolution into the Mott or superconducting states, we sidestepped these difficulties by first treating the problem directly in momentum space and defining a real-space model only at scales below the intervalley ordering (when the obstruction to a honeycomb representation is gone). We also contrasted the bilayer system with trilayer graphene where Mott insulators have recently been observed. In the trilayer system, it is reasonable to construct a real-space triangular-lattice two-orbital model, but the symmetries allow for complex hopping (with some restrictions). We argue that this system may offer a valuable platform to realize interesting quantum spin-orbital liquids.

ACKNOWLEDGMENTS

We thank Yuan Cao, Valla Fatemi, and Pablo Jarillo-Herrero for extensive discussions of their data and sharing their insights. We also thank Shiang Fang, Liang Fu, Gene Mele, Patrick Lee, Oskar Vafek, Feng Wang, Noah F. Q. Yuan, and Ya-Hui Zhang for interesting discussions or correspondence. T. S. is supported by U.S. Department of Energy Grant No. DE-SC0008739 and in part by a Simons Investigator award from the Simons Foundation. A. V. was supported by a Simons Investigator award and by NSF-DMR 1411343.

Note added.—Recently, Ref. [66] appeared, which has significant differences from the present paper; Refs. [67–69], which discuss, in particular, the symmetries and constructions of Wannier functions, also appeared. Note that these discussions on Wannier functions disregard the presence of the (emergent) sixfold rotation symmetry, and, consequentially, the Dirac points observed at charge neutrality are not symmetry-protected features of these models. A more detailed discussion can be found in our subsequent work reported in Ref. [70].

APPENDIX A: LATTICE AND SYMMETRIES

In this Appendix, we document some details on the conventions and the symmetry transformations.

Consider a monolayer of graphene. We let the primitive lattice vectors \mathbf{A} and reciprocal lattice vectors \mathbf{B} be

$$\begin{aligned} \mathbf{A}_1 &= a\hat{x}, & \mathbf{A}_2 &= a\left(-\frac{1}{2}\hat{x} + \frac{\sqrt{3}}{2}\hat{y}\right); \\ \mathbf{B}_1 &= \frac{4\pi}{\sqrt{3}a}\left(\frac{\sqrt{3}}{2}\hat{x} + \frac{1}{2}\hat{y}\right), & \mathbf{B}_2 &= \frac{4\pi}{\sqrt{3}a}\hat{y}, \end{aligned} \quad (\text{A1})$$

where $a = 2.46 \text{ \AA}$ is the lattice constant (some authors use a to denote the C—C bond length, which is a factor of $\sqrt{3}$ smaller than the lattice constant we are using here). In this choice, we can choose the basis of the honeycomb lattice sites to be

$$\mathbf{r}_1 = \frac{1}{3}\mathbf{A}_1 + \frac{2}{3}\mathbf{A}_2; \quad \mathbf{r}_2 = \frac{2}{3}\mathbf{A}_1 + \frac{1}{3}\mathbf{A}_2. \quad (\text{A2})$$

In momentum space, the K and K' points are given by $\pm(\mathbf{B}_1 + \mathbf{B}_2)/3$ or, for the equivalent ones lying on the x axis, $\pm(2\mathbf{B}_1 - \mathbf{B}_2)/3$. Note that $|\mathbf{K}| = 4\pi/(3a)$, as is well known. Furthermore, we take the Dirac speed v_F to be 10^6 ms^{-1} . Besides, we choose the moiré lattice vectors to be

$$\mathbf{a}_1 = \frac{a}{2\sin(\theta/2)}\left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right); \quad \mathbf{a}_2 = \frac{a}{2\sin(\theta/2)}\left(-\frac{\sqrt{3}}{2}, \frac{1}{2}\right). \quad (\text{A3})$$

In the main text, we list all the symmetries of the continuum theory (Table I). Here, we tabulate explicitly the symmetry transformations of the electron operators, which follow from that of the Dirac points in the monolayer problem:

$$\begin{aligned} \hat{t}_\rho \hat{\psi}_{\pm\mu; \mathbf{k}} \hat{t}_\rho^{-1} &\propto e^{i\mathbf{k}\cdot\rho} \hat{\psi}_{\pm\mu; \mathbf{k}}; \\ \hat{C}_6 \hat{\psi}_{\pm\mu; \mathbf{k}} \hat{C}_6^{-1} &= \sigma_1 e^{\mp i(2\pi/3)\sigma_3} \hat{\psi}_{\mp\mu; C_6 \mathbf{k}}; \\ \hat{M}_y \hat{\psi}_{\pm\mu; \mathbf{k}} \hat{M}_y^{-1} &= \sigma_1 \hat{\psi}_{\pm M_y[\mu]; M_y \mathbf{k}}; \\ \hat{T} \hat{\psi}_{\pm\mu; \mathbf{k}} \hat{T}^{-1} &= \hat{\psi}_{\mp\mu; -\mathbf{k}}, \end{aligned} \quad (\text{A4})$$

where $\mu = t, b$. Note that M_y is the only symmetry which flips the two layers, i.e., $M_y[t] = b$ and vice versa.

The symmetries listed in Eq. (A4) generate all the spatial symmetries of the continuum theory of the TBG [7,12,21] (in wallpaper group 17). In particular, we see that t_ρ and M_y preserve the valley index (K versus $-K$) whereas C_6 and \hat{T} do not. However, their (pairwise) nontrivial products leave valley invariant, and it is helpful to also document their symmetry action explicitly (which are fixed by the above):

$$\begin{aligned} \hat{C}_3 \hat{\psi}_{\pm\mu; \mathbf{k}} \hat{C}_3^{-1} &= e^{\mp i(2\pi/3)\sigma_3} \hat{\psi}_{\pm\mu; C_3 \mathbf{k}}; \\ (\hat{C}_6 \hat{T}) \hat{\psi}_{\pm\mu; \mathbf{k}} (\hat{C}_6 \hat{T})^{-1} &= \sigma_1 e^{\pm i(2\pi/3)\sigma_3} \hat{\psi}_{\pm\mu; -C_6 \mathbf{k}}; \\ (\hat{C}_2 \hat{T}) \hat{\psi}_{\pm\mu; \mathbf{k}} (\hat{C}_2 \hat{T})^{-1} &= \sigma_1 \hat{\psi}_{\pm\mu; \mathbf{k}}. \end{aligned} \quad (\text{A5})$$

Here, we make two remarks regarding the subtleties in the symmetry representation documented here: First, the momenta \mathbf{k} appearing above are defined as the deviation from the original Dirac points in the monolayer problem. Generally, they correspond to different momenta in the moiré BZ. For instance, the Dirac point labeled by $(+, t)$, i.e., that of the \mathbf{K} point in the top layer, is mapped to K_M , whereas $(-, t)$ is mapped to K'_M . Similarly, $(+, b)$ and $(-, b)$ are, respectively, mapped to K'_M and K_M . If desired, one can also rewrite Eqs. (A4) and (A5) using a common set of momentum coordinates defined with respect to the origin of the moiré BZ.

Second, the representation of the translation symmetry \hat{t}_ρ has a subtlety in its definition, which is because the

microscopic translation effectively becomes an internal symmetry for the slowly varying fields appearing in the continuum theory. As such, for a single layer one can deduce its representation only up to an undetermined phase, hence the appearance of α in Eq. (A4). However, the relative momentum across the different slowly varying fields, say, the operators corresponding to the + valley of the top and bottom layers, is a physical quantity. Consequentially, there is really only one common ambiguity across all the degrees of freedom appearing in the continuum theory.

APPENDIX B: VALLEY SYMMETRY AND WANNIER OBSTRUCTION

We argue here that the valley-symmetry-resolved band structure does not admit a Wannier representation. Note, since we ignore spin, that this is a two-band model, which will be crucial for what follows. If one were to include other bands, the arguments below would fail, although precisely what selection of bands would lead to localized Wannier functions (LWFs) remains to be determined. In some ways, it is not very surprising that a valley-resolved band structure does not admit a Wannier description, and a simple example is a single valley of monolayer graphene, which is just an isolated Dirac node. But in those cases the band structure does not terminate on raising the energy and, hence, does not form a band. In contrast, in our present problem for TBG, there is an isolated band, and so one may expect to capture the physics with LWFs. Nonetheless, we argue that there is an obstruction, as can be seen as follows.

We begin with three ingredients: (i) a two-band model; (ii) $C_2\mathcal{T}$ symmetry; and a third ingredient which will be specified shortly. The two ingredients above enforce the following form on the momentum-space Hamiltonian:

$$H = \epsilon_0(p) + \epsilon_1(p)\sigma_1 + \epsilon_2(p)\sigma_2, \quad (\text{B1})$$

where there is no condition on the function $\epsilon_i(p)$. Similar to the main text, we implement $C_2\mathcal{T}$ by $\sigma_1\mathcal{K}$, where \mathcal{K} denotes complex conjugation. Check that $C_2\mathcal{T}$ leaves the Hamiltonian above invariant. Now, if we are interested in the band wave functions, they are independent of the first term in the Hamiltonian, and we could pass to the following one by imposing a constraint:

$$H' = \epsilon_1(p)\sigma_1 + \epsilon_2(p)\sigma_2. \quad (\text{B2})$$

The obvious constraint is to demand

$$\sigma_3 H' \sigma_3 = -H', \quad (\text{B3})$$

which is nothing but the chiral condition that specifies class AIII, and one can show that this condition can be implemented as an on-site symmetry for a two-band system. Now, we introduce the third ingredient: (iii) The

two Dirac points at the middle of this band structure have the same chirality. This ingredient allows us to write down the following effective Hamiltonian close to neutrality:

$$H'_{\text{low}} = -iv_F[\partial_x\sigma_1 + \partial_y\sigma_2] \otimes \mathbb{1}, \quad (\text{B4})$$

where we now have a four-component structure to include the two Dirac nodes. Note that there is no mass term that will gap out these nodes and also preserve the chiral condition (B3); hence, this Hamiltonian corresponds to the surface of a three-dimensional topological phase in class AIII, with index $\nu = 2$, corresponding to the two Dirac nodes. Since this Hamiltonian is the surface state of a nontrivial 3D topological phase, it does not admit a Wannier representation. However, when combined with the opposite valley band structure, together the pair of band structures does admit LWFs but at the price of losing valley conservation symmetry.

Finally, let us address a conundrum that the careful reader may be puzzled by. The valley-resolved bands are stated to be the anomalous surface states of a 3D topological phase; nevertheless, they appear as isolated bands, which seems to contradict the usual expectation that such anomalous bands cannot be separated in energy. The way this conundrum is resolved in the present case is through the two-band condition, which further allows us to map the problem to one with particle-hole symmetry (class AIII). The later problem can have anomalous surface states that are disconnected from the bulk bands, because they are forced to stick at a zero chemical potential, and hence cannot be deformed into the bulk. This mapping to class AIII holds only for the two-band model; hence, if we add bands or fold the Brillouin zone from translation symmetry breaking, the presented arguments no longer hold.

APPENDIX C: WANNIER FUNCTIONS

We construct Wannier functions using the projection method [57]. The method proceeds by first specifying a collection of well-localized, symmetric trial wave functions in real space, which serves as the seed for constructing a smooth gauge required in obtaining well-localized Wannier functions for the problem of interest.

Let us begin by considering the symmetry properties of a real-space wave function in our present problem. For simplicity, we let β be a collective index for the valleys and layers, i.e., $\beta = (+, t), (-, t), (+, b), (-, b)$. Define the real-space electron operators

$$\hat{\psi}_{\beta;r}^\dagger \propto \int_0^\Lambda d^2\bar{k} e^{-i\bar{k}\cdot r} \hat{\psi}_{\beta;\bar{k}}^\dagger, \quad (\text{C1})$$

where we do not keep track of the overall normalization of the operator. Here, \bar{k} is defined as the momentum measured with respect to the origin of the moiré BZ. Note that $\hat{\psi}_{\beta;r}^\dagger$ inherits symmetry transformation from that of $\hat{\psi}_{\beta;\bar{k}}^\dagger$.

To construct a collection of well-localized, symmetric trial wave functions, one can follow the standard discussion concerning the symmetry representation associated with such a real-space basis, say, as reviewed in Supplemental Materials of Ref. [30]. We briefly sketch the main ideas below. Let $W_{\mathbf{h}_0}^\beta(\mathbf{r})$ be a two-component (column) vector localized to \mathbf{h}_0 (the two components here originate from the sublattice degree of freedom in the microscopic problem). Define

$$\hat{W}_{\mathbf{h}_0}^\dagger \equiv \sum_\beta \int d^2\mathbf{r} \hat{\psi}_{\beta,\mathbf{r}}^\dagger W_{\mathbf{h}_0}^\beta(\mathbf{r}) \quad (\text{C2})$$

and its associated momentum-space operator

$$\hat{\Gamma}_{\mathbf{h}_0;\mathbf{k}}^\dagger \equiv \frac{1}{V} \sum_{\mathbf{a}} e^{i\mathbf{k}\cdot\mathbf{a}} \hat{W}_{\mathbf{h}_0+\mathbf{a}}^\dagger = \int d^2\mathbf{k} \hat{\psi}_{\beta,\mathbf{k}}^\dagger \Gamma_{\mathbf{h}_0}^\beta(\mathbf{k}), \quad (\text{C3})$$

where \mathbf{a} is a moiré lattice vector and we assume a periodic system with V moiré unit cells.

For our purpose, we want $\hat{W}_{\mathbf{h}_0}^\dagger$ to serve as our seed for the construction of symmetric Wannier functions. To this end, suppose $\mathbf{h}_0 = (\mathbf{a}_1 - \mathbf{a}_2)/3$, which corresponds to a honeycomb site in the moiré potential, i.e., an AB/BA region. We demand $\hat{W}_{\mathbf{h}_0}^\dagger$ to be invariant under time reversal, the mirror M_y , and the threefold rotation about \mathbf{h}_0 , which we denote by \bar{C}_3 . In addition, recall that the previously predicted charge density profile [8,9,11,14,23] suggests that the Wannier functions take the shape shown in Fig. 3(a). Therefore, it is natural to consider a trial $\hat{W}_{\mathbf{h}_0}^\dagger$ taking the form

$$\hat{W}_{\mathbf{h}_0}^\dagger = \hat{w}_0^\dagger + \hat{C}_3 \hat{w}_0^\dagger \hat{C}_3^\dagger + \hat{C}_3^2 \hat{w}_0^\dagger \hat{C}_3^{\dagger 2}, \quad (\text{C4})$$

where \hat{w}_0^\dagger is localized to the unit-cell origin (an AA site). By definition, $\hat{W}_{\mathbf{h}_0}^\dagger$ transforms trivially under \bar{C}_3 , which one can verify leads to the correct C_3 representations for the nearly flat bands.

It remains to ensure that \hat{w}_0^\dagger is symmetric under \mathcal{T} and M_y . In the spirit of Eq. (C2), we may write $\hat{w}_0^\dagger = \sum_\beta \hat{w}_0^{\beta\dagger}$ for $\beta = (+, t), (-, t), (+, b), (-, b)$. From the symmetry transformation in Eq. (A4), we set

$$\hat{w}_0^{-t\dagger} = \hat{\mathcal{T}} \hat{w}_0^{+t\dagger} \hat{\mathcal{T}}^{-1}; \quad \hat{w}_0^{-b\dagger} = \hat{\mathcal{T}} \hat{w}_0^{+b\dagger} \hat{\mathcal{T}}^{-1}. \quad (\text{C5})$$

Similarly, to respect M_y symmetry, we set

$$\hat{w}_0^{+b\dagger} = \zeta_{M_y} \hat{M}_y \hat{w}_0^{+t\dagger} \hat{M}_y^{-1}, \quad (\text{C6})$$

where $\zeta_{M_y} = \pm 1$. As such, we reduce our degree of freedom on the specification of the trial wave function $\hat{W}_{\mathbf{h}_0}^\dagger$ down to our choice of $\hat{w}_0^{+t\dagger} = \int d^2\mathbf{r} \hat{\psi}_{\beta,\mathbf{r}}^\dagger w_0^{+t}(\mathbf{r})$. Our

only condition on $w_0^{+t}(\mathbf{r})$ is that it is a two-component wave function well localized to 0; however, we simply assume a Gaussian form:

$$w_0^{+t}(\mathbf{r}) = e^{-|\mathbf{r}|^2/2\xi^2} \phi_0^{+t}, \quad (\text{C7})$$

where ξ is the localization length and ϕ_0 is a constant two-component vector. Correspondingly, all the other two-component wave functions take a similar form, although they can be localized to a different point, say, $\bar{C}_3\mathbf{0}$. Such a choice is particularly convenient, as its Fourier transform can be readily evaluated:

$$\gamma_c^\beta(\mathbf{k}) = \int d^2\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} e^{-|\mathbf{r}-\mathbf{c}|^2/2\xi^2} \phi_c^\beta \propto e^{-i\mathbf{k}\cdot\mathbf{c}} e^{-|\mathbf{k}|^2\xi^2/2} \phi_c^\beta, \quad (\text{C8})$$

which enables an efficient computation of the overlap between our trial and the Bloch wave functions.

Thus far, we have focused on only one well-localized wave function in real space. In our problem, the two sites of the effective honeycomb lattice are related by C_6 ; i.e., we simply construct the wave function localized to $\mathbf{h}_1 \equiv C_6\mathbf{h}_0$ through

$$\hat{W}_{\mathbf{h}_1}^\dagger \equiv \hat{C}_6 \hat{W}_{\mathbf{h}_0}^\dagger \hat{C}_6^{-1}. \quad (\text{C9})$$

Besides, to describe a four-band problem, we should have two orbitals on each of the honeycomb sites. These two orbitals are not symmetry related. However, to reproduce the M_y representation in momentum space, we have to take the two orbitals to, respectively, correspond to $\zeta_{M_y} = +1$ and $\zeta_{M_y} = -1$.

In our numerical construction of the Wannier functions, we take the localization length to be $0.15|\mathbf{a}_1|$ and the two-component vectors

$$\begin{aligned} \phi_0^{+t, \zeta_{M_y}=+1} &= \begin{pmatrix} -0.416 + 0.168i \\ 0.820 + 0.356i \end{pmatrix}; \\ \phi_0^{+t, \zeta_{M_y}=-1} &= \begin{pmatrix} 0.296 - 0.380i \\ 0.776 + 0.407i \end{pmatrix}. \end{aligned} \quad (\text{C10})$$

These choices are found simply by a search of the parameter space to optimize the minimum overlap between the trial and the Bloch wave functions. We check the overlap for approximately 1180 momenta along the high-symmetry line $M_M - K_M - \Gamma_M - M_M$, as well as for an additional 1000 randomly sampled points in the BZ. The minimum and maximum over the overlap are, respectively, found to be 0.38 and 3.80, indicating a satisfactory construction of the Wannier functions. In Fig. 3 in the main text, we present the results for the two symmetry-related Wannier functions labeled by $\zeta_{M_y} = +1$; the

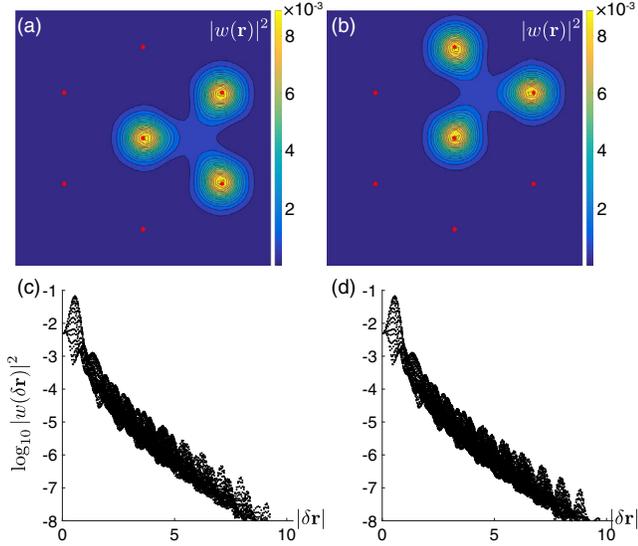


FIG. 5. Localization of the other set of constructed Wannier functions.

corresponding results for the pair with $\zeta_{M_y} = -1$ are shown in Fig. 5. Remarkably, the localization properties of the two sets are essentially identical.

APPENDIX D: TIGHT-BINDING MODEL

In this Appendix, we provide an explicit tabulation of the bonds in the H^{WF} corresponding to Figs. 4(b) and 4(g) in the main text, as well as the associated I_z^{WF} , whose eigenvalues are plotted in Fig. 4(e).

TABLE II. Effective Hamiltonian H^{WF} and valley-charge operator I_z^{WF} . We denote a lattice vector $\mathbf{a} = l_1\mathbf{a}_1 + l_2\mathbf{a}_2$ by (l_1, l_2) .

(a) H^{WF}			
To	Fr	\mathbf{a}	$t(\mu\text{eV})$
1	1	(0,0)	213.8
2	2	(0,0)	213.8
3	3	(0,0)	-213.8
4	4	(0,0)	-213.8
2	3	(0,1)	-76.2
4	1	(0,1)	-76.2
1	1	(0,2)	81.2
1	4	(0,2)	76.2
2	2	(0,2)	81.2
3	2	(0,2)	76.2
3	3	(0,2)	-67.2
4	4	(0,2)	-67.2
2	3	(1,-1)	-76.2
3	3	(1,-1)	66.4
4	1	(1,-1)	-76.2
4	4	(1,-1)	66.4
2	3	(1,1)	76.2

(Table continued)

TABLE II. (Continued)

(a) H^{WF}			
To	Fr	\mathbf{a}	$t(\mu\text{eV})$
4	1	(1,1)	76.2
3	3	(1,2)	66.4
4	4	(1,2)	66.4
1	1	(2,0)	81.2
2	2	(2,0)	81.2
3	3	(2,0)	-67.2
4	4	(2,0)	-67.2
1	4	(2,1)	76.2
3	2	(2,1)	76.2
3	3	(2,1)	66.4
4	4	(2,1)	66.4
1	1	(2,2)	81.2
1	4	(2,2)	-76.2
2	2	(2,2)	81.2
3	2	(2,2)	-76.2
3	3	(2,2)	-67.2
4	4	(2,2)	-67.2
(b) I_z^{WF}			
To	Fr	\mathbf{a}	t
1	2	(0,0)	0.451 <i>i</i>
3	4	(0,0)	-0.443 <i>i</i>
1	2	(0,1)	0.451 <i>i</i>
1	3	(0,1)	-0.217 <i>i</i>
2	4	(0,1)	-0.217 <i>i</i>
3	1	(0,1)	-0.217 <i>i</i>
3	4	(0,1)	-0.443 <i>i</i>
4	2	(0,1)	-0.217 <i>i</i>
1	2	(1,0)	-0.172 <i>i</i>
1	3	(1,0)	-0.217 <i>i</i>
2	1	(1,0)	0.172 <i>i</i>
2	4	(1,0)	-0.217 <i>i</i>
3	1	(1,0)	-0.217 <i>i</i>
3	4	(1,0)	0.181 <i>i</i>
4	2	(1,0)	-0.217 <i>i</i>
4	3	(1,0)	-0.181 <i>i</i>
1	2	(1,1)	0.451 <i>i</i>
1	3	(1,1)	0.217 <i>i</i>
2	4	(1,1)	0.217 <i>i</i>
3	1	(1,1)	0.217 <i>i</i>
3	4	(1,1)	-0.443 <i>i</i>
4	2	(1,1)	0.217 <i>i</i>
1	2	(1,2)	-0.172 <i>i</i>
3	4	(1,2)	0.181 <i>i</i>

In the following, we parametrize a “bond” by

$$t\hat{c}_{\mathbf{r}_{\text{To}}+\mathbf{a}}\hat{c}_{\mathbf{r}_{\text{Fr}}} + \text{H.c.}, \quad (\text{D1})$$

where \mathbf{a} is a moiré lattice vector and To, Fr = 1, ..., 4 labels the four Wannier functions localized to each unit cell. Physically, orbital 1 corresponds to the one localized to \mathbf{h}_0 with $\zeta_{M_y} = +1$; orbital 2 is the one localized to \mathbf{h}_1

symmetry related to orbital 1; orbital 3 is localized to \mathbf{h}_0 with $\zeta_{M_y} = -1$; and orbital 4 is symmetry related to 3.

The bonds in H^{WF} with $t_c = 63\mu\text{eV}$ are tabulated in Table II(a). Note that we tabulated only half of the bonds, in the sense that the Hermitian conjugates of the listed bonds are not included. In particular, for consistency we also halve the coefficient of the on-site chemical potential $\sim t\hat{c}_r^\dagger\hat{c}_r$, which is Hermitian by itself. We in addition subtract the ‘‘trace part’’ of the chemical potential (i.e., we remove a constant energy offset) from the model. The effective valley-charge operator defined with $t'_c = 0.15$ is similarly tabulated in Table II(b).

APPENDIX E: HARTREE-FOCK THEORY FOR SELECTION OF IVC ORDERING

Here, we discuss a simple mean-field treatment to illustrate that an IVC state is favored by the system at $\nu = -2$, which is described by the following simplified Hamiltonian of Eq. (12):

$$H = H_0 + V, \quad (\text{E1})$$

where the free Hamiltonian is

$$H_0 = \sum_{an\mathbf{k}} \epsilon_{an}(\mathbf{k}) c_{ana}^\dagger(\mathbf{k}) c_{ana}(\mathbf{k}) \quad (\text{E2})$$

with a the valley index, n the band index, and α the spin index. Notice that the dispersion $\epsilon_{an}(\mathbf{k})$ is independent of the spin and, due to time reversal, $\epsilon_{an}(\mathbf{k}) = \epsilon_{-an}(-\mathbf{k})$.

$$\begin{aligned} V_{\text{MF}} = & \frac{g}{N} \sum_{\mathbf{k}_1\mathbf{k}_2\mathbf{q}} [\langle c_{ana}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{ana}(\mathbf{k}_1) \rangle c_{a'n'\alpha'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{a'n'\alpha'}(\mathbf{k}_2) + \langle c_{a'n'\alpha'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{a'n'\alpha'}(\mathbf{k}_2) \rangle c_{ana}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{ana}(\mathbf{k}_1) \\ & - \langle c_{ana}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{a'n'\alpha'}(\mathbf{k}_2) \rangle c_{a'n'\alpha'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{ana}(\mathbf{k}_1) - \langle c_{a'n'\alpha'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{ana}(\mathbf{k}_1) \rangle c_{ana}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{a'n'\alpha'}(\mathbf{k}_2) \\ & - \langle c_{ana}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{ana}(\mathbf{k}_1) \rangle \langle c_{a'n'\alpha'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{a'n'\alpha'}(\mathbf{k}_2) \rangle + \langle c_{ana}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{a'n'\alpha'}(\mathbf{k}_2) \rangle \langle c_{a'n'\alpha'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{ana}(\mathbf{k}_1) \rangle]. \end{aligned} \quad (\text{E6})$$

The first, second, and fifth terms are the Hartree contributions, while the other terms are the Fock contributions. The Hartree contribution is determined by the local total electron density alone and independent of ordering, so for our purposes they can be simply dropped. We thus focus on the Fock terms.

We would like to compare the energies of a spin-polarized state (SP), a valley-Z-polarized state (IZP), and a valley-XY-polarized state (IVC). In state SP, we assume

$$\begin{aligned} & \langle c_{ana}^\dagger(\mathbf{k}_1) c_{a'n'\alpha'}(\mathbf{k}_2) \rangle \\ & = \delta_{aa'} \delta_{nn'} [n_{1n}(\mathbf{k}_1) \delta_{\alpha\alpha'} + \phi_{1n}(\mathbf{k}_1) \sigma_{\alpha\alpha'}^z] \delta_{\mathbf{k}_1\mathbf{k}_2}. \end{aligned} \quad (\text{E7})$$

The corresponding macroscopic quantities are defined as

We assume a simple form of interaction:

$$V = \frac{g}{N} \sum_{\mathbf{k}_1\mathbf{k}_2\mathbf{q}} c_{ana}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{ana}(\mathbf{k}_1) c_{a'n'\alpha'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{a'n'\alpha'}(\mathbf{k}_2), \quad (\text{E3})$$

where $n_{a\alpha}(x)$ is the electron density with flavor a and spin α . Repeated indices are summed over here. This interaction has an $SU(8)$ symmetry. As discussed in the main text, the more complete form of the interaction that takes into account the form factors arising from projecting the interactions onto the nearly flat bands should be

$$V = \frac{g}{N} \sum_{\mathbf{k}_1\mathbf{k}_2\mathbf{q}} \Lambda_{nn'}^a(\mathbf{k}_1 + \mathbf{q}, \mathbf{k}_1) \Lambda_{mm'}^{a'}(\mathbf{k}_2 - \mathbf{q}, \mathbf{k}_2) \cdot c_{ana}^\dagger(\mathbf{k}_1 + \mathbf{q}) c_{an'a}(\mathbf{k}_1) \cdot c_{a'm\alpha'}^\dagger(\mathbf{k}_2 - \mathbf{q}) c_{a'm'\alpha'}(\mathbf{k}_2) \quad (\text{E4})$$

with the form factors given by the Bloch wave functions of the states in the nearly flat bands via

$$\Lambda_{nn'}^a(\mathbf{k}_1, \mathbf{k}_2) = \langle u_{an}(\mathbf{k}_1) | u_{an'}(\mathbf{k}_2) \rangle, \quad (\text{E5})$$

where $|u_{an}(\mathbf{k})\rangle$ is the Bloch wave function of a state in the nearly flat bands labeled by valley index a , band index n , and momentum \mathbf{k} (it has no dependence on the spin indices). However, for simplicity, we first present the result from analyzing the simplified interaction (E3) and comment on the preliminary result from analyzing (E4) at the end of this Appendix.

We factorize the interaction in a Hartree-Fock mean-field manner:

$$n_{1n} = \frac{1}{N} \sum_{\mathbf{k}} n_{1n}(\mathbf{k}), \quad \phi_{1n} = \frac{1}{N} \sum_{\mathbf{k}} \phi_{1n}(\mathbf{k}). \quad (\text{E8})$$

In state IZP, we assume

$$\begin{aligned} & \langle c_{ana}^\dagger(\mathbf{k}_1) c_{a'n'\alpha'}(\mathbf{k}_2) \rangle \\ & = \delta_{nn'} \delta_{\alpha\alpha'} [n_{2n}(\mathbf{k}_1) \delta_{aa'} + \phi_2(\mathbf{k}_1) \tau_{aa'}^z] \delta_{\mathbf{k}_1\mathbf{k}_2}. \end{aligned} \quad (\text{E9})$$

The corresponding macroscopic quantities are defined as

$$n_{2n} = \frac{1}{N} \sum_{\mathbf{k}} n_{2n}(\mathbf{k}), \quad \phi_2 = \frac{1}{N} \sum_{\mathbf{k}} \phi_2(\mathbf{k}) \quad (\text{E10})$$

In the IVC state, we assume

$$\begin{aligned} & \langle c_{ana}^\dagger(\mathbf{k}_1) c_{a'n'\alpha'}(\mathbf{k}_2) \rangle \\ & = \delta_{nn'} \delta_{aa'} [n_{3an}(\mathbf{k}_1) \delta_{aa'} + \phi_{3n}(\mathbf{k}_1) \tau_{aa'}^x] \delta_{\mathbf{k}_1 \mathbf{k}_2}. \quad (\text{E11}) \end{aligned}$$

Notice here that ϕ_{3n} is a complex number, and changing its phase implies changing the valley order in the valley- XY plane. For simplicity, we take ϕ_{3n} to be positive for both n . The corresponding macroscopic quantities are defined as

$$n_{3an} = \frac{1}{N} \sum_{\mathbf{k}} n_{3an}(\mathbf{k}), \quad \phi_{3n} = \frac{1}{N} \sum_{\mathbf{k}} \phi_{3n}(\mathbf{k}). \quad (\text{E12})$$

Below, we calculate the energies of these states by the mean-field approximation.

(1) *Spin-polarized state.*—We start with the SP state. In this case, the interaction Hamiltonian is replaced by its mean-field representative, which reads

$$\begin{aligned} V_{\text{MF}} & = g \sum_{\mathbf{k}} [-2\phi_{1n} \sigma_{aa'}^z c_{ana}^\dagger(\mathbf{k}) c_{ana}(\mathbf{k}) \\ & \quad + 4(\phi_{11}^2 + \phi_{12}^2 - n_{11}^2 - n_{12}^2)]. \quad (\text{E13}) \end{aligned}$$

The total mean-field Hamiltonian is then

$$\begin{aligned} H_{\text{MF}} & = H_0 + V_{\text{MF}} \\ & = \sum_{an\mathbf{k}} \{ [\epsilon_{an}(\mathbf{k}) - 2g\phi_{1n}] c_{an+}^\dagger(\mathbf{k}) c_{an+}(\mathbf{k}) \\ & \quad + [\epsilon_{an}(\mathbf{k}) + 2g\phi_{1n}] c_{an-}^\dagger(\mathbf{k}) c_{an-}(\mathbf{k}) \} \\ & \quad + 4gN(\phi_{11}^2 + \phi_{12}^2 - n_{11}^2 - n_{12}^2). \quad (\text{E14}) \end{aligned}$$

Consider the limit where g is much larger than the bandwidth. In this limit, we expect the spin is fully polarized. Without loss of generality, assume $\phi_{1n} \geq 0$ for both n . In this case, all electrons are in the state with $\alpha = +$ and $n = 1$. Self-consistency requires that

$$\begin{aligned} \frac{1}{N} \sum_{\mathbf{k}} \langle c_{a1+}^\dagger(\mathbf{k}) c_{a'1+}(\mathbf{k}) \rangle & = \delta_{aa'} = \delta_{aa'}(n_{11} + \phi_{11}), \\ \frac{1}{N} \sum_{\mathbf{k}} \langle c_{a1-}^\dagger(\mathbf{k}) c_{a'1-}(\mathbf{k}) \rangle & = 0 = \delta_{aa'}(n_{11} - \phi_{11}), \\ \frac{1}{N} \sum_{\mathbf{k}} \langle c_{a2+}^\dagger(\mathbf{k}) c_{a'2+}(\mathbf{k}) \rangle & = 0 = \delta_{aa'}(n_{12} + \phi_{12}), \\ \frac{1}{N} \sum_{\mathbf{k}} \langle c_{a2-}^\dagger(\mathbf{k}) c_{a'2-}(\mathbf{k}) \rangle & = 0 = \delta_{aa'}(n_{12} - \phi_{12}), \end{aligned} \quad (\text{E15})$$

which yields

$$n_{11} = \phi_{11} = \frac{1}{2}, \quad n_{12} = \phi_{12} = 0. \quad (\text{E16})$$

This is indeed a fully polarized state.

In this fully polarized state, and further assuming that the lower flat band is strictly lower than the higher flat band, the total energy of this state is

$$E_1 = \sum_{a\mathbf{k}} \epsilon_{a1}(\mathbf{k}) - 2gN. \quad (\text{E17})$$

(2) *Valley I^z polarized.*—Next, we turn to the state IZP. In this case, the interaction Hamiltonian is replaced by

$$\begin{aligned} V_{\text{MF}} & = g \sum_{\mathbf{k}} [-2\phi_2 \tau_{aa'}^z c_{ana}^\dagger(\mathbf{k}) c_{a'na}(\mathbf{k}) \\ & \quad + 8\phi_2^2 - 4(n_{21}^2 + n_{22}^2)]. \quad (\text{E18}) \end{aligned}$$

The total mean-field Hamiltonian is

$$\begin{aligned} H_{\text{MF}} & = H_0 + V_{\text{MF}} \\ & = \sum_{na\mathbf{k}} \{ [\epsilon_{+n}(\mathbf{k}) - 2g\phi_2] c_{+na}^\dagger(\mathbf{k}) c_{+na}(\mathbf{k}) \\ & \quad + [\epsilon_{-n}(\mathbf{k}) + 2g\phi_2] c_{-na}^\dagger(\mathbf{k}) c_{-na}(\mathbf{k}) \} \\ & \quad + gN[8\phi_2^2 - 4(n_{21}^2 + n_{22}^2)]. \quad (\text{E19}) \end{aligned}$$

Suppose g is much larger than the bandwidth, the system wants to fully polarize along $a = +$, and both bands with $n = 1$ and $n = 2$ are occupied for $a = +$. Self-consistency requires that

$$\begin{aligned} \frac{1}{N} \sum_{\mathbf{k}} \langle c_{+1\alpha}^\dagger(\mathbf{k}) c_{+1\alpha'}(\mathbf{k}) \rangle & = \delta_{\alpha\alpha'} = \delta_{aa'}(n_{21} + \phi_2), \\ \frac{1}{N} \sum_{\mathbf{k}} \langle c_{+2\alpha}^\dagger(\mathbf{k}) c_{+2\alpha'}(\mathbf{k}) \rangle & = \delta_{\alpha\alpha'} = \delta_{aa'}(n_{22} + \phi_2), \\ \frac{1}{N} \sum_{\mathbf{k}} \langle c_{-1\alpha}^\dagger(\mathbf{k}) c_{-1\alpha'}(\mathbf{k}) \rangle & = 0 = \delta_{aa'}(n_{21} - \phi_2), \\ \frac{1}{N} \sum_{\mathbf{k}} \langle c_{-2\alpha}^\dagger(\mathbf{k}) c_{-2\alpha'}(\mathbf{k}) \rangle & = 0 = \delta_{aa'}(n_{22} - \phi_2), \end{aligned} \quad (\text{E20})$$

which yields

$$n_{21} = n_{22} = \phi_2 = \frac{1}{2}. \quad (\text{E21})$$

This is indeed a fully valley- Z -polarized state.

In this fully polarized case, the total energy of this state is

$$E_2 = 2 \sum_{\mathbf{k}} \epsilon_{+1}(\mathbf{k}) - 2gN = E_1. \quad (\text{E22})$$

- (3) *IVC state*.—Finally, we discuss the IVC state. In this case, the interaction Hamiltonian is replaced by

$$V_{\text{MF}} = g \sum_{\mathbf{k}} \{ -2\phi_{3n} [c_{+n\alpha}^\dagger(\mathbf{k})c_{-n\alpha}(\mathbf{k}) + c_{-n\alpha}^\dagger(\mathbf{k})c_{+n\alpha}(\mathbf{k})] + 2(2\phi_{3n}^2 - n_{3an}^2) \}. \quad (\text{E23})$$

The total mean-field Hamiltonian is

$$\begin{aligned} H_{\text{MF}} &= H_0 + V_{\text{MF}} \\ &= \sum_{an\mathbf{k}} \epsilon_{an}(\mathbf{k}) c_{an\alpha}^\dagger(\mathbf{k}) c_{an\alpha}(\mathbf{k}) \\ &\quad - 2g \sum_{na\mathbf{k}} \phi_{3n} [c_{+n\alpha}^\dagger(\mathbf{k})c_{-n\alpha}(\mathbf{k}) + c_{-n\alpha}^\dagger(\mathbf{k})c_{+n\alpha}(\mathbf{k})] \\ &\quad + 2gN(2\phi_{3n}^2 - n_{3an}^2). \end{aligned} \quad (\text{E24})$$

Denote $\bar{\epsilon}_n(\mathbf{k}) = \{[\epsilon_{+n}(\mathbf{k}) + \epsilon_{-n}(\mathbf{k})]/2\}$ and $\delta\epsilon_n(\mathbf{k}) = \{[\epsilon_{+n}(\mathbf{k}) - \epsilon_{-n}(\mathbf{k})]/2\}$, and the spectrum of the above Hamiltonian is

$$E_{\pm n}(\mathbf{k}) = \bar{\epsilon}_n(\mathbf{k}) \pm \sqrt{\delta\epsilon_n(\mathbf{k})^2 + 4g^2\phi_{3n}^2}. \quad (\text{E25})$$

Denote the eigenmodes corresponding to $E_{\pm n}(\mathbf{k})$ by $d_{\pm n}(\mathbf{k})$, and they satisfy

$$\begin{pmatrix} c_{+n}(\mathbf{k}) \\ c_{-n}(\mathbf{k}) \end{pmatrix} = \begin{pmatrix} \cos\frac{\theta_n(\mathbf{k})}{2} & -\sin\frac{\theta_n(\mathbf{k})}{2} \\ \sin\frac{\theta_n(\mathbf{k})}{2} & \cos\frac{\theta_n(\mathbf{k})}{2} \end{pmatrix} \begin{pmatrix} d_{+n}(\mathbf{k}) \\ d_{-n}(\mathbf{k}) \end{pmatrix} \quad (\text{E26})$$

with

$$\begin{aligned} \cos\theta_n(\mathbf{k}) &= \frac{\delta\epsilon_n(\mathbf{k})}{\sqrt{\delta\epsilon_n(\mathbf{k})^2 + 4g^2\phi_{3n}^2}}, \\ \sin\theta_n(\mathbf{k}) &= \frac{-2g\phi_{3n}}{\sqrt{\delta\epsilon_n(\mathbf{k})^2 + 4g^2\phi_{3n}^2}}. \end{aligned} \quad (\text{E27})$$

Again, consider the limit where g is much larger than the bandwidth; then the system tends to occupy the bands with energies E_{-1} . Self-consistency requires that

$$\begin{aligned} \phi_{3n} &= \frac{1}{N} \sum_{\mathbf{k}} \langle c_{+n\alpha}^\dagger(\mathbf{k})c_{-n\alpha}(\mathbf{k}) \rangle = \sum_{\mathbf{k}} \frac{\sin\theta_n(\mathbf{k})}{2N} \langle d_{+n\alpha}^\dagger(\mathbf{k})d_{+n\alpha}(\mathbf{k}) - d_{-n\alpha}^\dagger(\mathbf{k})d_{-n\alpha}(\mathbf{k}) \rangle \\ &= \frac{1}{N} \sum_{\mathbf{k}} \frac{g\phi_{3n}}{\sqrt{\delta\epsilon_n(\mathbf{k})^2 + 4g^2\phi_{3n}^2}} \langle d_{-n\alpha}^\dagger(\mathbf{k})d_{-n\alpha}(\mathbf{k}) - d_{+n\alpha}^\dagger(\mathbf{k})d_{+n\alpha}(\mathbf{k}) \rangle, \\ n_{3+n} &= \frac{1}{N} \sum_{\mathbf{k}} \left(\frac{1 + \cos\theta_n(\mathbf{k})}{2} \langle d_{+n\alpha}^\dagger(\mathbf{k})d_{+n\alpha}(\mathbf{k}) \rangle + \frac{1 - \cos\theta_n(\mathbf{k})}{2} \langle d_{-n\alpha}^\dagger(\mathbf{k})d_{-n\alpha}(\mathbf{k}) \rangle \right), \\ n_{3-n} &= \frac{1}{N} \sum_{\mathbf{k}} \left(\frac{1 + \cos\theta_n(\mathbf{k})}{2} \langle d_{-n\alpha}^\dagger(\mathbf{k})d_{-n\alpha}(\mathbf{k}) \rangle + \frac{1 - \cos\theta_n(\mathbf{k})}{2} \langle d_{+n\alpha}^\dagger(\mathbf{k})d_{+n\alpha}(\mathbf{k}) \rangle \right). \end{aligned} \quad (\text{E28})$$

In the limit where g is much larger than the bandwidth, $\phi_{3n} \rightarrow 1/2$, which means the system tends to fully polarize along the valley- XY direction. At the same time, $n_{an} \rightarrow 1/2$. The total energy of this state in this case is

$$E_3 = 2 \sum_{\mathbf{k}} [\bar{\epsilon}_1(\mathbf{k}) - \sqrt{\delta\epsilon_1(\mathbf{k})^2 + g^2}]. \quad (\text{E29})$$

This energy is lower than E_2 in Eq. (E22):

$$\begin{aligned} E_3 - E_2 &= 2 \sum_{\mathbf{k}} [\bar{\epsilon}_1(\mathbf{k}) - \sqrt{\delta\epsilon_1(\mathbf{k})^2 + g^2} - \epsilon_{+1}(\mathbf{k}) + g] \\ &= 2 \sum_{\mathbf{k}} [g - \delta\epsilon_1(\mathbf{k}) - \sqrt{\delta\epsilon_1(\mathbf{k})^2 + g^2}] < 0, \end{aligned} \quad (\text{E30})$$

where time-reversal symmetry of the noninteracting band structure is used in the last step: $\sum_{\mathbf{k}} \delta\epsilon_1(\mathbf{k}) = 0$.

Therefore, if the interaction strength is much larger than the bandwidth, the system tends to be fully polarized. Within this analysis, among different fully polarized states, the valley- XY -ordered state (IVC) has a lower energy than a spin-polarized state and a valley- Z -ordered state.

Finally, we comment on the effect of taking into account the form factors into the interaction term, as in Eq. (E4). It turns out that in this case which state has the lowest energy can potentially be changed from the IVC state if bands are too flat, and our preliminary calculations show that the spin-polarized state and the valley- Z -ordered state have a lower energy compared to the IVC state when the interaction strength is around 10 times the bandwidth of the nearly flat bands. A more reliable way to settle down the real ground state is to do a numerical calculation formulated

in momentum space, using the Hamiltonian given by Eqs. (E1), (E2), and (E4).

APPENDIX F: SPIN-ORBITAL MODEL FOR MOTT INSULATORS IN TRILAYER GRAPHENE

In this Appendix, we derive an effective spin-orbital model applicable to the trilayer graphene described in Sec. IX, in the limit of $U \gg t$. This effective model is applicable for both the cases of $\nu = -1$ and $\nu = -2$, and it is obtained by a systematic expansion in the large- U limit to the order of t^3/U^2 .

Besides the translation symmetry of the triangular lattice, the system is assumed to have $U(2) \times U(2)$ symmetries, corresponding to charge and flavor $U(1)$ conservations, as well as spin conservation. In addition, there is a C_6 symmetry that maps one flavor into the other and a time-reversal symmetry that also maps one flavor into the other (while leaving the spin unchanged, so $T^2 = 1$ for this time reversal).

We start from a Hubbard model with nearest-neighbor hopping and on-site Hubbard interactions. Consider three nearby sites (A , B , and C) forming an elementary triangle of this triangular lattice. The kinetic Hamiltonian is taken as

$$H_0 = -t \sum_{\alpha\alpha'} e^{i\eta_\alpha \phi} [c_{\alpha\alpha'}^\dagger(B) c_{\alpha\alpha}(A) + c_{\alpha\alpha'}^\dagger(C) c_{\alpha\alpha}(B) + c_{\alpha\alpha'}^\dagger(A) c_{\alpha\alpha}(C)] + \text{H.c.}, \quad (\text{F1})$$

where $a = \pm$ labels the flavor, α labels the spin, and $\eta_a = \pm 1$ if $a = \pm$. The interaction Hamiltonian on each site is taken as

$$V = \frac{U}{2} (N - N_0)^2, \quad (\text{F2})$$

where $N = \sum_{\alpha\alpha'} c_{\alpha\alpha}^\dagger(\mathbf{r}_i) c_{\alpha\alpha}(\mathbf{r}_i)$, and we are interested in both the cases with $N_0 = 2$ and $N_0 = 1$. All other terms of the many-body system can be generated by applying various symmetries.

We are looking for an effective spin-orbital model at large U to the order of t^3/U^2 . We first present the result and discuss some simple physical consequences before presenting the details of the derivation. The final result is

$$H_{\text{eff}} = H^{(2)} + H^{(3)} \quad (\text{F3})$$

with

$$H^{(2)} = \frac{2t^2}{U} \frac{1}{n^2} (n^2 + \mathbf{S}_A \cdot \mathbf{S}_B)(n^2 + \mathbf{I}_A \cdot \mathbf{I}_B) - \frac{4nt^2}{U} \quad (\text{F4})$$

and

$$H^{(3)} = \frac{3t^3}{U^2} \left[-\frac{1}{4n^3} h^{(3,1)} + \frac{1}{4n^2} h^{(3,2)} \right] - \frac{12nt^3}{U^2}, \quad (\text{F5})$$

where

$$\begin{aligned} h^{(3,1)} &= 8n^2 \cos 3\phi (n^2 + \mathbf{S}_A \cdot \mathbf{S}_B + \mathbf{S}_B \cdot \mathbf{S}_C + \mathbf{S}_C \cdot \mathbf{S}_A) (n^2 + I_A^z I_B^z + I_B^z I_C^z + I_C^z I_A^z) \\ &\quad + 4n^2 (n^2 + \mathbf{S}_A \cdot \mathbf{S}_B + \mathbf{S}_B \cdot \mathbf{S}_C + \mathbf{S}_C \cdot \mathbf{S}_A) \cdot [e^{i\phi} (I_A^+ I_B^- + I_B^+ I_C^- + I_C^+ I_A^-) + \text{H.c.}] \\ &\quad + 8 \sin 3\phi [I_A^z I_B^z I_C^z + n^2 (I_A^z + I_B^z + I_C^z)] \cdot [\mathbf{S}_A \cdot (\mathbf{S}_B \times \mathbf{S}_C)] \\ &\quad - 4i \mathbf{S}_A \cdot (\mathbf{S}_B \times \mathbf{S}_C) \cdot [I_A^z (e^{i\phi} I_B^+ I_C^- - e^{-i\phi} I_B^- I_C^+) + (A \rightarrow B \rightarrow C \rightarrow A)] \end{aligned} \quad (\text{F6})$$

and

$$\begin{aligned} h^{(3,2)} &= T_B^{kj} T_C^{jk} e^{2i\phi(I_B^z + 2I_C^z)} + e^{-2i\phi(I_B^z + 2I_C^z)} T_B^{jk} T_C^{kj} + (B \rightarrow C \rightarrow A \rightarrow B) \\ &= 4 \cdot (n^2 + \mathbf{S}_A \cdot \mathbf{S}_B) \cdot (2n^2 \cos 3\phi + 2I_A^z I_B^z \cos 3\phi + e^{-i\phi} I_A^+ I_B^- + e^{i\phi} I_A^- I_B^+) + (B \rightarrow C \rightarrow A \rightarrow B). \end{aligned} \quad (\text{F7})$$

In the above, $n = \frac{1}{2}$ for $N_0 = 1$, and $n = 1$ for $N_0 = 2$.

We briefly comment on these results before turning to the detailed derivation. First, we notice the effective model indeed has the same set of symmetries as the original Hubbard model. Second, we note that $H^{(2)}$ is actually $SU(4)$ invariant, and $H^{(3)}$ is also $SU(4)$ invariant at $\phi = 0$, which is most easily seen by inspecting Eqs. (F19) and (F30). These $SU(4)$ symmetric interactions can potentially make the system an $SU(4)$ antiferromagnet. Third, there is a spin chirality term for the two valleys with opposite

coefficients, which can potentially lead to interesting kinds of topological order, such as a double-semion state. Last, we note the system may develop an $SU(4)$ antiferromagnetic order in the limit where $U \gg t$ at $N_0 = 2$, although there is evidence that the system is disordered at $N_0 = 1$. Suppose the $SU(4)$ antiferromagnetic Heisenberg model indeed results in an $SU(4)$ -broken state; we would like to understand how the $SU(4)$ -breaking terms in the Hamiltonian affects the ground state.

To this end, we expand $H^{(3)}$ for small ϕ and obtain

$$\begin{aligned}
h^{(3,1)} = & 8n^2(n^2 + \mathbf{S}_A \cdot \mathbf{S}_B + \mathbf{S}_B \cdot \mathbf{S}_C + \mathbf{S}_C \cdot \mathbf{S}_A)(n^2 + \mathbf{I}_A \cdot \mathbf{I}_B + \mathbf{I}_B \cdot \mathbf{I}_C + \mathbf{I}_C \cdot \mathbf{I}_A) - 8[\mathbf{S}_A \cdot (\mathbf{S}_B \times \mathbf{S}_C)] \cdot [\mathbf{I}_A \cdot (\mathbf{I}_B \times \mathbf{I}_C)] \\
& + 24\phi[I_A^z I_B^z I_C^z + n^2(I_A^z + I_B^z + I_C^z)] \cdot [\mathbf{S}_A \cdot (\mathbf{S}_B \times \mathbf{S}_C)] \\
& - 8\phi n^2(n^2 + \mathbf{S}_A \cdot \mathbf{S}_B + \mathbf{S}_B \cdot \mathbf{S}_C + \mathbf{S}_C \cdot \mathbf{S}_A)[I_A^x I_B^y - I_A^y I_B^x + (A \rightarrow B \rightarrow C \rightarrow A)] \\
& + 4\phi \mathbf{S}_A \cdot (\mathbf{S}_B \times \mathbf{S}_C) \cdot [I_A^z(I_B^+ I_C^- + I_B^- I_C^+) + (A \rightarrow B \rightarrow C \rightarrow A)] + \mathcal{O}(\phi^2)
\end{aligned} \tag{F8}$$

and

$$\begin{aligned}
h^{(3,2)} = & 8(n^2 + \mathbf{S}_A \cdot \mathbf{S}_B)(n^2 + \mathbf{I}_A \cdot \mathbf{I}_B) \\
& + (A \rightarrow B \rightarrow C \rightarrow A) + 8\phi(n^2 + \mathbf{S}_A \cdot \mathbf{S}_B)(I_A^x I_B^y - I_A^y I_B^x) \\
& + (A \rightarrow B \rightarrow C \rightarrow A) + \mathcal{O}(\phi^2).
\end{aligned} \tag{F9}$$

As we can see, the main effect of $SU(4)$ -breaking interactions to the leading order of ϕ , besides giving rise to the spin chirality terms, is to introduce terms roughly in the following form to the Hamiltonian:

$$\delta H = J\phi[I_A^y I_B^x - I_A^x I_B^y + (A \rightarrow B \rightarrow C \rightarrow A)]. \tag{F10}$$

Consider the I 's as classical spins on the XY plane and parametrize $I_A^x = \cos \theta_A$ and $I_A^y = \sin \theta_A$; the above Hamiltonian becomes

$$\delta H = J\phi \sin(\theta_A - \theta_B) + (A \rightarrow B \rightarrow C \rightarrow A). \tag{F11}$$

This term tends to introduce some canting of the I order to gain energy.

Below, we derive effective spin-orbital Hamiltonians of such systems in the large- U limit to the order of t^3/U^2 , for $N_0 = 1$ and $N_0 = 2$ separately. We use the standard Van Vleck perturbation theory to derive the effective Hamiltonian, which involves two steps: writing down the matrix elements of the effective Hamiltonian within the low-energy manifold and expressing these matrix elements in terms of charge-neutral operators.

In this problem, the effective Hamiltonian can be written as

$$H_{\text{eff}} = H^{(2)} + H^{(3)} + \dots \tag{F12}$$

with

$$H^{(2)} = \mathcal{P}H_0\mathcal{D}H_0\mathcal{P} \tag{F13}$$

and

$$H^{(3)} = \mathcal{P}H_0\mathcal{D}H_0\mathcal{D}H_0\mathcal{P}, \tag{F14}$$

where \mathcal{P} is the projector into the ground-state manifold of V and

$$\mathcal{D} = \frac{1 - \mathcal{P}}{E_0 - V}. \tag{F15}$$

1. Mott insulator at $N_0 = 1$

We first discuss the effective Hamiltonian for $N_0 = 1$. We first write the results in terms of some $SU(4)$ generators and convert them into a form in terms of the \mathbf{S} and \mathbf{I} operators later. The final result in terms of $SU(4)$ generators is

$$H^{(2)} = \frac{2t^2}{U} \sum_{i,j=1}^4 T_A^{ij} T_B^{ji} - \frac{2t^2}{U}, \tag{F16}$$

where

$$T^{ij} = c_i^\dagger c_j, \tag{F17}$$

with $i = 1, 2, 3, 4$, and $|1\rangle = |+\uparrow\rangle$, $|2\rangle = |+\downarrow\rangle$, $|3\rangle = |-\uparrow\rangle$, and $|4\rangle = |-\downarrow\rangle$. In terms of these states, the action of T^{ij} is

$$T^{ij} = \delta_{jk}|i\rangle, \tag{F18}$$

and

$$\begin{aligned}
H^{(3)} = & -\frac{6t^3}{U^2} \sum_{i,j,k=1}^4 [T_A^{ji} T_B^{kj} T_C^{ik} e^{-2i\phi(I_A^z + I_B^z + I_C^z)} \\
& + e^{2i\phi(I_A^z + I_B^z + I_C^z)} T_A^{ij} T_B^{jk} T_C^{ki}] \\
& + \frac{3t^3}{U^2} \sum_{j,k=1}^4 [T_B^{kj} T_C^{jk} e^{2i\phi(I_B^z + 2I_C^z)} + e^{-2i\phi(I_B^z + 2I_C^z)} T_B^{jk} T_C^{kj} \\
& + (B \rightarrow C \rightarrow A \rightarrow B)] - \frac{6t^3}{U^2} \cos 3\phi.
\end{aligned} \tag{F19}$$

The details of the calculations are below.

a. Order t^2/U

To get the effective Hamiltonian at the order t^2/U , it is sufficient to consider a single bond of the triangular lattice, and all other terms can be generated by applying symmetries.

Now we first calculate the matrix elements of the effective Hamiltonian at the order t^2/U . Denote the two sites linked by this bond by A and B , and then $|i, j\rangle = c_i^\dagger(A)c_j^\dagger(B)|0\rangle$.

There are only two types of nonzero matrix elements:

$$\begin{aligned}\langle i, j|H^{(2)}|i, j\rangle &= -\frac{2t^2}{U}, \\ \langle j, i|H^{(2)}|i, j\rangle &= \frac{2t^2}{U}\end{aligned}\quad (\text{F20})$$

for $i \neq j$.

These matrix elements can be recast into the effective Hamiltonian

$$H^{(2)} = \frac{2t^2}{U} \sum_{i,j=1}^4 T_A^{ij} T_B^{ji} - \frac{2t^2}{U}, \quad (\text{F21})$$

where

$$T^{ij} = c_i^\dagger c_j \quad (\text{F22})$$

such that

$$T^{ij}|k\rangle = \delta_{jk}|i\rangle. \quad (\text{F23})$$

These operators satisfy the commutation relation of $SU(4)$ generators:

$$[T^{ij}, T^{kl}] = \delta_{jk} T^{il} - \delta_{il} T^{kj}. \quad (\text{F24})$$

b. Order t^3/U^2

Now we turn to the order t^3/U^2 . A simple inspection of the model shows that, to calculate the effective Hamiltonian at the order of t^3/U^2 in this problem, we need to consider only a single elementary triangle; then all other terms can be obtained by symmetries.

Now we first calculate the matrix elements of the effective Hamiltonian to the order t^3/U^2 . To this end, we need to first have a systematic way to label the states in the ground-state manifold of the three-site problem of a single elementary triangle. It turns out that we can classify the states into three classes: $|i, j, k\rangle$, $|i, i, k\rangle$, and $|i, i, i\rangle$ with $i \neq j \neq k \neq i$, such that the matrix elements between states from two different classes always vanish. Now we need only to calculate the matrix elements between states within the same classes.

The results are

$$\langle i, i, i|H^{(3)}|i, i, i\rangle = 0, \quad (\text{F25})$$

$$\begin{aligned}\langle i, i, j|H^{(3)}|i, i, j\rangle &= -\frac{t^3}{U^2}(-2\cos\eta_i\phi + 2\cos\eta_j\phi) = 0, \\ \langle i, j, i|H^{(3)}|i, i, j\rangle &= \frac{3t^3}{U^2}(e^{i\eta_i\phi+2i\eta_j\phi} - e^{-i\eta_j\phi-2i\eta_i\phi}),\end{aligned}\quad (\text{F26})$$

$$\begin{aligned}\langle i, j, k|H^{(3)}|i, j, k\rangle &= -\frac{2t^3}{U^2}(\cos 3\eta_i\phi + \cos 3\eta_j\phi + \cos 3\eta_k\phi) \\ &= -\frac{6t^3}{U^2}\cos 3\phi,\end{aligned}$$

$$\langle i, k, j|H^{(3)}|i, j, k\rangle = \frac{3t^3}{U^2}(e^{2i\eta_k\phi+i\eta_j\phi} + e^{-2i\eta_j\phi-i\eta_k\phi}),$$

$$\langle j, k, i|H^{(3)}|i, j, k\rangle = -\frac{6t^3}{U^2}e^{-i\eta_i\phi-i\eta_j\phi-i\eta_k\phi},$$

$$\langle k, i, j|H^{(3)}|i, j, k\rangle = -\frac{6t^3}{U^2}e^{i\eta_i\phi+i\eta_j\phi+i\eta_k\phi}. \quad (\text{F27})$$

Recasting these matrix elements into a compact form yields

$$\begin{aligned}H^{(3)} &= -\frac{6t^3}{U^2} \sum_{i,j,k=1}^4 [T_A^{ji} T_B^{kj} T_C^{ik} e^{-2i\phi(I_A^z + I_B^z + I_C^z)} \\ &\quad + e^{2i\phi(I_A^z + I_B^z + I_C^z)} T_A^{ij} T_B^{jk} T_C^{ki}] \\ &\quad + \frac{3t^3}{U^2} \sum_{j,k=1}^4 [T_B^{kj} T_C^{jk} e^{2i\phi(I_B^z + 2I_C^z)} + e^{-2i\phi(I_B^z + 2I_C^z)} T_B^{jk} T_C^{kj} \\ &\quad + (B \rightarrow C \rightarrow A \rightarrow B)] - \frac{6t^3}{U^2} \cos 3\phi.\end{aligned}\quad (\text{F28})$$

2. Mott insulator at $N_0 = 2$

Now we discuss the effective Hamiltonian for $N_0 = 2$. In this case, there are six states on each site, which can be denoted as $|ij\rangle \equiv c_i^\dagger c_j^\dagger |0\rangle = -|ji\rangle$.

Again, we first write the result in terms of some $SU(4)$ generators and then convert it to a form in terms of the original S and I operators. The final result is

$$H^{(2)} = \frac{2t^2}{U} \sum_{ij} T_A^{ij} T_B^{ji} - \frac{4t^2}{U}, \quad (\text{F29})$$

where $T^{ij} = c_i^\dagger c_j$. And

$$\begin{aligned}H^{(3)} &= -\frac{6t^3}{U^2} [T_A^{lj} T_B^{jk} T_C^{kl} e^{2i\phi(\tilde{I}_A^z + \tilde{I}_B^z + \tilde{I}_C^z)} \\ &\quad + e^{-2i\phi(\tilde{I}_A^z + \tilde{I}_B^z + \tilde{I}_C^z)} T_A^{jl} T_B^{kj} T_C^{lk}] \\ &\quad + \frac{3t^3}{U^2} [T_B^{ij} T_C^{ji} e^{2i\phi(\tilde{I}_B^z + 2\tilde{I}_C^z)} + \text{H.c.} + (B \rightarrow C \rightarrow A \rightarrow B)] \\ &\quad - \frac{12t^3}{U^2} \cos 3\phi,\end{aligned}\quad (\text{F30})$$

where \tilde{I}_A^z gives the flavor of the particle that is acted by the T operators. For example, $T_A^{lj} T_B^{jk} T_C^{kl} e^{2i\phi(\tilde{I}_A^z + \tilde{I}_B^z + \tilde{I}_C^z)} = T_A^{lj} T_B^{jk} T_C^{kl} e^{i\phi(\eta_j + \eta_k + \eta_l)} = e^{2i\phi(\tilde{I}_A^z + \tilde{I}_B^z + \tilde{I}_C^z)} T_A^{lj} T_B^{jk} T_C^{kl}$.

The details of the calculations are below.

a. Order t^2/U

To get the effective Hamiltonian at the order t^2/U , it is sufficient to consider a single bond of the triangular lattice, and all other terms can be generated by applying symmetries.

Now we first calculate the matrix elements of the effective Hamiltonian at the order of t^2/U . Denote the two sites linked by this bond by A and B , and then $|ij, kl\rangle = c_i^\dagger(A)c_j^\dagger(A)c_k^\dagger(B)c_l^\dagger(B)|0\rangle$.

It is useful to distinguish three types of states: $|ij, kl\rangle$, $|ij, ik\rangle$, and $|ij, ij\rangle$, where different letters denote different states. Clearly, there are no matrix elements between two states from two different types, and all we need is to calculate the matrix elements between states within the same type.

The independent matrix elements include

$$\langle ij, ij|H^{(2)}|ij, ij\rangle = 0, \quad (\text{F31})$$

$$\langle ij, ik|H^{(2)}|ij, ik\rangle = -\frac{2t^2}{U},$$

$$\langle ij, ik|H^{(2)}|ik, ij\rangle = \frac{2t^2}{U}, \quad (\text{F32})$$

$$\langle ij, kl|H^{(2)}|ij, kl\rangle = -\frac{4t^2}{U},$$

$$\langle ik, jl|H^{(2)}|ij, kl\rangle = \frac{2t^2}{U}. \quad (\text{F33})$$

All other matrix elements at this order either vanish or can be obtained from the above by permutations.

From these matrix elements, we obtain $H^{(2)}$:

$$H^{(2)} = \frac{2t^2}{U} \sum_{ij} T_A^{ij} T_B^{ji} - \frac{4t^2}{U}. \quad (\text{F34})$$

As in the case of $N_0 = 1$, $T^{ij} = c_i^\dagger c_j$ that satisfies Eq. (F24), the commutation relations of the generators of $SU(4)$. Now acting on two-particle states on each site, the actions of these operators are

$$T^{ij}|kl\rangle = \delta_{jk}|il\rangle + \delta_{jl}|ki\rangle. \quad (\text{F35})$$

b. Order t^3/U^2

Now we turn to the order t^3/U^2 . A simple inspection of the model shows that, to calculate the effective Hamiltonian at the order of t^3/U^2 in this problem, we need to consider only a single elementary triangle, and then all other terms can be obtained by symmetries.

Now we first calculate the matrix elements of the effective Hamiltonian to the order t^3/U^2 . To this end, we need to first have a systematic way to label the states in the ground-state manifold of the three-site problem of a single elementary triangle.

It turns out there are five types of states: $|ij, ij, ij\rangle$, $|ij, jk, ij\rangle$, $|ij, jk, ik\rangle$, $\{|ij, kl, il\rangle, |ij, kl, ij\rangle\}$, and $|ij, ik, il\rangle$, where, for example,

$$|ij, jk, ki\rangle = c_i^\dagger(A)c_j^\dagger(A)c_j^\dagger(B)c_k^\dagger(B)c_k^\dagger(C)c_i^\dagger(C)|0\rangle. \quad (\text{F36})$$

All other states can be related to these states by certain permutations.

Now we calculate the matrix elements of the effective model between these states at the order of t^3/U^2 . The matrix elements between different types of the above states always vanish. So we need to calculate only the matrix elements between states within the same type:

$$\langle ij, ij, ij|H^{(3)}|ij, ij, ij\rangle = 0, \quad (\text{F37})$$

$$\langle ij, jk, ij|H^{(3)}|ij, jk, ij\rangle = -\frac{t^3}{U^2} [2 \cos(3\eta_k \phi) - 2 \cos(3\eta_i \phi)] = 0,$$

$$\langle ij, ij, jk|H^{(3)}|ij, jk, ij\rangle = -\frac{3t^3}{U^2} (e^{i\eta_k \phi + 2i\eta_i \phi} - e^{-2i\eta_k \phi - i\eta_i \phi}), \quad (\text{F38})$$

$$\langle ij, jk, ik|H^{(3)}|ij, jk, ik\rangle = -\frac{t^3}{U^2} [-2 \cos(3\eta_i \phi) - 2 \cos(3\eta_j \phi) - 2 \cos(3\eta_k \phi)] = \frac{6t^3}{U^2} \cos 3\phi,$$

$$\langle ij, ik, jk|H^{(3)}|ij, jk, ik\rangle = -\frac{3t^3}{U^2} (e^{2i\eta_i \phi + i\eta_j \phi} + e^{-2i\eta_j \phi - i\eta_i \phi}), \quad (\text{F39})$$

$$\langle ij, kl, il|H^{(3)}|ij, kl, il\rangle = -\frac{t^3}{U^2} [2 \cos(3\eta_j \phi) + 2 \cos(3\eta_k \phi) - 2 \cos(3\eta_l \phi) - 2 \cos(3\eta_i \phi)] = 0,$$

$$\langle ij, kl, il|H^{(3)}|ij, il, kl\rangle = \frac{3t^3}{U^2} (e^{2i\eta_k \phi + i\eta_i \phi} - e^{-i\eta_k \phi - 2i\eta_i \phi}),$$

$$\langle ij, kl, ij|H^{(3)}|ij, kl, ij\rangle = -\frac{t^3}{U^2} [2 \cos(3\eta_k \phi) + 2 \cos(3\eta_l \phi) - 2 \cos(3\eta_j \phi) - 2 \cos(3\eta_i \phi)] = 0,$$

$$\langle ij, kl, ij|H^{(3)}|ij, ij, kl\rangle = 0, \quad (\text{F40})$$

$$\begin{aligned}
\langle ij, kl, ij | H^{(3)} | ik, lj, ij \rangle &= -\frac{3t^3}{U^2} (e^{i\eta_k\phi + 2i\eta_j\phi} - e^{-2i\eta_k\phi - i\eta_j\phi}), \\
\langle ij, ik, il | H^{(3)} | ij, ik, il \rangle &= -\frac{t^3}{U^2} [2 \cos(3\eta_j\phi) + 2 \cos(3\eta_k\phi) + 2 \cos(3\eta_l\phi)] = -\frac{6t^3}{U^2} \cos 3\phi, \\
\langle ij, il, ik | H^{(3)} | ij, ik, il \rangle &= \frac{3t^3}{U^2} (e^{i\eta_k\phi + 2i\eta_l\phi} + e^{-i\eta_l\phi - 2i\eta_k\phi}), \\
\langle il, ij, ik | H^{(3)} | ij, ik, il \rangle &= -\frac{6t^3}{U^2} e^{i\eta_j\phi + i\eta_k\phi + i\eta_l\phi}.
\end{aligned} \tag{F41}$$

Recasting these matrix elements into a compact form yields

$$\begin{aligned}
H^{(3)} &= -\frac{6t^3}{U^2} [T_A^{lj} T_B^{jk} T_C^{kl} e^{2i\phi(\tilde{I}_A^z + \tilde{I}_B^z + \tilde{I}_C^z)} + e^{-2i\phi(\tilde{I}_A^z + \tilde{I}_B^z + \tilde{I}_C^z)} T_A^{jl} T_B^{kj} T_C^{lk}] \\
&\quad + \frac{3t^3}{U^2} [T_B^{ij} T_C^{ji} e^{2i\phi(\tilde{I}_B^z + 2\tilde{I}_C^z)} + \text{H.c.} + (B \rightarrow C \rightarrow A \rightarrow B)] - \frac{12t^3}{U^2} \cos 3\phi,
\end{aligned} \tag{F42}$$

where \tilde{I}_A^z gives the flavor of the particle that is acted by the T operators.

3. Effective models in terms of spin and orbital operators

As seen in the above, the effective Hamiltonian expressed in terms the operators T^{ij} is relatively concise, and they are the same for both $N_0 = -1$ and $N_0 = -2$ up to some constants. However, to gain more intuition, it is helpful to express these effective Hamiltonians in terms of spin operators \mathbf{S} and valley operator \mathbf{I} , where

$$\begin{aligned}
S^+ &= c_{+\uparrow}^\dagger c_{+\downarrow} + c_{-\uparrow}^\dagger c_{-\downarrow} = T^{12} + T^{34}, & S^- &= c_{+\downarrow}^\dagger c_{+\uparrow} + c_{-\downarrow}^\dagger c_{-\uparrow} = T^{21} + T^{43}, \\
S^z &= \frac{1}{2} (c_{+\uparrow}^\dagger c_{+\uparrow} + c_{-\uparrow}^\dagger c_{-\uparrow} - c_{+\downarrow}^\dagger c_{+\downarrow} - c_{-\downarrow}^\dagger c_{-\downarrow}) = \frac{1}{2} (T^{11} + T^{33} - T^{22} - T^{44}), \\
I^+ &= c_{+\uparrow}^\dagger c_{-\uparrow} + c_{+\downarrow}^\dagger c_{-\downarrow} = T^{13} + T^{24}, & I^- &= c_{-\uparrow}^\dagger c_{+\uparrow} + c_{-\downarrow}^\dagger c_{+\downarrow} = T^{31} + T^{42}, \\
I^z &= \frac{1}{2} (c_{+\uparrow}^\dagger c_{+\uparrow} + c_{+\downarrow}^\dagger c_{+\downarrow} - c_{-\uparrow}^\dagger c_{-\uparrow} - c_{-\downarrow}^\dagger c_{-\downarrow}) = \frac{1}{2} (T^{11} + T^{22} - T^{33} - T^{44}),
\end{aligned} \tag{F43}$$

as well as filling fraction

$$n = \frac{1}{2} (c_{+\uparrow}^\dagger c_{+\uparrow} + c_{+\downarrow}^\dagger c_{+\downarrow} + c_{-\uparrow}^\dagger c_{-\uparrow} + c_{-\downarrow}^\dagger c_{-\downarrow}) = \frac{1}{2} (T^{11} + T^{22} + T^{33} + T^{44}). \tag{F44}$$

Here, $n = 1/2$ means $N_0 = 1$ and $n = 1$ means $N_0 = 2$. In the above, \mathbf{S} and \mathbf{I} form two decoupled $SU(2)$ algebras, and n commutes with all others.

a. Effective Hamiltonian for $N_0 = 1$

To this end, we first reexpress the operators T^{ij} in terms of \mathbf{S} , \mathbf{I} , and n . For $n = 1/2$ ($N_0 = 1$),

$$\begin{aligned}
T^{11} &= (n + S^z)(n + I^z), & T^{12} &= S^+(n + I^z), & T^{13} &= I^+(n + S^z), & T^{14} &= S^+ I^+, \\
T^{22} &= (n - S^z)(n + I^z), & T^{23} &= S^- I^+, & T^{24} &= I^+(n - S^z), \\
T^{33} &= (n + S^z)(n - I^z), & T^{34} &= S^+(n - I^z), \\
T^{44} &= (n - S^z)(n - I^z).
\end{aligned} \tag{F45}$$

Substituting these into Eqs. (F16) and (F19) yields

$$\begin{aligned}
\left(H^{(2)} + \frac{2t^2}{U}\right) \cdot \frac{U}{2t^2} &= \sum_{ij} T_A^{ij} T_B^{ji} \\
&= \frac{1}{2} (S_A^+ S_B^- + I_A^+ I_B^- + \text{H.c.}) + (S_A^+ S_B^- + S_A^- S_B^+) (I_A^+ I_B^- + I_A^- I_B^+) + 2I_A^z I_B^z (S_A^+ S_B^- + S_A^- S_B^+) \\
&\quad + 2S_A^z S_B^z (I_A^+ I_B^- + I_A^- I_B^+) + 4(n^2 + S_A^z S_B^z) (n^2 + I_A^z I_B^z) \\
&= 4 \cdot (n^2 + \mathbf{S}_A \cdot \mathbf{S}_B) \cdot (n^2 + \mathbf{I}_A \cdot \mathbf{I}_B)
\end{aligned} \tag{F46}$$

and

$$\left(H^{(3)} + \frac{6t^3}{U^2} \cos 3\phi\right) \frac{U^2}{3t^3} = -2h^{(3,1)} + h^{(3,2)} \tag{F47}$$

with

$$\begin{aligned}
h^{(3,1)} &= 8n^2 \cos 3\phi (n^2 + \mathbf{S}_A \cdot \mathbf{S}_B + \mathbf{S}_B \cdot \mathbf{S}_C + \mathbf{S}_C \cdot \mathbf{S}_A) (n^2 + I_A^z I_B^z + I_B^z I_C^z + I_C^z I_A^z) \\
&\quad + 4n^2 (n^2 + \mathbf{S}_A \cdot \mathbf{S}_B + \mathbf{S}_B \cdot \mathbf{S}_C + \mathbf{S}_C \cdot \mathbf{S}_A) \cdot [e^{i\phi} (I_A^+ I_B^- + I_B^+ I_C^- + I_C^+ I_A^-) + \text{H.c.}] \\
&\quad + 8 \sin 3\phi [I_A^z I_B^z I_C^z + n^2 (I_A^z + I_B^z + I_C^z)] \cdot [\mathbf{S}_A \cdot (\mathbf{S}_B \times \mathbf{S}_C)] \\
&\quad - 4i \mathbf{S}_A \cdot (\mathbf{S}_B \times \mathbf{S}_C) \cdot [I_A^z (e^{i\phi} I_B^+ I_C^- - e^{-i\phi} I_B^- I_C^+) + (A \rightarrow B \rightarrow C \rightarrow A)]
\end{aligned} \tag{F48}$$

and

$$\begin{aligned}
h^{(3,2)} &= T_B^{kj} T_C^{jk} e^{2i\phi(I_B^z + 2I_C^z)} + e^{-2i\phi(I_B^z + 2I_C^z)} T_B^{jk} T_C^{kj} + (B \rightarrow C \rightarrow A \rightarrow B) \\
&= 4 \cdot (n^2 + \mathbf{S}_A \cdot \mathbf{S}_B) \cdot (2n^2 \cos 3\phi + 2I_A^z I_B^z \cos 3\phi + e^{-i\phi} I_A^+ I_B^- + e^{i\phi} I_A^- I_B^+) + (B \rightarrow C \rightarrow A \rightarrow B).
\end{aligned} \tag{F49}$$

b. Effective Hamiltonian for $N_0 = 2$

For $n = 1$ ($N_0 = 2$), it is useful to first consider the general relation $T^{ij} T^{kl} = \delta_{jk} T^{il} - c_i^\dagger c_k^\dagger c_j c_i$. Restricting to two-particle states, we can use this general relation to write down

$$\begin{aligned}
S^+ I^+ &= 2T^{14}, & S^+ I^- &= 2T^{32}, & S^+ I^z &= T^{12} - T^{34}, \\
S^- I^+ &= 2T^{23}, & S^- I^- &= 2T^{41}, & S^- I^z &= T^{21} - T^{43}, \\
S^z I^+ &= T^{13} - T^{24}, & S^z I^- &= T^{31} - T^{42}, & S^z I^z &= \frac{1}{2} (T^{11} - T^{22} - T^{33} + T^{44}).
\end{aligned} \tag{F50}$$

Using these, we can convert the relations and get

$$\begin{aligned}
T^{11} &= \frac{(n + S^z)(n + I^z)}{2}, & T^{12} &= \frac{S^+(n + I^z)}{2}, & T^{13} &= \frac{I^+(n + S^z)}{2}, & T^{14} &= \frac{S^+ I^+}{2}, \\
T^{22} &= \frac{(n - S^z)(n + I^z)}{2}, & T^{23} &= \frac{S^- I^+}{2}, & T^{24} &= \frac{I^+(n - S^z)}{2}, \\
T^{33} &= \frac{(n + S^z)(n - I^z)}{2}, & T^{34} &= \frac{S^+(n - I^z)}{2}, \\
T^{44} &= \frac{(n - S^z)(n - I^z)}{2},
\end{aligned} \tag{F51}$$

which differs from Eq. (F46) only by factors of 2.

Substituting these into Eqs. (F29) and (F30) yields

$$\left(H^{(2)} + \frac{4t^2}{U}\right) \frac{U}{2t^2} = \sum_{ij} T_A^{ij} T_B^{ji} \\ = (n^2 + \mathbf{S}_A \cdot \mathbf{S}_B) \cdot (n^2 + \mathbf{I}_A \cdot \mathbf{I}_B) \quad (\text{F52})$$

and

$$\left(H^{(3)} + \frac{12t^3}{U^2} \cos 3\phi\right) \cdot \frac{U^2}{3t^3} = -2h^{(3,1)} + h^{(3,2)} \quad (\text{F53})$$

with $8h^{(3,1)}$ and $4h^{(3,2)}$ given by the same expressions as in Eqs. (F48) and (F49), respectively, but notice that the value of n is changed from $1/2$ to 1 .

-
- [1] P. A. Lee, N. Nagaosa, and X.-G. Wen, *Doping a Mott Insulator: Physics of High-Temperature Superconductivity*, *Rev. Mod. Phys.* **78**, 17 (2006).
- [2] B. J. Powell and R. H. McKenzie, *Quantum Frustration in Organic Mott Insulators: From Spin Liquids to Unconventional Superconductors*, *Rep. Prog. Phys.* **74**, 056501 (2011).
- [3] Y. Iwasa and T. Takenobu, *Superconductivity, Mott-Hubbard States, and Molecular Orbital Order in Intercalated Fullerides*, *J. Phys. Condens. Matter* **15**, R495 (2003).
- [4] Y. Song, Z. Yamani, C. Cao, Y. Li, C. Zhang, J. S. Chen, Q. Huang, H. Wu, J. Tao, Y. Zhu, W. Tian, S. Chi, H. Cao, Y.-B. Huang, M. Dantz, T. Schmitt, R. Yu, A. H. Nevidomskyy, E. Morosan, Q. Si, and P. Dai, *A Mott Insulator Continuously Connected to Iron Pnictide Superconductors*, *Nat. Commun.* **7**, 13879 (2016).
- [5] Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, R. C. Ashoori, and P. Jarillo-Herrero, *Correlated Insulator Behaviour at Half-Filling in Magic-Angle Graphene Superlattices*, *Nature (London)* **556**, 80 (2018).
- [6] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, *Unconventional Superconductivity in Magic-Angle Graphene Superlattices*, *Nature (London)* **556**, 43 (2018).
- [7] 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).
- [8] G. Li, A. Luican, J. M. B. Lopes dos Santos, A. H. Castro Neto, A. Reina, J. Kong, and E. Y. Andrei, *Observation of van Hove Singularities in Twisted Graphene Layers*, *Nat. Phys.* **6**, 109 (2010).
- [9] G. T. de Laissardière, D. Mayou, and L. Magaud, *Localization of Dirac Electrons in Rotated Graphene Bilayers*, *Nano Lett.* **10**, 804 (2010).
- [10] E. J. Mele, *Commensuration and Interlayer Coherence in Twisted Bilayer Graphene*, *Phys. Rev. B* **81**, 161405 (2010).
- [11] A. Luican, G. Li, A. Reina, J. Kong, R. R. Nair, K. S. Novoselov, A. K. Geim, and E. Y. Andrei, *Single-Layer Behavior and Its Breakdown in Twisted Graphene Layers*, *Phys. Rev. Lett.* **106**, 126802 (2011).
- [12] R. Bistritzer and A. H. MacDonald, *Moiré Bands in Twisted Double-Layer Graphene*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12233 (2011).
- [13] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, *Continuum Model of the Twisted Graphene Bilayer*, *Phys. Rev. B* **86**, 155449 (2012).
- [14] D. Wong, Y. Wang, J. Jung, S. Pezzini, A. M. DaSilva, H.-Z. Tsai, H. S. Jung, R. Khajeh, Y. Kim, J. Lee, S. Kahn, S. Tollabimazraehno, H. Rasool, K. Watanabe, T. Taniguchi, A. Zettl, S. Adam, A. H. MacDonald, and M. F. Crommie, *Local Spectroscopy of Moiré-Induced Electronic Structure in Gate-Tunable Twisted Bilayer Graphene*, *Phys. Rev. B* **92**, 155409 (2015).
- [15] K. Kim, A. DaSilva, S. Huang, B. Fallahzad, S. Larentis, T. Taniguchi, K. Watanabe, B. J. LeRoy, A. H. MacDonald, and E. Tutuc, *Tunable Moiré Bands and Strong Correlations in Small-Twist-Angle Bilayer Graphene*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 3364 (2017).
- [16] S. Huang, K. Kim, D. K. Efimkin, T. Lovorn, T. Taniguchi, K. Watanabe, A. H. MacDonald, E. Tutuc, and B. J. LeRoy, *Topologically Protected Helical States in Minimally Twisted Bilayer Graphene*, *Phys. Rev. Lett.* **121**, 037702 (2018).
- [17] P. Rickhaus, J. Wallbank, S. Slizovskiy, R. Pisoni, H. Overweg, Y. Lee, M. Eich, M.-H. Liu, K. Watanabe, T. Taniguchi, V. Fal'ko, T. Ihn, and K. Ensslin, *Transport through a Network of Topological States in Twisted Bilayer Graphene*, *arXiv:1802.07317*.
- [18] A. Ramires and J. L. Lado, *Electrically Tunable Gauge Fields in Tiny-Angle Twisted Bilayer Graphene*, *arXiv:1803.04400*.
- [19] In the literature, a distinction is often made between Mott and Slater insulators. The former is used to denote situations in which there is a wide intermediate energy regime with insulating charge transport and well-developed localized magnetic moments. In the twisted bilayer system, it is not clear that such a regime exists. However, it is clear that the insulating behavior is due to electron-electron interaction effects. Thus, to avoid potential confusion, it may be safer to refer to the insulator as a "correlated insulator" to emphasize the importance of interaction effects without wading into the Mott versus Slater debate.
- [20] G. Chen, L. Jiang, S. Wu, B. Lv, H. Li, K. Watanabe, T. Taniguchi, Z. Shi, Y. Zhang, and F. Wang, *Gate-Tunable Mott Insulator in Trilayer Graphene-Boron Nitride Moiré Superlattice*, *arXiv:1803.01985*.
- [21] E. J. Mele, *Band Symmetries and Singularities in Twisted Multilayer Graphene*, *Phys. Rev. B* **84**, 235439 (2011).
- [22] Note that, due to a different choice of basis, our parameterization is slightly different from that of Refs. [12,21]. We therefore define w_1 with an additional minus sign to account for the difference in basis choice.
- [23] S. Fang and E. Kaxiras, *Electronic Structure Theory of Weakly Interacting Bilayers*, *Phys. Rev. B* **93**, 235153 (2016).
- [24] *International Tables for Crystallography*, Vol. A: Space-Group Symmetry, edited by T. Hahn, 5th ed. (Springer, New York, 2006).
- [25] D. B. Litvin, *Magnetic Group Tables: 1-, 2- and 3-Dimensional Magnetic Subperiodic Groups and Magnetic*

- Space Groups* (International Union of Crystallography, Chester, England, 2013).
- [26] Y. Kim, B. J. Wieder, C. L. Kane, and A. M. Rappe, *Dirac Line Nodes in Inversion-Symmetric Crystals*, *Phys. Rev. Lett.* **115**, 036806 (2015).
- [27] This can be understood by noting that, in the case of Bernal stacking, the system has an effective C_2 symmetry which interchanges the top and bottom layers. A perpendicular electric field will therefore break this symmetry, and hence the band degeneracy at charge neutrality is lifted.
- [28] M. Goerbig and G. Montambaux, in *Dirac Matter*, edited by B. Duplantier, V. Rivasseau, and J.-N. Fuchs (Springer International Publishing, New York, 2017), pp. 25–53.
- [29] Y. Cao, J. Y. Luo, V. Fatemi, S. Fang, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, *Superlattice-Induced Insulating States and Valley-Protected Orbits in Twisted Bilayer Graphene*, *Phys. Rev. Lett.* **117**, 116804 (2016).
- [30] H. C. Po, A. Vishwanath, and H. Watanabe, *Symmetry-Based Indicators of Band Topology in the 230 Space Groups*, *Nat. Commun.* **8**, 50 (2017).
- [31] B. Bradlyn, L. Elcoro, J. Cano, M. G. Vergniory, Z. Wang, C. Felser, M. I. Aroyo, and B. A. Bernevig, *Topological Quantum Chemistry*, *Nature (London)* **547**, 298 (2017).
- [32] H. Watanabe, H. C. Po, and A. Vishwanath, *Structure and Topology of Band Structures in the 1651 Magnetic Space Groups*, *Sci. Adv.* **4**, eaat8685 (2018).
- [33] 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).
- [34] A more exotic option to explain the reduced Landau level degeneracy should also be kept in mind. Instead of losing the valley quantum number by symmetry breaking, we lose it through fractionalization. For instance, the electron could split into a fermion that carries its charge and spin but not the valley quantum number and a charge-0, spinless boson that carries its valley quantum number. If the boson is gapped while the fermion forms a Fermi surface in the doped state, we get the reduced Landau level degeneracy. Of course, such fractionalization comes hand in hand with an emergent gauge field.
- [35] K. Nomura and A. H. MacDonald, *Quantum Hall Ferromagnetism in Graphene*, *Phys. Rev. Lett.* **96**, 256602 (2006).
- [36] J. Alicea and M. P. A. Fisher, *Graphene Integer Quantum Hall Effect in the Ferromagnetic and Paramagnetic Regimes*, *Phys. Rev. B* **74**, 075422 (2006).
- [37] A. F. Young, C. R. Dean, L. Wang, H. Ren, P. Cadden-Zimansky, K. Watanabe, T. Taniguchi, J. Hone, K. L. Shepard, and P. Kim, *Spin and Valley Quantum Hall Ferromagnetism in Graphene*, *Nat. Phys.* **8**, 550 (2012).
- [38] H. Tasaki, *From Nagaoka's Ferromagnetism to Flat-Band Ferromagnetism and Beyond: An Introduction to Ferromagnetism in the Hubbard Model*, *Prog. Theor. Phys.* **99**, 489 (1998).
- [39] In principle, as discussed earlier, the $U_v(1)$ symmetry is only approximate and will be weakly broken by, say, small corrections to the band structures or by three-body interaction terms, which will modify the physics very close to the phase transition.
- [40] C. Li and Z. Wang, *Mott and Wigner-Mott Transitions in Doped Correlated Electron Systems: Effects of Superlattice Potential and Intersite Correlation*, *Phys. Rev. B* **80**, 125130 (2009).
- [41] A. Camjayi, K. Haule, V. Dobrosavljević, and G. Kotliar, *Coulomb Correlations and the Wigner-Mott Transition*, *Nat. Phys.* **4**, 932 (2008).
- [42] O. I. Motrunich and T. Senthil, *Exotic Order in Simple Models of Bosonic Systems*, *Phys. Rev. Lett.* **89**, 277004 (2002).
- [43] O. I. Motrunich, *Bosonic Model with Z_3 Fractionalization*, *Phys. Rev. B* **67**, 115108 (2003).
- [44] G. Chen, H.-Y. Kee, and Y. B. Kim, *Cluster Mott Insulators and Two Curie-Weiss Regimes on an Anisotropic Kagome Lattice*, *Phys. Rev. B* **93**, 245134 (2016).
- [45] Strictly speaking, we should allow for all four bands per spin and work with a four-component ψ and a corresponding 4×4 Hamiltonian. However, for the present discussion, we eventually only are interested in the modes in the vicinity of the Fermi surface after the flavor ordering. It is thus sufficient to focus on the two lower bands that are split off by the flavor ordering. We therefore focus on just these two right from the start.
- [46] Z.-C. Gu, H.-C. Jiang, D. N. Sheng, H. Yao, L. Balents, and X.-G. Wen, *Time-Reversal Symmetry Breaking Superconducting Ground State in the Doped Mott Insulator on the Honeycomb Lattice*, *Phys. Rev. B* **88**, 155112 (2013).
- [47] T. Senthil, J. B. Marston, and M. P. A. Fisher, *Spin Quantum Hall Effect in Unconventional Superconductors*, *Phys. Rev. B* **60**, 4245 (1999).
- [48] G. E. Volovik, *On Edge States in Superconductors with Time Inversion Symmetry Breaking*, *JETP Lett.* **66**, 522 (1997).
- [49] R. Nandkishore, L. S. Levitov, and A. V. Chubukov, *Chiral Superconductivity from Repulsive Interactions in Doped Graphene*, *Nat. Phys.* **8**, 158 (2012).
- [50] R. Nandkishore, R. Thomale, and A. V. Chubukov, *Superconductivity from Weak Repulsion in Hexagonal Lattice Systems*, *Phys. Rev. B* **89**, 144501 (2014).
- [51] I. Kimchi, S. A. Parameswaran, A. M. Turner, F. Wang, and A. Vishwanath, *Featureless and Nonfractionalized Mott Insulators on the Honeycomb Lattice at 1/2 Site Filling*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16378 (2013).
- [52] P. Kim, H. Lee, S. Jiang, B. Ware, C.-M. Jian, M. Zaletel, J. H. Han, and Y. Ran, *Featureless Quantum Insulator on the Honeycomb Lattice*, *Phys. Rev. B* **94**, 064432 (2016).
- [53] B. Ware, I. Kimchi, S. A. Parameswaran, and B. Bauer, *Topological Crystalline Bose Insulator in Two Dimensions via Entanglement Spectrum*, *Phys. Rev. B* **92**, 195105 (2015).
- [54] M. A. Metlitski and R. Thorngren, *Intrinsic and Emergent Anomalies at Deconfined Critical Points*, *Phys. Rev. B* **98**, 085140 (2018).
- [55] C.-M. Jian, Z. Bi, and C. Xu, *Lieb-Schultz-Mattis Theorem and Its Generalizations from the Perspective of the Symmetry-Protected Topological Phase*, *Phys. Rev. B* **97**, 054412 (2018).
- [56] J. Y. Lee, A. Turner, and A. Vishwanath, *Building Symmetry Enriched Topological Phases from a Bipartite Lattice Construction and Anyon Condensation*, [arXiv:1802.02155](https://arxiv.org/abs/1802.02155).

- [57] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, *Maximally Localized Wannier Functions: Theory and Applications*, *Rev. Mod. Phys.* **84**, 1419 (2012).
- [58] A. A. Soluyanov and D. Vanderbilt, *Wannier Representation of \mathbb{Z}_2 Topological Insulators*, *Phys. Rev. B* **83**, 035108 (2011).
- [59] Y.-h. Zhang, D. Mao, Y. Cao, P. Jarillo-Herrero, and T. Senthil, [arXiv:1805.08232](https://arxiv.org/abs/1805.08232).
- [60] M. Koshino and E. McCann, *Trigonal Warping and Berry's Phase $n\pi$ in *abc*-Stacked Multilayer Graphene*, *Phys. Rev. B* **80**, 165409 (2009).
- [61] A. Kumar and R. Nandkishore, *Flat Bands with Berry Curvature in Multilayer Graphene*, *Phys. Rev. B* **87**, 241108 (2013).
- [62] Y.-H. Zhang and T. Senthil, *Bridging Hubbard Model Physics and Quantum Hall Physics in Trilayer Graphene/h-BN Moiré Superlattice*, [arXiv:1809.05110](https://arxiv.org/abs/1809.05110).
- [63] P. Corboz, M. Lajkó, A. M. Läuchli, K. Penc, and F. Mila, *Spin-Orbital Quantum Liquid on the Honeycomb Lattice*, *Phys. Rev. X* **2**, 041013 (2012).
- [64] R. K. Kaul, *Spin Nematics, Valence-Bond Solids, and Spin Liquids in $SO(n)$ Quantum Spin Models on the Triangular Lattice*, *Phys. Rev. Lett.* **115**, 157202 (2015).
- [65] M. G. Yamada, M. Oshikawa, and G. Jackeli, *Emergent $SU(4)$ Symmetry in α - $ZrCl_3$ and Crystalline Spin-Orbital Liquids*, *Phys. Rev. Lett.* **121**, 097201 (2018).
- [66] C. Xu and L. Balents, *Topological Superconductivity in Twisted Multilayer Graphene*, *Phys. Rev. Lett.* **121**, 087001 (2018).
- [67] N. F. Q. Yuan and L. Fu, *Model for Metal-Insulator Transition in Graphene Superlattices and Beyond*, *Phys. Rev. B* **98**, 045103 (2018).
- [68] J. Kang and O. Vafek, preceding paper, *Symmetry, Maximally Localized Wannier States, and Low-Energy Model for the Twisted Bilayer Graphene Narrow Bands*, *Phys. Rev. X* **8**, 031088 (2018).
- [69] M. Koshino, N. F. Q. Yuan, M. Ochi, K. Kuroki, and L. Fu, this issue, *Maximally Localized Wannier Orbitals and the Extended Hubbard Model for the Twisted Bilayer Graphene*, *Phys. Rev. X* **8**, 031087 (2018).
- [70] L. Zou, H. C. Po, A. Vishwanath, and T. Senthil, *Band Structure of Twisted Bilayer Graphene: Emergent Symmetries, Commensurate Approximants, and Wannier Obstructions*, *Phys. Rev. B* **98**, 085435 (2018).