

## Nearly flat Chern bands in moiré superlattices

Ya-Hui Zhang, Dan Mao, Yuan Cao, Pablo Jarillo-Herrero, and T. Senthil\*

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



(Received 18 July 2018; revised manuscript received 3 December 2018; published 12 February 2019)

Topology and electron interactions are two central themes in modern condensed matter physics. Here, we propose graphene-based systems where both the band topology and interaction effects can be simply controlled with electric fields. We study a number of systems of twisted double layers with small twist angle where a moiré superlattice is formed. Each layer is chosen to be either AB-stacked bilayer graphene, ABC-stacked trilayer graphene, or hexagonal boron nitride. In these systems, a vertical applied electric field enables control of the bandwidth, and interestingly also the Chern number. We find that the Chern numbers of the bands associated with each of the two microscopic valleys can be  $\pm 0, \pm 1, \pm 2, \pm 3$  depending on the specific system and vertical electrical field. We show that these graphene moiré superlattices are promising platforms to realize a number of fascinating many-body phenomena, including (fractional) quantum anomalous Hall effects. We also discuss conceptual similarities and implications for modeling twisted bilayer graphene systems.

DOI: [10.1103/PhysRevB.99.075127](https://doi.org/10.1103/PhysRevB.99.075127)

### I. INTRODUCTION

Van der Waals heterostructures have recently emerged as a highly tunable platform to study correlated electron physics. A striking recent example is the observation of correlated insulating and superconducting states in twisted bilayer graphene encapsulated by hexagonal boron nitride (*h*-BN) near “magic” twist angles [1,2]. The twisting produces a long period moiré pattern, and reconstructs the electronic band structure. Near the magic twist angle, the bands active near the Fermi energy become very narrow [3]. The effects of Coulomb interaction become important, and interesting correlation effects are seen. Correlated insulating states have also been reported in trilayer graphene on *h*-BN [4]. In other experiments, magnetic field induced fractional Chern insulators have been observed in moiré superlattices of bilayer graphene [5]. Further, an interesting recent paper proposes to realize the triangular lattice Hubbard model in twisted heterostructures of transition metal dichalcogenides [6].

Here, we describe graphene-based moiré superlattice systems to realize model systems with two nearly flat bands which each carry equal and opposite Chern number  $C$ , the most basic topological index of a band [7]. Previously, Ref. [8] demonstrated that when monolayer graphene is stacked on top of *h*-BN to form a moiré superlattice, isolated  $\pm C = 1$  Chern bands may potentially occur. Other proposals for bands with Chern number  $\pm C$  in various graphene systems can be found in Refs. [8–11]. The width of these bands is large compared to the expected Coulomb energy and, hence, correlation effects may be expected to be weak. Here, we show that modifications of the proposal of Ref. [8] naturally enable engineering  $\pm C \geq 1$  Chern bands which, moreover, are nearly flat. We dub such bands  $\pm C$  Chern bands. We find systems with various values of  $C = 0, 1, 2, 3$ . The near

flatness implies that correlation effects from Coulomb interaction are potentially important. Remarkably,  $C$  can be controlled simply by applying a transverse electric field. Such an electric field also allows control of the bandwidth of the nearly flat band as shown experimentally [4]. This allows electric control of correlation effects, the band topology, and the band filling and provides a possibly unique opportunity to study many fascinating phenomena. Systems with flat Chern bands have also been studied theoretically for many years [12–20] as a potential realization of novel interacting topological phases of matter such as fractional Chern insulators. Fractional Chern insulators have been realized in strong magnetic field experiments with Harper-Hofstadter bands [5] of bilayer graphene. However, interacting topological phases at zero magnetic field remain elusive in experiment.

As examples, we describe some natural and simple phases at integer fillings  $\nu_T$  (which is defined to be the number of electrons/moiré unit cell including spin and valley degrees of freedom) where, we expect that the spin/valley degree of freedom will spontaneously uniformly polarize (“spin/valley ferromagnet”) so as to fully fill the band. For instance, at  $\nu_T = 1$  there is a spin- and valley-polarized insulating state which will show an electrical quantum Hall effect with  $\sigma_{xy} = \frac{Ce^2}{h}$ . This is an integer quantum anomalous Hall (QAH) insulator. QAH effects have been reported in magnetically doped three-dimensional topological insulator (TI) systems [21–24]. However, QAH effect in a true two-dimensional system with a Chern band has never been achieved before.

At noninteger but rational fillings  $\nu_T$ , a number of topological ordered states, both Abelian and non-Abelian, seem possible. We discuss some of these, specifically fractional quantum anomalous Hall (FQAH) states, as well with a focus on states that have been previously established in numerical calculations on correlated flat Chern bands. Our proposed realization of nearly flat  $\pm C$  Chern bands with tunable interactions calls for a detailed exploration of the possible many-body phases. We will mostly leave this for future work but remark that

\*senthil@mit.edu

correlated  $\pm C$  Chern bands may also be an excellent platform for the appearance of fractional topological insulators.

For trilayer graphene/ $h$ -BN correlated insulating states have already been reported in Ref. [4]. A model for one sign of the electric field when  $C = 0$  was described in Ref. [25]. For the other sign of the electric field we get  $\pm C = 3$  Chern bands, and we propose possible explanations of the observed insulating states in this side.

Although the details are different, there is nevertheless a conceptual similarity between the moiré systems with  $\pm C$  Chern bands discussed in this paper and the twisted bilayer graphene system. In particular, due to the nonzero Chern number of valley filtered bands, it is obvious that it is not possible to write maximally localized Wannier functions while preserving the valley  $U(1)$  symmetry as a local “onsite” symmetry. Exactly the same feature appears, albeit in a more subtle form, in the theoretical analysis of Ref. [25] of the twisted bilayer graphene system. Thus, both the systems discussed in this paper and the twisted bilayer graphene present an experimental context for a novel theoretical problem where strong correlations combine with topological aspects of band structure. This makes their modeling different from more conventional correlated solids.

## II. MODELS FOR MOIRÉ MINIBAND

At low energies, the electronic physics of graphene is dominated by states near each of two valleys which we denote  $\pm$ . We start from a continuum model of these low-energy states:

$$H_0 = \sum_a \sum_{\mathbf{k}} \psi_a^\dagger(\mathbf{k}) h_a(\mathbf{k}) \psi_a(\mathbf{k}), \quad (1)$$

where  $a = +, -$  is the valley index.  $\psi_a(\mathbf{k})$  is a multicomponent vector with sublattice and layer index.

Time reversal transforms one valley to the other one, and requires that  $h_+(\mathbf{k}) = h_-^*(-\mathbf{k})$ . In the following, we focus on valley  $+$ . The model for valley  $-$  can be easily generated by time-reversal transformation.

We restrict attention to graphene with chiral stacking patterns, such as AB-stacked bilayer graphene and ABC-stacked trilayer graphene. The effective Hamiltonian for one valley  $+$  of such an  $n$ -layer graphene is

$$h_+(\mathbf{k}) = \begin{pmatrix} \frac{U}{2} & A(k_x - ik_y)^n \\ A(k_x + ik_y)^n & -\frac{U}{2} \end{pmatrix}, \quad (2)$$

where  $U$  is a mass term.  $A = \frac{v^n}{\gamma_1^{n-1}}$ , where  $v \approx 10^6$  m/s is the velocity of Dirac cone of each graphene layer and  $\gamma_1 \approx 380$  meV is the interlayer coupling parameter. For monolayer graphene with  $n = 1$ ,  $U$  is the energy difference between two sublattices. For AB-stacked bilayer graphene (BG) with  $n = 2$  and ABC-stacked trilayer graphene (TG) with  $n = 3$ ,  $U$  is the energy difference between top layer and bottom layer. Therefore,  $U$  can be easily controlled by perpendicular electric field in BG and TG, which are the main focus in this paper.

When two lattices with lattice constant  $a_1$  and  $a_2$  are stacked with a small twist angle  $\theta$  or when they have slightly different lattice constants with  $\xi = \frac{a_1 - a_2}{a_2}$ , there is a large

moiré superlattice with lattice constant  $a_M = \frac{a}{\sqrt{\xi^2 + \theta^2}}$  at small twist angle  $\theta$ , where  $a = \frac{a_1 + a_2}{2}$ . The moiré lattice reconstructs the original band into a small moiré Brillouin zone (MBZ) which is a hexagon with size  $|K| = \frac{4\pi}{3a_M}$ . The calculation of band structure [3] is very similar to the classic example of free-electron approximation in elementary solid state textbooks [26]. Details can be found in the Appendix. A. When  $a_M \gg a$ , valley mixing terms can be ignored, and two valleys can be treated separately.

We discuss two different categories of moiré systems in this paper: (1)  $n$ -layer graphene on top of  $h$ -BN with small twist angle. Each valley maps to the  $\Gamma$  point of the MBZ. (2)  $n_1$ -layer graphene on top of  $n_2$ -layer graphene with a relative twist angle  $\theta$  close to a “magic” value  $\approx 1^\circ$ . In the MBZ, the valley  $+$  of top and bottom graphene layers maps to the  $K$  and  $K'$  points, respectively. Valley  $-$  is related by time-reversal transformation. We assume that the two graphenes have the same chirality of stacking pattern. For example, we assume both BG are AB stacked in the BG/BG system. Systems with opposite chirality are discussed in Appendix C.

In these systems, generically only  $C_3$  rotation symmetry is preserved while the inversion symmetry or equivalently  $C_6$  rotation symmetry is broken. Time-reversal symmetry flips the valley index and only requires that the Chern numbers for the bands of two valleys are opposite.

In the next two sections, we demonstrate that these systems show the following two interesting features: (1) nearly flat bands; (2) nonzero Chern number of the narrow bands for each valley. The small bandwidth implies that strong correlation effects may be present and play an important role in determining the physics. The nonzero  $\pm$  Chern number further renders the physics different from traditional Hubbard-type lattice models which are only appropriate for topologically trivial bands.

## III. NEARLY FLAT BANDS

### A. Twisted graphene/ $h$ -BN systems

As shown experimentally for the TG/ $h$ -BN system in Ref. [4], the width of the valence and conduction bands can be tuned by applying perpendicular electric field  $\propto U$  (please see Fig. 1). We further demonstrate this theoretically for the BG/ $h$ -BN system in Fig. 2.

### B. Twisted graphene/graphene systems

For twisted graphene/graphene systems, there exist “magic angles” at which the bandwidth is reduced to almost zero. This demonstrates that the existence of magic angles is a common feature for graphene/graphene systems and is not special to the twisted bilayer graphene system studied in Ref. [3]. Given the discovery of correlated insulating and superconducting states in twisted bilayer graphene, we expect these other systems will also show interesting correlation driven physics.

For BG/BG and TG/BG systems, the magic angle is the same as the twisted bilayer graphene:  $\theta_M = 1.08^\circ$ . For TG/TG, we have  $\theta_M = 1.24^\circ$ . A plot of bandwidth vs twist angle is shown in Fig. 3 for the BG/BG system. Details for the other two systems can be found in Appendix C.

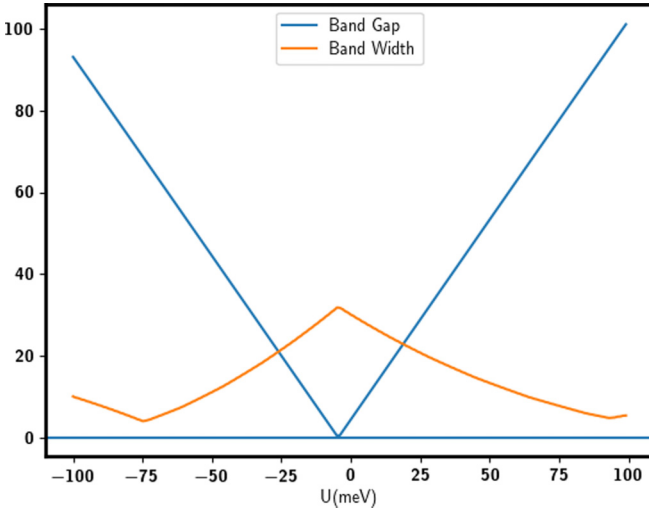


FIG. 1. Bandwidth of valence band with the applied voltage  $U$  for the TG/h-BN system.

#### IV. ANALYSIS OF CHERN NUMBER FOR MOIRÉ MINIBAND

Our results for Chern numbers at small  $|U|$  are summarized in Table I. We only show the absolute value  $|C|$  of the Chern number because the two valleys always have opposite Chern numbers. This table is one of the key results of this paper. Following, we explain these results.

For  $n$ -layer graphene/h-BN systems, the “mass” term  $U$  in Eq. (2) gaps out the band crossing at the  $\Gamma$  point, as shown in Fig. 2. Typically, different directions of perpendicular electric field give very similar band structure. When the sign of  $U$  changes, there is a band inversion which changes the contribution to the Berry curvature from states near the band touching points. In ordinary  $n$ -layer graphene the two valleys are connected in momentum space and their opposite Berry curvatures imply that the net Chern number is zero. The moiré potential, however, detaches the minibands in each valley from the rest of the spectrum, and their individual Chern numbers  $\pm C$  become well defined. The exact value

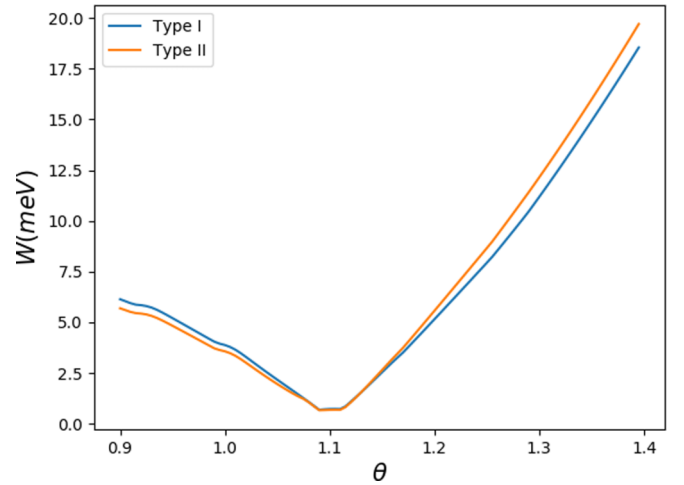


FIG. 3. Bandwidth of conduction band with twist angle for the BG/BG system.

of  $C$  of each miniband depends on contributions away from the band touching points, which do not change with sign of  $U$ . For  $n$ -layer graphene, Eq. (2) readily gives  $\Delta C = C(U) - C(-U) = n$ . Detailed calculations show that  $C(U < 0) = 0$  and  $C(U > 0) = n$  for the valence band. We have verified that the Chern number is quite stable to different parameters used for the moiré superlattice potential.

For the graphene/graphene system with  $n_1$  and  $n_2$  layers, we found that there is already a small gap  $\Delta_0 \sim 2$  meV for the band crossing point at  $K$  and  $K'$  points. We only discuss the case  $|U| > \Delta_0$ . With the same argument as in the graphene/h-BN system, we get  $\Delta C = n_1 + n_2$  for the valence band when  $U$  goes from negative to positive. For the BG/BG and TG/TG systems with  $n_1 = n_2 = n$ , there is a mirror reflection symmetry which requires  $C(U) = -C(-U)$ . Then, we can show analytically that  $C = \pm n$ . For the TG/BG system, we rely on numerical calculations to get the results in Table I.

The results above are based on the assumption that  $|U|$  is small. For graphene/graphene systems, there is a phase boundary  $U_c(\theta)$  across which the Chern number  $|C|$  drops

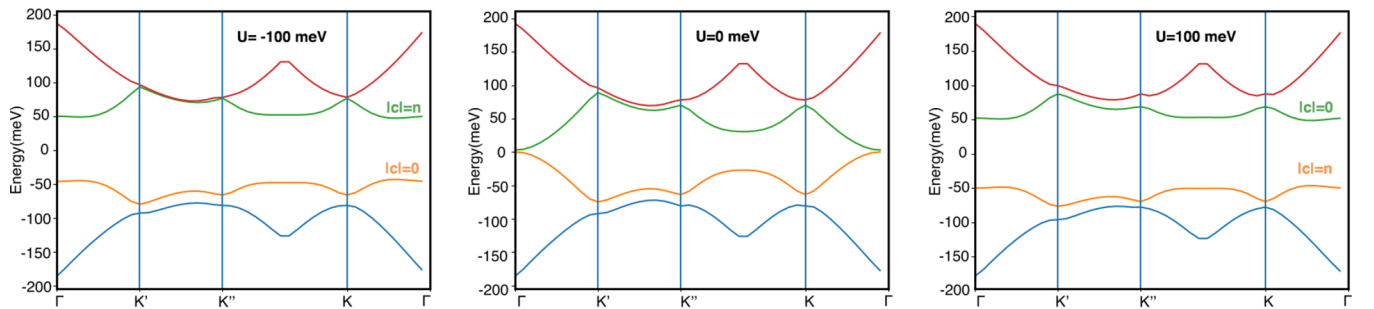


FIG. 2. Possibility of gate-driven nearly flat Chern band for twisted graphene/h-BN systems.  $U$  is the potential difference between the top and bottom layers, controlled by applied vertical electric field. Figures are generated for BG/h-BN. TG/h-BN shows quite similar behavior.  $n = 2$  for BG/h-BN and  $n = 3$  for TG/h-BN. Two sides with opposite  $U$  have similar band dispersions with Chern number  $|C| = 0$  and  $n$ , respectively, which enables to switch between Hubbard model physics and quantum Hall physics within one sample. Twist angle  $\theta = 0$  and  $a_M \approx 15$  nm.  $K' = (0, \frac{4\pi}{3a_M})$ ,  $K'' = (\frac{2\pi}{\sqrt{3}a_M}, \frac{2\pi}{3a_M})$ , and  $K = (0, -\frac{4\pi}{3a_M})$  are equivalent in the MBZ.

TABLE I. Chern number  $|C|$  of conduction/valence bands for twisting graphene/ $h$ -BN ( $\Delta_0 = 0$ ) and graphene/graphene systems ( $\Delta_0 \lesssim 2$  meV).

Systems	$U < -\Delta_0$	$U > \Delta_0$
BG/ $h$ -BN, valence	0	2
TG/ $h$ -BN, valence	0	3
BG/ $h$ -BN, conduction	2	0
TG/ $h$ -BN, conduction	3	0
BG/BG	2	2
TG/TG	3	3
TG/BG, conduction	2	3
TG/BG, valence	3	2

by 1. The phase diagram for the BG/BG system is shown in Fig. 4.  $U_c(\theta)$  goes to almost zero at the magic angle  $\theta_M$  in our simple continuum model. At  $U_c$ , the gap between the valence band and the band below closes. However, it is likely that lattice relaxation and interaction effects will enhance this hybridization gap and, as a consequence,  $U_c(\theta_M)$  may remain large even at the magic angle in realistic experimental systems. We leave it to experiment to measure the true hybridization gap.<sup>1</sup>

## V. HAMILTONIAN OF CHERN BANDS

Having shown that nearly flat Chern bands are possible in graphene-based moiré systems, we now discuss the Hamiltonian to model these  $\pm$  Chern bands. The model of course needs to have a global  $U_c(1)$  charge conservation symmetry, and spin  $SU(2)$  rotation symmetry. In addition, the valley  $U_v(1)$  symmetry is an excellent approximate symmetry and we will treat it as exact.

The *microscopic* Hamiltonian is

$$H = H_{\text{band}} + H_{\text{int}}, \quad (3)$$

where  $H_{\text{band}}$  is the band dispersion and  $H_{\text{int}}$  is the electron-electron interaction

$$H_{\text{int}} = \sum_{a_1, a_2; \sigma_1, \sigma_2} \sum_{\mathbf{q}} \rho_{a_1, \sigma_1}(\mathbf{q}) V(\mathbf{q}) \rho_{a_2, \sigma_2}(-\mathbf{q}), \quad (4)$$

where summation of  $\mathbf{q}$  is in the whole  $R^2$  space instead of the mini Brillouin zone. We use the Coulomb potential  $V(\mathbf{q}) \sim \frac{1}{|\mathbf{q}|}$ .

A low-energy Hamiltonian is obtained by projecting this microscopic Hamiltonian to the active bands near the Fermi energy. The kinetic term is then restricted to

$$H_0 = \sum_{a\sigma} \sum_k c_{a\sigma}^\dagger(k) \xi_a(k) c_{a\sigma}(k), \quad (5)$$

where  $a = +, -$  is valley index and  $\sigma = \uparrow, \downarrow$  is spin index.  $\xi_+(k) = \xi_-(-k)$  is guaranteed by time-reversal symmetry.

The *leading term* in the interaction part of the effective Hamiltonian will again be the density-density repulsion,

<sup>1</sup>For a clean sample, the valley contrasting Chern number may also be extracted from the conductance of the helical edge modes in the fully filled quantum valley Hall insulator at  $\nu_T = 4$ .

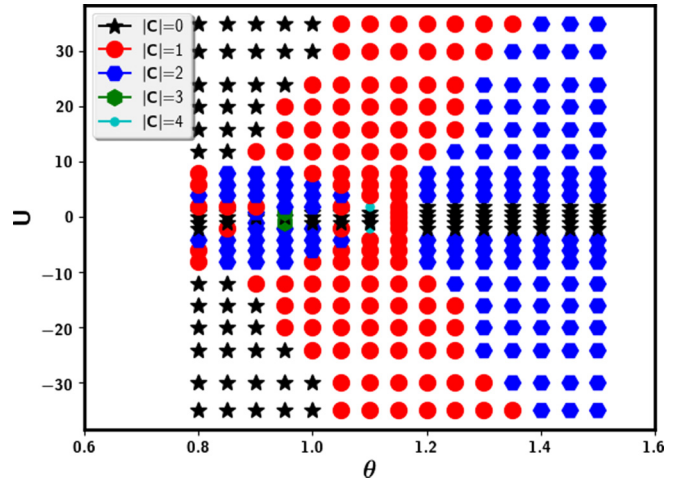


FIG. 4. Chern number  $|C|$  of BG/BG for conduction band.  $U$  is in units of meV. There is a clear phase boundary  $U_c(\theta)$ , at which Chern number jumps from  $|C| = 2$  to 1 because the hybridization gap between conduction band and the band above closes at  $\Gamma$  point.  $U_c(\theta_M) = 0$  at magic angle  $\theta_M$  in our simple model. As explained in the main text, our method may underestimate  $U_c(\theta)$  and  $U_c(\theta)$  may remain large even at magic angle in realistic systems.

but we must use density operators projected to the active bands. For concreteness, we will describe this below for the valence band. Valence band annihilation operator  $c(\mathbf{k})$  can be expressed in terms of the microscopic electron operator  $f_{\alpha;a,\sigma}(\mathbf{k})$ :

$$c_{a,\sigma}(\mathbf{k}) = \sum_{m,n} \mu_{mn,\alpha;a}(\mathbf{k}) f_{\alpha;a,\sigma}(\mathbf{k} + m\mathbf{G}_1 + n\mathbf{G}_2), \quad (6)$$

where  $\mu_{mn,\alpha;a}(\mathbf{k})$  is the Bloch wave function of valence band.  $m, n$  is the label of the momentum points of  $M \times M$  grid in free-electron approximation.  $\alpha$  is label of orbital for the microscopic electron  $f_{\alpha;a,\sigma}(k)$ , for example,  $\alpha = t, b$  labels top layer and bottom layers of BG. In terms of  $f_{\alpha;a,\sigma}(\mathbf{k})$ , we have  $\rho_a(\mathbf{q}) = \sum_{\sigma} f_{\alpha;a,\sigma}^\dagger(\mathbf{k} + \mathbf{q}) f_{\alpha;a,\sigma}(\mathbf{k})$ . We emphasize that here we distinguish  $\mathbf{q}$  and  $\mathbf{k} + m\mathbf{G}_1 + n\mathbf{G}_2$  as different momenta.

From Eq. (6) we get

$$f_{\alpha;a}(\mathbf{k} + m\mathbf{G}_1 + n\mathbf{G}_2) = \mu_{mn,\alpha;a}^*(\mathbf{k}) c_a(\mathbf{k}) + \dots, \quad (7)$$

where the ellipsis denotes other bands and we have suppressed spin index.

We can now represent the projected density operator  $\tilde{\rho}_a(\mathbf{q})$  in terms of  $c_a(\mathbf{k})$ :

$$\tilde{\rho}_{a,\sigma}(\mathbf{q}) = \sum_{\mathbf{k}} \lambda_a(\mathbf{k}, \mathbf{k} + \mathbf{q}) c_{a,\sigma}^\dagger(P(\mathbf{k} + \mathbf{q})) c_{a,\sigma}(\mathbf{k}), \quad (8)$$

where  $P(\mathbf{k}) = \mathbf{k}_0$  projects  $\mathbf{k}$  to  $\mathbf{k}_0$  in the moiré Brillouin zone (MBZ) if  $\mathbf{k} = \mathbf{k}_0 + m\mathbf{G}_1 + n\mathbf{G}_2$ .

The form factor  $\lambda_a$  is defined through the Bloch wave function  $\mu_a(\mathbf{k})$ :

$$\lambda_a(\mathbf{k}, \mathbf{k} + \mathbf{q}) = \langle \mu_a(\mathbf{k}) | \mu_a(\mathbf{k} + \mathbf{q}) \rangle. \quad (9)$$

If  $\mathbf{k} = \mathbf{k}_0 + m_0\mathbf{G}_1 + n_0\mathbf{G}_2$  is not in the MBZ,  $\mu_{mn,\alpha;a}(\mathbf{k})$  is defined as

$$\mu_{m,n,\alpha;a}(\mathbf{k}) = \mu_{m+m_0, n+n_0, \alpha;a}(\mathbf{k}_0). \quad (10)$$

We emphasize again that  $V_{a_1, a_2}(\mathbf{q} + m\mathbf{G}_1 + n\mathbf{G}_2) \neq V_{a_1, a_2}(\mathbf{q})$  unless  $m, n = 0$ . Correspondingly,  $\lambda_a(\mathbf{k}, \mathbf{k} + \mathbf{q}) \neq \lambda_a(\mathbf{k}, \mathbf{k} + \mathbf{q} + m\mathbf{G}_1 + n\mathbf{G}_2)$  unless  $m, n = 0$ . This is the reason we need to sum over  $\mathbf{q} \in R^2$  in Eq. (12). In practice, one can always truncate  $\mathbf{q} = P(\mathbf{q}) + m\mathbf{G}_1 + n\mathbf{G}_2$  with  $|m|, |n| < M$ .  $M = 1$  corresponds to summation over only MBZ.

$c_a(k)$ ,  $\mu_a(\mathbf{k})$ , and  $\lambda_a(\mathbf{k}, \mathbf{k} + \mathbf{q})$  are all not gauge invariant. There is a gauge degree of freedom:

$$\begin{aligned} c_a(k) &\rightarrow c_a(k)e^{i\theta_a(k)}, \\ \mu_a(\mathbf{k}) &\rightarrow \mu_a(\mathbf{k})e^{i\theta_a(k)}, \\ \lambda_a(\mathbf{k}, \mathbf{k} + \mathbf{q}) &\rightarrow \lambda_a(\mathbf{k}, \mathbf{k} + \mathbf{q})e^{i[\theta_a(\mathbf{k}+\mathbf{q})-\theta_a(\mathbf{k})]}. \end{aligned} \quad (11)$$

One can easily check that  $\tilde{\rho}_a(\mathbf{q})$  is gauge invariant. For each valley  $a$ , if the band is topologically trivial, we can fix the gauge  $\theta_a(\mathbf{k})$  globally. However, if the band for this valley has nonzero Chern number, there is no way to fix a global gauge in the whole Brillouin zone. For this situation, the density operator  $\tilde{\rho}_a(\mathbf{q})$  is shown to satisfy Girvin-MacDonald-Platzman (GMP) algebra [27] (also called  $W_\infty$  algebra) in the  $q \rightarrow 0$  limit [28], which is familiar from the physics of a single Landau level for quantum Hall systems in a high magnetic field.

The effective density-density interaction takes the form

$$H_V = \sum_{a_1, a_2, \sigma_1, \sigma_2} \sum_{\mathbf{q}} \tilde{\rho}_{a_1, \sigma_1}(\mathbf{q}) V(\mathbf{q}) \tilde{\rho}_{a_2, \sigma_2}(-\mathbf{q}), \quad (12)$$

where summation of  $\mathbf{q}$  is in the whole  $R^2$  space instead of the moiré Brillouin zone. We use the Coulomb potential  $V(\mathbf{q}) \sim \frac{1}{|\mathbf{q}|}$ .

Due to the nontrivial algebra satisfied by the  $\tilde{\rho}$  operators, interaction effects are qualitatively different from the case with zero Chern number. Actually, it has been shown that for a single Chern band, interactions may drive the system into exotic topological ordered states similar to fractional quantum Hall states [20]. We will address this possibility in subsequent sections.

The terms  $H_0 + H_V$  [from Eqs. (5) and (12)] define an approximate effective Hamiltonian for the partially filled valence band. At this level of approximation, the electron charge and the electron spin in each valley are separately conserved. With just these terms, the effective Hamiltonian thus has a  $U(2) \times U(2)$  symmetry. This will be further broken down to the assumed physical symmetries of just charge  $U_c(1)$ , valley  $U_v(1)$ , and total spin  $SU(2)$  by weaker terms. The pertinent such weaker interaction is a Hund's coupling:

$$H_J = -J_H \int d^2x \sum_{a_1, a_2} \mathbf{S}_{a_1}(\mathbf{q}) \cdot \mathbf{S}_{a_2}(-\mathbf{q}), \quad (13)$$

where  $\mathbf{S}_a(q)$  is the spin operator associated with electrons in valley  $a$  at a momentum  $q$ . Like the density operator itself,  $\mathbf{S}_a(q)$  is also projected to the valence band and takes the form

$$\mathbf{S}_a(q) = \frac{1}{2} \sum_{\mathbf{k}, \sigma_1, \sigma_2} \lambda_a(\mathbf{k}, \mathbf{k} + \mathbf{q}) c_{a, \sigma_1}^\dagger(P(\mathbf{k} + \mathbf{q})) \sigma_{\sigma_1 \sigma_2} c_{a, \sigma_2}(\mathbf{k}). \quad (14)$$

We expect that  $J_H$  is weaker than the Coulomb scale by roughly a factor  $\frac{a}{aM}$ . We will therefore initially ignore it and will include its effects as needed later.

For completeness, we note that we can also define the valley density or, more generally, a  $SU(2)_v$  ‘‘valley pseudospin’’ operator projected to the valence band:

$$I_\mu(\mathbf{q}) = \sum_{ab; \mathbf{k}\sigma} \lambda_{ab}(\mathbf{k}, \mathbf{k} + \mathbf{q}) c_{a, \sigma}^\dagger(P(\mathbf{k} + \mathbf{q})) \tau_{ab}^\mu c_{b, \sigma}(\mathbf{k}), \quad (15)$$

where

$$\lambda_{ab}(\mathbf{k}, \mathbf{k} + \mathbf{q}) = \langle \mu_a(\mathbf{k}) | \mu_b(\mathbf{k} + \mathbf{q}) \rangle. \quad (16)$$

Only the  $I_z(q=0)$  operator is conserved.

Our final approximate effective Hamiltonian for the partially filled valence band takes the form

$$H_{\text{eff}} = H_0 + H_V + H_J \quad (17)$$

with each of the three terms on the right given by Eqs. (5), (12), and (13), respectively.

## VI. QUANTUM ANOMALOUS HALL EFFECT AND QUANTUM VALLEY HALL EFFECT

We now discuss the possible insulating phases at integer filling  $\nu_T$  of valence bands with nonzero Chern number.  $\nu_T$  is defined as the total density for each moiré unit cell.  $\nu_T = 4$  corresponds to filling four bands completely. Here, we only focus on the case  $|C| \neq 0$  and  $\nu_T = 1, 2, 3$ . For  $C = 0$ , there is a localized Wannier orbital for each valley and the physics is governed by an extended Hubbard model. The  $C = 0$  side of the ABC trilayer graphene/h-BN system is discussed in details elsewhere [29]. In this paper, we focus on the topologically nontrivial side. For the  $C \neq 0$  side, Wannier localization is not possible for each valley. Due to the Wannier obstruction, a simple Hubbard-type lattice model is impossible. Physically, we can not think of correlated insulating phases at partial band filling in the traditional ‘‘charge frozen’’ picture of a Hubbard model. The systems discussed in this paper thus pose the novel theoretical problem of correlation effects in a partially filled topological band. A general framework to deal with this kind of system is currently absent. If we initially ignore  $H_J$ , the model has two energy scales:  $V$  for the Coulomb interaction and  $W$  for the bandwidth. Here, we restrict ourselves to the strong interaction limit  $\frac{V}{W} \rightarrow 0$ . Guided by previous studies on multicomponent quantum Hall systems, we propose that the insulating phases at  $\nu_T = 1, 2, 3$  are ‘‘ferromagnetic’’ fully polarized in spin-valley space so as to spontaneously yield completely filled bands. Hartree-Fock calculation is enough to get the ground state. Following, we will explore this proposal within a simple Hartree-Fock calculation.

### A. Hartree-Fock calculation of quantum Hall ferromagnetism

Similar to quantum Hall systems of graphene in a high magnetic field, in the flat-band limit, Coulomb interaction tends to drive the system to uniformly polarize spin or valley and fill band of the same species. Because of full  $SU(2)$  spin rotation symmetry, different directions of spin polarization are degenerate. The valley degree of freedom only has  $U(1)$  rotation symmetry in the  $z$  direction. Then, we need to decide

whether  $I_z$  or  $I_x$  (or equivalently  $I_y$  because of the rotation symmetry generated by  $I_z$ ) is polarized. In the remaining part of this section we perform a Hartree-Fock calculation to show that the system prefers to polarize the valley in the  $z$  direction instead of in the  $xy$  direction.

We consider the general repulsive interaction in Eq. (12), and study the Hartree-Fock energy of fully polarized (in spin-valley space) states at integer  $\nu_T$ . We assume that  $V \gg W$  but do not simply take the exact flat-band limit  $W = 0$ . The Hartree energy is always the same for the state with the same density. The only difference comes from the Fock exchange energy:

$$H_F = - \sum_{a_1\sigma_1; a_2\sigma_2; \mathbf{k}, \mathbf{q}} V(\mathbf{q}) \lambda_{a_1}(k, q) \lambda_{a_2}(k+q, -q) \times \langle c_{a_1, \sigma_1}^\dagger(\mathbf{k} + \mathbf{q}) c_{a_2, \sigma_2}(\mathbf{k} + \mathbf{q}) \rangle \langle c_{a_2, \sigma_2}^\dagger(\mathbf{k}) c_{a_1, \sigma_1}(\mathbf{k}) \rangle. \quad (18)$$

First, we consider a Hartree-Fock state which does not mix different  $a_1$ ,  $\sigma_1$  and  $a_2$ ,  $\sigma_2$ :

$$\langle c_{a_1, \sigma_1}^\dagger(k) c_{a_2, \sigma_2}(k) \rangle = \delta_{a_1 a_2} \delta_{\sigma_1 \sigma_2} F_{a_1 \sigma_1}(\mathbf{k}), \quad (19)$$

where  $F_{a\sigma}(\mathbf{k}) = \theta[\mu - \xi_a(\mathbf{k})]$  is Fermi-Dirac distribution at zero-temperature limit. If the Fermi energy contour is a circle, we can use  $F_{a\sigma}(\mathbf{k}) = \theta(k_F - k)$ . We will assume this in the following. For the case where the full band is filled for integer  $\nu_T$ , this assumption trivially holds and we can simply remove the  $F_{a\sigma}(\mathbf{k})$  factor.

For a state for which valley  $a$  is polarized,<sup>2</sup> the Fock exchange energy is simplified to

$$E_F = - \sum_{\mathbf{k}, \mathbf{q}} V(\mathbf{q}) |\lambda_a(\mathbf{k}, \mathbf{q})|^2, \quad (20)$$

where we have used  $\lambda(k, q) = \lambda(k+q, -q)^*$ .

The kinetic energy is

$$E_{K;0} = \sum_{\mathbf{k}} \xi_{a\sigma}(\mathbf{k}). \quad (21)$$

Next, we also calculate the energy of an inter-valley-coherent (IVC) state with full valley polarization in the  $I_x$  direction.<sup>3</sup>

Using Eq. (15), the IVC state can be written

$$|\Psi_{\text{IVC}}; \sigma\rangle = \prod_{|k| < k_F} \frac{1}{\sqrt{2}} [e^{i\theta_+(k)} c_{+; \sigma}^\dagger(k) + e^{i\theta_-(k)} c_{-; \sigma}^\dagger(k)] |0\rangle, \quad (22)$$

where  $\theta_a(\mathbf{k})$  is the phase of the Bloch wave function  $\mu_a(\mathbf{k})$ . The above state is independent of the gauge choice for the Bloch wave function.

Then, we calculate the expectation value of energy:

$$E_{\text{IVC}} = \langle \Psi_{\text{IVC}} | H_K + H_V | \Psi_{\text{IVC}} \rangle. \quad (23)$$

<sup>2</sup>Such a state can be thought as the ground state of an  $I_z$ -polarized mean field Hamiltonian:  $H_M = H_K - M I_z(\mathbf{q} = 0)$  with  $M \rightarrow \infty$ .

<sup>3</sup>Such a state can be viewed as the ground state of the following  $I_x$ -polarized mean field Hamiltonian  $H_M = H_K - M I_x(\mathbf{q} = 0)$  with  $M \rightarrow +\infty$ .

For simplicity, we drop out the spin index. The kinetic energy is straightforward:

$$E_K = \sum_{\mathbf{k}} \frac{1}{2} [\xi_+(\mathbf{k}) + \xi_-(\mathbf{k})] = \sum_{\mathbf{k}} \xi_+(\mathbf{k}), \quad (24)$$

which is the same as the kinetic energy of an  $I_z$ -polarized state.

The Fock energy for the IVC state is

$$E_{F; \text{IVC}} = - \frac{1}{4} \sum_{a_1, a_2; k, q} V(\mathbf{q}) \lambda_{a_1}(\mathbf{k}, \mathbf{q}) \lambda_{a_2}(\mathbf{k} + \mathbf{q}, -\mathbf{q}) \times e^{-i[\theta_{a_1}(\mathbf{k} + \mathbf{q}) - \theta_{a_1}(\mathbf{k})]} e^{i[\theta_{a_2}(\mathbf{k} + \mathbf{q}) - \theta_{a_2}(\mathbf{k})]} = - \frac{1}{4} \sum_{a_1, a_2; k, q} V(\mathbf{q}) |\lambda_{a_1}(\mathbf{k}, \mathbf{q})| |\lambda_{a_2}(\mathbf{k}, \mathbf{q})|, \quad (25)$$

where the phase of  $\lambda_a(\mathbf{k}, \mathbf{q})$  is removed by the additional phase factor  $e^{i\theta_a(\mathbf{k})}$ . As expected, the final expression does not depend on the phase of  $\lambda_a(\mathbf{k}, \mathbf{q})$ , which is not gauge invariant. We have used the identity  $\lambda_a(\mathbf{k} + \mathbf{q}, -\mathbf{q}) = \lambda_a^*(\mathbf{k}, \mathbf{q})$ .

The Fock energy from the  $I_z$ -polarized state is

$$E_{F;0} = - \sum_{\mathbf{k}, \mathbf{q}} V(\mathbf{q}) |\lambda_+(\mathbf{k}, \mathbf{q})|^2 = - \frac{1}{4} \sum_{k, q} V(\mathbf{q}) (2|\tilde{\lambda}_+(\mathbf{k}, \mathbf{q})|^2 + 2|\tilde{\lambda}_-(\mathbf{k}, \mathbf{q})|^2). \quad (26)$$

Finally, we can get the energy difference between IVC state and the  $I_z$ -polarized state:

$$\Delta E_{\text{IVC}} = E_{F; \text{IVC}} - E_{F;0} + E_{K; \text{IVC}} - E_{K;0} = \frac{1}{4} \sum_{\mathbf{k}, \mathbf{q}} V(\mathbf{q}) (|\lambda_+(\mathbf{k}, \mathbf{q})| - |\lambda_-(\mathbf{k}, \mathbf{q})|)^2 > 0. \quad (27)$$

Time reversal requires  $|\lambda_+(\mathbf{k}, \mathbf{q})| = |\lambda_-(\mathbf{k}, -\mathbf{q})|$ . However, generically  $|\lambda_+(\mathbf{k}, \mathbf{q})|$  is not equal to  $|\lambda_-(\mathbf{k}, \mathbf{q})|$  because inversion is broken for each valley. Therefore,  $I_z$  polarization is always favored in the flat-band limit for repulsive interaction  $V(\mathbf{q}) > 0$ . This conclusion is independent of the interaction used (as long as it is repulsive) and the details of the band structures. It is generically true for all of the spin-valley bands with opposite Chern numbers for the two valleys in the flat-band limit. Once the bandwidth is large, the IVC order ( $I_x$  order) may indeed be favored compared to the  $I_z$  polarization because of the kinetic terms, as discussed in Ref. [25]. In this paper, we focus on the limit that bandwidth is much smaller than the interaction. Then, we will only discuss the possibility of the  $I_z$  polarization.

Next, we discuss the possible states for different integer fillings based on the favored  $I_z$  polarization from the Hartree-Fock calculation. The TG/h-BN system has already been shown to have correlated insulators in  $\nu_T = 1$  and  $\nu_T = 2$  [4]. The following discussions can also apply to TG/h-BN system for one direction of the vertical displacement field.

At  $\nu_T = 1$ , we expect one band is filled completely. Let us assume that it is the band (+,  $\uparrow$ ). For the next band, density-density interaction gives the same energy to spin-polarized

state and valley-polarized state. Including the Hund's coupling, we expect spin-polarized state is selected. Therefore, we expect that the next band to be filled after  $\nu_T = 1$  is  $(-, \uparrow)$ . The third and the fourth bands to fill are either  $(-, \downarrow)$ ,  $(+, \downarrow)$  or  $(+, \downarrow)$ ,  $(-, \downarrow)$ . These two choices do not make too much difference for the properties of the insulator at filling  $\nu_T = 3$ . We therefore assume the four bands are filled in the following sequence:  $(+, \uparrow)$ ,  $(-, \uparrow)$ ,  $(+, \downarrow)$ ,  $(-, \downarrow)$ . This sequence is only meaningful at integer  $\nu_T$ . For fractional filling of  $\nu_T$ , spin-singlet or valley-singlet states are possible, which depend on energetics not captured by the simple Hartree-Fock calculation.

Following, we will discuss the property of insulators at these fillings  $\nu_T = 1, 2, 3$ .

### B. $\nu_T = 4$ : Quantum valley Hall insulator with $\sigma_{xy}^v = 4C \frac{e^2}{h}$

We start with the trivial case  $\nu_T = 4$  when we get a band insulator from fully filling all four bands. It has quantized integer "valley Hall conductivity"  $\sigma_{xy}^v = 4C \frac{e^2}{h}$ . If  $U_v(1)$  valley symmetry is preserved at the sample edge, there are  $4C$  number of helical edge modes:  $2C$  edge modes of one valley moving in one direction and  $2C$  edge modes of the other valley moving in the opposite direction. Ideally, there should be a quantized conductance  $G = 4C \frac{e^2}{h}$  coming from these  $4C$  edge channels. However, impurities at the edge may backscatter the two different valley modes and hence reduce the conductance. To reduce valley scattering, it may be easier to observe these edge modes at the domain wall between AB- (ABC-) stacked and BA- (CBA-) stacked BG (TG) on  $h$ -BN systems. It has been shown experimentally that such a smooth domain wall can preserve the valley index [30]. For BG/ $h$ -BN and TG/ $h$ -BN systems, at the domain wall (as opposed to the sample edge), we expect to observe  $4C$  edge modes for one valley and  $4C$  edge modes going in the opposite direction for the other valley.

### C. $\nu_T = 2$ : Spin-polarized quantum valley Hall insulator with $\sigma_{xy}^v = 2C \frac{e^2}{h}$

At  $\nu_T = 2$ , the spin-valley-polarized state will have two filled bands. We expect that the Hund coupling selects the spin-polarized state. Then, we have a quantum valley Hall insulator with  $\sigma_{xy}^v = 2C \frac{e^2}{h}$ . Properties are quite similar to  $\nu_T = 4$  IQVHE effect, except that the number of helical edge modes is  $C$  (one half of the  $\nu_T = 4$  case). Besides, as usual, there will be a skyrmion defect of the ferromagnetic order parameter that carries a valley charge  $Q_v = 2C$  but is electrically neutral.

### D. $\nu_T = 1$ and $3$ : Quantum anomalous Hall insulator with $\sigma_{xy}^c = C \frac{e^2}{h}$

For  $\nu_T = 1$  and  $3$ , valley and spin are both polarized to form a filled band. Then, there is an IQHE effect with Hall conductivity  $\sigma_{xy}^v = C \frac{e^2}{h}$ . This is a quantum anomalous Hall insulator, and depending on the system, has Chern number  $C > 1$  for which there is no previous experimental realization in a solid state system. At  $\nu_T = 3$ , the quantum Hall insulator is built on top of the quantum valley Hall insulator at  $\nu_T = 2$ .

The IQHE state should have  $C$  chiral edge modes which contribute a Hall conductance  $G = C \frac{e^2}{h}$ . For  $\nu_T = 3$ , these additional  $C$  chiral modes are robust, compared to the  $2C$  helical modes from IQVHE effect. Also, the system may form domains with opposite valley polarization which have opposite Chern number and Hall conductivity. It may be possible to align the domains by cooling in a magnetic field. We expect typical behavior of hysteresis loops of Hall resistance for QAHE to be observed in a magnetic field [31].

For both  $\nu_T = 1$  and  $3$ , when the state is both spin and valley polarized, there is a skyrmion defect in the spin ferromagnetic order parameter. Following standard arguments, such a skyrmion carries charge  $Q_c = C$ . Valley polarization breaks time reversal but preserves  $U(1)_v$  symmetry. So, there is no meron excitation associated with valley. For  $C = 2$ , the skyrmion is a charge- $2e$  boson while for  $C = 3$  it is a charge- $3e$  fermion. Similar to quantum Hall systems, the skyrmion excitation is actually the cheapest charged excitation for this insulating state in some parameter range (see Appendix D for detailed calculation). Thus, for  $C = 2$  these QAH insulating states at  $\nu_T = 1, 3$  have pair binding: we can regard the bosonic charge- $2e$  skyrmion as a Cooper pair. It is interesting to consider the effects of doping away from  $\nu_T = 1$  or  $3$ . For small dopings, the excess charge will be accommodated by introducing skyrmions. A natural possibility, which is well established in  $\nu = 1$  quantum Hall systems [32], is that the skyrmions form a lattice. It is interesting, however, to consider the alternate possibility that, for  $C = 2$  systems, the bosonic skyrmions condense into a liquid, instead of forming skyrmion lattice. Due to the charge  $2e$  on the skyrmion, this is a superconducting phase. It is interesting to search for this kind of skyrmion superconductor close to  $\nu_T = 1$  or  $3$  in BG/ $h$ -BN system or twisted BG/BG system. If such a superconductor is indeed found close to the ferromagnetic quantum Hall insulator, the phase diagram may be quite interesting. There may be two energy scales:  $T_c$ , the critical temperature of superconductor, and  $T^*$ , the energy scale for single-electron gap. Below  $T^*$  there are only bosonic excitations.  $T_c$  should be proportional to the skyrmion density (thus doping level), while  $T^*$  decreases with doping. At small doping, we expect  $T_c \ll T^*$ . At intermediate temperatures  $T_c < T < T^*$ , we may find a skyrmion metal with "gapped" fermionic excitation. Skyrmion condensation provides a mechanism to get superconductivity starting from an insulator, instead of normal Fermi liquid.

## VII. FRACTIONAL QUANTUM ANOMALOUS HALL EFFECT

We now briefly consider the physics at fractional fillings  $\nu_T$ . Based on some simple observations, we suggest situations which are ripe for realization of a fractional quantum anomalous Hall effect (FQAHE), although a very clean sample and strong correlation may be necessary.

Upon electron doping, the spin- and valley-polarized states at  $\bar{\nu}_T = 1$  or  $3$ , the next band to be filled is likely the one with the same spin and opposite valley because of Hund's coupling. This also gives fractional filling of a Chern band with effective filling  $\nu_{\text{eff}} = \nu_T - \bar{\nu}_T$ . Therefore, it is promising to search for valley-polarized FQAHE states at these fillings.

It has been established that there are a series of Abelian states at filling  $\nu_{\text{eff}} = \frac{m}{2km|C|+1}$  with  $k, m$  integers for flat Chern band with Chern number  $C$  [33–36]. For example, if  $k = 1, m = -1$ , we can have an Abelian FQHE state at  $\nu_{\text{eff}} = \frac{1}{2|C|-1}$ . For  $k = 1, m = 1$ , FQHE state happens at  $\nu_{\text{eff}} = \frac{1}{2|C|+1}$ . These states will have a quantized electrical Hall conductivity  $\sigma_{xy}^c = (\bar{\nu}_T - \nu_{\text{eff}})C$  where we also include the contribution from QAHE at filling  $\bar{\nu}_T = 1$  or 3. The fully filled and partially filled bands reside in opposite valleys which have opposite Chern numbers, thus, their Hall conductivities subtract.

Other topological ordered states may also be possible in these nearly flat  $\pm$  Chern bands, including those with non-Abelian quasiparticles and fractional topological insulators.

### VIII. SIMILARITIES WITH TWISTED BILAYER GRAPHENE

When the bands have zero Chern number, it is possible to build a tight-binding model by constructing Wannier orbitals for each valley in triangular lattice. A discussion of TLG/ $h$ -BN in this case was already given in Ref. [25] and the resulting theory is a Hubbard model on triangular lattice with both spin and valley degrees of freedom at each site. In this paper, our focus is on the cases with nonzero Chern number for each valley. For this case, it is clearly not possible to construct maximally localized Wannier orbital for each valley. This feature is also present, in a more subtle form, in the treatment in Ref. [25] of the small-angle twisted bilayer graphene system. In that case there are two bands per valley which have Dirac crossings. The obstruction to constructing maximally localized Wannier functions can be traced to the same sign chirality of the two Dirac points within a single valley. The systems with  $\pm C$  Chern bands studied in this paper thus provide a conceptually simpler version where the Wannier obstruction of valley-filtered bands is “obvious.” These systems lend further credence to this aspect of the theoretical treatment of Ref. [25] of the twisted bilayer graphene.

However, just like in Ref. [25], it should be possible to build Wannier orbitals for two valleys together because the total Chern number is 0 for two valleys. The cost is that  $U_v(1)$  valley symmetry can not be easily implemented in the resulting tight-binding models. We leave the construction of this kind of tight-binding model, using the methods introduced in Ref. [25], to future work. In this paper, we have instead tried to understand the resulting states phenomenologically by working in momentum space directly.

Theoretically, all these systems present an experimental context in which the physics is controlled by correlation effects on bands with nontrivial topological aspects. This makes them likely different from “ordinary” correlated solids which can be faithfully modeled through interacting lattice tight-binding models with ordinary action of all physical symmetries.

### IX. CONCLUSION

In conclusion, we propose graphene-based moiré superlattice systems to get nearly flat  $\pm$  Chern bands with Chern number  $|C| = 2, 3$ . Remarkably, the Chern number can be

controlled simply by a vertical electric field which further also tunes the bandwidth, and enhances interaction effects. This should enable not only a study of correlated  $\pm$  Chern bands, but also to cleanly experimentally isolate the effects of band Chern number from other effects in an interacting system. We proposed  $\nu_T = 2, 4$  and  $\nu_T = 1, 3$  insulating phases to be quantum valley Hall insulator and quantum anomalous Hall insulator, respectively, in the limit of strong interaction. For other rational fractional fillings, we propose to search for FQAE states and FQVHE states. Realization of such exotic topological ordered states at zero field could provide a powerful platform to have more control over anyons, for example, to explore non-Abelian anyons bounded to lattice defects [37]. For  $C = \pm 2$  system, we also suggest the possibility of an exotic superconductor from condensation of charge- $2e$  bosonic skyrmions close to filling  $\nu_T = 1$  or 3. Finally, we note conceptual similarities between the systems with  $\pm C$  Chern bands studied here and the twisted bilayer graphene system.

### ACKNOWLEDGMENTS

We thank V. Fatemi and J. Checkelsky for many inspiring discussions. T.S. is grateful to A. Po, L. Zou, and A. Vishwanath for a previous collaboration which paved the way for this paper. Y.-H.Z. thanks S. Chen for very helpful explanations of experimental details. Y.-H.Z., D.M., and T.S. were supported by NSF Grant No. DMR-1608505, and partially through a Simons Investigator Award from the Simons Foundation to S. Todadri. Y.C. and P.J.H. were primarily supported by the Gordon and Betty Moore Foundations EPiQS Initiative through Grant No. GBMF4541 and the STC Center for Integrated Quantum Materials (NSF Grant No. DMR-1231319).

### APPENDIX A: METHODS FOR MODELING MOIRÉ MINIBANDS

#### 1. Band structure

For the calculation of band structure, we follow the continuum model approach in Refs. [38] and [3]. If the two layers have different lattice constants with  $\xi = \frac{a_1 - a_2}{a_2}$  or a twist angle  $\theta$ , there is a moiré superlattice with lattice constant  $a_M \approx \frac{a}{\sqrt{\xi^2 + \theta^2}}$ . For the systems considered here,  $a_M \sim 50\text{--}60 \sim 10\text{--}15$  nm. In the limit  $a_M \gg a$ , it is appropriate to work in the continuum limit. We ignore intervalley scatterings and treat the two valleys separately. This is an excellent approximation for small twist angles. The two valleys are related by time-reversal transformation. Therefore, we can do calculations for only one valley, for example, valley +.

We start from a continuum model of graphene for the valley +:

$$H_0 = \sum_{\mathbf{k}} c_a^\dagger(\mathbf{k}) h_{ab}(k) c_b(\mathbf{k}), \quad (\text{A1})$$

where  $a, b$  are collection of layer and orbital indices. Then, moiré lattice gives superlattice potentials:

$$H_M = \sum_{ab; \mathbf{k}, \mathbf{G}_j} c_a^\dagger(\mathbf{k}) V_{ab}(\mathbf{G}_j) c_b(\mathbf{k} + \mathbf{G}_j) + \text{H.c.}, \quad (\text{A2})$$



where  $G_j$  is the moiré superlattice reciprocal vector. We choose  $\mathbf{G}_1 = (0, \frac{4\pi}{\sqrt{3}a_M})$  and  $\mathbf{G}_2 = (-\frac{2\pi}{\sqrt{3}a_M}, \frac{2\pi}{a_M})$  for the moiré Brillouin zone (MBZ) of the moiré superlattice. The form of the  $H_M$  is model dependent and details are shown in Appendices B and C.

The moiré potential reconstructs the original band into a small moiré Brillouin zone (MBZ) which is a hexagon with size  $|K| = \frac{4\pi}{3a_M}$ . The calculation of band structure is very similar to the classic example of free-electron approximation in elementary solid state textbooks. For each momentum  $\mathbf{k}$  in the MBZ, we build a large matrix  $H(k)$  whose bases include states for  $M * M$  momentum points:  $\mathbf{k} + m\mathbf{G}_1 + n\mathbf{G}_2$ ,  $m, n = -M/2, \dots, M/2$ . Typically, we use  $M = 11$  for calculation of band structure and  $M = 7$  for calculation of Berry curvature. Diagonalization of  $H(k)$  gives both the energy dispersion and the Bloch wave function  $|\mu(\mathbf{k})\rangle$ .

We focus on the conduction and valence bands close to the neutrality point. In contrast to twisted bilayer graphene (i.e., monolayer on monolayer), here there is no  $C_6$  (and hence inversion symmetry) rotation symmetry even approximately. Hence, we do not expect any generic protected crossings between the valence and conduction bands.

## 2. Numerical calculation of Berry curvature and Chern number

In addition to the band structure, we can also easily calculate the Berry curvature and Chern number of the valence band. Our calculation of Berry curvature and Chern number follows Ref. [39]. We discretize the moiré Brillouin zone (MBZ) with an  $N_x \times N_y$  grid and then we have a lattice in momentum space. For each momentum  $\mathbf{k}$  in MBZ, we diagonalize the large matrix  $H(\mathbf{k})$  and extract the eigenvector  $\mu(\mathbf{k})$  corresponding to the valence band. We calculate the following two complex quantities:

$$U_{\hat{x}}(\mathbf{k}) = \langle \mu(\mathbf{k} + \delta k_x \hat{k}_x) | \mu(\mathbf{k}) \rangle \quad (\text{A3})$$

and

$$U_{\hat{y}}(\mathbf{k}) = \langle \mu(\mathbf{k} + \delta k_y \hat{k}_y) | \mu(\mathbf{k}) \rangle, \quad (\text{A4})$$

where  $\hat{k}_x$  and  $\hat{k}_y$  are unit vectors in momentum space in  $k_x$  and  $k_y$  directions.  $\delta k_x$  and  $\delta k_y$  are lattice constants in  $k_x$  and  $k_y$  directions for our momentum lattice. The phases of  $U_{\hat{x}}$  and  $U_{\hat{y}}$  are Berry connections on the links of our momentum space lattice. They are not gauge invariant under a gauge transformation  $\mu(\mathbf{k}) \rightarrow \mu(\mathbf{k})e^{i\theta(\mathbf{k})}$ . However, we can calculate the Wilson loop for one plaquette at  $\mathbf{k}$ :

$$W(\mathbf{k}) = U_{\hat{x}}(\mathbf{k})U_{\hat{y}}(\mathbf{k} + \delta k_x \hat{k}_x)U_x^*(\mathbf{k} + \delta k_y \hat{k}_y)U_y^*(\mathbf{k}). \quad (\text{A5})$$

$W(\mathbf{k})$  is gauge invariant and Berry curvature at  $\mathbf{k}$  can be approximated as

$$\mathbf{B}(\mathbf{k}) = \frac{\arg W(\mathbf{k})}{\delta k_x \delta k_y}. \quad (\text{A6})$$

We emphasize that we need to guarantee that  $|\arg W(\mathbf{k})| < \pi$  for any  $\mathbf{k}$ . This condition is crucial to avoid ambiguity in the calculation. So long as Berry curvature has no singularity, we can always fulfill this condition by reducing the size of the plaquette. The error of this approximation can be reduced by increasing  $N_x$  and  $N_y$ , which decreases  $\delta k_x$  and  $\delta k_y$ . The Chern

number is calculated by

$$C = \frac{1}{2\pi} \sum_{\mathbf{k}} \mathbf{B}(\mathbf{k}) \delta k_x \delta k_y. \quad (\text{A7})$$

To get accurate results in our continuum model, we also need  $M \rightarrow \infty$ . Therefore, the calculation of Chern number is exact at the limit  $N_x, N_y, M \rightarrow \infty$ . Fortunately, we found that  $M = 3-5$ ,  $N_x, N_y = 40$  is enough to get an accurate Chern number.

## APPENDIX B: MODELS FOR BILAYER GRAPHENE ON BORON NITRIDE

We give some details of our continuum models for calculation of BG/h-BN or TG/h-BN systems, following Ref. [38].

### 1. Moiré Hamiltonians

We label the microscopic real-space electron operators as  $c_{A_1}, c_{B_1}, c_{A_2}, c_{B_2}$ . We assume  $B_1$  is on top of  $A_2$ . For convenience, we label  $\psi_a = \begin{pmatrix} c_{A_a} \\ c_{B_a} \end{pmatrix}$  where  $a = 1, 2$  is layer index. The Hamiltonian is

$$H = H_0 + H_M, \quad (\text{B1})$$

where

$$\begin{aligned} H_0 = & t \sum_a \sum_k \psi_a^\dagger(k) (k_x \sigma_x + k_y \sigma_y) \psi_a(k) \\ & + \gamma_1 \left( \sum_k \psi_1^\dagger(k) \sigma_- \psi_2(k) + \text{H.c.} \right) \\ & + t_3 \left( \sum_k (k_x + ik_y) \psi_1^\dagger(k) \sigma_+ \psi_2(k) + \text{H.c.} \right), \end{aligned} \quad (\text{B2})$$

where  $t = 2676$  meV and  $\gamma_1 = 340$  meV.  $t_3 = 0.051t$  is the trigonal warping term.

The moiré superlattice potential acts only on the first layer, which is closest to h-BN:

$$H_M = \sum_{\mathbf{G}} \psi_1^\dagger(\mathbf{k} + \mathbf{G}) H_V(\mathbf{G}) \psi_1(\mathbf{k}) + U \psi_1^\dagger(\mathbf{k}) \psi_1(\mathbf{k}). \quad (\text{B3})$$

$U$  is the energy difference between top layer and bottom layer, which is controlled by perpendicular electric field.  $H_V(\mathbf{G})$  can be expanded in terms of  $I, \sigma_x, \sigma_y, \sigma_z$ :

$$\begin{aligned} H_V(\mathbf{G}) = & H_0(\mathbf{G}) + H_z(\mathbf{G}) \sigma_z + \text{Re}[H_{AB}(\mathbf{G})] \sigma_x \\ & + \text{Im}[H_{AB}(\mathbf{G})] \sigma_y. \end{aligned} \quad (\text{B4})$$

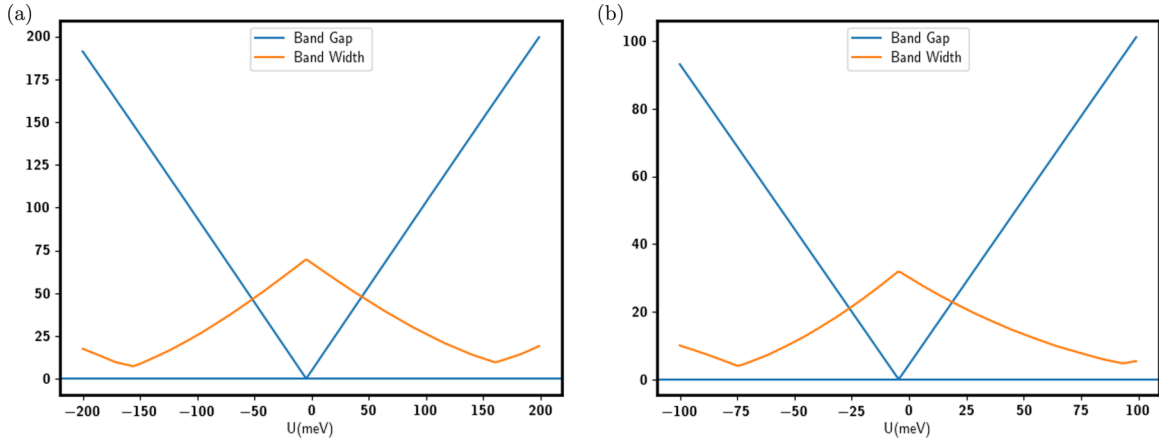
In the following, we list parameters for  $H_V(\mathbf{G})$ . They are fit from the *ab initio* calculation in Ref. [38]:

$$\begin{aligned} H_a(\mathbf{G}_1) = H_a(\mathbf{G}_3) = H_a(\mathbf{G}_5) = & C_a e^{i\varphi_a}, \\ H_a(\mathbf{G}_2) = H_a(\mathbf{G}_4) = H_a(\mathbf{G}_6) = & C_a e^{-i\varphi_a}, \end{aligned} \quad (\text{B5})$$

where  $a = 0, z$ . We have  $C_0 = -10.13$  meV,  $\varphi_0 = 86.53^\circ$  and  $C_z = -9.01$  meV,  $\varphi_z = 8.43^\circ$ .

For  $H_{AB}$ , we have

$$\begin{aligned} H_{AB}(\mathbf{G}_1) = H_{AB}^*(\mathbf{G}_4) = & C_{AB} e^{i(\frac{2\pi}{3} - \varphi_{AB})}, \\ H_{AB}(\mathbf{G}_3) = H_{AB}^*(\mathbf{G}_2) = & C_{AB} e^{-i\varphi_{AB}}, \\ H_{AB}(\mathbf{G}_5) = H_{AB}^*(\mathbf{G}_6) = & C_{AB} e^{i(-\frac{2\pi}{3} - \varphi_{AB})}, \end{aligned} \quad (\text{B6})$$


 FIG. 5. Bandwidth and band gap (in units of meV) with applied voltage difference  $U$ . (a) BG/h-BN; (b) TG/h-BN.

where  $C_{AB} = 11.34$  meV and  $\varphi_{AB} = 19.60^\circ$ . All possible  $\mathbf{G}$  can be written as

$$\mathbf{G} = m\mathbf{G}_1 + n\mathbf{G}_2, \quad (\text{B7})$$

where  $\mathbf{G}_1 = (0, \frac{4\pi}{\sqrt{3}a_M})$  and  $\mathbf{G}_2 = (-\frac{2\pi}{a_M}, \frac{2\pi}{\sqrt{3}a_M})$ .

ABC trilayer graphene on boron nitride can be modeled similarly. Because we can assume that only the layer closest to the  $h$ -BN layer is influenced, we can use the same parameters as above.

### 2. Gate tunable band dispersions

Qualitatively, different direction of applied vertical field does not change band dispersion, as is obvious in Fig. 5. Increasing  $|U|$  can reduce the bandwidth and increase the band gap between the conduction and the valence bands.

### 3. Chern number of BG/h-BN and TG/h-BN systems

The Chern number for valence band of valley  $+$  jumps by  $\Delta C = n$  when changing  $U$  from negative to positive. To

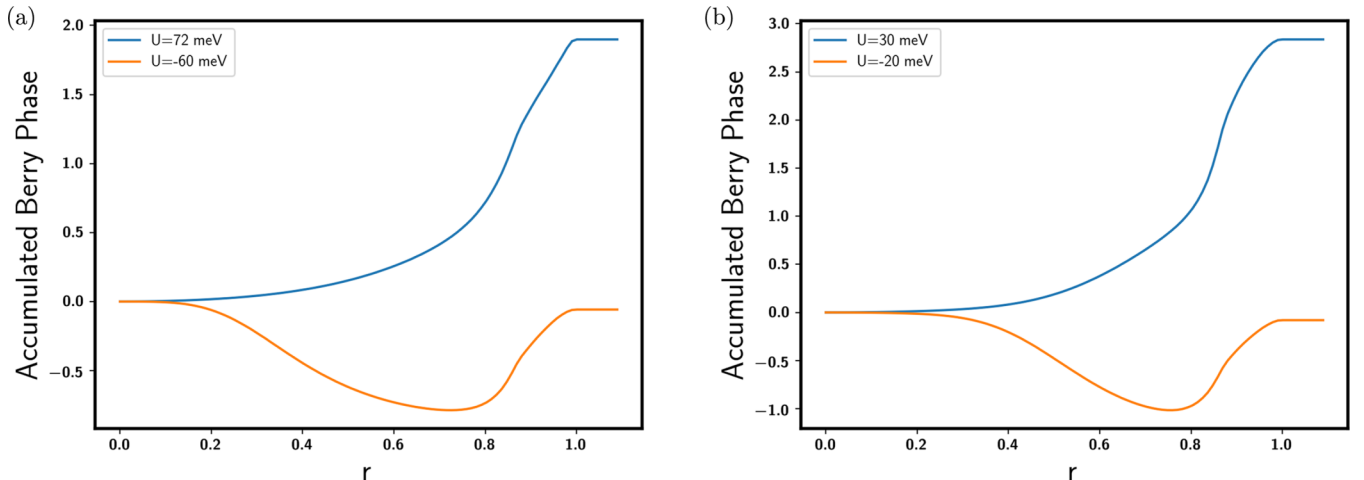
extract the value of Chern number of each side, we perform numerical calculation as explained in Appendix A.2. As shown in Fig. 6, Chern number ( $C(U < 0)$ ,  $C(U > 0)$ ) is (0,2) for BG/h-BN and (0,3) for TG/h-BN.

We also want to discuss the stability of Chern number at the band structure level, i.e., whether it changes for other parameters of moiré superlattice potential terms. For this purpose, we can use the two-band continuum model in Eq. (2) for BG and TG before adding moiré potentials. The main effect from the moiré pattern is a moiré superlattice potential for the first graphene layer. It can always be written as

$$H_M = \sum_{\mathbf{G}} c_{a1}^\dagger H_{ab}(\mathbf{G}) c_{b1}, \quad (\text{B8})$$

where  $a, b = A, B$  is sublattice index for first graphene layer.

At low energy for BG, only  $A_1$  and  $B_2$  need to be kept. Projecting to these lowest bands  $A_1$  and  $B_2$ , only  $H_{AA}(\mathbf{G})$  in Eq. (B8) is kept for layer 1. The low-energy mode can be expressed as a spinor  $\psi = \begin{pmatrix} c_t \\ c_b \end{pmatrix}$  with  $c_t = c_{A_1}$  and  $c_b = c_{B_2}$ . For ABC-stacked TG, similarly  $(B_1, A_2)$  and  $(B_2, A_3)$  are dimerized. We only keep  $c_t = c_{A_1}$  and  $c_b = c_{B_3}$ .  $h$ -BN is a


 FIG. 6. Integration of Berry curvature in region within MBZ with distance smaller than  $r$  to  $\Gamma$  point.  $r$  has been normalized by  $|\mathbf{K}_M| = \frac{4\pi}{3a_M}$ . The value at  $r = 1$  is the Chern number. (a) BG/h-BN; (b) TG/h-BN.

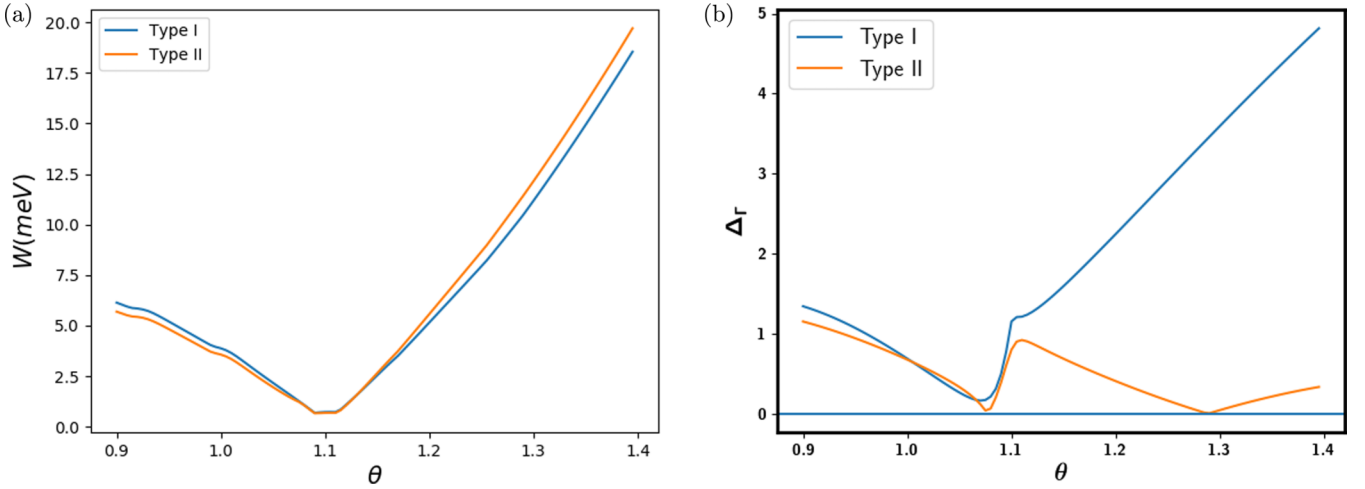


FIG. 7. BG/BG system with both type-I and -II stackings. At magic angle  $\theta \approx 1.08^\circ$ , bandwidth reduces to almost zero. Meanwhile, the hybridization gap between conduction band and the band above also goes to zero. Roughly,  $U_c(\theta) \sim \Delta_\Gamma(\theta)$ . AB on BA stacking has very different  $U_c(\theta)$  when  $\theta > \theta_M$ .  $t_M = 110$  meV in these calculations. (a) Bandwidth of conduction band with twist angle; (b) the gap  $\Delta_\Gamma$  (in units of meV) between conduction band and the band above at  $\Gamma$  point.

strong insulator and therefore all of electronic degrees of freedom come from the graphene. We consider the systems for which only  $h$ -BN layer on top of the graphene is aligned. As a result, the superlattice potential only acts on the the first layer of BG or TG. For both BG/ $h$ -BN and TG/ $h$ -BN, the only important term from the moiré superlattice is

$$H_M = \sum_{\mathbf{G}_j} \sum_{\mathbf{k}} \psi^\dagger(\mathbf{k} + \mathbf{G}_j) \begin{pmatrix} V_0 e^{i\theta_j} & 0 \\ 0 & 0 \end{pmatrix} \psi(\mathbf{k}) + \text{H.c.} \quad (\text{B9})$$

In summary, for both BG and TG, essential properties of band structure are well described by Eqs. (2) and (B9). Only two parameters  $V_0$  and  $\theta$  are important. In this system,  $C_2$  symmetry is broken and the only point group symmetry remaining is  $C_3$  rotation. Then, with similar methods as in previous sections, we calculate Berry curvature by keeping  $M^2$  momentum points and  $2M^2$  bands. With this simple model, we study the full phase diagram as a function of  $V_0$  and  $\theta_0$ . The result is that Chern number is quite stable. Chern number in one direction of perpendicular electric field ( $E_z < 0$ ) is robustly  $C = 2$  for BG/ $h$ -BN and  $C = 3$  for TG/ $h$ -BN.

### APPENDIX C: TWISTED GRAPHENE ON GRAPHENE SYSTEMS

In this Appendix, we discuss the possibility of nearly flat Chern bands in twisted graphene + graphene systems close to the magic angle ( $\theta_M = 1.08^\circ$  for BG/BG and TG/TG;  $\theta_M = 1.24^\circ$  for TG/TG). Each graphene layer is either AB-stacked bilayer graphene or ABC-stacked trilayer graphene. We will study the following three different systems: BG/BG, TG/TG, and TG/BG. Typically, there are two different patterns of stacking, depending on whether two graphenes have the same chirality or not. We define the stacking pattern with the same chirality as type I and the other one with opposite chirality as type II. Taking BG/BG as an example, we call AB on AB as type I and AB on BA as type II. In the following, we give a short summary of our results for BG/BG system with type-I

stacking. The other two systems have similar magic angles. For each specific valley, we place the original quadratic band touching point of one BG at  $K$  point of the MBZ, and place the band touching point of the other BG at  $K'$  of the MBZ. We use the same model for moiré hopping terms between the middle two layers as in Ref. [3]. There is only one parameter  $t_M$  in this model. We use  $t_M = 110$  meV which is fit from *ab initio* numerics [38].

First, at  $U = 0$ , we found that the original band touching point is gapped out. The gap is quite small:  $\Delta \lesssim 1.5$  meV. Similar to the twisted monolayer graphene on monolayer graphene system [3], there is magic angle where bandwidth goes to below 1 meV, as shown in Fig. 7(a). For  $t_M = 110$  meV, magic angle is at  $\theta \approx 1.08^\circ$ . For type-I BG/BG, we show the band structure at  $\theta = 1.08^\circ$  for several perpendicular  $U$  in Fig. 8. At  $U = 0$ , two cones at  $K$  and  $K'$  are gapped with different sign of mass. After we add positive  $U = 1.8$  meV, the cone at  $K$  with negative mass closes its gap and changes to positive mass. Similarly, if we add negative  $U = -1.8$  meV, the cone at  $K'$  with positive mass closes its gap and changes mass to negative.

Let us focus on the region far away from the magic angle. If we neglect the moiré hopping term  $t_M$ , there is a band crossing described by Eq. (2) with chirality  $n_1$  at  $K$  point and there is another band crossing with chirality  $n_2$  at  $K'$  point of MBZ. For BG/BG with AB on AB stacking pattern,  $n_1 = n_2 = 2$ . When we change  $U$  from negative to positive, both band crossings change mass from negative to positive. The sign change of mass can be seen in Fig. 8.

There is another feature associated with the magic angle. For each twist angle  $\theta$ , there is a critical  $U_c(\theta)$  through which the hybridization gap between conduction band and the band above closes. At magic angle  $U_c(\theta)$  goes to zero. To see  $U_c(\theta)$ , we plot the dispersion around  $\Gamma$  point for type-I BG/BG system at twist angle  $\theta = 1.17^\circ$  in Fig. 9. Clearly, at  $U_c = 12$  meV the hybridization gap between conduction band and the band above closes to a Dirac cone and reopens at  $\Gamma$  point. We have shown results of Chern numbers for

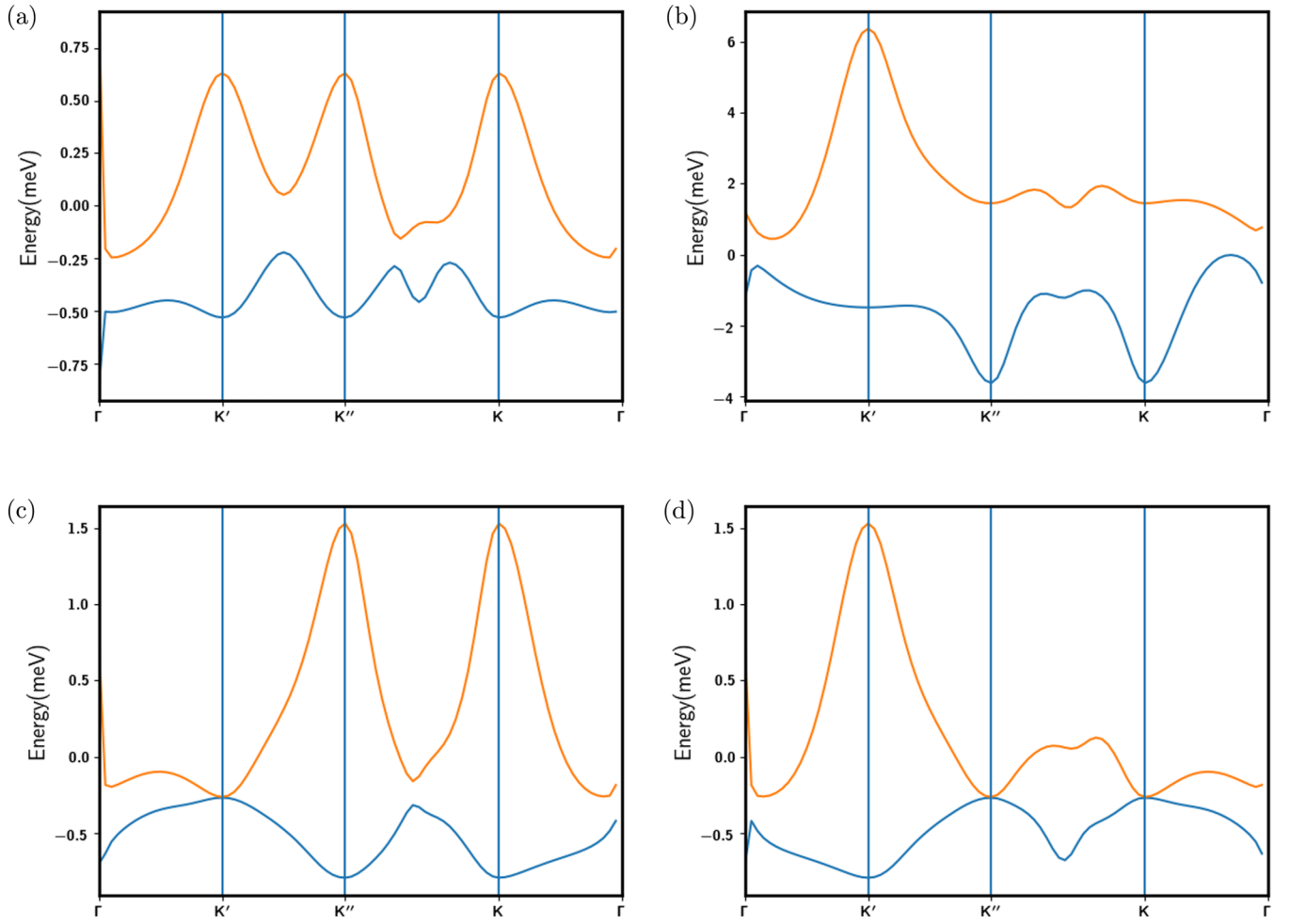


FIG. 8. Band structures for type-I BG/BG system. Twist angle is  $\theta = 1.08^\circ$ .  $K' = (0, \frac{4\pi}{3a_M})$ ,  $K'' = (\frac{2\pi}{\sqrt{3}a_M}, \frac{2\pi}{3a_M})$  and  $K = (0, -\frac{4\pi}{3a_M})$  are equivalent. At  $U = \pm 1.8$  meV, mass term of one cone at  $K$  or  $K'$  changes sign. (a)  $U = 0$  meV; (b)  $U = 10$  meV; (c)  $U = -1.8$  meV; (d)  $U = 1.8$  meV.

BG/BG system in the main text. Here, we list some results for TG/TG and TG/BG systems. We focus on type-I stacking for these systems. For the TG/TG system, there is a mirror reflection symmetry  $M_x$  which acts in the same way as the BG/BG system. We show the bandwidth for type-I TG/TG and TG/BG systems in Fig. 10. Clearly, the first magical angle is at  $\theta_M \approx 1.08^\circ$  for TG/BG while it is at  $\theta_M \approx 1.24^\circ$  for TG/TG.

We also list Chern number  $C$  of valley  $+$  for BG/BG systems with both stacking patterns and type-I TG/TG and TG/BG systems in Fig. 11. For type-I BG/BG and TG/TG systems, there is a reflection symmetry which requires that  $C(U) = -C(-U)$  and  $C(U = 0) = 0$ .

Following, we give a more detailed description of models and symmetries for twisted graphene/graphene systems using BG/BG as an example.

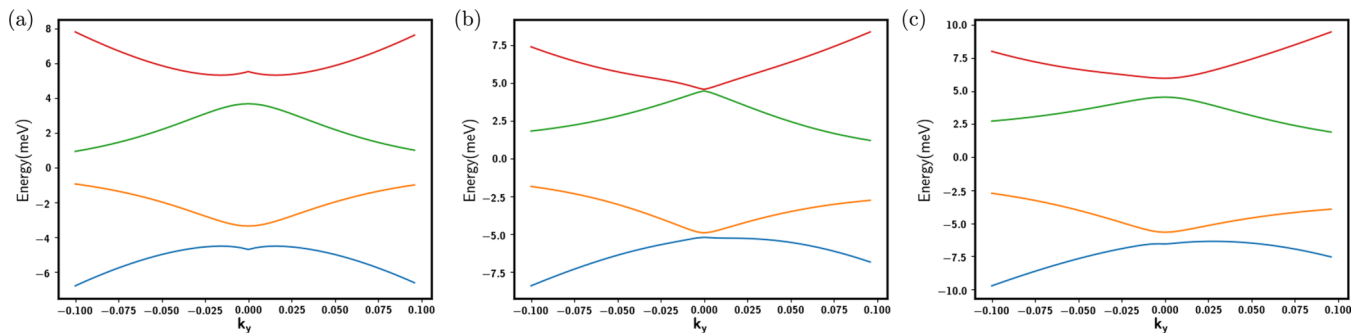


FIG. 9. Dispersion around  $\Gamma$  point for type-I BG/BG system. Twist angle is  $\theta = 1.17^\circ$ . Gap closing is obvious at  $U_c(\theta) = 12$  meV. (a)  $U = 0$  meV; (b)  $U = 12$  meV; (c)  $U = 20$  meV.

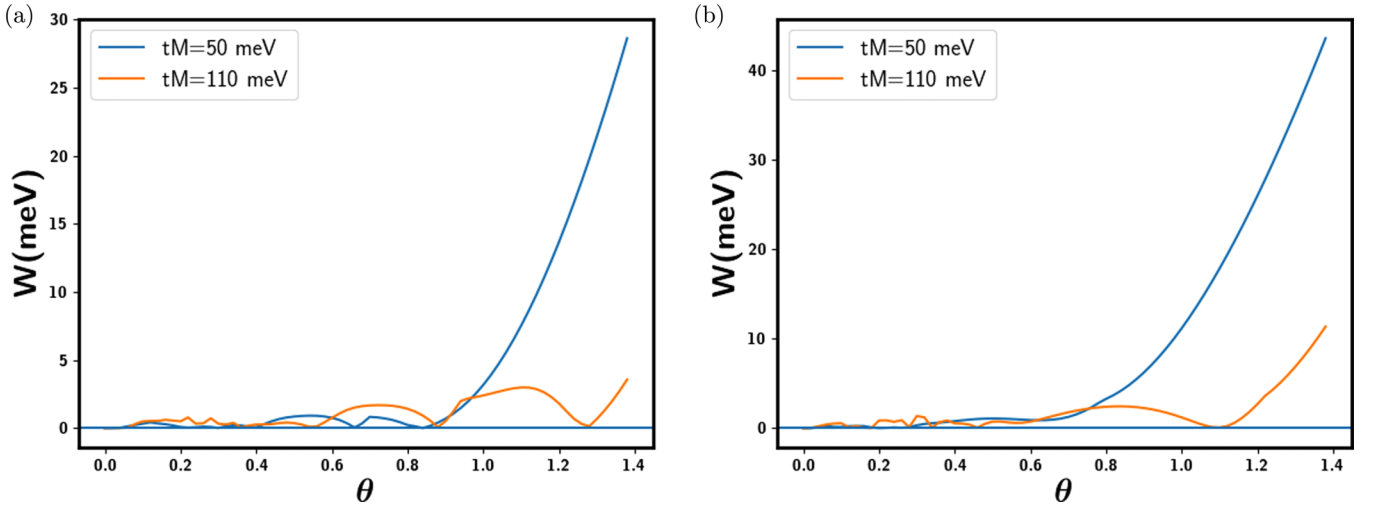


FIG. 10. Magic angle for type-I TG/TG and TG/BG systems.  $t_M$  are hopping parameters for modeling moiré hopping term. Real system corresponds to  $t_M = 110$  meV. (a) TG/TG; (b) TG/BG.

### 1. Model and symmetry of BG/BG system

We first show details of our models for the following two different stacking patterns of BG/BG system: AB on AB (type I), AB on BA (type II).

#### a. Type-I BG/BG

Here, we perform a numerical simulation of continuum models. Again, we focus only on one valley +; valley - is just its time-reversal partner. For top and bottom BG, we

define a four-component spinor  $\psi_a = \begin{pmatrix} c_{A_1}^a \\ c_{B_1}^a \\ c_{A_2}^a \\ c_{B_2}^a \end{pmatrix}$ , where  $a = t, b$  labels top BG or bottom BG layer. For valley +,  $k^t$  of  $\psi_t(k^t)$  is defined relative to the band touching point of the original large Brillouin zone of top layer and similarly  $k^b$  of  $\psi_b(k^b)$  is defined relative to the band touching point of bottom layer. In our setup,  $k^t = 0$  is at  $K = (-\frac{2\pi}{\sqrt{3}a_M}, -\frac{2\pi}{3})$  point of MBZ while  $k^b = 0$  is at  $K' = (-\frac{2\pi}{\sqrt{3}a_M}, \frac{2\pi}{3})$  point of MBZ. We use  $k$  instead of  $k^t$  or  $k^b$  in the following just for simplicity.

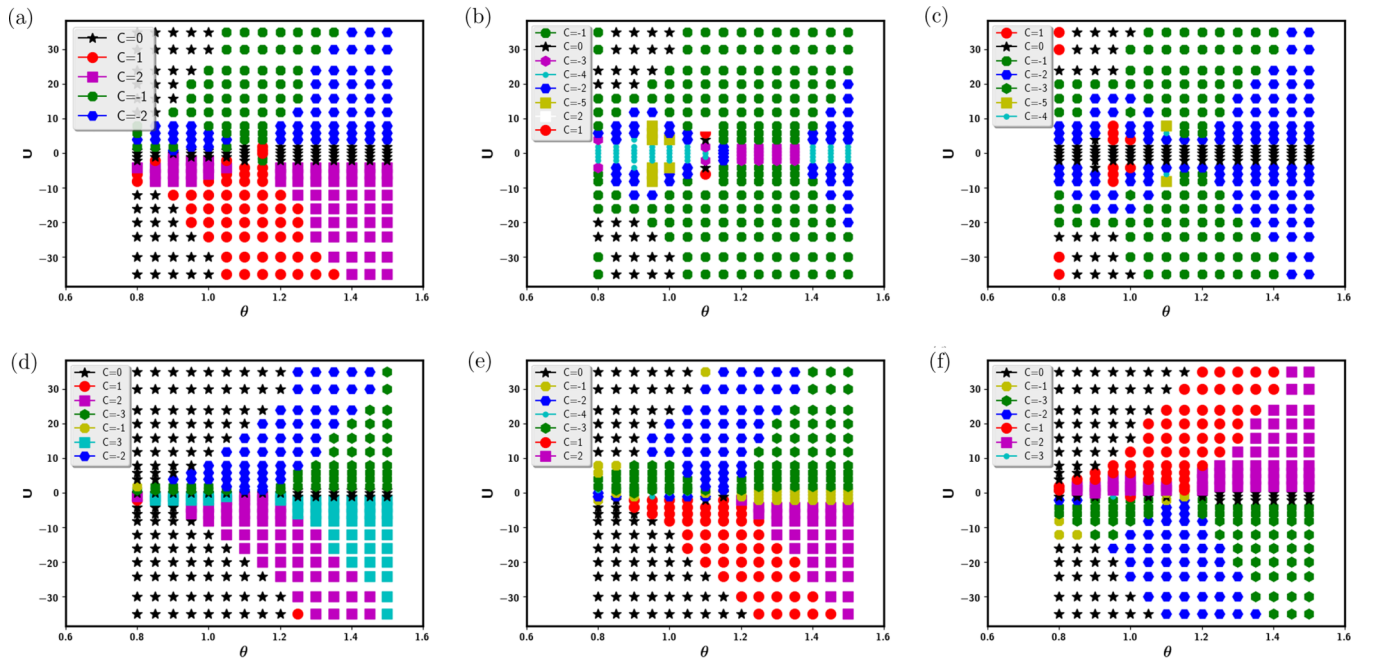


FIG. 11. Chern number  $C$  for valley + of twisted graphene on graphene systems. (a) Type-I BG/BG: conduction band; (b) type-II BG/BG: conduction band; (c) type-II BG/BG: valence band; (d) type-I TG/TG: conduction band; (e) type-I TG/BG: conduction band; (f) type-I TG/BG: valence band.

The Hamiltonian of valley + for twisted BG/BG system is

$$H = \sum_{a, \mathbf{k}} \psi_a^\dagger(\mathbf{k}) H_a^0(\mathbf{k}) \psi_a(\mathbf{k}) + \sum_{\mathbf{k}, \mathbf{Q}_j} (\psi_t^\dagger(\mathbf{k}) T_j \psi_b(\mathbf{k} + \mathbf{Q}_j) + \text{H.c.}), \quad (\text{C1})$$

where

$$H_a^0(\mathbf{k}) = \begin{pmatrix} U_1^a & v(k_x - ik_y) & 0 & t_3(k_x + ik_y) \\ v(k_x + ik_y) & U_1^a & \gamma_1 & 0 \\ 0 & \gamma_1 & U_2^a & v(k_x - ik_y) \\ t_3(k_x - ik_y) & 0 & v(k_x + ik_y) & U_2^a \end{pmatrix}. \quad (\text{C2})$$

Here,  $U_1^a$  and  $U_2^a$  mean the energy of first and second layers of top or bottom BG. This term comes from the applied voltage difference.  $\gamma_1$  is the vertical hopping term between  $B_1$  and  $A_2$ .  $t_3$  is the trigonal warping term. Moiré hopping matrix  $T_j$  only hybridize  $(c_{A_2}^t, c_{B_2}^t)$  with  $(c_{A_1}^b, c_{B_1}^b)$ . It is

$$T_j = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ t_M & t_M e^{-ij\varphi} & 0 & 0 \\ t_M e^{ij\varphi} & t_M & 0 & 0 \end{pmatrix}, \quad (\text{C3})$$

where  $\varphi = \frac{2\pi}{3}$  and  $j = 0, 1, 2$ . One can check that this Hamiltonian is invariant under  $C_3$  rotation with  $\mathbf{k} \rightarrow C_3 \mathbf{k}$ , and

$$\begin{aligned} c_{A_1}^t(\mathbf{k}) &\rightarrow c_{A_1}^t(C_3 \mathbf{k}) e^{i\varphi}, \\ c_{B_1}^t(\mathbf{k}) &\rightarrow c_{B_1}^t(C_3 \mathbf{k}) e^{-i\varphi}, \\ c_{A_2}^t(\mathbf{k}) &\rightarrow c_{A_2}^t(C_3 \mathbf{k}) e^{-i\varphi}, \\ c_{B_2}^t(\mathbf{k}) &\rightarrow c_{B_2}^t(C_3 \mathbf{k}), \\ c_{A_1}^b(\mathbf{k}) &\rightarrow c_{A_1}^b(C_3 \mathbf{k}) e^{-i\varphi}, \\ c_{B_1}^b(\mathbf{k}) &\rightarrow c_{B_1}^b(C_3 \mathbf{k}), \\ c_{A_2}^b(\mathbf{k}) &\rightarrow c_{A_2}^b(C_3 \mathbf{k}), \\ c_{B_2}^b(\mathbf{k}) &\rightarrow c_{B_2}^b(C_3 \mathbf{k}) e^{i\varphi}. \end{aligned} \quad (\text{C4})$$

In addition to the  $C_3$  symmetry, we have a mirror reflection symmetry  $M_x$  which exchanges top and bottom graphene layers without flipping the valley index. The action of  $M_x$  is

$$\begin{aligned} c_{A_1}^t(k) &\rightarrow c_{B_2}^b(M_x k), \\ c_{B_1}^t(k) &\rightarrow c_{A_2}^b(M_x k), \\ c_{A_2}^t(k) &\rightarrow c_{B_1}^b(M_x k), \\ c_{B_2}^t(k) &\rightarrow c_{A_1}^b(M_x k), \end{aligned} \quad (\text{C5})$$

where  $M_x(k_x^t, k_y^t) = (k_x^b, -k_y^b)$ . It is easy to check that  $H_0$  is invariant under the above transformation. For the terms in  $T_j$ , we have

$$\begin{aligned} c_{A_2}^t(k) c_{A_1}^b(k + \mathbf{Q}_j) &\rightarrow c_{B_1}^{b\dagger}(M_x k) c_{B_2}^t(M_x(k + \mathbf{Q}_j)), \\ c_{B_2}^t(k) c_{B_1}^b(k + \mathbf{Q}_j) &\rightarrow c_{A_1}^{b\dagger}(M_x k) c_{A_2}^t(M_x(k + \mathbf{Q}_j)), \\ c_{A_2}^t(k) c_{B_1}^b(k + \mathbf{Q}_j) e^{-ij\varphi} &\rightarrow c_{B_1}^{b\dagger}(M_x k) c_{A_2}^t(M_x(k + \mathbf{Q}_j)) e^{-ij\varphi}, \\ c_{B_2}^t(k) c_{A_1}^b(k + \mathbf{Q}_j) e^{ij\varphi} &\rightarrow c_{A_1}^{b\dagger}(M_x k) c_{B_2}^t(M_x(k + \mathbf{Q}_j)) e^{ij\varphi}. \end{aligned} \quad (\text{C6})$$

Note that  $M_x Q_{0,1,2} \rightarrow -Q_{0,2,1}$ , so the Hamiltonian  $H(k) \rightarrow H(M_x k)$ .

In experiments, an energy difference  $U$  between top layer and bottom layer can be induced by a perpendicular electric field. We simply model this energy difference by adding an onsite energy  $U_1^t = U$ ,  $U_2^t = \frac{2U}{3}$ ,  $U_1^b = \frac{U}{3}$ , and  $U_2^b = 0$ . Basically, we assume voltage increase  $V_{12} = V_{23} = V_{34} = \frac{U}{3e}$ , where 1,2,3,4 label the four layers from top to bottom. It is easy to see that  $M_x$  maps  $U$  to  $-U$  (up to a constant energy shift). Therefore, for each band, Chern number satisfies the following equation:  $C(U) = -C(U)$  which naturally implies that  $C(U = 0) = 0$ .

### b. Type-II BG/BG

There is a different system setup: AB-stacked BG on BA-stacked BG. To model this different system, we label  $\psi_t =$

$\begin{pmatrix} c_{A_1}^t \\ c_{B_1}^t \\ c_{A_2}^t \\ c_{B_2}^t \end{pmatrix}$  for top layer and  $\psi_b = \begin{pmatrix} c_{B_1}^b \\ c_{A_1}^b \\ c_{B_2}^b \\ c_{A_2}^b \end{pmatrix}$  for bottom layer. The Hamiltonian of valley + for this system is

$$H = \sum_{a, \mathbf{k}} \psi_a^\dagger(\mathbf{k}) H_a^0(\mathbf{k}) \psi_a(\mathbf{k}) + \sum_{\mathbf{k}, \mathbf{Q}_j} (\psi_t^\dagger(\mathbf{k}) T_j \psi_b(\mathbf{k} + \mathbf{Q}_j) + \text{H.c.}). \quad (\text{C7})$$

$H_t^0$  should be the same as AB-stacked BG on AB-stacked BG case, while  $H_b^0$  should change to

$$H_b^0(\mathbf{k}) = \begin{pmatrix} U_1^b & v(k_x + ik_y) & 0 & t_3(k_x - ik_y) \\ v(k_x - ik_y) & U_1^b & \gamma_1 & 0 \\ 0 & \gamma_1 & U_2^b & v(k_x + ik_y) \\ t_3(k_x + ik_y) & 0 & v(k_x - ik_y) & U_2^b \end{pmatrix}. \quad (\text{C8})$$

The moiré hopping matrix  $T_j$  also should change to

$$T_j = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ t_M e^{-ij\varphi} & t_M & 0 & 0 \\ t_M & t_M e^{ij\varphi} & 0 & 0 \end{pmatrix}, \quad (\text{C9})$$

where  $\varphi = \frac{2\pi}{3}$  and  $j = 0, 1, 2$ .

One can check that this Hamiltonian is invariant under  $C_3$  rotation with  $\mathbf{k} \rightarrow C_3\mathbf{k}$ , and

$$\begin{aligned} c_{A_1}^t(\mathbf{k}) &\rightarrow c_{A_1}^t(C_3\mathbf{k})e^{i\varphi}, \\ c_{B_1}^t(\mathbf{k}) &\rightarrow c_{B_1}^t(C_3\mathbf{k})e^{-i\varphi}, \\ c_{A_2}^t(\mathbf{k}) &\rightarrow c_{A_2}^t(C_3\mathbf{k})e^{-i\varphi}, \\ c_{B_2}^t(\mathbf{k}) &\rightarrow c_{B_2}^t(C_3\mathbf{k}), \\ c_{A_1}^b(\mathbf{k}) &\rightarrow c_{A_1}^b(C_3\mathbf{k})e^{-i\varphi}, \\ c_{B_1}^b(\mathbf{k}) &\rightarrow c_{B_1}^b(C_3\mathbf{k}), \\ c_{A_2}^b(\mathbf{k}) &\rightarrow c_{A_2}^b(C_3\mathbf{k})e^{i\varphi}, \\ c_{B_2}^b(\mathbf{k}) &\rightarrow c_{B_2}^b(C_3\mathbf{k})e^{-i\varphi}. \end{aligned} \quad (\text{C10})$$

For AB on BA BG/BG, there is a mirror symmetry  $M_y$  which flips the valley index. The transformation is

$$\begin{aligned} c_{A_1}^t(k) &\rightarrow \tilde{c}_{A_2}^b(M_y k), \\ c_{B_1}^t(k) &\rightarrow \tilde{c}_{B_2}^b(M_y k), \\ c_{A_2}^t(k) &\rightarrow \tilde{c}_{A_1}^b(M_y k), \\ c_{B_2}^t(k) &\rightarrow \tilde{c}_{B_1}^b(M_y k), \end{aligned} \quad (\text{C11})$$

where  $\tilde{c}$  is for another valley in the large Brillouin zone.  $M_y(k_x^t, k_y^t) = (-\tilde{k}_x^b, \tilde{k}_y^b)$ , where  $\tilde{k}^b$  is defined as relative to the band touching point of valley – for bottom graphene at large BZ. For terms in  $T_j$ , we have

$$\begin{aligned} c_{A_2}^{t\dagger}(k)c_{A_1}^b(k+Q_j) &\rightarrow \tilde{c}_{A_1}^{b\dagger}(M_y k)\tilde{c}_{A_2}^t(M_y(k+Q_j)), \\ c_{B_2}^{t\dagger}(k)c_{B_1}^b(k+Q_j) &\rightarrow \tilde{c}_{B_1}^{b\dagger}(M_y k)\tilde{c}_{B_2}^t(M_y(k+Q_j)), \\ c_{A_2}^{t\dagger}(k)c_{B_1}^b(k+Q_j)e^{-ij\varphi} &\rightarrow \tilde{c}_{A_1}^{b\dagger}(M_y k)\tilde{c}_{B_2}^t(M_y(k+Q_j))e^{-ij\varphi}, \\ c_{B_2}^{t\dagger}(k)c_{A_1}^b(k+Q_j)e^{ij\varphi} &\rightarrow \tilde{c}_{B_1}^{b\dagger}(M_y k)\tilde{c}_{A_2}^t(M_y(k+Q_j))e^{ij\varphi}, \end{aligned} \quad (\text{C12})$$

where  $M_y Q_{0,1,2} = Q_{0,2,1}$ . Note that the matrix element for another valley is the time-reversal partner of the original valley. For example,  $\tilde{c}_{A_1}^{b\dagger}(k)\tilde{c}_{B_2}^t(k+Q_j)e^{ij\varphi}$ . The Hamiltonian  $H(k) \rightarrow \tilde{H}(M_y k)$ .  $\tilde{H}$  is the Hamiltonian of another valley.

Combination of  $M_y$  and time reversal  $T$  gives an antiunitary symmetry which preserves valley index. Because we have  $SU(2)$  spin rotation symmetry, time reversal can be implemented as simply complex conjugation without flipping the spin. Therefore, we have an antiunitary  $M_y T$  symmetry which maps  $(k_x^t, k_y^t)$  to  $(k_x^b, -k_y^b)$ .  $M_y T$  does not flip the sign of Berry curvature. It also reverses the direction of perpendicular electric field. Therefore, we have requirement of Chern number of each band:  $C(U) = C(-U)$ . As a result, there is no constraint for Chern number at  $U = 0$  and an arbitrary Chern number is possible at  $U = 0$ .

### c. Magic angle for BG/BG

Within the continuum model, it is known that the twisted bilayer graphene system has magic angles where velocity of Dirac cone is zero [3]. In this section, we follow Ref. [3] to derive an analytical expression for the renormalized mass at band touching point with twist angle  $\theta$  for BG/BG system. We assume that the stacking pattern is AB on AB. For an isolated (A-B)-stacked bilayer graphene, the Hamiltonian of one valley is

$$h_0(k, \theta) = \begin{pmatrix} 0 & t_3 k_\theta^* & 0 & -v k_\theta \\ t_3 k_\theta & 0 & -v k_\theta^* & 0 \\ 0 & -v k_\theta & 0 & \gamma_1 \\ -v k_\theta^* & 0 & \gamma_1 & 0 \end{pmatrix} \quad (\text{C13})$$

with basis  $(c_{A_1} c_{B_2} c_{A_2} c_{B_1})^T$  and  $k_\theta = (k_x + ik_y)e^{-i\theta/2}$ . Subscripts 1,2 are for layers 1 and 2, A, B are for different sublattices. For layers 3 and 4, replace 1 by 3 and 2 by 4. Here,  $\gamma_1$  is the hopping term between  $B_1$  and  $A_2$ .

We consider bilayer 1,2 on top of bilayer 3,4 and hopping only between layers 2 and 3. The hopping matrix is

$$\begin{aligned} T_{32}(k+Q_j, k) &= t_M [c_{A_3}^\dagger(k+Q_j) c_{B_3}^\dagger(k+Q_j)] \\ &\times \begin{pmatrix} 1 & e^{ij\varphi} \\ e^{-ij\varphi} & 1 \end{pmatrix} \begin{pmatrix} c_{A_2}(k) \\ c_{B_2}(k) \end{pmatrix}, \end{aligned} \quad (\text{C14})$$

where  $Q_j$  is moiré reciprocal lattice vector,  $Q_0 = K_\theta(0, -1)$ ,  $Q_1 = K_\theta(\sqrt{3}/2, 1/2)$ , and  $Q_2 = K_\theta(-\sqrt{3}/2, 1/2)$ .

Integrating the higher-energy orbitals  $B_1, A_2, B_3, A_4$ , the effective Hamiltonian is

$$h_{\text{eff}}(k+Q_j, k) = \begin{pmatrix} 0 & t_3 k_\theta^* - v^2 k_\theta^2 / \gamma_1 & t_M v k_\theta / \gamma_1 & t_M v^2 k_\theta (k+Q_j)_{-\theta} e^{ij\varphi} / \gamma_1^2 \\ \text{c.c.} & 0 & t_M e^{-ij\varphi} & t_M v (k+Q_j)_{-\theta} / \gamma_1 \\ \text{c.c.} & \text{c.c.} & 0 & t_3 (k+Q_j)_{-\theta}^* - v^2 (k+Q_j)_{-\theta}^2 / \gamma_1 \\ \text{c.c.} & \text{c.c.} & \text{c.c.} & 0 \end{pmatrix}, \quad (\text{C15})$$

where c.c. stands for complex conjugate. The basis is  $(c_{A_1} c_{B_2} c_{A_3} c_{B_4})^T$ .

We can write an eight-band model to gain some insight. We focus on the band touching point for valley + of the top graphene layer, which is at  $K$  point of the MBZ. In the limit  $k^t \rightarrow 0$ , each momentum  $k^t$  connects to three momentums of the bottom graphene layer:  $k^b = k^t + Q_j$ . Keeping these four momentum points, we can write an eight-band model.

At small  $|k|$ ,  $H(k) = H^{(0)} + H^{(1)}(k)$ , where

$$H^{(0)} = \begin{pmatrix} h_0 & T_1 & T_2 & T_3 \\ T_1^\dagger & h_1 & 0 & 0 \\ T_2^\dagger & 0 & h_2 & 0 \\ T_3^\dagger & 0 & 0 & h_3 \end{pmatrix}, \quad (\text{C16})$$

$h_j, T_j$  are  $2 \times 2$  matrix, for  $j = 1, 2, 3$ ,

$$h_j = \begin{pmatrix} 0 & -\frac{v^2}{\gamma_1}(Q_j)_{-\theta}^2 \\ -\frac{v^2}{\gamma_1}((Q_j)_{-\theta}^2)^* & 0 \end{pmatrix}, \quad (\text{C17})$$

$$T_j = \begin{pmatrix} 0 & 0 \\ t_M e^{-ij\phi} & \frac{t_M v}{\gamma_1}(Q_j)_{-\theta} \end{pmatrix}. \quad (\text{C18})$$

The zero-eigenvalue states can be written as  $\psi = (\psi_0, \psi_1, \psi_2, \psi_3)$ , where each entry is a two-component spinor. We can obtain the wave function of the zero energy state as

$$\psi_j = -h_j^{-1} T_j^\dagger \psi_0 = \frac{t_M}{v K_\theta^4} \begin{pmatrix} K_\theta^2(Q_j)_\theta \\ \frac{\gamma_1}{v} e^{ij\phi} (Q_j)_{-\theta}^{2*} \end{pmatrix} \psi_0^{(2)}, \quad (\text{C19})$$

and the square-norm of  $\psi$  is

$$|\psi|^2 = \psi_0^\dagger \begin{pmatrix} 1 & 0 \\ 0 & 1 + 3 \frac{\gamma_1^2 t_M^2}{v^4 K_\theta^4} \left(1 + \frac{v^2 K_\theta^2}{\gamma_1^2}\right) \end{pmatrix} \psi_0. \quad (\text{C20})$$

Note that  $\psi_0$  forms a two-dimensional Hilbert space. Therefore,  $\psi$  also forms a two-dimensional space. The block between  $k$  and  $k + Q_j$  in the perturbation term  $H^{(1)}(k)$  is

$$H^{(1)}(k + Q_j; k) = \begin{pmatrix} 0 & -v^2 k_\theta^2 / \gamma_1 & t_M v k_\theta / \gamma_1 & t_M (v k_\theta / \gamma_1)(v(k + Q_j)_{-\theta} / \gamma_1) e^{ij\phi} \\ \text{c.c.} & 0 & 0 & t_M v k_{-\theta} / \gamma_1 \\ \text{c.c.} & \text{c.c.} & 0 & -v^2 k_{-\theta}(k + 2Q_j)_{-\theta} / \gamma_1 \\ \text{c.c.} & \text{c.c.} & \text{c.c.} & 0 \end{pmatrix}, \quad (\text{C21})$$

where  $j = 1, \dots, 3$ ,  $H^{(1)}(k)$  is an  $8 \times 8$  matrix.

We can obtain the elements of  $H^{(1)}(k)$  in the  $\psi$  basis, where the  $\theta$  dependence of wave vectors is ignored:

$$H^{(1)}(k)_{11} = 0, \quad (\text{C22})$$

$$H^{(1)}(k)_{12} = -\frac{v^2}{\gamma_1} (k_x + ik_y)^2 \left(1 - 3 \frac{t_M^2}{v^2 K_\theta^2}\right) / \alpha, \quad (\text{C23})$$

$$H^{(1)}(k)_{22} = 0, \quad (\text{C24})$$

where the normalization factor  $\alpha = \sqrt{1 + 3 \frac{\gamma_1^2 t_M^2}{v^4 K_\theta^4} \left(1 + \frac{v^2 K_\theta^2}{\gamma_1^2}\right)}$ .

One can see that the mass of the original quadratic band is enhanced by a factor  $m^* = \alpha m_0 / \left(1 - 3 \frac{t_M^2}{v^2 K_\theta^2}\right)$ . Interestingly,

$m^* \rightarrow \infty$  at magic angle decided by  $1 - 3 \frac{t_M^2}{v^2 K_\theta^2} = 0$ . This equation gives  $\theta_M = 1.08^\circ$  which is equal to the magic angle value of twisted monolayer graphene on monolayer graphene system [3].

## 2. Analysis of Chern number

In the following, we give some insights about the phase diagram of Chern number for twisted graphene on graphene systems.

### a. Type I: Double graphenes with opposite chirality

For either BG or TG, we can only keep two bands. We label the number of layers for top graphene as  $n_t$  and the number of layers for bottom graphene as  $n_b$ . The Hamiltonian of each graphene is

$$H_a = \psi^\dagger(k) \begin{pmatrix} m_a & \frac{v^{n_a}}{\gamma_1^{n_a-1}} (k_x - ik_y)^{n_a} \\ \frac{v^{n_a}}{\gamma_1^{n_a-1}} (k_x + ik_y)^{n_a} & -m_a \end{pmatrix} \psi(k), \quad (\text{C25})$$

where  $\psi_a = \begin{pmatrix} c_1^a \\ c_2^a \end{pmatrix}$ .  $c_1^a$  and  $c_2^a$  are the first layer and the second layer for graphene  $a = t, b$ . Layer 1 is always fixed to be on top of layer 2. Therefore, mass  $m_a$  is proportional to energy difference between top layer and bottom layer, up to an offset. We put  $\psi_t$  at  $K_M$  of MBZ and  $\psi_b$  at  $K'_M$  of MBZ. Momentum  $\mathbf{k}$  in the above Hamiltonian is defined relative to  $K_M$  or  $K'_M$ .

Moiré hopping Hamiltonian is

$$H_M = \sum_{\mathbf{k}, \mathbf{Q}_j} (t_M e^{ij\phi} c_2^{\dagger}(\mathbf{k}) c_1^b(\mathbf{k} + \mathbf{Q}_j) + \text{H.c.}), \quad (\text{C26})$$

where  $\phi = \frac{2\pi}{3}$ . We have projected moiré hopping matrix in Eq. (C3) to low-energy band  $\psi_a$  here. There are also terms from higher-order perturbations when we integrate higher-energy degree of freedoms  $c'_{B_1}$ ,  $c'_{A_2}$ ,  $c^b_{B_1}$ , and  $c^b_{A_2}$ . The most important term is the following mass term:  $A c_2^{\dagger}(k) c_2^t(k) + A c_1^{b\dagger}(k) c_1^b(k)$ . Therefore, at  $U = 0$ , we have mass term  $m_t$  and  $m_b$  with opposite sign. For BG/BG and TG/TG systems,  $M_x$  symmetry requires  $m_t = -m_b$ . These mass terms gap out quadratic band touching at  $K$  and  $K'$  with opposite sign.

At  $U = 0$ , contributions to Chern number from  $K_M$  and  $K'_M$  are  $C_K = \frac{n_t}{2}$ ,  $C_{K'} = -\frac{n_b}{2}$ . For BG/BG and TG/TG systems, at  $U = 0$ , Chern number is  $C(U = 0) = 0$  as required by  $M_x$  symmetry. For TG/BG system, we find that  $C(U = 0) = -1$ , resulting from an additional contribution at the  $\Gamma$  point. Next, we can easily understand Chern numbers at nonzero  $|U|$ . When  $U > \Delta_0 > 0$ ,  $m_b$  changes to positive through a gap closing and reopening process. It is easy to figure out that Chern number of the conduction band need to change by  $-n_b$ . Therefore,  $C(U > \Delta_0) = -n_b$  for TG/TG and BG/BG systems. For TG/BG system,  $C(U > \Delta_0) = -1 - n_b = -3$ .

Similarly, when  $U < -\Delta_0$ ,  $m_t$  changes to negative and total Chern number of the conduction band needs to change by  $n_t$ . Therefore,  $C(U < -\Delta_0) = n_t$  for TG/TG and BG/BG systems. For TG/BG,  $C(U < -\Delta_0) = n_t - 1 = 2$ . When we



further increase  $|U|$ , the hybridization gap between the conduction band and the band above closes at  $U_c$  and reopens after  $|U| > U_c$ . Through this process, Chern number  $|C|$  drops by 1.  $U_c(\theta)$  vanishes at magic angle  $\theta_M$ .

### b. Type II: Double graphenes with opposite chirality

When the top and bottom graphenes have the opposite chirality, the Hamiltonians for them are

$$H_t = \psi^\dagger(k) \begin{pmatrix} m_t & \frac{\nu^{n_t}}{\gamma_1^{n_t-1}}(k_x - ik_y)^{n_t} \\ \frac{\nu^{n_t}}{\gamma_1^{n_t-1}}(k_x + ik_y)^{n_t} & -m_t \end{pmatrix} \psi(k) \quad (C27)$$

and

$$H_b = \psi^\dagger(k) \begin{pmatrix} m_b & \frac{\nu^{n_b}}{\gamma_1^{n_b-1}}(k_x + ik_y)^{n_b} \\ \frac{\nu^{n_b}}{\gamma_1^{n_b-1}}(k_x - ik_y)^{n_b} & -m_b \end{pmatrix} \psi(k). \quad (C28)$$

We still put  $\psi_t$  on  $K$  and  $\psi_b$  on  $K'$  of MBZ. The moiré hopping Hamiltonian is

$$H_M = \sum_{\mathbf{k}, \mathbf{Q}_j} (t_M c_2^{\dagger\dagger}(\mathbf{k}) c_1^b(\mathbf{k} + \mathbf{Q}_j) + \text{H.c.}), \quad (C29)$$

where  $\varphi = \frac{2\pi}{3}$ .

We also simulate the above Hamiltonian numerically. At  $U = 0$ , we still have  $m_t$  and  $m_b$  with opposite sign. However, quadratic band touching at  $K$  and  $K'$  has the opposite chirality. Therefore, contributions to Chern number are  $C_K = \frac{n_t}{2}$ ,  $C_{K'} = \frac{n_b}{2}$ . This is also consistent with symmetry requirements for BG/BG and TG/TG systems, for which  $M_y T$  symmetry requires that  $K$  and  $K'$  contribute the same Berry curvature. The contribution from region close to  $\Gamma$  point is arbitrary. Therefore, there is no constraint for Chern number at  $U = 0$ . We have  $C(U = 0) = C_0$  with  $C_0$  a general integer which depends on the system and twist angle.

When  $U > \Delta_0$ ,  $m_b$  becomes positive and Chern number should change by  $n_b$ . Therefore, conduction band  $C(U > U_c) = C_0 + n_b$ . When  $U < -\Delta_0$ ,  $m_t$  changes to negative and Chern number changes by  $n_t$ . Therefore, conduction band  $C(U < -U_c) = C_0 + n_t$ . The change of Chern number for the valence band across  $|\Delta_0|$  is opposite to the conduction band, although  $C_0$  of conduction and valence bands have no obvious correlation. When further increasing  $|U|$ , the hybridization gap between conduction band and the band above closes at  $U_c$  and reopens, resulting in a further drop of Chern number by 1, similar to type-I system.  $U_c(\theta)$  also vanishes at the magic angle.

## APPENDIX D: SKYRMION ENERGY OF IQAH INSULATOR

We try to estimate the energy cost of skyrmion excitation of the spin- and valley-polarized integer QAH state at total filling  $\nu_T = 1$  or 3. We focus on the  $C = 2$  case, though the following calculation can be easily generalized to other Chern numbers. We assume the band filled is  $(-, \downarrow)$ ; in the following we ignore valley index because we expect valley-flip excitation

is not important. The ground-state wave function is

$$|\Omega\rangle = \prod_{\mathbf{k}} c_{\downarrow}^{\dagger}(\mathbf{k})|0\rangle. \quad (D1)$$

To estimate the energy cost of skyrmion, we should estimate spin phase stiffness  $\rho_s$  first. This can be done by calculating the dispersion of spin wave. Spin wave with momentum  $\mathbf{q}$  is created by

$$S^{\dagger}(\mathbf{q}) = \sum_{\mathbf{k}} c_{\uparrow}^{\dagger}(\mathbf{k} + \mathbf{q}) c_{\downarrow}(\mathbf{k}) \lambda(\mathbf{k}, \mathbf{k} + \mathbf{q}), \quad (D2)$$

where we still need to include the form factor because we are projecting microscopic spin operator to the Chern band. Again, we only consider interaction energy

$$H_V = \sum_{\mathbf{q}} V(|\mathbf{q}|) \tilde{\rho}(\mathbf{q}) \tilde{\rho}(-\mathbf{q}), \quad (D3)$$

where we only consider density operator of valley  $-$ .

To make analytical calculation possible, we assume Berry curvature is uniform in BZ and then we can write the form factor in the following form:

$$\lambda(\mathbf{k}, \mathbf{k} + \mathbf{q}) = F(|\mathbf{q}|) e^{\frac{B}{2} i \mathbf{k} \wedge \mathbf{q}}, \quad (D4)$$

where  $F(|q|)$  is a real function which decays with  $|q|$ .  $B = C \frac{2\pi}{A_{\text{BZ}}}$  is the uniform Berry curvature. Then, it is easy to get the following commutation relation:

$$[\tilde{\rho}(\mathbf{q}_1), \tilde{\rho}(\mathbf{q}_2)] = \frac{F(|\mathbf{q}_1|)F(|\mathbf{q}_2|)}{F(|\mathbf{q}_1 + \mathbf{q}_2|)} 2i \sin\left(B \frac{\mathbf{q}_1 \wedge \mathbf{q}_2}{2}\right) \times \tilde{\rho}(\mathbf{q}_1 + \mathbf{q}_2), \quad (D5)$$

$$[\tilde{\rho}(\mathbf{q}_1), S^{\dagger}(\mathbf{q}_2)] = \frac{F(|\mathbf{q}_1|)F(|\mathbf{q}_2|)}{F(|\mathbf{q}_1 + \mathbf{q}_2|)} 2i \sin\left(B \frac{\mathbf{q}_1 \wedge \mathbf{q}_2}{2}\right) \times S^{\dagger}(\mathbf{q}_1 + \mathbf{q}_2), \quad (D6)$$

which resemble the GMP algebra in projected Landau level except an additional factor  $B$ .

Then, it is easy to calculate spin-wave dispersion following the standard approach

$$[H_V, S^{\dagger}(\mathbf{q})]|\Omega\rangle = 4 \sum_{\mathbf{p}} \sin^2\left(B \frac{\mathbf{p} \wedge \mathbf{q}}{2}\right) F^2(|\mathbf{p}|) \times V(|\mathbf{p}|) S^{\dagger}(\mathbf{q})|\Omega\rangle. \quad (D7)$$

During this derivation, we use the fact that  $\rho(\mathbf{q})|\Omega\rangle = 0$  for any  $\mathbf{q} \neq 0$ . Therefore, the dispersion of spin-wave excitation is

$$\xi(\mathbf{q}) = 4 \sum_{\mathbf{p}} \sin^2\left(B \frac{\mathbf{p} \wedge \mathbf{q}}{2}\right) V(|\mathbf{p}|) F^2(|\mathbf{p}|). \quad (D8)$$

At  $|\mathbf{q}| \rightarrow 0$  limit, we get dispersion of spin wave:

$$\xi(\mathbf{q}) = A |\mathbf{q}|^2 \quad (D9)$$

with

$$A = \frac{B^2}{2} \sum_{\mathbf{p}} |\mathbf{p}|^2 V(|\mathbf{p}|) F^2(|\mathbf{p}|), \quad (D10)$$

where we used the fact that  $\langle \sin^2 \theta \rangle = \frac{1}{2}$  when integrating over  $\theta \in [0, 2\pi)$ . From the spin-wave dispersion we can easily

extract the spin phase stiffness and then estimate the energy of skyrmion. Following Ref. [40], the phase stiffness is

$$\rho_s = \frac{nB^2}{4} \sum_{\mathbf{p}} |\mathbf{p}|^2 V(|\mathbf{p}|) F^2(|\mathbf{p}|). \quad (\text{D11})$$

We can define length in the units of  $a_M$  and momentum in the units of  $\frac{1}{a_M}$ . Therefore, density of electrons is  $n = 1$ . There is a  $\frac{1}{2}$  factor coming from spin  $S = \frac{1}{2}$ . In the following, we will keep  $n$  to make the dimension of the equation obvious.

Energy of skyrmion is

$$E_{\text{skyrmion}} = 4\pi \rho_s = n\pi B^2 \sum_{\mathbf{p}} |\mathbf{p}|^2 V(|\mathbf{p}|) F^2(|\mathbf{p}|) \quad (\text{D12})$$

and energy of widely separated skyrmion-antiskyrmion pair is

$$E_{\text{skyrmion pair}} = 2\pi n B^2 \sum_{\mathbf{p}} |\mathbf{p}|^2 V(|\mathbf{p}|) F^2(|\mathbf{p}|). \quad (\text{D13})$$

When  $|\mathbf{q}|$  is large, essentially  $S^\dagger(\mathbf{q}) = \sum_{\mathbf{k}} c_{\uparrow}^\dagger(\mathbf{k} + \mathbf{q}) c_{\downarrow}(\mathbf{k})$   $\lambda(\mathbf{k}, \mathbf{k} + \mathbf{q})$  creates an electron-hole pair with total momentum  $\mathbf{q}$ . Because of GMP algebra implicitly hidden in  $\lambda(\mathbf{k}, \mathbf{k} + \mathbf{q})$ , the distance  $r$  of this electron-hole pair is proportional to  $|\mathbf{q}|$ .  $\xi(|\mathbf{q}| \rightarrow \infty)$  actually corresponds to energy cost of infinitely separated charge- $e$  particle-hole pair:

$$E_{1e;\text{ph}} = 2 \sum_{\mathbf{p}} V(|\mathbf{p}|) F^2(|\mathbf{p}|), \quad (\text{D14})$$

where we approximate  $\sin^2$  term with its averaged value  $\frac{1}{2}$ . Correspondingly, energy cost of charge- $2e$  particle-hole pair should be

$$E_{2e;\text{ph}} = 4 \sum_{\mathbf{p}} V(|\mathbf{p}|) F^2(|\mathbf{p}|). \quad (\text{D15})$$

The energy cost is apparently sensitive to the form of  $F(|\mathbf{p}|)$ , which can only be calculated numerically.

To gain some intuition, we assume the following form of  $F(|\mathbf{p}|)$ :  $F^2(|\mathbf{p}|) = e^{-\frac{\alpha}{2}|\mathbf{p}|^2}$ .  $V(|\mathbf{p}|) = \frac{1}{|\mathbf{p}|}$  is Coulomb interaction. Then, it is easy to get

$$\frac{E_{\text{skyrmion pair}}}{E_{2e;\text{ph}}} = \frac{\pi n B^2}{2\alpha}, \quad (\text{D16})$$

where we add electron density  $n$  to make the above quantity dimensionless. For classic  $\nu_T = 1$  lowest Landau level problem, we have  $\alpha = 1$ ,  $B = 1$ ,  $n = \frac{1}{2\pi}$  for which the fundamental unit of length is  $l_B$ .  $\frac{E_{\text{skyrmion pair}}}{E_{2e;\text{ph}}} = \frac{1}{4}$  is in agreement with the well-known fact that  $\frac{E_{\text{skyrmion pair}}}{E_{1e;\text{ph}}} = \frac{1}{2}$  [32].

For the systems we studied in this paper, we use  $a_M$  as the unit of length and then we have hexagonal BZ with  $A_{\text{BZ}} = \frac{8\pi^2}{\sqrt{3}}$  and density  $n = 1$ . Then, the uniform Berry curvature is  $B = \frac{2\pi C}{A_{\text{BZ}}} = \frac{\sqrt{3}C}{4\pi}$ . Therefore, for  $C = 2$ ,

$$\frac{E_{\text{skyrmion pair}}}{E_{2e;\text{ph}}} = \frac{3}{8\pi\alpha}. \quad (\text{D17})$$

Therefore, when  $\alpha > \frac{3}{8\pi}$ , skyrmion is the cheapest charged excitation.

For the moiré superlattice system, we can numerically calculate  $F^2(|\mathbf{q}|) = |\lambda(0, \mathbf{q})|^2$ . We can also numerically do the integration in Eqs. (D13) and (D10) directly. For valence band of BG/h-BN at  $\theta = 0$ , we get  $\frac{E_{\text{skyrmion pair}}}{E_{2e;\text{ph}}} = 0.58$  by directly integrating Eqs. (D13) and (D10). Therefore, we expect finite densities of skyrmions emerge even in the ground state upon doping away from  $\nu_T = 1$ .

For conduction band of type-I BG/BG, at  $\theta = 1.17^\circ$  and  $U = 8$  meV, we get  $\frac{E_{\text{skyrmion pair}}}{E_{2e;\text{ph}}} = 1.35$ . At  $\theta = 0.9^\circ$  and  $U = 8$  meV,  $\frac{E_{\text{skyrmion pair}}}{E_{2e;\text{ph}}} = 0.87$ . Therefore, whether skyrmions emerge upon doping is quite sensitive to twist angle for this system. Smaller twist angle may be favorable to search for skyrmion superconductor.

## APPENDIX E: FRACTIONAL QUANTUM VALLEY HALL EFFECT

We focus on filling  $\nu_T = 2 + \nu_{\text{eff}}$ . As we argued in the main text,  $\nu_T = 2$  is a spin-polarized QVH state. Then, if we dope additional  $\nu_{\text{eff}}$  electrons, they must carry the other spin. Therefore, we expect to have spin-polarized topological ordered state with effective filling  $\nu_{\text{eff}} = \nu_T - 2$ . One option is that valley is also polarized and we get a FQAH state with effective filling  $\nu_{\text{eff}}$ . However, it may also be possible to have a state for which two valleys have equal density. This state is likely to be a fractional quantum valley Hall (FQVH) state with different valleys having opposite Hall conductivity. Based on our discussion of a FQAH state, there could be a spin-polarized FQVH state at filling  $\nu_{\text{eff}} = \frac{2m}{2km|C|+1}$ . These states have zero-charge Hall conductivity and valley Hall conductivity  $\sigma_{xy}^v = \nu_{\text{eff}} C$ . Taking  $|C| = 1$  as an example, we can have a FQVH state at effective filling  $\nu_{\text{eff}} = \frac{2}{3}$ , which is a  $\frac{1}{3}$  Laughlin state for one valley plus  $-\frac{1}{3}$  Laughlin state for another valley. For  $|C| > 1$  bands, we can have similar states. One interesting direction to explore in the future is to construct a new topological ordered state beyond two decoupled FQHE states with opposite Hall conductivities.

- [1] Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras *et al.*, *Nature (London)* **556**, 80 (2018).
- [2] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, *Nature (London)* **556**, 43 (2018).
- [3] R. Bistritzer and A. H. MacDonald, *Proc. Natl. Acad. Sci. USA* **108**, 12233 (2011).
- [4] G. Chen, L. Jiang, S. Wu, B. Lyu, H. Li, B. L. Chittari, K. Watanabe, T. Taniguchi, Z. Shi, J. Jung, Y. Zhang, and

F. Wang, *Nat. Phys.* (2019), <https://www.nature.com/articles/s41567-018-0387-2>.

- [5] E. M. Spanton, A. A. Zibrov, H. Zhou, T. Taniguchi, K. Watanabe, M. P. Zaletel, and A. F. Young, *Science* **360**, 62 (2018).
- [6] F. Wu, T. Lovorn, E. Tutuc, and A. MacDonald, *Phys. Rev. Lett.* **121**, 026402 (2018).
- [7] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982).

- [8] J. C. Song, P. Samutpraphoot, and L. S. Levitov, *Proc. Natl. Acad. Sci. USA* **112**, 10879 (2015).
- [9] Q. Tong, H. Yu, Q. Zhu, Y. Wang, X. Xu, and W. Yao, *Nat. Phys.* **13**, 356 (2017).
- [10] P. San-Jose, A. Gutiérrez-Rubio, M. Sturla, and F. Guinea, *Phys. Rev. B* **90**, 115152 (2014).
- [11] T. M. R. Wolf, O. Zilberberg, I. Levkivskiy, and G. Blatter, *Phys. Rev. B* **98**, 125408 (2018).
- [12] D. Sheng, Z.-C. Gu, K. Sun, and L. Sheng, *Nat. Commun.* **2**, 389 (2011).
- [13] T. Neupert, L. Santos, C. Chamon, and C. Mudry, *Phys. Rev. Lett.* **106**, 236804 (2011).
- [14] Y.-F. Wang, Z.-C. Gu, C.-D. Gong, and D. N. Sheng, *Phys. Rev. Lett.* **107**, 146803 (2011).
- [15] E. Tang, J.-W. Mei, and X.-G. Wen, *Phys. Rev. Lett.* **106**, 236802 (2011).
- [16] N. Regnault and B. A. Bernevig, *Phys. Rev. X* **1**, 021014 (2011).
- [17] K. Sun, Z. Gu, H. Katsura, and S. Das Sarma, *Phys. Rev. Lett.* **106**, 236803 (2011).
- [18] E. J. Bergholtz and Z. Liu, *Int. J. Mod. Phys. B* **27**, 1330017 (2013).
- [19] T. Neupert, C. Chamon, T. Iadecola, L. H. Santos, and C. Mudry, *Phys. Scr.* **2015**, 014005 (2015).
- [20] S. A. Parameswaran, R. Roy, and S. L. Sondhi, *C. R. Phys.* **14**, 816 (2013).
- [21] C.-Z. Chang, J. Zhang, X. Feng, J. Shen, Z. Zhang, M. Guo, K. Li, Y. Ou, P. Wei, L.-L. Wang *et al.*, *Science* **340**, 167 (2013).
- [22] J. Checkelsky, R. Yoshimi, A. Tsukazaki, K. Takahashi, Y. Kozuka, J. Falson, M. Kawasaki, and Y. Tokura, *Nat. Phys.* **10**, 731 (2014).
- [23] X. Kou, S.-T. Guo, Y. Fan, L. Pan, M. Lang, Y. Jiang, Q. Shao, T. Nie, K. Murata, J. Tang *et al.*, *Phys. Rev. Lett.* **113**, 137201 (2014).
- [24] A. Kandala, A. Richardella, S. Kempinger, C.-X. Liu, and N. Samarth, *Nat. Commun.* **6**, 7434 (2015).
- [25] H. C. Po, L. Zou, A. Vishwanath, and T. Senthil, *Phys. Rev. X* **8**, 031089 (2018).
- [26] N. Ashcroft and N. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976).
- [27] S. M. Girvin, A. H. MacDonald, and P. M. Platzman, *Phys. Rev. B* **33**, 2481 (1986).
- [28] S. A. Parameswaran, R. Roy, and S. L. Sondhi, *Phys. Rev. B* **85**, 241308(R) (2012).
- [29] Y.-H. Zhang and T. Senthil, [arXiv:1809.05110](https://arxiv.org/abs/1809.05110).
- [30] L. Ju, Z. Shi, N. Nair, Y. Lv, C. Jin, J. Velasco Jr, C. Ojeda-Aristizabal, H. A. Bechtel, M. C. Martin, A. Zettl, J. Analytis, and F. Wang, *Nature (London)* **520**, 650 (2015).
- [31] K. He, Y. Wang, and Q.-K. Xue, *Annu. Rev. Condens. Matter Phys.* **9**, 329 (2018).
- [32] S. L. Sondhi, A. Karlhede, S. A. Kivelson, and E. H. Rezayi, *Phys. Rev. B* **47**, 16419 (1993).
- [33] Y.-F. Wang, H. Yao, C.-D. Gong, and D. N. Sheng, *Phys. Rev. B* **86**, 201101(R) (2012).
- [34] Z. Liu, E. J. Bergholtz, H. Fan, and A. M. Läuchli, *Phys. Rev. Lett.* **109**, 186805 (2012).
- [35] A. Sterdyniak, C. Repellin, B. A. Bernevig, and N. Regnault, *Phys. Rev. B* **87**, 205137 (2013).
- [36] G. Möller and N. R. Cooper, *Phys. Rev. Lett.* **115**, 126401 (2015).
- [37] M. Barkeshli and X.-L. Qi, *Phys. Rev. X* **2**, 031013 (2012).
- [38] J. Jung, A. Raoux, Z. Qiao, and A. H. MacDonald, *Phys. Rev. B* **89**, 205414 (2014).
- [39] T. Fukui, Y. Hatsugai, and H. Suzuki, *J. Phys. Soc. Jpn.* **74**, 1674 (2005).
- [40] S. M. Girvin, in *Topological Aspects of Low Dimensional Systems* (Springer, Berlin, 1999), pp. 53–175.