

# Pseudomagnetic fields, particle-hole asymmetry, and microscopic effective continuum Hamiltonians of twisted bilayer graphene

Jian Kang<sup>1,\*</sup> and Oskar Vafek<sup>2,3,†</sup>

<sup>1</sup>*School of Physical Science and Technology, ShanghaiTech University, Shanghai 200031, China*

<sup>2</sup>*National High Magnetic Field Laboratory, Tallahassee, Florida 32310, USA*

<sup>3</sup>*Department of Physics, Florida State University, Tallahassee, Florida 32306, USA*



(Received 26 August 2022; revised 15 December 2022; accepted 19 December 2022; published 9 February 2023)

Using the method developed in the companion paper [O. Vafek and J. Kang, Continuum effective Hamiltonian for graphene bilayers for an arbitrary smooth lattice deformation from microscopic theories, *Phys. Rev. B* **107**, 075123 (2023)], we construct effective continuum theories for two different microscopic tight-binding models of twisted bilayer graphene at a twist angle of  $1.05^\circ$ , one Slater-Koster based and the other *ab initio* Wannier based. The energy spectra obtained from the continuum theory—either for rigid twist or including lattice relaxation—are found to be in nearly perfect agreement with the spectra from tight-binding models when the gradient expansion is carried out to second order, demonstrating the validity of the method. We also analyze the properties of the Bloch states of the resulting narrow bands, finding non-negligible particle-hole symmetry breaking near the  $\Gamma$  point in our continuum theory constructed for the *ab initio*-based microscopic model due to a term in the continuum theory that was previously overlooked. This reveals the difference with all existing continuum models where the particle-hole symmetry of the narrow band Hilbert space is nearly perfect.

DOI: [10.1103/PhysRevB.107.075408](https://doi.org/10.1103/PhysRevB.107.075408)

## I. INTRODUCTION

Recent discoveries of electronic correlations in twisted bilayer graphene (TBG), including correlated insulators [1], superconductivity [2], (quantum) anomalous Hall state [3,4], and others [5–27] have generated enthusiasm among both experimentalists and theorists. It is becoming clear that the interplay between band topology and strong electronic interactions plays an essential role in understanding the remarkable phenomena [28–59]. However, many of the key questions, such as the exact ground states and the mechanism of superconductivity, still remain open.

The most common theoretical approach to studying the correlated states is to start with a continuum effective Hamiltonian, often referred to as the Bistritzer-MacDonald (BM) model [60], which gives isolated narrow bands for a range of near-magic twist angles, and then to project the Coulomb interaction onto the wave functions of the narrow bands (sometimes including a few remote bands as well). The BM model [60] has achieved success in many respects. It correctly predicts the first magic angle where the bands around the charge neutrality point (CNP) become extremely narrow and captures their band topology. For relaxed structures, however, the BM model—which was originally derived for a rigid twist—does not include terms which are nominally of the same order in gradient expansion as the ones which are kept, such as the pseudomagnetic fields induced by the  $C_3$  symmetric strain from lattice relaxation. Moreover, next-order gradient terms are needed to accurately capture the narrow

bands near the magic angle due to the anomalously small noninteracting bandwidth obtained without such terms [61]. In addition, the narrow band wave functions of the BM model are nearly particle-hole (p-h) symmetric [58]. The presence of the p-h symmetry is known to play important role in choosing the correlated ground states [43,45,58,62]. Experimentally, it is also seen to be broken at low temperature in that various correlated states appear more stable on either the hole or the electron side of the CNP. This motivates development of a more accurate low energy effective continuum model for TBG.

The goal of this paper is to apply the general formulas developed in the previous companion paper [61] for an arbitrary smooth atomic displacement  $\mathbf{u}_j(\mathbf{r})$  to the specific case of TBG with the relative twist angle  $\theta = 1.05^\circ$ . The atomic displacement fields' configurations are computed by first fixing  $\theta$  and then minimizing the combination of the intralayer elastic terms and the interlayer adhesion terms computed using generalized stacking fault energy (GSFE) functions. We do so for two sets of GSFE parameters found in the literature [63,64]. In both cases, the regions of AB stacking in the moire pattern grow at the expense of the AA regions compared to just the rigid twist configuration, although the quantitative differences between the two models lead to smoother deformation fields for the set of parameters in Ref. [64]. For both models, we perform the Helmholtz decomposition of the displacement field due to the atomic relaxation [see Eq. (25)] and find that in both models it is dominated by the curl of an out-of-plane field  $\hat{z}\varepsilon^U(\mathbf{x})$ . The scalar field  $\varepsilon^U(\mathbf{x})$  is in turn spatially periodic with the triangular moire pattern and it is dominated by its first Fourier harmonic [see Eqs. (26) and (27) and Table I]. The  $\varepsilon^U(\mathbf{x})$  field for the set of relaxation parameters in Ref. [63] also obtains the contribution from higher Fourier

\*kangjian@shanghaitech.edu.cn

†vafek@magnet.fsu.edu

TABLE I. Parameters of the elastic theory and the lattice relaxation obtained from Refs. [63,64], where  $a$  is the lattice constant of the undistorted monolayer graphene.

Elastic/adhesion parameter	$\mathcal{K}$	$\mathcal{G}$	$c_1$	$c_2$	$c_3$
Ref. [63]	12.82 eV/Å <sup>2</sup>	9.57 eV/Å <sup>2</sup>	3.206 meV/Å <sup>2</sup>	0	0
Ref. [64]	13.265 eV/Å <sup>2</sup>	9.035 eV/Å <sup>2</sup>	0.7755 meV/Å <sup>2</sup>	-0.071 meV/Å <sup>2</sup>	-0.018 meV/Å <sup>2</sup>
Lattice relaxation	$\bar{\epsilon}_1/a^2$	$\bar{\epsilon}_2/a^2$	$\bar{\epsilon}_3/a^2$	$\bar{\epsilon}_4/a^2$	$\bar{\epsilon}_5/a^2$
Ref. [63]	0.4243	0.0222	0.0354	0.0039	0.0047
Ref. [64]	0.2270	0.0014	0.0064	-0.0002	0.0002

harmonics, leading to larger momentum transfer in the interlayer tunneling.

We next input the thus-determined atomic displacement fields into the formulas for the continuum Hamiltonian developed in the previous paper, expanding up to second order in gradients in the intralayer Hamiltonian and up to first order in gradients in the interlayer Hamiltonian. For the intralayer Hamiltonian,  $H_{\text{intra}}$ , we find an efficient way to compute the desired parameters of the continuum model from the microscopic tight-binding functions of Refs. [65,66] by Poisson resummation of the power-law decaying momentum space sums into real space, where they fall off exponentially fast. Each moment of the position vector weighted with the intralayer hopping function is accompanied by a gradient of either the fermion field or the atomic displacement field, and contributes a factor of  $|\mathbf{g}|a$ , where  $|\mathbf{g}| \sim |\mathbf{K}|\theta$  (with  $\theta$  in radians), i.e., a factor of  $\sim 0.08$ . For the interlayer Hamiltonian  $H_{\text{inter}}$ , the tunneling falls off fast in the momentum space once the wave vector significantly exceeds the inverse of the interlayer separation  $1/d_0$ , making the direct momentum summation

efficient. At the same time, each moment of the position vector weighted with the interlayer hopping function is also accompanied by a gradient of either the fermion field or the atomic displacement field, and thus contributes a factor of  $\sim |\mathbf{g}|d_0$ . Since  $d_0 \simeq 1.36a$ , the higher order gradient terms are suppressed by similar factors in the  $H_{\text{intra}}$  and  $H_{\text{inter}}$ .

At  $\theta = 1.05^\circ$ , the first-order gradient of fermion fields or of the atomic displacement fields in  $H_{\text{intra}}$  are of the same order as the contact terms in  $H_{\text{inter}}$  [33]. The second order in gradients intralayer terms are, in turn, of the same order as the first order in the gradient interlayer terms. This pattern continues for the higher order terms. As shown in Figs. 17 and 18, the disagreement between the first-order continuum Hamiltonian spectrum and the exact tight-binding spectrum is  $\sim 10$  meV, i.e., of the order of the narrow bandwidth. On the other hand, including the second-order terms in the continuum Hamiltonian improves the agreement significantly as seen in Fig. 1, with a nearly perfect agreement throughout the moire Brillouin zone; the largest disagreement is near the  $\Gamma$  point where there is at most 0.7 meV difference for the model of

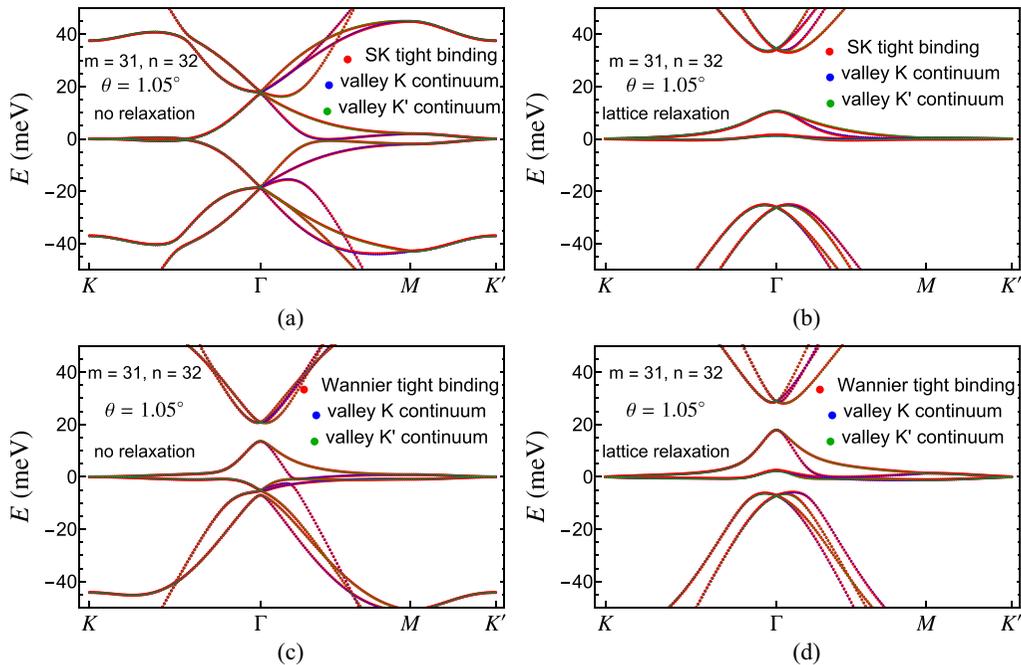


FIG. 1. Comparison of the energy spectra near the CNP obtained using the microscopic tight-binding model (red) and the continuum theory (blue for valley  $\mathbf{K}$  and green for valley  $\mathbf{K}'$ ) for the Slater-Koster (SK)-based model in Ref. [65] (above) and Wannier-based model of Ref. [66] (below) in the absence (left) and presence (right) of the lattice relaxation.

Ref. [65] and at most 0.3 meV difference for the model of Ref. [66].

Thus, the continuum Hamiltonian at valley  $\mathbf{K}$  is

$$H_{\text{eff}}^{\mathbf{K}} = H_{\text{intra}} + H_{\text{inter}}, \quad (1)$$

with  $H_{\text{intra}}$  given in Eq. (32) and  $H_{\text{inter}}$  in Eq. (40) together with the parameters in Tables II and III constitute a highly accurate continuum model for TBG at  $\theta = 1.05^\circ$  obtained directly from the *ab initio* microscopic tight-binding models, with or without relaxation, using the systematic gradient expansion. The effective Hamiltonian for valley  $\mathbf{K}'$  is readily obtained from  $H_{\text{eff}}^{\mathbf{K}}$  by the spinless time-reversal symmetry.

In addition to studying the energy spectra, we also analyze the wave functions for the resulting isolated narrow bands. First, we do so by computing the sublattice polarization as well as the Wilson loops [35,37,44]. Second, we quantify the degree of the p-h asymmetry in our continuum models for different momenta in the moire BZ by computing the deviation from unitarity of the momentum-resolved projected p-h operator. A momentum-averaged version of this operator was analyzed for the BM model in Ref. [58], where p-h symmetry was found to be nearly perfect. We define the deviation from unitarity as the difference of the smallest singular value from unity, confirming the finding of Ref. [58] of nearly perfect p-h symmetry in the BM model, where the p-h is broken by at most 1% near the  $\Gamma$  point. Further, we do this analysis for the continuum model obtained from the tight-binding model of Ref. [65], with relaxed lattice configuration, where we find at most 1.8% deviation from unitarity near the  $\Gamma$  point [see Fig. 11(a)]. On the other hand, for the continuum model obtained from the tight-binding model of Ref. [66], interestingly, the p-h symmetry is broken by  $\sim 16\%$  near the  $\Gamma$  point [see Fig. 11(b)]. The stronger p-h symmetry breaking in the model of Ref. [66] is due to the angle dependence of the microscopic interlayer hopping, resulting in a larger p-h symmetry-breaking contact interlayer tunneling term in the continuum theory, which we dubbed  $w_3$  previously (see the Supplemental Material of Ref. [62]). The significance of such sizable p-h symmetry breaking for the correlated states will be presented in a separate paper.

Finally, we analyze the effect of the atomic relaxation-induced pseudovector potential terms on the narrow bandwidth by studying the first-order model. Such terms appear already at the first order in gradient expansion, so there is no justification for dropping them in the BM model with the relaxation. Because the pseudovector potential terms are of the same order as the contact interlayer tunneling terms  $w_{0,1}$ , one may naively conclude that their effect is to broaden the bandwidth by a similar order and to prevent the magic-angle phenomenon. While they do increase the bandwidth at the ‘‘old’’ magic angle (i.e., without the periodic relaxation induced vector potential), we find that their effect can be compensated by a change of the twist angle, recovering the narrow band at a new (smaller) magic angle. We were able to demonstrate this by solving the problem analytically in the chiral limit including the pseudovector potential terms absent in Ref. [38]. We highlight the importance of  $C_3$  symmetry for this compensation.

This paper is organized as follows: In Sec. II, we calculate the lattice relaxation for two sets of GSFE parameters in

Refs. [63,64]. In Sec. III, we present our effective continuum theory of the TBG for two microscopic tight-binding models in Refs. [65,66], with the corresponding parameter values listed in Tables II, III, and V. We also plot the energy spectra of the continuum effective theories including the remote bands up  $\sim 200$  meV. Their nearly perfect agreement with the spectra from the tight-binding models demonstrates the validity of the constructed continuum theories. In Sec. IV, we investigate the properties of the Bloch states of the narrow bands, including the sublattice polarization, Wilson loops, and the p-h asymmetry. As shown in Fig. 19, the p-h asymmetry is dominated by  $w_3$ , a previously overlooked interlayer contact coupling. Section V studies the exactly flat-band limit when including the lattice relaxation-induced pseudomagnetic field. Finally, Sec. VI is devoted to the summary.

## II. RELAXED LATTICE DEFORMATION IN THE VICINITY OF THE FIRST MAGIC ANGLE

In this section, we follow the approach presented in Ref. [63] to obtain the lattice distortion when the twist angle is near the first magic angle. We assume that the lattice distortion is independent of the sublattice labeled by  $S = A$  or  $S = B$ , i.e.,  $U_{j,S}^{\parallel,\perp}(\mathbf{x}) = U_j^{\parallel,\perp}(\mathbf{x})$ , where  $j = t$  refers to the top layer and  $j = b$  refers to the bottom layer. We further neglect the lattice corrugation, so  $U_b^\perp = 0$  and  $U_t^\perp = d_0 \hat{z}$ . Under these assumptions, the intralayer elastic energy of the graphene system can be written as

$$U_E = \frac{1}{2} \sum_{j=t,b} \int d^2\mathbf{x} [\mathcal{K}(\partial_x U_{j,x}^\parallel + \partial_y U_{j,y}^\parallel)^2 + \mathcal{G}((\partial_x U_{j,x}^\parallel - \partial_y U_{j,y}^\parallel)^2 + (\partial_y U_{j,x}^\parallel + \partial_x U_{j,y}^\parallel)^2)], \quad (2)$$

where  $\mathcal{K}$  and  $\mathcal{G}$  are the bulk and shear modulus of the monolayer graphene; their values for two different models are given in Table I. It is more convenient to introduce symmetric and antisymmetric combinations

$$U^+ = \frac{1}{2}(U_t^\parallel + U_b^\parallel), \quad (3)$$

$$U^- = U_t^\parallel - U_b^\parallel. \quad (4)$$

The intralayer elastic energy then can be expressed as

$$U_E = \int d^2\mathbf{x} [\mathcal{K}(\partial_x U_x^+ + \partial_y U_y^+)^2 + \mathcal{G}((\partial_x U_x^+ - \partial_y U_y^+)^2 + (\partial_y U_x^+ + \partial_x U_y^+)^2)] + \frac{1}{4} \int d^2\mathbf{x} [\mathcal{K}(\partial_x U_x^- + \partial_y U_y^-)^2 + \mathcal{G}((\partial_x U_x^- - \partial_y U_y^-)^2 + (\partial_y U_x^- + \partial_x U_y^-)^2)], \quad (5)$$

with  $U^+$  and  $U^-$  decoupled.

In addition to the intralayer elastic energy, we also include the interlayer adhesion energy

$$U_B = \int d^2\mathbf{x} V[U^-(\mathbf{x})], \quad (6)$$

where  $V[U^-(\mathbf{x})]$  is a periodic and even function of the relative displacement  $U^-$ , i.e.,  $V[U^-] = V[-U^-]$  and  $V[U^-] =$

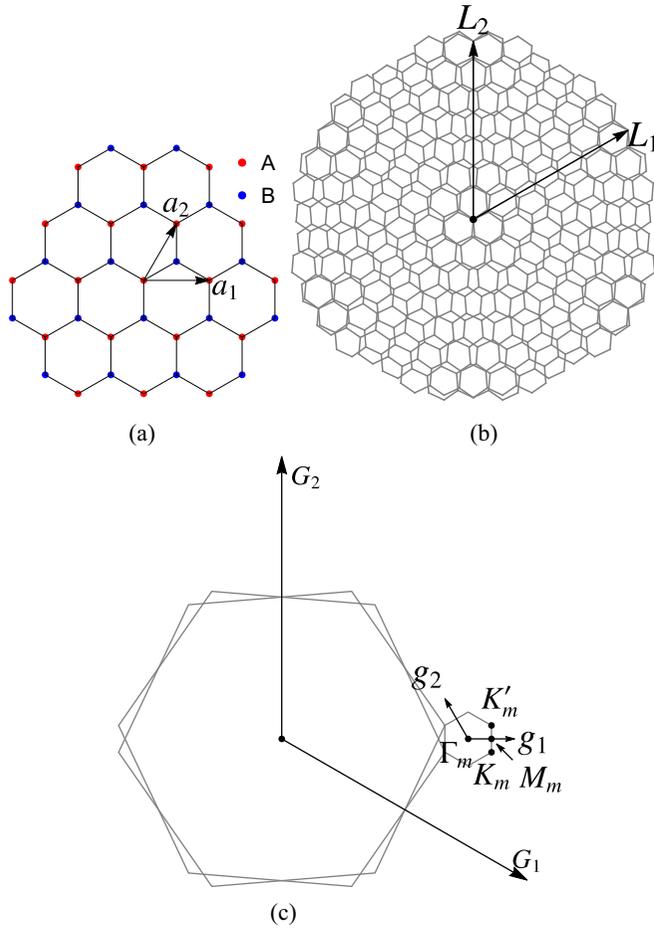


FIG. 2. Schematic plot for (a) the monolayer lattice vectors  $\mathbf{a}_{1,2}$ , (b) the moire lattice vectors  $\mathbf{L}_{1,2}$ , and (c) their associated reciprocal lattice vectors  $\mathbf{G}_{1,2}$  and  $\mathbf{g}_{1,2}$ .

$V[U^- + \mathbf{a}_i]$  ( $i = 1, 2$ ), where  $\mathbf{a}_1$  and  $\mathbf{a}_2$  are the primitive lattice vectors. As shown in Fig. 2, they are defined as

$$\mathbf{a}_1 = a(1, 0), \quad \mathbf{a}_2 = a\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right). \quad (7)$$

Therefore, the Fourier transform of  $V[U^-]$  can be expressed as

$$V[U^-(\mathbf{x})] = \sum_{\mathbf{G}} V_{\mathbf{G}} \cos(\mathbf{G} \cdot U^-(\mathbf{x})), \quad (8)$$

where  $\mathbf{G} = m\mathbf{G}_1 + n\mathbf{G}_2$  is a reciprocal lattice vector of the undistorted monolayer graphene, with integer  $m, n$  and  $\mathbf{G}_1 = \frac{2\pi}{a}(1, -\frac{1}{\sqrt{3}})$  and  $\mathbf{G}_2 = \frac{2\pi}{a}(0, \frac{2}{\sqrt{3}})$ . The Fourier coefficients  $V_{\mathbf{G}}$  fall off with large  $\mathbf{G}$ , so the sum can be truncated after a few shells. Furthermore, different  $V_{\mathbf{G}}$ s are related by symmetries. As a consequence, the adhesion potential has the form [64]

$$\begin{aligned} V[U^-(\mathbf{x})] = & c_0 + c_1(\cos(\mathbf{G}_1 \cdot U^-) + \cos(\mathbf{G}_2 \cdot U^-)) \\ & + \cos((\mathbf{G}_1 + \mathbf{G}_2) \cdot U^-) \\ & + c_2[\cos((\mathbf{G}_1 - \mathbf{G}_2) \cdot U^-) \\ & + \cos((2\mathbf{G}_1 + \mathbf{G}_2) \cdot U^-) \end{aligned}$$

$$\begin{aligned} & + \cos((\mathbf{G}_1 + 2\mathbf{G}_2) \cdot U^-)] \\ & + c_3[\cos(2\mathbf{G}_1 \cdot U^-) + \cos(2\mathbf{G}_2 \cdot U^-) \\ & + \cos(2(\mathbf{G}_1 + \mathbf{G}_2) \cdot U^-)]. \end{aligned} \quad (9)$$

The values of  $c_j$ 's are given in Table I.

For TBG, the displacement vector field  $U^-$  contains two parts, the relative twist between the two layers and the relative displacement due to the lattice relaxation or the heterostrain [61]:

$$U^-(\mathbf{x}) = \theta \hat{z} \times \mathbf{x} + \delta U(\mathbf{x}). \quad (10)$$

It follows that

$$\begin{aligned} \partial_x U_x^- \pm \partial_y U_y^- &= \partial_x \delta U_x \pm \partial_y \delta U_y, \\ \partial_x U_y^- + \partial_y U_x^- &= \partial_x \delta U_y^- + \partial_y \delta U_x^-. \end{aligned} \quad (11)$$

Therefore, as physically expected, the rigid twist term does not contribute to the intralayer elastic energy.

Introducing

$$\mathbf{g}_{\mathbf{G}} = -\theta \hat{z} \times \mathbf{G}, \quad (12)$$

if the twist angle  $\theta$  is small, then  $\mathbf{g}_{\mathbf{G}}$  becomes a reciprocal vector of the moire superlattice. Note that with this definition, we have a one-to-one mapping between the set of  $\mathbf{G}$ 's and the set of all moire reciprocal lattice vectors,  $\{\mathbf{g}\}$ . As seen in Fig. 2, the basis vectors of the set  $\{\mathbf{g}\}$  are

$$\mathbf{g}_1 = \frac{2\pi}{L} \left( \frac{2}{\sqrt{3}}, 0 \right), \quad \mathbf{g}_2 = \frac{2\pi}{L} \left( -\frac{1}{\sqrt{3}}, 1 \right), \quad (13)$$

where  $L = a/(2 \sin \frac{\theta}{2})$  is the length of  $\mathbf{L}_i$ , and the primitive moire lattice vectors are

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

Indeed, for  $\mathbf{G} = m_1\mathbf{G}_1 + m_2\mathbf{G}_2$ , with integer  $m, n$ ,

$$\mathbf{g}_{\mathbf{G}} = (m_2 - m_1)\mathbf{g}_1 - m_1\mathbf{g}_2 \quad (15)$$

because  $\mathbf{g}_{\mathbf{G}_1} = -(\mathbf{g}_1 + \mathbf{g}_2)$  and  $\mathbf{g}_{\mathbf{G}_2} = \mathbf{g}_1$ . Using Eqs. (12) and (10), we obtain

$$\cos(\mathbf{G} \cdot U^-(\mathbf{x})) = \cos(\mathbf{g}_{\mathbf{G}} \cdot \mathbf{x} + \mathbf{G} \cdot \delta U(\mathbf{x})). \quad (16)$$

In addition to the moire lattice constant  $L$  set by interatomic distance  $a$  and twist angle  $\theta$ , the combination of the intralayer elastic energy  $U_E$  and the interlayer adhesion potential  $U_B$  introduces another characteristic length scale  $l = a\sqrt{\frac{G+\kappa}{c_1}}$  [63]. If  $L \gg l$ , the interlayer adhesion dominates over the intralayer elastic energy and the relaxation maximizes the AB/BA stacking region while minimizing the AA stacking regions. As a consequence, the system breaks up into triangular domains of AB/BA stacking separated by domain walls [67]. On the other hand, if  $L \ll l$  then the lattice relaxation is weak and the structure is close to the one with rigid twist only. In this case, the size of AB/BA and AA stacking regions is about the same.

In the rest of this section, we focus on the bilayer system with the commensurate twist angle, i.e., the moire unit cell vectors  $\mathbf{L}_{1,2}$  satisfying

$$\mathbf{L}_1 = m\mathbf{a}_1 + n\mathbf{a}_2, \quad (17)$$

$$\mathbf{L}_2 = -n\mathbf{a}_1 + (m+n)\mathbf{a}_2, \quad (18)$$

where  $m$  and  $n$  are two integers, with the corresponding twist angle  $\theta = \cos^{-1}\left(\frac{m^2+4mn+n^2}{2(m^2+mn+n^2)}\right)$ .

At the first magic angle  $\theta = 1.05^\circ$  (with  $m = 31$  and  $n = 32$ ) and the parameters listed in Table I,  $L \approx 0.65l$  for Ref. [63] and  $L \approx 0.32l$  for Ref. [64]. Therefore, the lattice relaxation is expected to be stronger for Ref. [63], with a larger increase of AB/BA stacking regions and a larger decrease of AA stacking regions. As such, the contribution of higher Fourier harmonics to the relaxation is larger, as confirmed by the value of  $\tilde{\epsilon}_3$  in Table I, defined via Eqs. (25)–(27) and obtained from minimizing  $U_E + U_B$  defined in Eqs. (5) and (6). For Ref. [64], the lattice relaxation is weaker and smoother; it is dominated by the lowest harmonic terms, i.e.,  $\tilde{\epsilon}_1$ . As we show in the later sections, an important consequence of this lattice relaxation for the electronic structure is that one must go beyond the BM model [60] and include the interlayer tunneling terms with a larger momentum transfer than just the first shell to obtain an accurate description of the magic-angle narrow bands.

As mentioned, the lattice relaxation is obtained by minimizing  $U_E + U_B$  with respect to  $\delta\mathbf{U}(\mathbf{x})$ . This leads to the differential equation

$$-\frac{1}{2} \begin{pmatrix} (\mathcal{G} + \mathcal{K})\partial_x^2 + \mathcal{G}\partial_y^2 & \mathcal{K}\partial_x\partial_y \\ \mathcal{K}\partial_x\partial_y & (\mathcal{G} + \mathcal{K})\partial_y^2 + \mathcal{G}\partial_x^2 \end{pmatrix} \begin{pmatrix} \delta U_x \\ \delta U_y \end{pmatrix} = \sum_{\mathbf{G}} V_{\mathbf{G}} \sin(\mathbf{g}_{\mathbf{G}} \cdot \mathbf{x} + \mathbf{G} \cdot \delta\mathbf{U}(\mathbf{x})) \begin{pmatrix} G_x \\ G_y \end{pmatrix}. \quad (19)$$

Because the lattice relaxation field  $\delta\mathbf{U}(\mathbf{x})$  is a periodic function of  $\mathbf{x}$  that satisfies  $\delta\mathbf{U}(\mathbf{x}) = \delta\mathbf{U}(\mathbf{x} + \mathbf{L})$  where  $\mathbf{L} = n_1\mathbf{L}_1 + n_2\mathbf{L}_2$  is any moire superlattice vector, its Fourier transform can be written as

$$\delta\mathbf{U}(\mathbf{x}) = \sum_{\mathbf{g}} \delta\tilde{\mathbf{U}}(\mathbf{g}) e^{i\mathbf{g} \cdot \mathbf{x}}. \quad (20)$$

Here  $\sum_{\mathbf{g}}$  sums over all the reciprocal vectors of the moire superlattice, i.e., over the same set as in Eq. (15). Introducing the Fourier sum of  $\sin(\mathbf{g}_{\mathbf{G}} \cdot \mathbf{x} + \mathbf{G} \cdot \delta\mathbf{U})$ , we obtain

$$\frac{1}{2} \begin{pmatrix} (\mathcal{G} + \mathcal{K})g_x^2 + \mathcal{G}g_y^2 & \mathcal{K}g_x g_y \\ \mathcal{K}g_x g_y & (\mathcal{G} + \mathcal{K})g_y^2 + \mathcal{G}g_x^2 \end{pmatrix} \begin{pmatrix} \delta\tilde{U}_x(\mathbf{g}) \\ \delta\tilde{U}_y(\mathbf{g}) \end{pmatrix} = \sum_{\mathbf{G}} V_{\mathbf{G}} f_{\mathbf{g}}^{\delta\mathbf{U}}(\mathbf{G}) \begin{pmatrix} G_x \\ G_y \end{pmatrix}, \quad (21)$$

where

$$\sin(\mathbf{g}_{\mathbf{G}} \cdot \mathbf{x} + \mathbf{G} \cdot \delta\mathbf{U}(\mathbf{x})) = \sum_{\mathbf{g}} f_{\mathbf{g}}^{\delta\mathbf{U}}(\mathbf{G}) e^{i\mathbf{g} \cdot \mathbf{x}}. \quad (22)$$

The term  $\delta\tilde{\mathbf{U}}(\mathbf{g} = 0)$  corresponds to a uniform relative translation between two twisted and deformed layers. To show that we can set it to zero, we decompose  $\delta\mathbf{U}$  into two parts:  $\delta\mathbf{U}(\mathbf{x}) = \delta\mathbf{U}_0 + \delta\mathbf{U}_1(\mathbf{x})$ , where  $\delta\mathbf{U}_0 = \delta\tilde{\mathbf{U}}(\mathbf{g} = 0)$ , and  $\langle \delta\mathbf{U}_1(\mathbf{x}) \rangle_{\mathbf{x}} = 0$ , or, equivalently,  $\delta\tilde{\mathbf{U}}_1(\mathbf{g} = 0) = 0$ . Applying Eq. (12),

$$\begin{aligned} \cos(\mathbf{g}_{\mathbf{G}} \cdot \mathbf{x} + \mathbf{G} \cdot \delta\mathbf{U}) \\ = \cos(\mathbf{g}_{\mathbf{G}} \cdot (\mathbf{x} - \theta^{-1}\hat{z} \times \delta\mathbf{U}_0) + \mathbf{G} \cdot \delta\mathbf{U}_1(\mathbf{x})). \end{aligned} \quad (23)$$

The interlayer adhesion energy can now be written as a function that depends only on  $\delta\mathbf{U}_1$ :

$$\begin{aligned} U_B[\delta\mathbf{U}] &= \int d^2\mathbf{x} \sum_{\mathbf{G}} V_{\mathbf{G}} \\ &\quad \times \cos(\mathbf{g} \cdot (\mathbf{x} - \theta^{-1}\hat{z} \times \delta\mathbf{U}_0) + \mathbf{G} \cdot \delta\mathbf{U}_1(\mathbf{x})) \\ &= \int d^2\mathbf{x}' \sum_{\mathbf{G}} V_{\mathbf{G}} \\ &\quad \times \cos(\mathbf{g} \cdot \mathbf{x}' + \mathbf{G} \cdot \delta\mathbf{U}_1(\mathbf{x}' + \theta^{-1}\hat{z} \times \delta\mathbf{U}_0)) \\ &= U_B[\delta\mathbf{U}'_1], \end{aligned} \quad (24)$$

where  $\delta\mathbf{U}'_1(\mathbf{x}) = \delta\mathbf{U}_1(\mathbf{x} + \theta^{-1}\hat{z} \times \delta\mathbf{U}_0)$ . Because  $\langle \delta\mathbf{U}'_1 \rangle_{\mathbf{x}} = \langle \delta\mathbf{U}_1 \rangle_{\mathbf{x}} = 0$ , the interlayer adhesion energy of the configuration  $\delta\mathbf{U}$  is the same as the adhesion energy of  $\delta\mathbf{U}'_1$  whose spatial average vanishes. Additionally, the elastic energies of these two configurations are also the same since the energy depends only on the gradient of the lattice relaxation  $\delta\mathbf{U}$ . Therefore, we can set  $\delta\tilde{\mathbf{U}}(\mathbf{g} = 0) = 0$ . In addition, the parity of  $\delta\mathbf{U}$  is odd, i.e.,  $\delta\mathbf{U}(-\mathbf{x}) = -\delta\mathbf{U}(\mathbf{x})$ , leading to the odd parity of (the purely imaginary)  $\delta\tilde{\mathbf{U}}(\mathbf{g})$ .

Although the sums in Eq. (8), as well as Eqs. (19) and (21), formally include all the reciprocal lattice vectors  $\mathbf{G}$ , only the terms with small magnitude of  $\mathbf{G}$  contribute significantly. This is because we are not in the limit of  $L$  much larger than  $l$  at the first magic angle, which would cause variations of the displacements over a length scale much shorter than  $L$  (i.e., large gradients and therefore many  $\mathbf{g}$ 's across the domain wall separating AB and BA regions). Correspondingly, to numerically solve Eq. (21), the Fourier sum in Eq. (22) can be truncated by including  $\mathbf{g}$ s from just the first five shells as detailed below [see Eqs. (27)].

The nonlinear Eqs. (21) and (22) can be efficiently solved by the iteration method. It starts with a trial solution with  $\delta\mathbf{U}(\mathbf{x}) = 0$ , feeding into Eq. (22) to obtain  $f_{\mathbf{g}}^{\delta\mathbf{U}}(\mathbf{g})$ , and then updating  $\delta\mathbf{U}(\mathbf{x})$  by solving its Fourier components  $\delta\tilde{\mathbf{U}}(\mathbf{g})$  from Eq. (21). The iteration continues until  $\delta\mathbf{U}$  converges. Clearly, the solution is independent of the parameter  $c_0$  of the potential in Eq. (9). Since the out-of-plane corrugation is not included in this model,  $\delta\mathbf{U}(\mathbf{x})$  is a two-dimensional vector field. As such, by Helmholtz theorem it can be decomposed into a sum of a curl-free part (irrotational) and a divergence-free part (solenoidal) as

$$\delta\mathbf{U}(\mathbf{x}) = \nabla\varphi^U(\mathbf{x}) + \nabla \times (\hat{z}\varepsilon^U(\mathbf{x})). \quad (25)$$

As shown in Figs. 3 and 4, the numerically obtained lattice relaxation is dominated by the solenoidal part,  $\nabla \times (\hat{z}\varepsilon^U(\mathbf{x}))$ , for the lattice relaxation models of Refs. [63,64], with their choice for parameters listed in Table I. The resulting displacement field [68] is in qualitative agreement with the Bragg interferometry imaging of the strain fields in TBG [69]. In the following calculations, we will neglect the small irrotational part and include only the solenoidal part.

Because the vector field  $\delta\mathbf{U}(\mathbf{x})$  is spatially periodic and vanishes on average, the scalar field  $\varepsilon^U(\mathbf{x})$  is also periodic. It can therefore be written as

$$\varepsilon^U(\mathbf{x}) = \sum_{\mathbf{g}} \tilde{\varepsilon}_{\mathbf{g}}^U e^{i\mathbf{g} \cdot \mathbf{x}}, \quad (26)$$

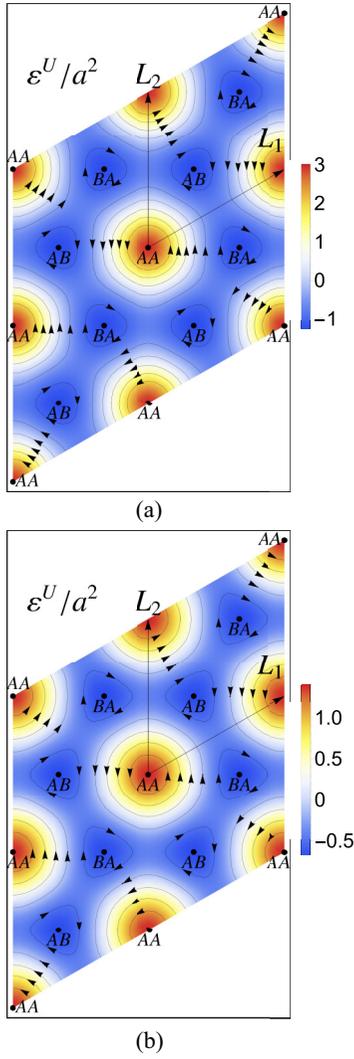


FIG. 3. The contour plot of the scalar field  $\varepsilon^U$  defining the solenoidal component of the atomic displacement field  $\delta\mathbf{U}$  in Eq. (25) for the two models of Ref. [63] (a) and Ref. [64] (b), where  $a$  is the monolayer graphene lattice constant. The arrows point along the divergence-free part of  $\delta\mathbf{U}$ . Near AA, it is in the same direction of the rigid twist of the uniform AA stacked configuration, leading to shrinking of AA stacking moiré region, while around AB/BA, it is in the opposite direction of the rigid twist, resulting in the increase of the AB/BA stacking region.

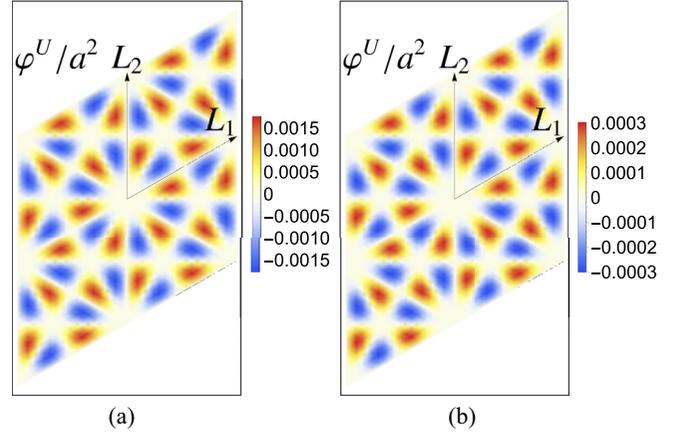


FIG. 4. The density plot of the scalar field  $\varphi^U$  defining the irrotational component of the displacement field  $\delta\mathbf{U}$  in Eq. (25) for the model used in Ref. [63] (a) and in Ref. [64] (b);  $a$  is the monolayer graphene lattice constant. Note that the irrotational component is negligibly small compared to the solenoidal component shown in Fig. 3.

with the  $\mathbf{g}$  sum truncated to five shells as for  $\delta\mathbf{U}(\mathbf{x})$ . Numerically, we found that in both models,  $\varepsilon^U(\mathbf{x})$  is dominated by the following components:

$$\begin{aligned}
 \tilde{\varepsilon}_{\mathbf{g}=0}^U &= 0, \\
 \tilde{\varepsilon}_{\pm\mathbf{g}_1}^U &= \tilde{\varepsilon}_{\pm\mathbf{g}_2}^U = \tilde{\varepsilon}_{\pm(\mathbf{g}_1+\mathbf{g}_2)}^U = \tilde{\varepsilon}_1, \\
 \tilde{\varepsilon}_{\pm(\mathbf{g}_1-\mathbf{g}_2)}^U &= \tilde{\varepsilon}_{\pm(2\mathbf{g}_1+\mathbf{g}_2)}^U = \tilde{\varepsilon}_{\pm(\mathbf{g}_1+2\mathbf{g}_2)}^U = \tilde{\varepsilon}_2, \\
 \tilde{\varepsilon}_{\pm 2\mathbf{g}_1}^U &= \tilde{\varepsilon}_{\pm 2\mathbf{g}_2}^U = \tilde{\varepsilon}_{\pm 2(\mathbf{g}_1+\mathbf{g}_2)}^U = \tilde{\varepsilon}_3, \\
 \tilde{\varepsilon}_{\pm(3\mathbf{g}_1+2\mathbf{g}_2)}^U &= \tilde{\varepsilon}_{\pm(3\mathbf{g}_1+\mathbf{g}_2)}^U = \tilde{\varepsilon}_{\pm(2\mathbf{g}_1+3\mathbf{g}_2)}^U \\
 &= \tilde{\varepsilon}_{\pm(2\mathbf{g}_1-\mathbf{g}_2)}^U = \tilde{\varepsilon}_{\pm(\mathbf{g}_1+3\mathbf{g}_2)}^U = \tilde{\varepsilon}_{\pm(\mathbf{g}_1-2\mathbf{g}_2)}^U = \tilde{\varepsilon}_4, \\
 \tilde{\varepsilon}_{\pm 3\mathbf{g}_1}^U &= \tilde{\varepsilon}_{\pm 3\mathbf{g}_2}^U = \tilde{\varepsilon}_{\pm 3(\mathbf{g}_1+\mathbf{g}_2)}^U = \tilde{\varepsilon}_5.
 \end{aligned} \tag{27}$$

The values of  $\tilde{\varepsilon}_i (i = 1, \dots, 5)$  are listed in Table I.

### III. ACCURATE EFFECTIVE MODEL NEAR THE FIRST MAGIC ANGLE

In the previous paper [61], we derived the effective continuum Hamiltonian using the gradient expansion of the slow envelope function of the fermions from  $\mathbf{K}$  and  $\mathbf{K}'$  points and the slowly varying atomic displacement fields  $\mathbf{U}$  to be

$$\begin{aligned}
 H_{\text{eff}}^{\mathbf{K}} &\simeq \frac{1}{A_{\text{mlg}}} \sum_{S,S'} \sum_{j,j'} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot(\boldsymbol{\tau}_S-\boldsymbol{\tau}_{S'})} \int d^2\mathbf{x} \mathcal{J}_j(\mathbf{x}) \mathcal{J}_{j'}(\mathbf{x}) e^{i(\mathbf{G}+\mathbf{K})\cdot(\mathbf{U}_j^{\parallel}(\mathbf{x})-\mathbf{U}_{j'}^{\parallel}(\mathbf{x}))} \int d^2\mathbf{y} e^{-i(\mathbf{G}+\mathbf{K})\cdot\mathbf{y}} \\
 &\times e^{i\frac{\mathbf{y}}{2}\cdot\nabla_{\mathbf{x}}(\mathbf{U}_j^{\parallel}(\mathbf{x})+\mathbf{U}_{j'}^{\parallel}(\mathbf{x}))\cdot(\mathbf{G}+\mathbf{K})} \left( t_{\text{sym}}^{jj'} [\mathbf{y} + \mathbf{U}_j^{\perp}(\mathbf{x}) - \mathbf{U}_{j'}^{\perp}(\mathbf{x}), \boldsymbol{\delta}_S, \boldsymbol{\delta}_{S'}] + t_{jj',S}^{(1)}(\mathbf{y}) \frac{1}{3} \sum_{\alpha=1}^3 \delta\theta_{j,S}^{(\alpha)} + t_{jj',S'}^{(2)}(\mathbf{y}) \frac{1}{3} \sum_{\alpha'=1}^3 \delta\theta_{j',S'}^{(\alpha')} \right) \\
 &\times \left[ \Psi_{j,S}^{\dagger}(\mathbf{x}) \Psi_{j',S'}(\mathbf{x}) + \frac{\mathbf{y}}{2} \cdot ((\nabla_{\mathbf{x}} \Psi_{j,S}^{\dagger}(\mathbf{x})) \Psi_{j',S'}(\mathbf{x}) - \Psi_{j,S}^{\dagger}(\mathbf{x}) \nabla_{\mathbf{x}} \Psi_{j',S'}(\mathbf{x})) \right] \\
 &+ \sum_{j,S} \int d^2\mathbf{x} (\epsilon_0 + \kappa \nabla \cdot \mathbf{U}_j^{\parallel}(\mathbf{x})) \Psi_{j,S}^{\dagger}(\mathbf{x}) \Psi_{j,S}(\mathbf{x}).
 \end{aligned} \tag{28}$$

In the above,  $S$  and  $S'$  sum over sublattices  $A$  and  $B$ , and  $j$  sums over the top and bottom layers. We also consider the possibility that the hopping constant depends not only on the displacement  $\mathbf{y} + \mathbf{U}_j^\perp(\mathbf{x}) - \mathbf{U}_{j'}^\perp(\mathbf{x})$  but also on the orientation of the nearest-neighbor bonds [66]. Without the lattice distortion, the hoppings are given by  $t_{\text{sym}}^{jj'}[\mathbf{y} + \mathbf{U}_j^\perp(\mathbf{x}) - \mathbf{U}_{j'}^\perp(\mathbf{x}), \delta_S, \delta_{S'}]$ . In the presence of the lattice distortion, the hoppings can be expanded to the first order of the change of bond angles  $\delta\theta_{j,S}^{(\alpha)}$  and  $\delta\theta_{j',S'}^{(\alpha')}$  where  $\alpha$  and  $\alpha'$  are the index of nearest-neighbor bonds, ranging from 1 to 3. Thus, the correction is  $t_{jj',S}^{(1)}(\mathbf{y}) \frac{1}{3} \sum_{\alpha=1}^3 \delta\theta_{j,S}^{(\alpha)} + t_{jj',S'}^{(2)}(\mathbf{y}) \frac{1}{3} \sum_{\alpha'=1}^3 \delta\theta_{j',S'}^{(\alpha')}$ , where  $t_{j \neq j', S}^{(1)}(\mathbf{y}) = \frac{\partial t_{\text{sym}}^{j \neq j'}}{\partial \theta_{j,S}}$  and  $t_{j \neq j', S'}^{(2)}(\mathbf{y}) = \frac{\partial t_{\text{sym}}^{j \neq j'}}{\partial \theta_{j',S'}}$ .

It is worth emphasizing that the electron-phonon coupling can be readily obtained from the formula of the effective continuum theory in Eq. (28). For this purpose, the lattice deformation is decomposed into the static and dynamic parts,  $\mathbf{U}_{j,S} = \mathbf{U}_{j,S}^{(0)} + \mathbf{U}_{j,S}^{(1)}$ .  $\mathbf{U}_{j,S}^{(0)}$  is the static configuration of the lattice deformation obtained by minimizing the sum of the intralayer elastic energy  $U_E$  in Eq. (5) and the interlayer adhesion energy  $U_B$  in Eq. (6). As argued in Sec. II, without external strains,  $\mathbf{U}_{j,S}^{(0)} = \pm \frac{1}{2} \delta \mathbf{U}$  [the signs  $+$  and  $-$  are for the top (bottom) layer], with  $\delta \mathbf{U}$  given by Eqs. (25)–(27). For the two models proposed in Refs. [63,64], the numerical values of  $\delta \mathbf{U}$  are presented in Table I.  $\mathbf{U}_{j,S}^{(1)}$  is the oscillation part of  $\mathbf{U}_{j,S}$ , i.e., the phonon in the bilayer system. Therefore, the expansion of Eq. (28) to a desired order of  $\mathbf{U}_{j,S}^{(1)}$  naturally leads to the coupling between phonons and electrons in such a system.

In what follows, we will apply this formula to derive the effective Hamiltonian of the magic angle TBG with and without the lattice relaxation which we obtained in Sec. II, and compare the energy spectra with the tight-binding models of Refs. [65,66].

As a first step, we consider the Jacobian factor

$$\mathcal{J}_{j,S}(\mathbf{x}) = \left| \det \left( \frac{\partial(\mathbf{x} - \mathbf{U}_j^\parallel(\mathbf{x}))}{\partial \mathbf{x}} \right) \right|^{1/2}. \quad (29)$$

$$H_{\text{intra}}^{(0)} = \int d^2\mathbf{x} \sum_{j=t,b} \sum_{SS'} \Psi_{j,S}^\dagger(\mathbf{x}) \left\{ \mu \delta_{SS'} + v_F \bar{\sigma}_{SS'} \cdot (\mathbf{p}^{(j)} + \gamma \mathcal{A}^{(j)}(\mathbf{x})) + \beta_0 \mathbf{p}^2 \delta_{SS'} + \frac{C_0}{2} (\mathbf{p} \cdot \mathcal{A}(\mathbf{x}) + \mathcal{A}(\mathbf{x}) \cdot \mathbf{p}) \delta_{SS'} \right. \\ \left. + \beta_1 ((p_x^2 - p_y^2) \sigma_1 + 2p_x p_y \sigma_2)_{SS'} + \frac{1}{2} \sum_{\mu} (p_{\mu} \xi_{\mu,SS'}(\mathbf{x}) + \xi_{\mu,SS'}(\mathbf{x}) p_{\mu}) \right\} \Psi_{j,S}(\mathbf{x}), \quad (33)$$

where  $j$  is summed over the top ( $t$ ) and bottom ( $b$ ) layers,  $S, S'$  are summed over the  $A$  and  $B$  sublattices, and  $\mu$  over  $x$  and  $y$  components. In the above, we split  $H_{\text{intra}}$  into two terms:  $H_{\text{intra}}^{(0)}$  and  $\delta H_{\text{intra}}$ . The first term,  $H_{\text{intra}}^{(0)}$ , contains all the contributions up to the second order in gradients whose energy scale is above 1 meV and dominates the second term  $\delta H_{\text{intra}}$ . The numerical values of the coefficients appearing in  $H_{\text{intra}}^{(0)}$  for the two different microscopic models [65,66] can be found in Table II. The definition of the lattice distortion-induced pseudovector fields  $\mathcal{A}(\mathbf{x})$  and the fields  $\xi_{SS'}(\mathbf{x})$  are discussed

below. The second term,  $\delta H_{\text{intra}}$ , contains other second-order gradient contributions that are smaller than 1 meV. We have checked numerically that inclusion of  $\delta H_{\text{intra}}$  does not improve the agreement between the spectra of the tight-binding Hamiltonians and  $H_{\text{eff}}$  in any significant way, as demonstrated in the Fig. 1 where  $\delta H_{\text{intra}}$  is omitted. Therefore, we will focus on  $H_{\text{intra}}^{(0)}$  in the main text. For completeness, we spell out the details of  $\delta H_{\text{intra}}$  in Appendix A.

Since  $\mathbf{U}_j^\parallel(\mathbf{x})$  varies smoothly in the real space, its gradient  $|\nabla \mathbf{U}_j^\parallel| \ll 1$  and the determinant can be approximated as

$$(\mathcal{J}_{j,S}(\mathbf{x}))^2 \approx 1 - \partial_{\mu} U_{j,\mu}^\parallel(\mathbf{x}). \quad (30)$$

Using  $\mathbf{U}_j^\parallel(\mathbf{x}) = \pm \frac{1}{2} (\theta \hat{z} \times \mathbf{x} + \delta \mathbf{U}(\mathbf{x}))$ , with the sign  $+$  ( $-$ ) for the top (bottom) layer, respectively, and applying Eq. (25), we obtain

$$(\mathcal{J}_{j,S}(\mathbf{x}))^2 \approx 1 \mp \frac{1}{2} \nabla^2 \varphi^U(\mathbf{x}), \\ \implies \mathcal{J}_{j,S}(\mathbf{x}) \approx 1 \mp \frac{1}{4} \nabla^2 \varphi^U(\mathbf{x}). \quad (31)$$

So to the linear order of  $\delta \mathbf{U}$ , the deviation of  $\mathcal{J}$  from 1 depends only on the curl-free part  $\varphi^U$ . As shown in Fig. 4, for both models of the lattice relaxation,  $|\varphi^U| \sim 10^{-3} a^2$  and it varies over the length scale much larger than  $a$ , leading to  $|\nabla^2 \varphi^U| \lesssim 10^{-5}$  over the whole real space. As a consequence, the factor  $\mathcal{J}$  is very close to 1. We have checked this numerically by directly computing  $\mathcal{J}_{j,S}(\mathbf{x})$  from Eq. (29) and confirmed that its value does not deviate from 1 by more than  $\sim 10^{-5}$ . Thus, any deviation from 1 in this factor can be safely neglected in the following calculations.

It is worth emphasizing here that  $\mathcal{J} \approx 1$  relies on the particular form of the lattice relaxation  $\delta \mathbf{U}$  that is dominated by its solenoidal part. In the more general case (not considered explicitly here), the presence of the position dependent  $\mathcal{J}$  in the intralayer terms can be interpreted as a spatial variation of the Fermi velocity of the massless Dirac fermion [70].

### A. Intralayer couplings

The effective continuum Hamiltonian  $H_{\text{eff}}^K$  can be decomposed into two parts: the intralayer  $H_{\text{intra}}$  and the interlayer tunneling  $H_{\text{inter}}$  as in Eq. (1). Expanding to the second-order gradients, we obtain

$$H_{\text{intra}} = H_{\text{intra}}^{(0)} + \delta H_{\text{intra}}, \quad (32)$$

In Eq. (33),  $\bar{\sigma} = (\sigma_x, -\sigma_y)$ ,  $\mathbf{p} = -i\hbar \nabla = -i\hbar \frac{\partial}{\partial \mathbf{x}}$  is the momentum operator, and  $\mathbf{p}^{(j)}$  is the operator  $\mathbf{p}$  rotated by  $\mp \theta/2$ .

TABLE II. Numerical values of the parameters entering in the intralayer Hamiltonian  $H_{\text{intra}}$  in Eq. (32) for two different microscopic models.

Intralayer	$\mu$ (eV)	$v_F/a$ (eV)	$\beta_0/a^2$ (eV)	$\beta_1/a^2$ (eV)	$v_F\gamma$ (eV)	$C_0/a$ (eV)	$D_0/a$ (eV)
Ref. [65]	0.7878	2.1256	-0.1846	-0.3714	-3.3644	0.9426	-0.7491
Ref. [66]	-0.3460	2.1790	-0.1305	-0.5673	-4.3195	-2.3724	-1.9308

When the angle  $|\theta| \ll 1$ ,

$$\mathbf{p}^{(j)} \approx \left( p_x + \frac{\theta_j}{2} p_y, p_y - \frac{\theta_j}{2} p_x \right), \quad (34)$$

where we introduced the notation  $\theta_t = -\theta_b = \theta$ . The pseudovector field  $\mathcal{A}(\mathbf{x})$  is induced by the lattice distortion, having the form of

$$\begin{aligned} \mathcal{A}(\mathbf{x}) &= (\partial_x \delta U_x - \partial_y \delta U_y, -(\partial_x \delta U_y + \partial_y \delta U_x)) \\ &\approx (2\partial_x \partial_y \varepsilon^U(\mathbf{x}), (\partial_x^2 - \partial_y^2) \varepsilon^U(\mathbf{x})), \end{aligned} \quad (35)$$

and  $\mathcal{A}^{(j)}(\mathbf{x})$  is defined as

$$\begin{aligned} \mathcal{A}_\mu^{(j)}(\mathbf{x}) &= \pm R \left( \frac{\theta_j}{2} \right)_{\mu\nu} \mathcal{A}_\nu \\ &\approx \pm \left( \mathcal{A}_x - \frac{\theta_j}{2} \mathcal{A}_y, \mathcal{A}_y + \frac{\theta_j}{2} \mathcal{A}_x \right)_\mu, \end{aligned} \quad (36)$$

where  $R(\theta) = \cos(\theta)I_2 - i\sin(\theta)\sigma_2$  is the  $2 \times 2$  matrix corresponding to the counterclockwise rotation along  $\hat{z}$  by the angle of  $\theta$ . When  $|\theta| \ll 1$ ,  $R(\theta)_{\mu\nu} \approx \delta_{\mu\nu} - \theta\epsilon_{\mu\nu}$ . The signs + and - are for  $j$  corresponding to the top and bottom layers, respectively, reflecting the fact that the lattice distortions on two layers are opposite.

The field  $\xi(\mathbf{x})$  is also induced by the lattice distortion. It is given by

$$\begin{aligned} \xi_{x,SS'}(\mathbf{x}) &= \left[ \left( \frac{v_F}{2} + 2D_0 \right) (\partial_x \delta U_x) \sigma_1 \right. \\ &\quad \left. - \left[ \left( \frac{v_F}{2} + D_0 \right) \partial_y \delta U_x + D_0 \partial_x \delta U_y \right] \sigma_2 \right]_{SS'}, \\ \xi_{y,SS'}(\mathbf{x}) &= \left( \left[ \left( \frac{v_F}{2} + D_0 \right) \partial_x \delta U_y + D_0 \partial_y \delta U_x \right] \sigma_1 \right. \\ &\quad \left. - \left( \frac{v_F}{2} + 2D_0 \right) (\partial_y \delta U_y) \sigma_2 \right)_{SS'}. \end{aligned} \quad (37)$$

The values of the constants  $\mu$ ,  $v_F$ ,  $\gamma$ ,  $C_0$ , and  $D_0$  are listed in Table II for two different microscopic tight-binding models. All of these constants can be expressed via the microscopic hopping function. Detailed formulas for their efficient evaluation are derived in Appendix A.

In Eq. (32), the term  $\mu \Psi_{j,S}^\dagger \Psi_{j,S}$  leads to an overall shift of the energy spectrum and thus is irrelevant in most calculations. Among other terms, the leading one is  $v_F \bar{\sigma} \cdot \mathbf{p}$ , that produces the Dirac cone of the monolayer graphene. At the first magic angle  $\theta = 1.05^\circ$ , this term in both models has the energy scale of  $v_F k_\theta \sim 160\text{--}170$  meV, where  $k_\theta = 4\pi\theta/(3a)$ . Using the values listed in Table II, we can estimate the coupling between the fermion and the pseudovector field  $\mathcal{A}(\mathbf{x})$ . We found  $|v_F \gamma \mathcal{A}^{(j)}(\mathbf{x})| \lesssim 100$  meV, the same order as  $v_F k_\theta$ , showing the necessity of including this term in the effective continuum Hamiltonian, even if we were to only keep the first-order gradients.

Since the  $\mathcal{A}$  is also a two-dimensional vector field and  $\langle \mathcal{A} \rangle = 0$  averaged over the whole space, it can also be decom-

posed into the irrotational and the solenoidal parts:

$$\mathcal{A} = \nabla \varphi^{\mathcal{A}} + \nabla \times (\hat{z} \varepsilon^{\mathcal{A}}). \quad (38)$$

As shown in Fig. 5, the solenoidal part of the pseudovector field  $\mathcal{A}$  is larger than its irrotational. Interestingly, the induced pseudomagnetic field resulting from  $\mathcal{A}$ , defined as

$$\mathcal{B}\hat{z} = \frac{c}{e} \nabla \times (\gamma \mathcal{A}), \quad (39)$$

is about 30 T around the AB/BA stacked regions, and can be as high as 75 T near the AA stacked region, as shown in Fig. 6.

## B. Interlayer tunnelings

Up to the first-order gradients, the interlayer tunneling part of the effective continuum Hamiltonian in Eq. (1) can be

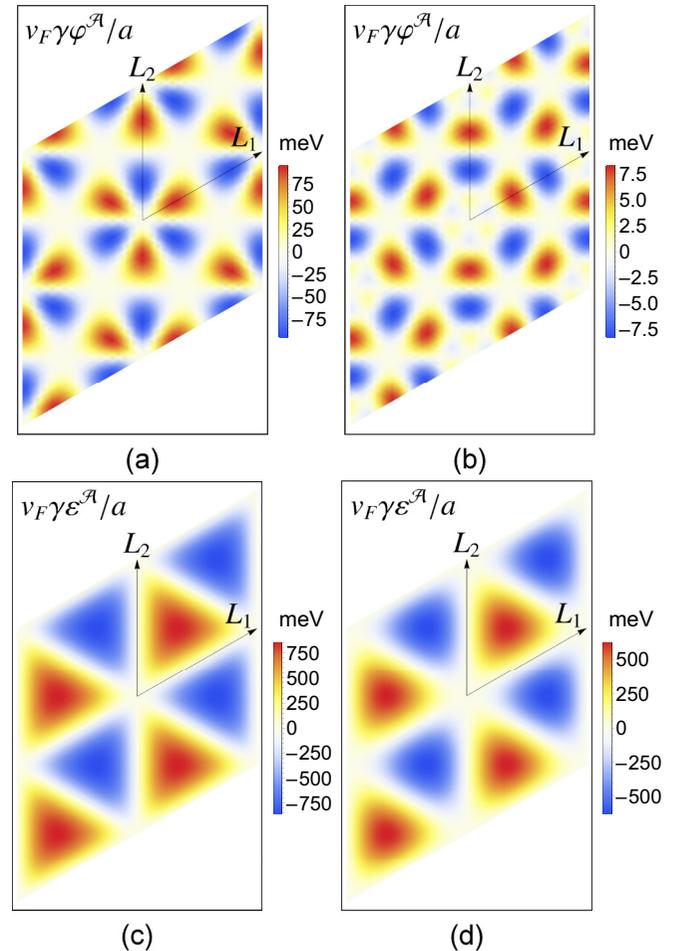


FIG. 5.  $v_F \gamma \varphi^{\mathcal{A}}$  (above) and  $v_F \gamma \varepsilon^{\mathcal{A}}$  (below) defined in Eq. (38) for the two models developed in Refs. [63] (left) and [64] (right), where  $v_F$  is the Fermi velocity of the undistorted monolayer Dirac cone.

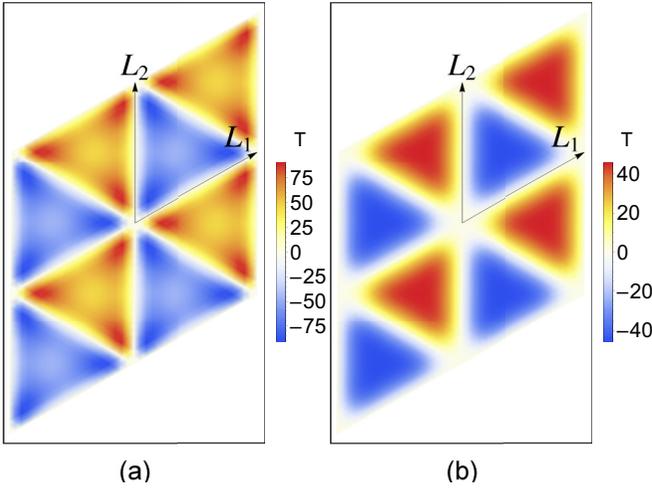


FIG. 6. The pseudomagnetic field  $\mathcal{B}$  induced by the lattice relaxation as defined in Eq. (39) for the two models developed in Ref. [63] (left) and [64] (right).

written as

$$H_{\text{inter}} = \sum_{SS'} \int d^2\mathbf{x} \Psi_{t,S}^\dagger(\mathbf{x}) \left( T_{SS'}(\mathbf{x}) + \frac{1}{2} \{ \mathbf{p}, \mathbf{\Lambda}_{SS'}(\mathbf{x}) \} \right) \Psi_{b,S'}(\mathbf{x}) + \text{H.c.} \quad (40)$$

The first and the second terms in the above parenthesis describe the contact and gradient interlayer couplings, respectively. As before,  $\mathbf{p} = -i\hbar\nabla$  is the momentum operator. The scalar field  $T_{SS'}$  and the vector field  $\mathbf{\Lambda}_{SS'}$  can be expanded as

$$T_{SS'}(\mathbf{x}) = \sum_{\mu,j} T_{SS'}^{(\mu,l)} e^{iq_{\mu,l}\cdot\mathbf{x}}, \quad (41)$$

$$\mathbf{\Lambda}_{SS'}(\mathbf{x}) = \sum_{\mu,l} \mathbf{\Lambda}_{SS'}^{(\mu,l)} e^{iq_{\mu,l}\cdot\mathbf{x}}, \quad (42)$$

where the vectors  $\mathbf{q}_{\mu,l}$  form shells in the extended moire BZ as illustrated in Fig. 7. Any vector  $\mathbf{q}_{\mu,l}$  can be decomposed as  $\mathbf{q}_{\mu,l} = \mathbf{q}_1 + \mathbf{g}$ , with  $\mathbf{q}_1 = -4\pi\hat{\mathbf{y}}/(3|L_1|)$  and  $\mathbf{g}$  being a reciprocal lattice vector of the moire superlattice, defined in Eqs. (13). Different  $\mathbf{q}$  vectors are distinguished by their subscript indices  $(\mu, l)$ , with  $\mu$  denoting the shell ordered by

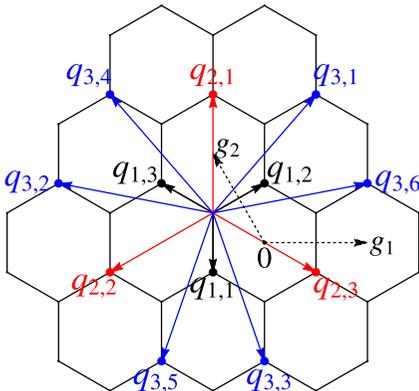


FIG. 7. The first and extended BZ of the moire superlattice. The vectors  $\mathbf{q}_{\mu,l}$  in the first three shells are also plotted here.

its radius  $|\mathbf{q}|$  from small to large, and  $l$  labeling different  $\mathbf{q}$  vectors inside the same shell.

Symmetries further constrain the form of  $T_{SS'}$  and  $\mathbf{\Lambda}_{SS'}$ . The lattice distortion  $\mathbf{U}(\mathbf{x})$  considered in this paper is invariant under  $C_2\mathcal{T}$ ,  $C_{2x}$ , and  $C_3$  transformations, and so is the effective continuum Hamiltonian. For example, under  $C_2\mathcal{T}$  transformation, the fermion fields  $\Psi_{j,S}(\mathbf{x}) \rightarrow \mathcal{K}\Psi_{j,\bar{S}}(-\mathbf{x})$ , where  $\bar{S}$  is the sublattice index different from  $S$  and  $\mathcal{K}$  is the complex conjugation. Therefore, the interlayer tunneling matrices must satisfy the constraints  $T(\mathbf{x}) = \sigma_x T^*(-\mathbf{x})\sigma_x$  and  $\mathbf{\Lambda}(\mathbf{x}) = \sigma_x \mathbf{\Lambda}^*(-\mathbf{x})\sigma_x$ . Correspondingly, their Fourier components must satisfy

$$\begin{aligned} T_{SS'}^{(\mu,l)} &= (\sigma_x (T^{(\mu,l)})^* \sigma_x)_{SS'}, \\ \mathbf{\Lambda}_{SS'}^{(\mu,l)} &= (\sigma_x (\mathbf{\Lambda}^{(\mu,l)})^* \sigma_x)_{SS'}. \end{aligned} \quad (43)$$

This implies that the above Fourier components can be written as

$$\begin{aligned} T_{SS'}^{(\mu,l)} &= (w_0^{(\mu,l)}\sigma_0 + w_1^{(\mu,l)}\sigma_1 + w_2^{(\mu,l)}\sigma_2 + iw_3^{(\mu,l)}\sigma_3)_{SS'}, \\ \mathbf{\Lambda}_{SS'}^{(\mu,l)} &= (\lambda_0^{(\mu,l)}\sigma_0 + \lambda_1^{(\mu,l)}\sigma_1 + \lambda_2^{(\mu,l)}\sigma_2 + i\lambda_3^{(\mu,l)}\sigma_3)_{SS'}, \end{aligned} \quad (44)$$

where  $w_i^{(\mu,l)}$  and the vectors  $\lambda_i^{(\mu,l)}$  are all real. The more detailed symmetry analysis, including  $C_{2x}$  and  $C_3$ , is presented in the Appendices [71]. Here, we only list several constraints from which all  $w_i^{(\mu,l)}$ s and  $\lambda_i^{(\mu,l)}$ s with  $\mu \leq 3$  can be obtained based on Table III. For  $1 \leq l \leq 3$ , due to  $C_3$  symmetry, the parameters  $w_i^{(\mu,l)}$  need to satisfy

$$w_i^{(\mu,l)} = w_i^{(\mu,1)}, \quad w_i^{(3,l+3)} = w_i^{(3,4)} \quad \text{for } i = 0, 3, \quad (45)$$

$$w_1^{(\mu,l)} - iw_2^{(\mu,l)} = e^{i\frac{2\pi}{3}(l-1)} (w_1^{(\mu,1)} - iw_2^{(\mu,1)}), \quad (46)$$

$$w_1^{(3,l+3)} - iw_2^{(3,l+3)} = e^{i\frac{2\pi}{3}(l-1)} (w_1^{(3,4)} - iw_2^{(3,4)}). \quad (47)$$

Similarly, for the vectors  $\lambda_i^{(\mu,l)}$ , if we restrict  $1 \leq l \leq 3$ , the  $C_3$  symmetry leads to the following constraints. For  $i = 0$  and 3,

$$(\lambda_i^{(\mu,l)})_\alpha = R\left(\frac{2\pi}{3}(l-1)\right)_{\alpha\beta} (\lambda_i^{(\mu,1)})_\beta, \quad (48)$$

$$(\lambda_i^{(3,l+3)})_\alpha = R\left(\frac{2\pi}{3}(l-1)\right)_{\alpha\beta} (\lambda_i^{(3,4)})_\beta. \quad (49)$$

In addition,

$$\begin{aligned} (\lambda_1^{(\mu,l)} - i\lambda_2^{(\mu,l)})_\alpha &= e^{i\frac{2\pi}{3}(l-1)} R\left(\frac{2\pi}{3}(l-1)\right)_{\alpha\beta} \\ &\quad \times (\lambda_1^{(\mu,1)} - i\lambda_2^{(\mu,1)})_\beta, \end{aligned} \quad (50)$$

$$\begin{aligned} (\lambda_1^{(3,l+3)} - i\lambda_2^{(3,l+3)})_\alpha &= e^{i\frac{2\pi}{3}(l-1)} R\left(\frac{2\pi}{3}(l-1)\right)_{\alpha\beta} \\ &\quad \times (\lambda_1^{(3,4)} - i\lambda_2^{(3,4)})_\beta \end{aligned} \quad (51)$$

where  $R(\theta) = \cos(\theta)I_{2 \times 2} - i\sin(\theta)\sigma_2$  is the  $2 \times 2$  matrix corresponding to the counterclockwise rotation along  $\hat{z}$  by the angle of  $\theta$ . Furthermore, the  $C_{2x}$  symmetry imposes the constraints on both  $w_i^{(\mu,l)}$  and  $\lambda_i^{(\mu,l)}$ . For  $i \neq 2$ ,

$$w_i^{(3,1)} = w_i^{(3,4)}, \quad (52)$$

$$(\lambda_i^{(3,1)})_\alpha = (\tau_3)_{\alpha\beta} (\lambda_i^{(3,4)})_\beta. \quad (53)$$

TABLE III. Parameters of the interlayer tunneling terms for two models in the absence/presence of the lattice relaxation.  $a$  is the magnitude of the primitive lattice vector and all numbers are in units of meV.

	$w_0^{(1,1)}$	$w_1^{(1,1)}$	$w_2^{(1,1)}$	$w_3^{(1,1)}$	$w_0^{(2,1)}$	$w_1^{(2,1)}$	$w_2^{(2,1)}$	$w_3^{(2,1)}$	$w_0^{(3,1)}$	$w_1^{(3,1)}$	$w_2^{(3,1)}$	$w_3^{(3,1)}$
Ref. [65], unrelaxed	110.9	110.9	0	0	1.6	1.6	0	0	negligible			
Ref. [65], relaxed	54.4	124.9	0	0	-6.9	9.0	0	0	17.5	-10.8	-18.8	0
Ref. [66], unrelaxed	104.0	104.0	0	-2.9	1.1	1.1	0	0	negligible			
Ref. [66], relaxed	78.6	113.1	0	-2.8	-0.3	3.4	0	-0.5	11.0	-5.6	-9.7	-0.6
	$\lambda_0^{(1,1)}/a$	$\lambda_1^{(1,1)}/a$	$\lambda_2^{(1,1)}/a$	$\lambda_3^{(1,1)}/a$	$\lambda_0^{(2,1)}/a$	$\lambda_1^{(2,1)}/a$	$\lambda_2^{(2,1)}/a$	$\lambda_3^{(2,1)}/a$	$\lambda_0^{(3,1)}/a$	$\lambda_1^{(3,1)}/a$	$\lambda_2^{(3,1)}/a$	$\lambda_3^{(3,1)}/a$
Ref. [65], unrelaxed	(-91.9, 0)	(-91.9, 0)	(0, 0)	(0, 0)	(1.8, 0)	(1.8, 0)	(0, 0)	(0, 0)	negligible			
Ref. [65], relaxed	(-102.0, 0)	(-74.45, 0)	(0, -27.6)	(0, 0)	(-5.4, 0)	(1.3, 0)	(0, -6.6)	(0, 0)	(8.5, -15.6)	(-3.3, 7.8)	(-7.8, 11.7)	(0, 0)
Ref. [66], unrelaxed	(-84.2, 0)	(-84.2, 0)	(0, -76.1)	(0.6, 0)	(2.0, 0)	(2.0, 0)	(0, 0.3)	(0, 0)	negligible			
Ref. [66], relaxed	(-90.8, 0)	(-90.0, 0)	(0, -83.3)	(0.7, 0)	(-0.2, 0)	(0, 0)	(0, -1.7)	(0, 0)	(4.5, -7.7)	(3.8, 7.3)	(-7.3, 4.7)	(0, 0)

For  $w_2^{(\mu,l)}$  and  $\lambda_2^{(\mu,l)}$ , we obtain

$$w_2^{(3,1)} = -w_2^{(3,4)}, \quad (54)$$

$$(\lambda_2^{(3,1)})_\alpha = -(\tau_3)_{\alpha\beta} (\lambda_2^{(3,4)})_\beta, \quad (55)$$

where the superscripts  $\alpha$  and  $\beta$  label the components of the vectors field  $\lambda_i^{(\mu,l)}$ . We should also emphasize that the constraints listed in Eqs. (45)–(55) are not complete. For example, by  $C_{2x}$  symmetry, we can also derive  $w_2^{(1,1)} = 0$ . The more detailed and complete discussion on symmetry constraints are presented in Appendix C.

If we keep only the innermost  $\mathbf{q}$  shell, then the contact term in the interlayer Hamiltonian of our theory limits to

$$T_{SS'}^{(1)}(\mathbf{x}) = \sum_{j=1}^3 e^{i\mathbf{q}_j \cdot \mathbf{x}} (w_0 I_{2 \times 2} + iw_3 \sigma_3 + w_1 \left( \cos\left(\frac{2\pi(j-1)}{3}\right) \sigma_1 - \sin\left(\frac{2\pi(j-1)}{3}\right) \sigma_2 \right))_{SS'}, \quad (56)$$

since  $\mathbf{q}_1 = \mathbf{q}_{1,1}$ ,  $\mathbf{q}_2 = \mathbf{q}_{1,2}$ , and  $\mathbf{q}_{1,3}$  and

$$w_0 = w_0^{(1,1)}, \quad w_1 = w_1^{(1,1)}, \quad w_3 = w_3^{(1,1)}. \quad (57)$$

Therefore, our theory recovers the interlayer term in the BM continuum model [30,58,60] if we set  $w_3 = 0$  and keep only  $\mathbf{q}_s$  in the first shell and neglect the gradient couplings  $\Lambda_{SS'}$ . Furthermore, in the absence of the lattice relaxation, as derived in the Appendices [71], our theory gives  $w_0 = w_1$  [60]. As shown in the next section, the  $w_3$  term is responsible for non-negligible p-h asymmetry for the model of Ref. [66].

Figures 15 and 16 show the comparison of the spectrum obtained from  $H_{\text{eff}}^{\mathbf{K}}$ , truncating to a different number of  $\mathbf{q}$  shells in the interlayer tunneling terms. For the rigid twist (i.e., when the lattice relaxation is absent), the approximation of including only the innermost  $\mathbf{q}$  shell gives the spectrum that is almost identical with the one produced by the tight-binding model in most of the moire BZ, with the mismatch of only  $\sim 2$  meV around the center of the moire Brillouin zone point  $\Gamma$ ; the bandwidth of the narrow bands for the rigid twist is about 40 meV and 20 meV for the models in Refs. [65,66], respectively, with at least one of the band gaps to the remote bands vanishing. All these features have been well reproduced by including only one  $\mathbf{q}$  shell in  $H_{\text{eff}}^{\mathbf{K}}$ . To further improve the agreement, we include the first two  $\mathbf{q}$  shells and achieve the accuracy presented in Figs. 1(a) and 1(c). The excellent agreement obtained with only two shells reflects the fact that the Fourier transform of the interlayer hopping quickly decays as a function of the momentum [60,71].

On the other hand, in the presence of the lattice relaxation, we need to include more shells to achieve the comparable accuracy. This is demonstrated in the Figs. 15 and 16. The increase of needed shells results from the factor  $e^{i(\mathbf{G}+\mathbf{K}) \cdot (\mathbf{U}_{j,s}^{\parallel}(\mathbf{x}) - \mathbf{U}_{j,s'}^{\parallel}(\mathbf{x}))}$  in Eq. (28). The Fourier transform of the interlayer hopping (i.e., for  $j \neq j'$ ) is the largest for  $\mathbf{G}_s$ , satisfying  $|\mathbf{G} + \mathbf{K}| = |\mathbf{K}|$ . Because of the spatial inhomogeneity of the lattice relaxation  $\delta U$  [Eq. (20)], the mentioned exponential factor induces the interlayer scattering with the momentum transfer of all possible  $\mathbf{q}_{\mu,l}$ ; the strength of the scattering is proportional to the Fourier transform of the exponential factor. For the lattice relaxation in Ref. [64],  $\delta U$  is dominated by the lowest wave vectors  $\pm \mathbf{g}_1$ ,  $\pm \mathbf{g}_2$ , and  $\pm(\mathbf{g}_1 + \mathbf{g}_2)$ . As a consequence,  $H_{\text{inter}}$  should include, at least, the terms with the momentum transfer of  $\mathbf{q}_{1,l} \pm \mathbf{g}_1$ ,  $\mathbf{q}_{1,l} \pm \mathbf{g}_2$ , and  $\mathbf{q}_{1,l} \pm (\mathbf{g}_1 + \mathbf{g}_2)$ , i.e., all the  $\mathbf{q}$  vectors in the first three shells. While all the values of  $w_i^{(\mu,l)}$  and  $\lambda_i^{(\mu,l)}$  in the first three

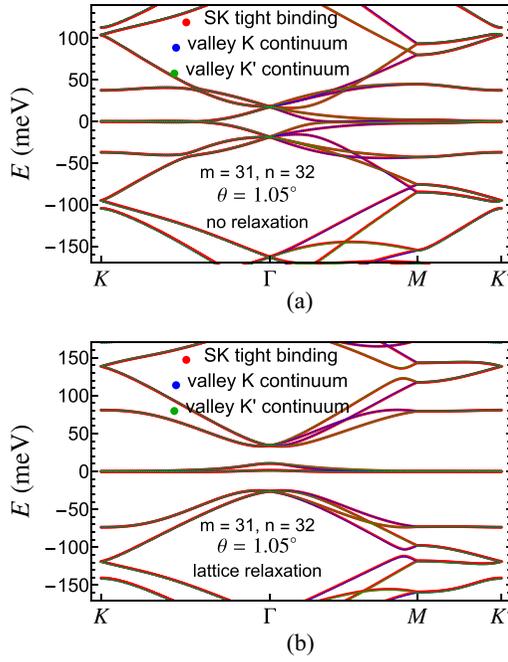


FIG. 8. Comparison between the microscopic tight-binding model (red) in Ref. [65] and the corresponding continuum model (blue for valley  $\mathbf{K}$  and green for valley  $\mathbf{K}'$ ) in the absence (above) and presence (below) of the lattice relaxation.

shells can be obtained from Table III and the formula listed in Eqs. (45)–(55), the values in the next three shells can be calculated in the same way based on Table V and Eqs. (C10)–(C12) in the Appendices. Figure 16(b) shows that the first four shells are needed to achieve the accuracy of 0.3 meV at  $\Gamma$  for the lattice configuration of Ref. [64].

For Ref. [63], both the first,  $\tilde{\epsilon}_1$ , and the third,  $\tilde{\epsilon}_3$ , harmonics of  $\delta U$  are sizable (see Table I). The wave vectors of the third harmonic are  $\pm 2\mathbf{g}_1$ ,  $\pm 2\mathbf{g}_2$ , and  $\pm 2(\mathbf{g}_1 + \mathbf{g}_2)$ . Following the argument in the above paragraph, we therefore expect that  $H_{\text{inter}}$  should include the terms with the momentum transfer of all  $\mathbf{q}$ s in the first six shells. Indeed, as demonstrated numerically in Fig. 15(b), we achieve the accuracy of 0.8 meV around  $\Gamma$  with six  $\mathbf{q}$  shells.

Having obtained both the intralayer and interlayer parts of the continuum model for the moire periodic distortions, utilizing the Bloch theorem, we diagonalize  $H_{\text{eff}}$  in the moire momentum space [72,73]. As shown in Figs. 8 and 9, the spectra of  $H_{\text{eff}}$  (for both valleys) and the microscopic tight-binding model agree with each other beyond the narrow band regime. We have found that both spectra are consistent with each other until the energy reaches  $\sim \pm 0.7$  eV, where significant deviations start to rapidly grow.

#### IV. ANALYSIS OF THE NARROW BAND HILBERT SPACE: SUBLATTICE POLARIZATION, P-H SYMMETRY, AND WILSON LOOPS

Having obtained the energy spectrum of  $H_{\text{eff}}^{\mathbf{K}}$  presented in the previous section, we now turn to the properties of the Hilbert space spanned by the narrow bands. While narrow bands appear in both models near the CNP when the twist

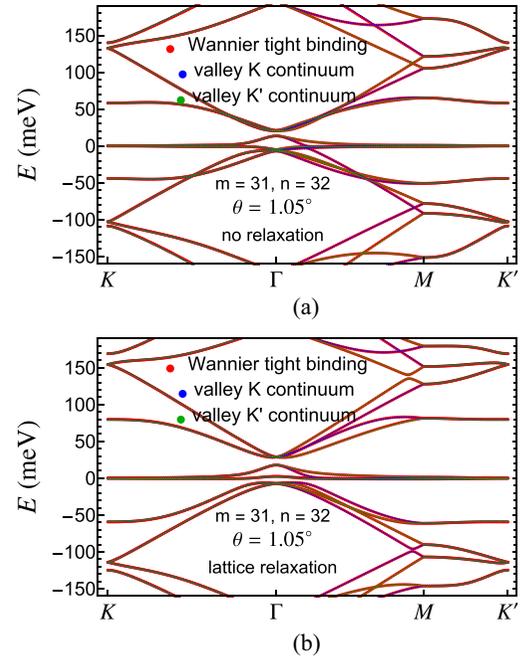


FIG. 9. Comparison between the microscopic tight-binding model (red) in Ref. [66] and the corresponding continuum model (green and blue) in the absence (above) and presence (below) of the lattice relaxation.

angle is  $1.05^\circ$ , the corresponding states are found to be notably different. In this section, we consider three properties of the narrow band Hilbert space at valley  $\mathbf{K}$ : the sublattice polarization, the deviation from the p-h symmetry, and the Wilson loop.

The sublattice polarization of the narrow bands is defined via the eigenvalues of the  $2 \times 2$  projected sublattice matrix  $S_{ij}(\mathbf{k}) = \langle \Psi_i(\mathbf{k}) | \sigma_z | \Psi_j(\mathbf{k}) \rangle$ , where  $\Psi_i(\mathbf{k})$  is the Bloch state with the momentum of  $\mathbf{k}$  in band  $i$  and  $\sigma_z$  is the sublattice polarization operator. Because of the  $C_2\mathcal{T}$  symmetry, the two eigenvalues of the projected  $\sigma_z$  have the same magnitude with opposite signs,  $\pm 1$ , corresponding to the perfect polarization obtained in the chiral limit [38]. The sublattice polarization calculated based on the  $H_{\text{eff}}^{\mathbf{K}}$  is shown in Fig. 10 for the two microscopic models. While the narrow band Hilbert space in either model is not perfectly polarized, the one proposed in Ref. [65] has larger sublattice polarization than the one in Ref. [66], implying the former model is closer to the chiral limit than the latter.

The p-h symmetry [35] plays an important role in that it leads to the  $U(4)$  symmetry of the projected Coulomb interaction, and helps in identifying the ground state in the strong coupling limit. The p-h transformation  $\hat{\mathcal{P}}$  acts within a valley and is defined as  $i\mu_y$ —the interchange of the two layers and changing the sign of the top layer—followed by the in-plane inversion  $\mathbf{r} \rightarrow -\mathbf{r}$ . When keeping only the first-order intralayer gradient terms and only the contact interlayer terms in the BM model, we have  $\hat{\mathcal{P}}^\dagger H_{\text{BM}}(\mathbf{k}) \hat{\mathcal{P}} = -H_{\text{BM}}(-\mathbf{k})$ . However,  $\hat{\mathcal{P}}$  is only an approximate symmetry as it is generally broken by the higher order gradient terms. For example, it is broken by the  $O(k^2)$  and  $O(k\partial U)$  terms in  $H_{\text{intra}}$ , as well as  $w_3$  and the vector couplings  $\lambda_{i \neq 3}$  in the interlayer tunnelings.

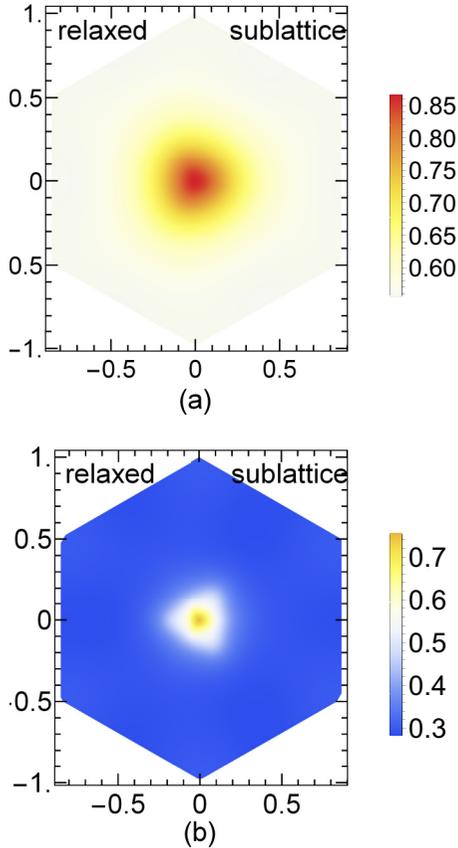


FIG. 10. The sublattice polarization of the narrow bands for two microscopic models in (a) Ref. [65] and (b) Ref. [66].

To quantify the degree of the p-h symmetry violation within the narrow bands, we define the  $2 \times 2$  projected p-h matrix as  $\mathcal{P}_{ij}(\mathbf{k}) = \langle \Psi_i(-\mathbf{k}) | \hat{\mathcal{P}} | \Psi_j, \mathbf{k} \rangle$ . If the p-h symmetry is exact, it is expected that the matrix  $\mathcal{P}_{ij}(\mathbf{k})$  is unitary for arbitrary  $\mathbf{k}$ , and thus both the singular values,  $\lambda_1(\mathbf{k})$  and  $\lambda_2(\mathbf{k})$ , are 1. Otherwise,  $\lambda_1(\mathbf{k})$  and  $\lambda_2(\mathbf{k})$  are smaller than 1 and therefore the deviation of  $\lambda_i$  ( $i = 1, 2$ ) from 1 measures the p-h asymmetry of the Hilbert space. Figure 11 illustrates the two singular values  $\lambda_i$  for both models. While the narrow bands in Ref. [65] are almost perfectly p-h symmetric, those in Ref. [66] shows significant p-h asymmetry. As we demonstrate in Fig. 19, the dominant source of the p-h asymmetry in the model of Ref. [66] comes from the interlayer contact coupling  $w_3$  and the subdominant contribution comes from the gradient coupling  $\lambda$ .

Finally, we consider the eigenvalues of the Wilson-loop operator as another property of the Hilbert space of the narrow bands. This operator  $\hat{O}$  is defined as  $P e^{-i \frac{1}{N_1} \mathbf{g}_1 \cdot \mathbf{r}} P$ , where  $N_1$  is the number of unit cells along the direction of  $\mathbf{L}_1$  in the entire lattice with periodic boundary conditions, and  $P$  is the projection operator onto the Bloch states of the narrow bands. Since this operator commutes with the momentum operator along  $\mathbf{g}_2$ , its eigenstate is labeled by the momentum  $k$  along  $\mathbf{g}_2$ . In the BM model, the phase of the eigenvalues of  $\hat{O}$ , labeled as  $\langle x_{\pm} \rangle$ , has the winding number of  $\pm 1$  as  $k$  runs from 0 to 1, illustrating the nontrivial topological properties of the narrow band system [35,37,44].

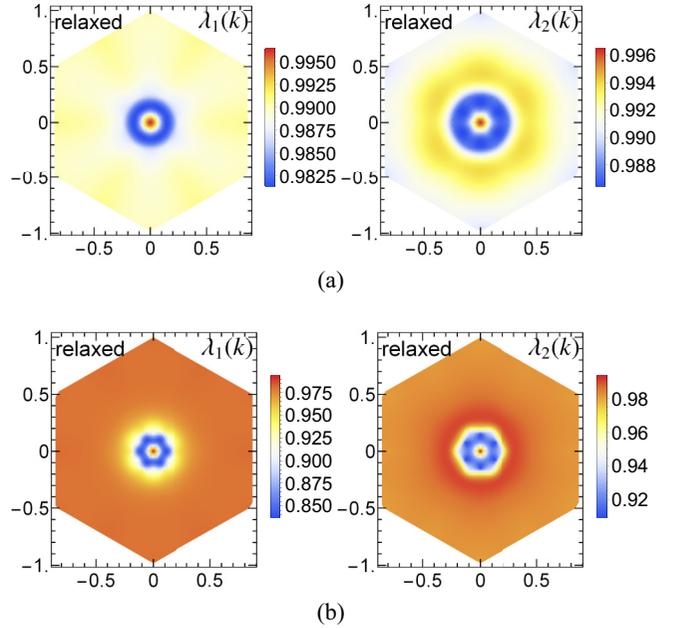


FIG. 11. The two singular values (left)  $\lambda_1(\mathbf{k})$  and (right)  $\lambda_2(\mathbf{k})$  of the p-h matrix  $\mathcal{P}(\mathbf{k})$  for two models in Refs. [65,66]. Their deviations from 1 measure the p-h asymmetry of the narrow bands.

In addition, if  $w_0 = 0$ , the system is in the chiral limit and  $\langle x_{\pm} \rangle$  is almost a linear function of  $\mathbf{k}$ , but becomes quite flat when  $\mathbf{k} \sim 0.5$  if the system is far away from the chiral limit ( $w_0 \lesssim w_1$ ) [35] (also see Fig. 1 in Ref. [44]). This behavior is also qualitatively reproduced in our constructed  $H_{\text{eff}}^{\mathbf{K}}$ , as shown in Fig. 12. For comparison, Fig. 12 also shows the winding of the phase  $\langle x_{\pm} \rangle$  obtained from the BM model, with the values of the parameters  $v_F$ ,  $w_0$ , and  $w_1$  taken from Table II. It is found that the curve for the BM model is straighter, suggesting that the terms neglected in the BM model but present in  $H_{\text{eff}}^{\mathbf{K}}$  drive the system further away from the chiral limit. Moreover, the curve for the model in Ref. [65] is straighter than the one for the model in Ref. [66], and thus consistent with the former model being closer to the chiral limit than the latter.

## V. EXACTLY FLAT BAND LIMIT WITH RELAXATION INDUCED PSEUDOMAGNETIC FIELDS

The inclusion of the strain is believed to greatly increase the bandwidth at the magic angle [50,54,74]. Having seen that the relaxed atomic configuration of the TBG obtained in Sec. II expands the AB/BA stacked regions and shrinks the AA stacked regions relative to just a rigid twist, thus intrinsically inducing strain, it is interesting to ask whether the bandwidth undergoes the increase as well. Motivated by this question, the goal of this section is to generalize the chiral limit introduced and analyzed in Ref. [38] for the BM model, including the relaxation-induced pseudomagnetic vector potential  $\mathcal{A}$ . While the relaxation-induced  $\mathcal{A}$  indeed increases the vanishing bandwidth at the magic angle found in Ref. [38] without  $\mathcal{A}$ , we demonstrate below that decreasing the twist angle can compensate the effect of  $\mathcal{A}$  on the bandwidth, resulting in exactly flat bands at the CNP at a new (smaller) magic angle. Throughout our analysis, we pay particular attention to

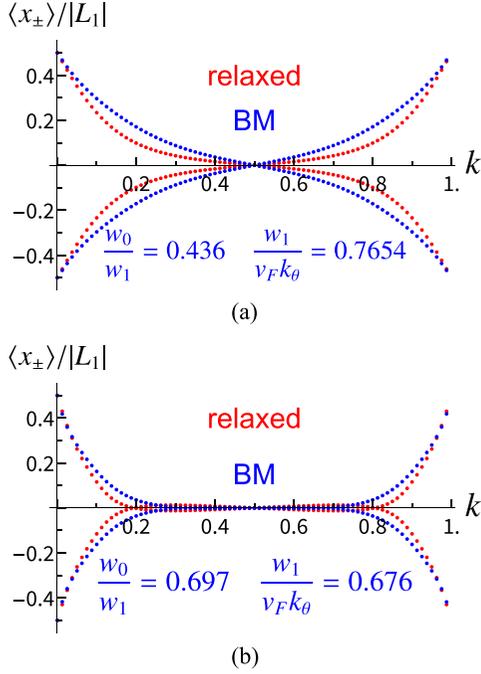


FIG. 12. The phase of the eigenvalues of the Wilson loop operator of the two valley-polarized narrow bands (red) with different continuum model constructed for two microscopic tight-binding models proposed by (a) Ref. [65] and (b) Ref. [66]. For comparison, the same phase of the eigenvalues has also been calculated for the BM model (blue) that contains only  $v_F$ ,  $w_0$ , and  $w_1$  with their numerical values taken from Tables II and III.

the importance of  $C_3$  symmetry (preserved by  $\mathcal{A}$ ) in making the compensation possible, noting that extrinsically induced strain generally breaks  $C_3$ .

For the purposes of this section, we start from the Hamiltonian

$$H_{\text{chiral}} = \begin{pmatrix} v_F \bar{\sigma}_{\frac{\theta}{2}} \cdot (\mathbf{p} + \gamma \mathcal{A}) & T(\mathbf{x}) \\ T^\dagger(\mathbf{x}) & v_F \bar{\sigma}_{-\frac{\theta}{2}} \cdot (\mathbf{p} - \gamma \mathcal{A}) \end{pmatrix} \quad (58)$$

that acts on the four-component spinor  $\Psi(\mathbf{x}) = (\psi_{t,A}(\mathbf{x}), \psi_{t,B}(\mathbf{x}), \psi_{b,A}(\mathbf{x}), \psi_{b,B}(\mathbf{x}))^T$ , where the subscripts  $t/b$  label the top/bottom layers and  $A/B$  labels the sublattice.

In the equation above,  $\bar{\sigma}_{\theta/2} = e^{-i\frac{\theta}{4}\sigma_3} \bar{\sigma} e^{i\frac{\theta}{4}\sigma_3}$  and  $\mathcal{A}(\mathbf{x})$  is the real inhomogeneous pseudomagnetic vector potential induced by the lattice relaxation, as calculated by Eq. (35); just as before,  $\bar{\sigma} = (\sigma_x, -\sigma_y)$ . This  $\mathcal{A}$  field is invariant under all the symmetry transformations discussed in Sec. II, such as  $C_3$ ,  $C_2\mathcal{T}$ , etc. Compared with the full  $H_{\text{intra}}$  in Eq. (32), all the second-order terms have been neglected in Eq. (58). In addition, the interlayer tunnelings also neglect the gradient coupling  $\Lambda$ , as well as the  $w_0$  and  $w_3$  terms in the contact coupling. Thus, the interlayer tunneling  $T(\mathbf{x})$  in  $H_{\text{chiral}}$  can be written as

$$T(\mathbf{x}) = \sum_{\mu,l} (w_1^{(\mu,l)} \sigma_1 + w_2^{(\mu,l)} \sigma_2) e^{i\mathbf{q}_{\mu,l} \cdot \mathbf{x}}. \quad (59)$$

Since  $T(\mathbf{x})$  contains only  $\sigma_1$  and  $\sigma_2$ ,  $e^{-i\frac{\theta}{4}\sigma_3} T(\mathbf{x}) e^{-i\frac{\theta}{4}\sigma_3} = T(\mathbf{x})$ . Introducing the unitary diagonal matrix:  $\mathcal{U} = \text{diag}(e^{-i\frac{\theta}{4}}, e^{i\frac{\theta}{4}}, e^{i\frac{\theta}{4}}, e^{-i\frac{\theta}{4}})$ ,  $H_{\text{chiral}}$  can be simplified by applying

the unitary transformation

$$H'_{\text{chiral}} = \mathcal{U}^\dagger H_{\text{chiral}} \mathcal{U} = \begin{pmatrix} v_F \bar{\sigma} \cdot (\mathbf{p} + \gamma \mathcal{A}) & T(\mathbf{x}) \\ T^\dagger(\mathbf{x}) & v_F \bar{\sigma} \cdot (\mathbf{p} - \gamma \mathcal{A}) \end{pmatrix}, \quad (60)$$

while the transformed spinor is labeled as  $\Phi(\mathbf{x}) = \mathcal{U}^\dagger \Psi(\mathbf{x}) = (\phi_{t,A}(\mathbf{x}), \phi_{t,B}(\mathbf{x}), \phi_{b,A}(\mathbf{x}), \phi_{b,B}(\mathbf{x}))^T$ . Again, note that each  $2 \times 2$  block of  $H'_{\text{chiral}}$  contains only  $\sigma_1$  and  $\sigma_2$ , thus  $\{\sigma_3 \otimes I, H'_{\text{chiral}}\} = 0$ , i.e., the chiral Hamiltonian is antisymmetric under the chiral p-h transformation  $\sigma_3 \otimes I$ .

First, we consider the states near the CNP at the corner of the moire BZ, i.e.,  $\mathbf{K}_m$  or  $\mathbf{K}'_m$ . Turning off the interlayer tunneling  $T(\mathbf{x})$ , we will show that two zero modes still exist even in the presence of the  $\mathcal{A}$  field. To prove it, consider the equation for the zero modes at  $\mathbf{K}_m$ :

$$\bar{\sigma} \cdot (\mathbf{p} + \gamma \mathcal{A}) \begin{pmatrix} \phi_{t,A}(\mathbf{x}) \\ \phi_{t,B}(\mathbf{x}) \end{pmatrix} = 0, \quad (61)$$

$$\rightarrow \begin{cases} ((-i\partial_1 + \partial_2) + \gamma(\mathcal{A}_1 + i\mathcal{A}_2))\phi_{t,B} = 0 \\ (-i\partial_1 - \partial_2) + \gamma(\mathcal{A}_1 - i\mathcal{A}_2)\phi_{t,A} = 0. \end{cases} \quad (62)$$

Using the Helmholtz decomposition in Eq. (38),  $\mathcal{A} = (\partial_1 \varphi^A + \partial_2 \varepsilon^A, \partial_2 \varphi^A - \partial_1 \varepsilon^A)$ , we immediately obtain the two independent solutions to Eq. (62):

$$\begin{cases} \phi_{t,A} = e^{-i\gamma\varphi^A} e^{\gamma\varepsilon^A} \\ \phi_{t,B} = 0 \end{cases} \quad \begin{cases} \phi_{t,A} = 0 \\ \phi_{t,B} = e^{-i\gamma\varphi^A} e^{-\gamma\varepsilon^A}. \end{cases} \quad (63)$$

The pseudovector field  $\mathcal{A}$  is periodic and its average over space  $\langle \mathcal{A} \rangle = 0$ , and so are  $\varphi^A$  and  $\varepsilon^A$ . Therefore, the two solutions in Eq. (63) are also bounded and periodic, giving the two zero modes at  $\mathbf{K}_m$ .

Under  $C_3$  transformation, the spinor  $\Phi(\mathbf{x}) \rightarrow e^{i\frac{2\pi}{3}\sigma_z} \Phi(\mathbf{x}')$ , where  $\mathbf{x}'$  is the position  $\mathbf{x}$  rotated clockwise by  $2\pi/3$  [38]. Therefore, the two zero modes in Eq. (63) carry the extra phases of  $e^{i2\pi/3}$  and  $e^{-i2\pi/3}$ , respectively, and thus transform differently under  $C_3$ . Furthermore, the chiral p-h transformation  $\sigma_z \otimes I$  commutes with  $C_3$ . As the interlayer tunneling  $T(\mathbf{x})$  is gradually turned on [38], each of these two modes at  $\mathbf{K}_m$  must transform to itself under  $\sigma_z \otimes I$ , and therefore each still has zero energy. As a consequence, the two bands around the CNP touch at the Dirac cone at  $\mathbf{K}_m$  even when  $\mathcal{A}$  is included.

Following the arguments presented in Ref. [38], we can also express the Fermi velocity of the Dirac cone in terms of the wave function at  $\mathbf{K}_m$ . For this purpose, we choose the basis  $\Phi' = (\phi_{t,A}, \phi_{b,A}, \phi_{t,B}, \phi_{b,B})$ , and the zero modes at  $\mathbf{K}_m$  satisfy the equation

$$\begin{pmatrix} 0 & \mathcal{D}(\mathbf{x}) \\ \mathcal{D}^\dagger(\mathbf{x}) & 0 \end{pmatrix} \begin{pmatrix} \Phi_{\mathbf{K}_m,A}(\mathbf{x}) \\ \Phi_{\mathbf{K}_m,B}(\mathbf{x}) \end{pmatrix} = 0, \quad (64)$$

where  $\Phi_{\mathbf{K}_m,A} = (\phi_{\mathbf{K}_m,t,A}, \phi_{\mathbf{K}_m,b,A})$  and  $\Phi_{\mathbf{K}_m,B} = (\phi_{\mathbf{K}_m,t,B}, \phi_{\mathbf{K}_m,b,B})$  are two component spinors.  $\mathcal{D}(\mathbf{x})$  is a  $2 \times 2$  matrix differential operator of the form

$$\mathcal{D}(\mathbf{x}) = \begin{pmatrix} v_F \pi_+(\mathbf{x}) & U(\mathbf{x}) \\ U(-\mathbf{x}) & v_F \pi_-(\mathbf{x}) \end{pmatrix}, \quad (65)$$

where

$$U(\mathbf{x}) = \sum_{\mu,l} (w_1^{(\mu,l)} - iw_2^{(\mu,l)}) e^{iq_{\mu,l}\mathbf{x}}, \quad (66)$$

$$\pi_{\pm} = p_1 + ip_2 \pm \gamma(\mathcal{A}_1 + i\mathcal{A}_2). \quad (67)$$

Since  $\mathcal{A}(\mathbf{x}) = \mathcal{A}(-\mathbf{x})$  due to the  $C_2\mathcal{T}$  symmetry,  $\mathcal{D}^\dagger(\mathbf{x}) = \mathcal{D}^*(-\mathbf{x})$ . Thus, if the two-component spinor  $\Psi(\mathbf{x})$  satisfies  $\mathcal{D}(\mathbf{x})\Psi(\mathbf{x}) = 0$  so(0,  $\Psi(\mathbf{x})$ ) is a zero mode at  $\mathbf{K}_m$ , the spinor  $\Psi^*(-\mathbf{x})$  satisfies  $\mathcal{D}^\dagger(\mathbf{x})\Psi^*(-\mathbf{x}) = 0$ , and therefore  $(\Psi^*(-\mathbf{x}), 0)$  is another zero mode. As the momentum  $\mathbf{p}$  slightly deviates from  $\mathbf{K}_m$ ,  $\boldsymbol{\sigma} \cdot \mathbf{p}$  can be treated as perturbation and thus the Fermi velocity of the Dirac cone at  $\mathbf{K}_m$  is

$$v_{\text{Dirac}} = v_F \frac{|\langle \Psi^*(-\mathbf{x}) | \Psi(\mathbf{x}) \rangle|}{\langle \Psi | \Psi \rangle}. \quad (68)$$

Because  $v_{\text{Dirac}}$  is a real number, in principle, it can vanish by tuning the interlayer coupling constants.

For the BM model in the chiral limit, Ref. [38] showed that the bands around the CNP become exactly flat as long as  $v_{\text{Dirac}} = 0$ . Their argument is still valid when the relaxation-induced pseudovector field  $\mathcal{A}$  is present. To prove this statement, we consider the equation

$$\mathcal{D}(\mathbf{x})\Psi_{k,B}(\mathbf{x}) = 0, \quad (69)$$

where  $\mathbf{k}$  is the Bloch momentum. Introducing the complex coordinates  $z = x_1 + ix_2$  and  $\bar{z} = x_1 - ix_2$ , we find

$$\mathcal{D}(\mathbf{x}) = v_F \begin{pmatrix} -2i\bar{\partial} + \gamma(\mathcal{A}_1 + i\mathcal{A}_2) & v_F^{-1}U(\mathbf{x}) \\ v_F^{-1}U(-\mathbf{x}) & -2i\bar{\partial} - \gamma(\mathcal{A}_1 + i\mathcal{A}_2) \end{pmatrix}, \quad (70)$$

where  $\bar{\partial} = \partial_{\bar{z}} = \frac{1}{2}(\partial_1 + i\partial_2)$ . Note that the differential operators in  $\mathcal{D}(\mathbf{x})$  contain only  $\bar{\partial}$ , i.e.,  $\partial_z$  is absent. Since we already showed that  $\Phi_{\mathbf{K}_m,B}$  is the solution of Eq. (69),  $f(z)\Phi_{\mathbf{K}_m,B}$  is also a solution as long as  $f(z)$  is a holomorphic function because then  $\bar{\partial}f(z) = 0$ . To construct the solution with the Bloch boundary conditions at an arbitrary momentum  $\mathbf{k}$ , we introduce [38]

$$\eta(z = x_1 + ix_2) = \frac{\vartheta_{\frac{kL_1}{2\pi} - \frac{1}{6}, -\frac{1}{6} - \frac{k(L_2-L_1)}{2\pi}}\left(\frac{z}{|L_1|}e^{-i\frac{\pi}{6}}, e^{i\frac{2\pi}{3}}\right)}{\vartheta_{-\frac{1}{6}, \frac{1}{6}}\left(\frac{\bar{z}}{|L_1|}e^{-i\frac{\pi}{6}}, e^{i\frac{2\pi}{3}}\right)}, \quad (71)$$

where  $\vartheta$  is the theta function, defined as

$$\vartheta_{a,b}(z, \tau) = \sum_{n=-\infty}^{\infty} e^{i\pi\tau(n+a)^2} e^{2\pi i(n+a)(z+b)}. \quad (72)$$

Note that  $\eta(z)$  satisfies the boundary condition  $\eta(\mathbf{x} + \mathbf{L}_{1,2}) = e^{i\mathbf{k}\cdot\mathbf{L}_{1,2}}\eta(\mathbf{x})$ . Then, the Bloch wave function

$$\Psi_{k,B}(\mathbf{x}) = \eta(\mathbf{x})\Phi_{\mathbf{K}_m,B}(\mathbf{x}) \quad (73)$$

is a solution of Eq. (69). However,  $\Psi_{k,B}$  contains singular points because  $\eta(z)$  is not an analytical function. The denominator in Eq. (71) vanishes at  $\mathbf{x}_0 + m\mathbf{L}_1 + n\mathbf{L}_2$ , where  $\mathbf{x}_0 = -\frac{1}{3}(\mathbf{L}_1 + \mathbf{L}_2)$ , and  $m$  and  $n$  are arbitrary integers, and, in general,  $\Psi_{k,B}$  is not normalizable [38].

The solution in Eq. (73) can still be physical if  $\Phi_{\mathbf{K}_m,B}(\mathbf{x}_0) = 0$ , thus canceling the zero in the denominator. Without the pseudovector potential  $\mathcal{A}$ , it can be achieved when  $v_{\text{Dirac}}$  vanishes [38]. The same argument also applies if  $\mathcal{A}$  is

present. To prove it, consider the zero mode  $\Phi_{\mathbf{K}_m,B}(\mathbf{x}) = (\Phi_{\mathbf{K}_m,t,B}(\mathbf{x}), \Phi_{\mathbf{K}_m,b,B}(\mathbf{x}))$ . It satisfies

$$\begin{aligned} 0 &= \Phi_{\mathbf{K}_m,B}^T(-\mathbf{x})\mathcal{D}(\mathbf{x})\Phi_{\mathbf{K}_m,B}(\mathbf{x}) \\ &= \Phi_{\mathbf{K}_m,B}^T(\mathbf{x})\mathcal{D}(-\mathbf{x})\Phi_{\mathbf{K}_m,B}(-\mathbf{x}). \end{aligned} \quad (74)$$

Using  $\mathcal{A}(\mathbf{x}) = \mathcal{A}(-\mathbf{x})$ , we obtain

$$\bar{\partial}(\Phi_{\mathbf{K}_m,B}^T(-\mathbf{x})\Phi_{\mathbf{K}_m,B}(\mathbf{x})) = 0. \quad (75)$$

For notational convenience, we introduce  $v(\mathbf{x}) = \Phi_{\mathbf{K}_m,B}^T(-\mathbf{x})\Phi_{\mathbf{K}_m,B}(\mathbf{x})$ . By the above formula,  $v(\mathbf{x})$  is a constant in space. If  $v_{\text{Dirac}} = 0$ , by Eq. (68),  $v(\mathbf{x})$  must vanish everywhere, including  $\mathbf{x} = \mathbf{x}_0$ . Note that

$$\begin{aligned} v(\mathbf{x}_0) &= \Phi_{\mathbf{K}_m,t,B}(-\mathbf{x}_0)\Phi_{\mathbf{K}_m,t,B}(\mathbf{x}_0) \\ &\quad + \Phi_{\mathbf{K}_m,b,B}(-\mathbf{x}_0)\Phi_{\mathbf{K}_m,b,B}(\mathbf{x}_0). \end{aligned} \quad (76)$$

Due to the symmetry of  $C_3$  rotation around  $\mathbf{x}_0$ ,  $\Phi_{\mathbf{K}_m,b,B}(\pm\mathbf{x}_0) = 0$  [38]. Therefore, if  $v_{\text{Dirac}} = 0$ , then either  $\Phi_{\mathbf{K}_m,t,B}(\mathbf{x}_0) = 0$  or  $\Phi_{\mathbf{K}_m,t,B}(-\mathbf{x}_0) = 0$ . In the former case, both components of  $\Phi_{\mathbf{K}_m,B}$  vanish at  $\mathbf{x}_0$ , and therefore Eq. (73) is a normalizable solution if  $v_{\text{Dirac}} = 0$ . If the latter is the case, redefine

$$\eta(z) = \frac{\vartheta_{\frac{kL_1}{2\pi} + \frac{1}{6}, -\frac{1}{6} - \frac{k(L_2-L_1)}{2\pi}}\left(\frac{z}{|L_1|}e^{-i\frac{\pi}{6}}, e^{i\frac{2\pi}{3}}\right)}{\vartheta_{\frac{1}{6}, -\frac{1}{6}}\left(\frac{z}{|L_1|}e^{-i\frac{\pi}{6}}, e^{i\frac{2\pi}{3}}\right)} \quad (77)$$

so the denominator vanishes at  $-\mathbf{x}_0 + m\mathbf{L}_1 + n\mathbf{L}_2$ , and thus all the above arguments follow.

In the rest of this section, we consider the pseudovector field  $\mathcal{A}$  induced by the lattice relaxation of Ref. [64]. As illustrated in Fig. 5,  $\varphi^A \ll \varepsilon^A$ , suggesting that  $\varphi^A$  can be neglected, and thus  $\mathcal{A} \approx \nabla \times (\hat{\mathbf{z}}\varepsilon^A)$ . In addition, Table I shows that the Fourier series of the lattice relaxation is dominated by the lowest six  $\mathbf{g}$ s. Therefore, we can keep only these six terms and neglect others. Furthermore, because  $\varepsilon^A$  is real and odd, its lowest Fourier components are purely imaginary. By  $C_3$  symmetry, they satisfy the relation

$$\begin{aligned} \tilde{\varepsilon}_{\mathbf{g}_1}^A &= \tilde{\varepsilon}_{\mathbf{g}_2}^A = \tilde{\varepsilon}_{-(\mathbf{g}_1+\mathbf{g}_2)}^A \\ &= -\tilde{\varepsilon}_{-\mathbf{g}_1}^A = -\tilde{\varepsilon}_{-\mathbf{g}_2}^A = -\tilde{\varepsilon}_{\mathbf{g}_1+\mathbf{g}_2}^A = i\tilde{\varepsilon}_1^A. \end{aligned} \quad (78)$$

In addition, the interlayer contact coupling field is set to be

$$T(\mathbf{x}) = w_1 \sum_{l=1}^3 \begin{pmatrix} 0 & e^{-i\frac{2\pi}{3}(l-1)} \\ e^{i\frac{2\pi}{3}(l-1)} & 0 \end{pmatrix} e^{iq_{1,l}\mathbf{x}}, \quad (79)$$

where only the inner most  $\mathbf{q}$  shell is included. Introducing the dimensionless parameters  $\alpha = w_1/(v_F k_\theta)$  and  $\gamma\tilde{\varepsilon}_1^A$ , the Fermi velocity of the Dirac cone at  $\mathbf{K}_m$  and  $\mathbf{K}'_m$  can be approximated as [71]

$$v_{\text{Dirac}} \approx v_F \frac{1 - 6(\gamma\tilde{\varepsilon}_1^A)^2 - 3\alpha^2 + 14\sqrt{3}\alpha^2\gamma\tilde{\varepsilon}_1^A}{1 + 3\alpha^2 + 6(\gamma\tilde{\varepsilon}_1^A)^2}. \quad (80)$$

From Tables I and II,  $\gamma\tilde{\varepsilon}_1^A \approx 0.06$ , leading to  $\alpha \approx 0.79$  when  $v_{\text{Dirac}}$  vanishes.

We also numerically checked the existence of the exactly flat bands in the presence of the  $\mathcal{A}$  field induced by the lattice relaxation of Ref. [64]. As demonstrated in Fig. 13, the bandwidth  $W_{\text{band}}$  vanishes when the interlayer coupling constant  $w_1$  is tuned to be around  $0.7857v_F k_\theta$ , very close to the value obtained from the approximate formula in Eq. (80). For the BM model where the pseudovector field  $\mathcal{A}$  is absent,

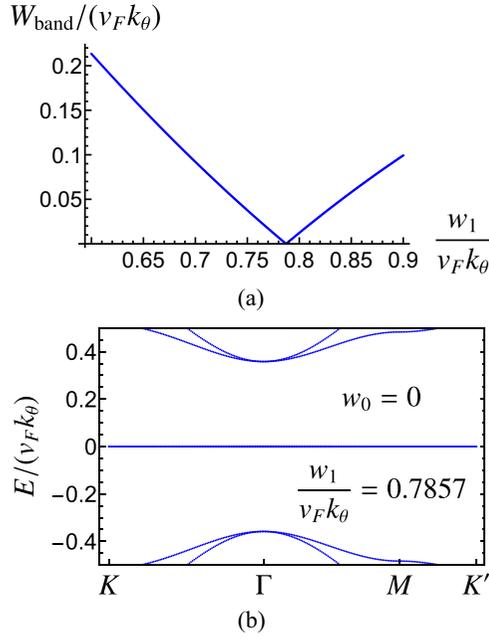


FIG. 13. The existence of the exactly flat band for  $H_{\text{chiral}}$  in the presence of the pseudovector field  $\mathcal{A}$  induced by the relaxation that is obtained from Ref. [64]. (a) The bandwidth  $W_{\text{band}}$  of the narrow bands around the CNP as a function of  $w_1$ . (b) The dispersion of both the narrow and remote bands when  $w_1/(v_F k_\theta) = 0.7857$ .

the exactly flat bands occur when  $w_1/(v_F k_\theta) = 0.586$  (see Ref. [38]); if  $w_1$  and  $v_F$  are set to the values listed in Tables II and III for the model in Ref. [66], the corresponding twist angle is  $1.07^\circ$ . However, this angle decreases to  $0.83^\circ$  when the pseudovector field  $\mathcal{A}$  induced by the lattice relaxation [66] is included in the chiral limit.

## VI. SUMMARY

In this paper, we constructed and analyzed the effective continuum theories corresponding to the microscopic tight-binding models proposed in Refs. [65,66] based on the systematic method proposed in Ref. [61]. The nearly perfect agreement between the dispersion of the tight-binding models and the dispersion of the effective continuum theories demonstrates the correctness of the constructed continuum theories and the validity of the method. We therefore envision that the experimentally measured  $\mathbf{u}_j(\mathbf{r})$  can be plugged into our effective Hamiltonian, and the resulting energy spectra and eigenfunctions can then be used to directly compare with the scanning tunneling spectroscopy measurement of the electronic local density of states. This may pave the way for a more quantitative comparison between the theoretical predictions and the experimental results. In addition, our theory

provides electron-phonon couplings as a byproduct, which are important to fully understand the role of phonons in superconductivity of TBG.

Our continuum model goes beyond the BM model in several aspects. First, the p-h symmetry of the narrow bands is only weakly broken within the BM model, while it is much more strongly broken in our continuum theory constructed for the tight-binding model of Ref. [66]. While the p-h symmetry of the energy spectrum is broken in both models [65,66], we focused on the p-h asymmetry of the narrow band *Hilbert space*, because it is more important in determining the correlated ground states near the magic angle. As shown in Fig. 19, the p-h asymmetry is dominated by the contribution from the interlayer contact term  $w_3$  that has been overlooked in previous works. Another source of the p-h asymmetry is the interlayer gradient terms  $\Lambda$  [50,75], whose numerical value listed in Table III is about two times larger than the value given in Refs. [50,75]. As a consequence, compared with the BM model and other continuum theories, our effective theory for the microscopic model of Ref. [66] leads to a much larger p-h asymmetry of the wave functions in the narrow bands.

Second, the interlayer tunneling in the BM model contains terms only with the minimal momentum transfer, i.e., the tunneling with three  $\mathbf{q}$ s in the first shell and neglects all other  $\mathbf{q}$ s. This approximation works quite well if the lattice relaxation is absent. In the presence of the lattice relaxation, however, Figs. 15 and 16 have demonstrated the necessity to include more  $\mathbf{q}$ s to even qualitatively match the dispersion.

We also investigated the existence of the exactly flat bands near the CNP when the lattice-induced pseudomagnetic fields are present. As long as the pseudovector potentials respect the  $C_3$  symmetry, our theoretical analysis and numerical calculations found exactly flat bands in the chiral limit, but at a smaller twist angle ( $0.83^\circ$ ) than without the relaxation-induced pseudovector fields ( $1.07^\circ$ ). In other words, despite the relaxation-induced strain fields, the bands can be exactly flat in the chiral limit due to the compensation from lowering the twist angle. Our analysis demonstrates the importance of the  $C_3$  symmetry in making this compensation possible.

## ACKNOWLEDGMENTS

J.K. acknowledges support from NSFC Grant No. 12074276, the Double First-Class Initiative Fund of ShanghaiTech University, and the start-up grant of ShanghaiTech University. O.V. is supported by NSF No. DMR-1916958 and is partially funded by the Gordon and Betty Moore Foundation's EPiQS Initiative Grant No. GBMF11070, National High Magnetic Field Laboratory through NSF Grant No. DMR-1644779, and the State of Florida. Part of this work was performed at the Aspen Center for Physics, which is supported by National Science Foundation Grant No. PHY-1607611.

## APPENDIX A: INTRALAYER DISPERSION

In this Appendix, we derive the intralayer part of the effective continuum Hamiltonian  $H_{\text{eff}}^K$  from the microscopic tight-binding model. The microscopic tight-binding model has the general form of

$$H_{\text{tb}} = \sum_{SS'} \sum_{jj'} \sum_{r_S, r_{S'}} t(\mathbf{X}_{j,S} - \mathbf{X}'_{j',S'}) c_{j,S,r_S}^\dagger c_{j',S',r_{S'}} \quad (\text{A1})$$

For intralayer hoppings, the hopping displacement  $\mathbf{X}_{j,S} - \mathbf{X}'_{j',S'}$  contains only in-plane components. In both models considered in this paper, the intralayer hopping is isotropic, depends only on  $|\mathbf{X}_{j,S} - \mathbf{X}'_{j',S'}|$ , and thus the intralayer hopping  $t(\delta\mathbf{X}) = t(|\delta\mathbf{X}|)$ . In addition, we only consider the lattice relaxation with which  $\mathbf{U}_{j,S} = \mathbf{U}_j^\parallel(\mathbf{x})$  is independent of the sublattice. We also neglect the corrugation so  $\mathbf{U}_{i/b}^\perp(\mathbf{x}) = \pm \frac{d_0}{2} \hat{z}$ . As a consequence,

$$\begin{aligned} H_{\text{intra}}^{\mathbf{K}} &\simeq \frac{1}{A_{\text{mlg}}} \sum_{S,S'} \sum_j \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot(\tau_S - \tau_{S'})} \int d^2\mathbf{x} \int d^2\mathbf{y} e^{-i(\mathbf{G}+\mathbf{K})\cdot\mathbf{y}} e^{i\frac{\mathbf{y}}{2}\cdot\nabla_{\mathbf{x}} 2\mathbf{U}_j^\parallel(\mathbf{x})\cdot(\mathbf{G}+\mathbf{K})} \\ &\times t(\mathbf{y}) \left[ \Psi_{j,S}^\dagger(\mathbf{x}) \Psi_{j,S'}(\mathbf{x}) + \frac{\mathbf{y}}{2} \cdot ((\nabla_{\mathbf{x}} \Psi_{j,S}^\dagger(\mathbf{x})) \Psi_{j,S'}(\mathbf{x}) - \Psi_{j,S}^\dagger(\mathbf{x}) \nabla_{\mathbf{x}} \Psi_{j,S'}(\mathbf{x})) \right. \\ &\left. + \frac{1}{8} y^\mu y^\nu ((\partial_\mu \partial_\nu \Psi_{j,S}^\dagger(\mathbf{x})) \Psi_{j,S'}(\mathbf{x}) - 2(\partial_\mu \Psi_{j,S}^\dagger(\mathbf{x})) (\partial_\nu \Psi_{j,S'}(\mathbf{x})) + \Psi_{j,S}^\dagger(\mathbf{x}) (\partial_\mu \partial_\nu \Psi_{j,S'}(\mathbf{x}))) \right]. \end{aligned} \quad (\text{A2})$$

The lattice displacement

$$\mathbf{U}_{j,\mu}^\parallel = \left( R\left(\frac{\theta_j}{2}\right) - I_{2\times 2} \right)_{\mu\nu} x_\nu \pm \frac{1}{2} \delta U_\mu.$$

where  $\theta_t = -\theta_b = \theta$ , and  $R(\theta/2)$  is the  $2 \times 2$  matrix corresponding to the counterclockwise rotation around  $z$  axis with the angle of  $\theta/2$ , and  $+$  ( $-$ ) sign is for the top (bottom) layer, respectively. Note that  $\nabla_{\mathbf{x}} \hat{z} \times \mathbf{x} \cdot (\mathbf{G} + \mathbf{K}) = -\hat{z} \times (\mathbf{G} + \mathbf{K})$ . Thus,

$$e^{-i(\mathbf{G}+\mathbf{K})\cdot\mathbf{y}} e^{i\frac{\mathbf{y}}{2}\cdot\nabla_{\mathbf{x}} 2\mathbf{U}^\parallel(\mathbf{x})\cdot(\mathbf{G}+\mathbf{K})} = e^{-iR(-\theta_j/2)\mathbf{y}\cdot(\mathbf{G}+\mathbf{K})} e^{\pm i\mathbf{y}/2\cdot\nabla_{\mathbf{x}} \delta\mathbf{U}(\mathbf{x})\cdot(\mathbf{G}+\mathbf{K})}.$$

In the main text, we consider only the lattice relaxation proposed in Refs. [63,64], in which the lattice distortion is dominated by the solenoid part. However, in this section, for completeness, we consider a more general  $\mathbf{U}_j^\parallel$ , whose irrotational part may also be important, and only in the last step, we set  $\nabla \cdot \delta\mathbf{U} = 0$ . As mentioned in the main text, up to the second order of the derivatives, the intralayer part is

$$H_{\text{intra}} = H_{\text{intra}}^{(0)} + \delta H_{\text{intra}}, \quad (\text{A3})$$

$$\begin{aligned} H_{\text{intra}}^{(0)} &= \int d^2\mathbf{x} \sum_j \sum_{SS'} \Psi_{j,S}^\dagger(\mathbf{x}) \left\{ \mu \delta_{SS'} + v_F \bar{\sigma}_{SS'} \cdot (\mathbf{p}^{(j)} + \gamma \mathcal{A}^{(j)}) + \alpha_{dp} \phi^{(j)} \delta_{SS'} + \beta_0 \mathbf{p}^2 \delta_{SS'} - \frac{C_0}{2} (\mathbf{p} \cdot \mathcal{A}(\mathbf{x}) + \mathcal{A}(\mathbf{x}) \cdot \mathbf{p}) \delta_{SS'} \right. \\ &\left. + \beta_1 ((p_x^2 - p_y^2) \sigma_1 + 2p_x p_y \sigma_2)_{SS'} \pm \frac{1}{2} (p_\mu \xi_{\mu,SS'}(\mathbf{x}) + \xi_{\mu,SS'}(\mathbf{x}) p_\mu) + 2D_0 \{ \phi^{(j)}, \bar{\sigma}_{SS'} \cdot \mathbf{p} \} \right\} \Psi_{j,S}(\mathbf{x}), \end{aligned} \quad (\text{A4})$$

$$\begin{aligned} \delta H_{\text{intra}} &= \int d^2\mathbf{x} \sum_j \sum_{SS'} \Psi_{j,S}^\dagger(\mathbf{x}) \left\{ \sum_{\mu\nu} [C_1 ((\nabla \cdot \delta\mathbf{U})^2 + (\partial_\mu \delta U_\nu) (\partial_\nu \delta U_\mu)) + C_2 (\partial_\mu \delta U_\nu) (\partial_\mu \delta U_\nu)] \delta_{SS'} \right. \\ &\left. + \alpha \left( -\frac{\theta}{2} (\nabla \times \delta\mathbf{U})_z - \frac{1}{2} (\nabla \cdot \delta\mathbf{U})^2 \right) \delta_{SS'} + \zeta_{SS'}(\mathbf{x}) \right\} \Psi_{j,S'}(\mathbf{x}). \end{aligned} \quad (\text{A5})$$

Here, we list all the expressions of the coefficients and fields that appear in the above formula:

$$\mu = \sum_a e^{-i\mathbf{K}\cdot\mathbf{a}} t(|\mathbf{a}|), \quad (\text{A6})$$

$$v_F = -i \sum_a e^{-i\mathbf{K}\cdot(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})} (\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_x t(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|), \quad (\text{A7})$$

$$v_F \gamma = \frac{1}{2} \sum_a e^{-i\mathbf{K}\cdot(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})} [(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_x]^2 \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|}, \quad (\text{A8})$$

$$\phi^{(j)} = \pm \frac{1}{2} \nabla \cdot \delta\mathbf{U}, \quad (\text{A9})$$

$$\alpha = \frac{1}{4} \sum_a e^{-i\mathbf{K}\cdot\mathbf{a}} |\mathbf{a}| t'(|\mathbf{a}|) \quad \text{and} \quad \alpha_{dp} = \frac{\sqrt{3}a}{2} \frac{\partial \epsilon}{\partial |\delta_{\mathbf{S}}^a|} + \alpha, \quad (\text{A10})$$

$$\mathcal{A}_\mu^{(j)}(\mathbf{x}) = \pm R\left(\frac{\theta_j}{2}\right)_{\mu\nu} \mathcal{A}_\nu \approx \pm \left( \mathcal{A}_x + \frac{\theta_j}{2} \mathcal{A}_y, \mathcal{A}_y - \frac{\theta_j}{2} \mathcal{A}_x \right)_\mu, \quad (\text{A11})$$

$$\text{with } \mathcal{A}(\mathbf{x}) = (\partial_x \delta U_x - \partial_y \delta U_y, -(\partial_x \delta U_y + \partial_y \delta U_x)) \approx (2\partial_x \partial_y \varepsilon^U(\mathbf{x}), (\partial_x^2 - \partial_y^2) \varepsilon^U(\mathbf{x})), \quad (\text{A12})$$

$$\beta_0 = -\frac{1}{4} \sum_a e^{-i\mathbf{K}\cdot\mathbf{a}} |\mathbf{a}|^2 t(|\mathbf{a}|), \quad (\text{A13})$$

$$C_0 = -\frac{i}{2} \sum_a e^{-i\mathbf{K}\cdot\mathbf{a}} (\mathbf{a}_x)^3 \frac{t'(|\mathbf{a}|)}{|\mathbf{a}|}, \quad (\text{A14})$$

$$\beta_1 = -\frac{1}{2} \sum_a e^{-i\mathbf{K}\cdot(\mathbf{a}+\delta\boldsymbol{\tau}_{AB})} [(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_x]^2 t(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|), \quad (\text{A15})$$

$$\xi_{x,SS'}(\mathbf{x}) = \left(\frac{v_F}{2} + 2D_0\right) (\partial_x \delta U_x) (\sigma_1)_{SS'} - \left[\left(\frac{v_F}{2} + D_0\right) \partial_x \delta U_x + D_0 \partial_x \delta U_y\right] (\sigma_2)_{SS'}, \quad (\text{A16})$$

$$\xi_{y,SS'}(\mathbf{x}) = \left[\left(\frac{v_F}{2} + D_0\right) \partial_x \delta U_y + D_0 \partial_y \delta U_x\right] (\sigma_1)_{SS'} - \left(\frac{v_F}{2} + 2D_0\right) (\partial_y \delta U_y) (\sigma_2)_{SS'}, \quad (\text{A17})$$

$$D_0 = -\frac{i}{6} \sum_a e^{i\mathbf{K}\cdot(\mathbf{a}+\delta\boldsymbol{\tau}_{AB})} ((\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_x)^3 \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|}, \quad (\text{A18})$$

where we have introduced the notation  $\delta\boldsymbol{\tau}_{SS'} = \boldsymbol{\tau}_S - \boldsymbol{\tau}_{S'}$ , so  $\delta\boldsymbol{\tau}_{AB} = -\frac{1}{3}(\mathbf{a}_1 + \mathbf{a}_2)$ .  $\mathbf{a}$  is an arbitrary lattice vector.  $\theta_t = -\theta_b = \theta$ . The signs  $+$  and  $-$  in Eq. (A11) are for the top and bottom layers, respectively. Note that for the lattice relaxation considered in the main text,  $|\nabla \cdot \delta\mathbf{U}| \lesssim 10^{-5}$  is tiny and thus the pseudoscalar field  $\phi^{(j)}$  can be safely neglected, as well as the term proportional to  $\alpha_{dp}$ . The detailed discussion on the coefficient  $\alpha_{dp}$  can be found in Appendix A6.

As mentioned in the text, although the terms in  $\delta H_{\text{intra}}$  are also second order, they are numerically small compared with other second-order terms in  $H_{\text{intra}}^{(0)}$ . Here, we express the fields and coefficients in  $\delta H_{\text{intra}}$  in terms of the lattice distortion  $\delta\mathbf{U}(\mathbf{x})$  and the microscopic hopping function:

$$C_1 = \frac{1}{8} \sum_a e^{-i\mathbf{K}\cdot\mathbf{a}} \left( \frac{7}{8} |\mathbf{a}| t'(|\mathbf{a}|) + \frac{1}{8} |\mathbf{a}|^2 t''(|\mathbf{a}|) \right), \quad (\text{A19})$$

$$C_2 = \frac{1}{8} \sum_a e^{-i\mathbf{K}\cdot\mathbf{a}} \left( \frac{3}{8} |\mathbf{a}| t'(|\mathbf{a}|) + \frac{1}{8} |\mathbf{a}|^2 t''(|\mathbf{a}|) \right). \quad (\text{A20})$$

The formula of the field  $\zeta_{SS'}(\mathbf{x})$  is listed in Eqs. (A58)–(A60).

### 1. Expansion of the Jacobian factor

In the main text, we have expanded the the Jacobian factor  $\mathcal{J}$  to the first order of  $\partial\mathbf{U}^{\parallel}$ , and argued that it depends only on the divergence of  $\mathbf{U}^{\parallel}$ . Since  $\nabla \cdot \mathbf{U}^{\parallel} \approx 10^{-5}$ , its deviation from 1 can be safely neglected. In this subsection, we will go to the second order of the derivatives, and derive its corresponding terms in the effective continuum  $H_{\text{eff}}^{\mathbf{K}}$ .

Since we will expand to the second order of  $\partial\mathbf{U}^{\parallel}$ , we write  $\mathbf{U}^{\parallel}$  as

$$U_{t,\mu}^{\parallel} = \left( I - R\left(-\frac{\theta}{2}\right) \right)_{\mu\nu} x_\nu + \frac{1}{2} \delta U_\nu = -\frac{\theta}{2} \epsilon_{\mu\nu} x_\nu + \frac{1}{8} \theta^2 x_\mu + \frac{1}{2} \delta U_\mu \longrightarrow \frac{\partial U_{t,\mu}^{\parallel}}{\partial x_\nu} = -\frac{\theta}{2} x_\nu + \frac{1}{2} \frac{\partial \delta U_\mu}{\partial x_\nu}.$$

If all the elements of the matrix  $M$  has the property  $|M_{\mu\nu}| \ll 1$ , its determinant can be expanded as

$$\det(I + M) = \exp(\text{Tr}(\ln M)) \approx 1 + \text{Tr}(M) + \frac{1}{2}((\text{Tr}(M))^2 - \text{Tr}(M^2)) + \mathcal{O}(M^3).$$

Therefore, we obtain the expansion of the determinant up to the second order of the derivative

$$\begin{aligned} \mathcal{J}_t^2 &= 1 - \frac{1}{2} \nabla \cdot \delta\mathbf{U} + \frac{\theta}{4} (\nabla \times \delta\mathbf{U})_z + \frac{1}{8} (\nabla \cdot \delta\mathbf{U})^2 - \frac{1}{8} (\partial_\mu \delta U_\nu) (\partial_\nu \delta U_\mu), \\ \mathcal{J}_b^2 &= 1 + \frac{1}{2} \nabla \cdot \delta\mathbf{U} + \frac{\theta}{4} (\nabla \times \delta\mathbf{U})_z + \frac{1}{8} (\nabla \cdot \delta\mathbf{U})^2 - \frac{1}{8} (\partial_\mu \delta U_\nu) (\partial_\nu \delta U_\mu). \end{aligned} \quad (\text{A21})$$

For the lattice relaxation considered in the main text,  $|\nabla \cdot \delta\mathbf{U}| \sim 10^{-5}$  and can be safely neglected.

### 2. Expansion by the order of derivatives

Since  $\delta\mathbf{U} \lesssim 0.3a$  varies over the moire unit cell, its gradient  $|\nabla_x \delta\mathbf{U}| \ll 1$ . We can expand the exponential by the order of  $\nabla_x \delta\mathbf{U}$ , i.e.,

$$e^{i\frac{1}{2} \nabla_x \delta\mathbf{U} \cdot (\mathbf{G} + \mathbf{K})} \approx 1 + i\frac{y_\mu}{2} \partial_\mu \delta U_\rho (\mathbf{G} + \mathbf{K})_\rho - \frac{1}{8} y_\mu y_\nu (\partial_\mu \delta U_\rho) (\partial_\nu \delta U_\sigma) (\mathbf{G} + \mathbf{K})_\rho (\mathbf{G} + \mathbf{K})_\sigma.$$

In the rest of this section, we will derive each term in the expansion and express the coefficient in terms of the microscopic hopping function. Before doing this, we define the Fourier transformation of the hopping functions as

$$\tilde{t}(\mathbf{q}) = A_{m|g}^{-1} \int d^2y e^{-iq\cdot y} t(\mathbf{y}) \implies i\partial_{q_\mu} \tilde{t}(\mathbf{q}) = A_{m|g}^{-1} \int d^2y e^{-iq\cdot y} y_\mu t(\mathbf{y}). \quad (\text{A22})$$

### 3. Leading term

First, we consider the leading term that in  $H_{\text{intra}}$ ,

$$\sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\delta\boldsymbol{\tau}_{SS'}} \tilde{t}(\mathbf{G} + \mathbf{K}) \Psi_{j,S}^\dagger \Psi_{j,S'} = \sum_{\mathbf{a}} e^{-i\mathbf{K}\cdot(\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})} t(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|) \Psi_{j,S}^\dagger \Psi_{j,S'}, \quad (\text{A23})$$

where  $\mathbf{a}$  is an arbitrary lattice vector. In the last formula above, we have used the Poisson summation formula to transform the summation over  $\mathbf{G}$  to the summation over the lattice vectors. Due to  $C_3$  symmetry, it is easy to show that the summation above vanishes when  $S \neq S'$ . When  $S = S'$ , the above summation leads to the term

$$\mu \sum_{j,S} \int d^2\mathbf{x} \Psi_{j,S}^\dagger(\mathbf{x}) \Psi_{j,S}(\mathbf{x}) \quad \text{with } \mu = \sum_{\mathbf{a}} e^{-i\mathbf{K}\cdot\mathbf{a}} t(|\mathbf{a}|). \quad (\text{A24})$$

Combined with Eq. (A21), to the second order of  $\nabla U_j^\parallel$ , this terms leads to

$$\mu \sum_{j,S} \int d^2\mathbf{x} \left( 1 \mp \frac{1}{2} \nabla \cdot \delta\mathbf{U} + \frac{\theta}{4} (\nabla \times \delta\mathbf{U})_z + \frac{1}{8} (\nabla \cdot \delta\mathbf{U})^2 - \frac{1}{8} (\partial_\mu \delta U_\nu) (\partial_\nu \delta U_\mu) \right) \Psi_{j,S}^\dagger(\mathbf{x}) \Psi_{j,S}(\mathbf{x}). \quad (\text{A25})$$

### 4. First-order derivative

Next, we consider the next leading term, i.e., the terms containing the first-order derivative of either  $\delta\mathbf{U}$  or  $\Psi_{j,S}(\mathbf{x})$ .

#### a. Fermi velocity

For the terms containing  $\partial_\mu \Psi_{j,S}(\mathbf{x})$ , we have

$$\int d^2\mathbf{x} v_{j,SS'}^\mu \Psi_{j,S}^\dagger(\mathbf{x}) p_\mu \Psi_{j,S'}(\mathbf{x}), \quad \text{with} \quad (\text{A26})$$

$$v_{j,SS'}^\mu = \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\delta\boldsymbol{\tau}_{SS'}} \int d^2\mathbf{y} e^{-i(\mathbf{G}+\mathbf{K})\cdot\boldsymbol{\rho}R(\mp\theta/2)_{\rho\nu}y_\nu} (-iy_\mu) t(|\mathbf{y}|) = \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\delta\boldsymbol{\tau}_{SS'}} \partial_{q_\nu} \tilde{t}(\mathbf{q})|_{\mathbf{q}=\mathbf{G}+\mathbf{K}R\left(\pm\frac{\theta}{2}\right)_{\mu\nu}}. \quad (\text{A27})$$

Applying the Poisson summation formula, we obtain that

$$\sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\delta\boldsymbol{\tau}_{SS'}} \partial_{q_\nu} \tilde{t}(\mathbf{q})|_{\mathbf{q}=\mathbf{G}+\mathbf{K}} = -i \sum_{\mathbf{a}} e^{-i\mathbf{K}\cdot(\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})} (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu t(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|).$$

For  $S = S'$ , it can be shown that the above summation vanishes because of  $C_3$  symmetry. For  $S \neq S'$ , due to  $C_3$  and  $m_y$  (mirror reflection over the  $yz$  plane), the above summation leads to  $v_F(\sigma_1\hat{x} - \sigma_2\hat{y})_{SS'}$ , with

$$v_F = -i \sum_{\mathbf{a}} e^{-i\mathbf{K}\cdot(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})} (\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_x t(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|),$$

where  $\delta\boldsymbol{\tau}_{AB} = -\frac{1}{3}(\mathbf{a}_1 + \mathbf{a}_2)$ .

Thus, making the approximation that  $R(\theta)_{\mu\nu} = \delta_{\mu\nu} - \theta\epsilon_{\mu\nu}$ , this gradient term can be written as

$$v_F \sum_j \sum_{SS'} \Psi_{j,S}^\dagger(\mathbf{x}) \left( \left( p_x + \frac{\theta_j}{2} p_y \right) \sigma_1 - \left( p_y - \frac{\theta_j}{2} p_x \right) \sigma_2 \right) \Psi_{j,S'}(\mathbf{x}). \quad (\text{A28})$$

For notational convenience, we can first define the layer-dependent momentum operator:

$$\mathbf{p}^{(j)} = \left( p_x + \frac{\theta_j}{2} p_y, p_y - \frac{\theta_j}{2} p_x \right),$$

so this term can be written in a simpler form:

$$v_F \sum_j \sum_{SS'} \Psi_{j,S}^\dagger(\mathbf{x}) \bar{\sigma}_{SS'} \cdot \mathbf{p}^{(j)} \Psi_{j,S'}(\mathbf{x}). \quad (\text{A29})$$

Now, combined with the expansion of the Jacobi factor in Eq. (A21) and expanded to the second order of the derivatives, we obtain

$$v_F \sum_j \sum_{SS'} \Psi_{j,S}^\dagger(\mathbf{x}) \left( 1 \mp \frac{1}{2} \nabla \cdot \delta\mathbf{U} \right) \bar{\sigma}_{SS'} \cdot \mathbf{p}^{(j)} \Psi_{j,S'}(\mathbf{x}). \quad (\text{A30})$$

### b. Pseudovector and pseudoscalar fields

To the first-order derivative of  $\delta\mathbf{U}$ ,  $H_{\text{intra}}$  also contains a term that couples to  $\Psi_{j,S}^\dagger \Psi_{j,S'}$ ,

$$\begin{aligned} & \pm \sum_{\mathbf{G}} e^{i\mathbf{G} \cdot \delta\boldsymbol{\tau}_{SS'}} \int d^2\mathbf{y} e^{-i(\mathbf{G}+\mathbf{K}) \cdot \mathbf{R}(-\theta_j/2)\mathbf{y}} t(|\mathbf{y}|) \frac{i}{2} y_\mu (\partial_\mu \delta U_\rho)(\mathbf{G} + \mathbf{K})_\rho \\ & = \pm \frac{i}{2} \sum_{\mathbf{G}} e^{i\mathbf{G} \cdot \delta\boldsymbol{\tau}_{SS'}} \int d^2\mathbf{z} e^{-i(\mathbf{G}+\mathbf{K}) \cdot \mathbf{z}} t(|\mathbf{z}|) R\left(\frac{\theta_j}{2}\right)_{\mu\nu} z_\nu (\mathbf{G} + \mathbf{K})_\rho (\partial_\mu \delta U_\rho) \end{aligned} \quad (\text{A31})$$

where by Poisson summation formula,

$$\begin{aligned} & \frac{i}{2} \sum_{\mathbf{G}} e^{i\mathbf{G} \cdot \delta\boldsymbol{\tau}_{SS'}} \int d^2\mathbf{z} e^{-i(\mathbf{G}+\mathbf{K}) \cdot \mathbf{z}} t(\mathbf{z}) z_\nu (\mathbf{G} + \mathbf{K})_\rho \\ & = \frac{1}{2} \sum_{\mathbf{a}} e^{-i\mathbf{K} \cdot (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})} \partial_\rho [(\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu t(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)] \\ & = \frac{1}{2} \sum_{\mathbf{a}} e^{-i\mathbf{K} \cdot (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})} \left[ \delta_{\rho\nu} t(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|) + (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\rho \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|} \right]. \end{aligned} \quad (\text{A32})$$

This summation, when  $S = S'$ , leads to

$$\frac{1}{2} \sum_{\mathbf{a}} e^{-i\mathbf{K} \cdot \mathbf{a}} \left[ t(|\mathbf{a}|) + \frac{1}{2} |\mathbf{a}| t'(|\mathbf{a}|) \right] \delta_{\nu\rho} = \left( \frac{\mu}{2} + \frac{1}{4} \sum_{\mathbf{a}} |\mathbf{a}| t'(|\mathbf{a}|) \right) \delta_{\nu\rho}. \quad (\text{A33})$$

For notational convenience, we introduce  $\alpha = \frac{1}{4} \sum_{\mathbf{a}} |\mathbf{a}| t'(|\mathbf{a}|)$ . When  $S = S'$ , Eq. (A31) leads to

$$\pm \left( \frac{\mu}{2} + \alpha \right) \delta_{\rho\nu} R\left(\frac{\theta_j}{2}\right)_{\mu\nu} \partial_\mu \delta U_\rho \approx \left( \frac{\mu}{2} + \alpha \right) \left( \pm \nabla \cdot \delta \mathbf{U} - \frac{\theta}{2} (\nabla \times \delta \mathbf{U})_z \right). \quad (\text{A34})$$

Now, combined with the expansion of the Jacobi determinant, we see that the terms that couple to  $\Psi_{j,S}^\dagger \Psi_{j,S}$  are

$$\begin{aligned} & \left( \frac{\mu}{2} + \alpha \right) \left( \pm \nabla \cdot \delta \mathbf{U} - \frac{\theta}{2} (\nabla \times \delta \mathbf{U})_z \right) + \left( \frac{\mu}{2} + \alpha \right) (\pm \nabla \cdot \delta \mathbf{U}) \left( \mp \frac{1}{2} \nabla \cdot \delta \mathbf{U} \right) \\ & = \left( \frac{\mu}{2} + \alpha \right) \left( \pm \nabla \cdot \delta \mathbf{U} - \frac{\theta}{2} (\nabla \times \delta \mathbf{U})_z - \frac{1}{2} (\nabla \cdot \delta \mathbf{U})^2 \right). \end{aligned} \quad (\text{A35})$$

Note the main text, where we consider only the lattice relaxation proposed in Refs. [63,64]. The divergence of the atomic displacement field  $|\nabla \cdot \delta \mathbf{U}| \lesssim 10^{-5}$  is tiny and can be safely neglected.

When  $S \neq S'$ , by  $C_3$  and  $m_y$  mirror symmetries, the summation in Eq. (A32) produces

$$\begin{aligned} & \frac{1}{2} \sum_{\mathbf{a}} e^{-i\mathbf{K} \cdot (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})} \left[ \delta_{\rho\nu} t(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|) + (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\rho \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|} \right] \\ & = \frac{1}{2} \sum_{\mathbf{a}} e^{-i\mathbf{K} \cdot (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})} (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\rho \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|} \\ & = v_F \gamma ((\tau_3)_{\rho\nu} (\sigma_1)_{SS'} + (\tau_1)_{\rho\nu} (\sigma_2)_{SS'}), \end{aligned} \quad (\text{A36})$$

$$\text{with } v_F \gamma = \frac{1}{2} \sum_{\mathbf{a}} e^{-i\mathbf{K} \cdot (\mathbf{a} + \delta\boldsymbol{\tau}_{AB})} [(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_x]^2 \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|}. \quad (\text{A37})$$

To further simplify the notation, we introduce the layer-dependent pseudovector field  $\mathcal{A}^{(j)}$  as

$$\begin{aligned} & \mathcal{A}_\mu^{(j)}(\mathbf{x}) = \pm R\left(\frac{\theta_j}{2}\right)_{\mu\nu} \mathcal{A}_\nu \approx \pm \left( \mathcal{A}_x - \frac{\theta_j}{2} \mathcal{A}_y, \mathcal{A}_y + \frac{\theta_j}{2} \mathcal{A}_x \right)_\mu, \\ & \text{with } \mathcal{A} = (\partial_x \delta U_x - \partial_y \delta U_y, -(\partial_x \delta U_y + \partial_y \delta U_x)) \approx (2\partial_x \partial_y \varepsilon^U(\mathbf{x}), (\partial_x^2 - \partial_y^2) \varepsilon^U(\mathbf{x})), \end{aligned} \quad (\text{A38})$$

where the signs + and - are for the top and bottom layers, respectively.

Now, combining with the expansion of the Jacobi factor and keeping the terms up to the second order of derivatives:

$$v_F \gamma \int d^2\mathbf{x} \Psi_{j,S}^\dagger(\mathbf{x}) \left( \bar{\boldsymbol{\sigma}}_{SS'} \cdot \mathcal{A}^{(j)} - (\bar{\boldsymbol{\sigma}}_{SS'} \cdot \mathcal{A}) \frac{1}{2} \nabla \cdot \delta \mathbf{U} \right) \Psi_{j,S'}(\mathbf{x}). \quad (\text{A39})$$

**5. Second-order derivative terms**

In this section, we considered the next order terms, i.e., the second order of combined derivative of  $\delta U$  and gradient of fermion fields.

**a. Second-order gradient of fermion field**

First, consider the term

$$\sum_G e^{iG \cdot \delta \tau_{SS'}} \int d^2 y e^{-i(G+K) \cdot y} t(|y|) \frac{1}{2} (-iy_\mu) (-iy_\nu) \Psi_{j,S}^\dagger(\mathbf{x}) p_\mu p_\nu \Psi_{j,S'}(\mathbf{x}).$$

Applying the Poisson summation formula,

$$\begin{aligned} & -\frac{1}{2} \sum_G e^{iG \cdot \delta \tau_{SS'}} \int d^2 y e^{-i(G+K) \cdot y} t(|y|) y_\mu y_\nu \\ & = -\frac{1}{2} \sum_G e^{iG \cdot \delta \tau_{SS'}} \left. \frac{\partial^2}{\partial q_\mu \partial q_\nu} \tilde{t}(\mathbf{q}) \right|_{\mathbf{q}=G+K} = -\frac{1}{2} \sum_a e^{-iK \cdot (a + \delta \tau_{SS'})} (a + \delta \tau_{SS'})_\mu (a + \delta \tau_{SS'})_\nu t(|a + \delta \tau_{SS'}|). \end{aligned} \quad (A40)$$

For  $S = S'$ ,  $\delta \tau = 0$ . We obtain

$$\beta_0 = -\frac{1}{2} \sum_a e^{-iK \cdot a} a_\mu a_\nu t(|a|) = -\frac{1}{4} \sum_a e^{-iK \cdot a} |a|^2 t(|a|).$$

For  $S \neq S'$ , by  $C_3$  and  $m_y$  symmetry, we obtain

$$-\frac{1}{2} \sum_a e^{-iK \cdot (a + \delta \tau_{SS'})} (a + \delta \tau_{SS'})_\mu (a + \delta \tau_{SS'})_\nu t(|a + \delta \tau_{SS'}|) = \beta_1 [(\tau_3)_{\mu\nu} \sigma_1 + (\tau_1)_{\mu\nu} \sigma_2]_{SS'}, \quad (A41)$$

$$\text{with } \beta_1 = -\frac{1}{2} \sum_a e^{-iK \cdot (a + \delta \tau_{AB})} [(a + \delta \tau_{AB})_x]^2 t(|a + \delta \tau_{AB}|). \quad (A42)$$

Thus, combining together, we obtain

$$\int d^2 x \sum_j \sum_{SS'} \Psi_{j,S}^\dagger(\mathbf{x}) [\beta_0 p^2 \delta_{SS'} + \beta_1 ((p_x^2 - p_y^2) \sigma_1 + 2p_x p_y \sigma_2)_{SS'}] \Psi_{j,S}(\mathbf{x}). \quad (A43)$$

**b. Cross terms between atomic displacement gradients and gradients of the fermion field**

Next, consider the cross term between the first-order derivative of  $\delta U$  and the gradient of the fermion field:

$$A_{mlg}^{-1} \sum_G e^{iG \cdot \delta \tau_{SS'}} \int d^2 y e^{-i(G+K) \cdot y} t(|y|) \frac{i}{2} y_\mu \partial_\mu \delta U_\rho(\mathbf{G} + \mathbf{K})_\rho (-iy_\nu) \frac{1}{2} ((p_\nu \Psi_{j,S}(\mathbf{x}))^\dagger \Psi_{j,S'} + \Psi_{j,S}^\dagger(\mathbf{x}) p_\nu \Psi_{j,S'}(\mathbf{x})). \quad (A44)$$

Applying the Poisson summation formula, we can obtain

$$\begin{aligned} & \frac{1}{2} \sum_G e^{iG \cdot \delta \tau_{SS'}} \int d^2 y e^{-i(G+K) \cdot y} t(|y|) y_\mu y_\nu (\mathbf{G} + \mathbf{K})_\rho \\ & = -\frac{1}{2} \sum_G e^{iG \cdot \delta \tau_{SS'}} \left. \frac{\partial^2}{\partial q_\mu \partial q_\nu} \right|_{\mathbf{q}=G+K} (\mathbf{G} + \mathbf{K})_\rho \\ & = -\frac{i}{2} \sum_a e^{-iK \cdot (a + \delta \tau_{SS'})} \left[ \delta_{\mu\rho} (a + \delta \tau_{SS'})_\nu t(|a + \delta \tau_{SS'}|) + \delta_{\nu\rho} (a + \delta \tau_{SS'})_\mu t(|a + \delta \tau_{SS'}|) \right. \\ & \quad \left. + (a + \delta \tau_{SS'})_\mu (a + \delta \tau_{SS'})_\nu (a + \delta \tau_{SS'})_\rho \frac{t'(|a + \delta \tau_{SS'}|)}{|a + \delta \tau_{SS'}|} \right]. \end{aligned} \quad (A45)$$

For  $S = S'$ , the above formula leads to

$$C_{\mu\nu\rho} = -\frac{i}{2} \sum_a e^{-iK \cdot a} a_\mu a_\nu a_\rho \frac{t'(|a|)}{|a|}. \quad (A46)$$

By  $C_3$  and  $m_y$  symmetries, we can prove that the only nonzero components of the tensor  $C_{\mu\nu\rho}$  are

$$C_{xxx} = -C_{xyy} = -C_{yyx} = -C_{yxy} = C_0 \quad \text{with} \quad C_0 = -\frac{i}{2} \sum_a e^{-iK \cdot a} (a_x)^3 \frac{t'(|a|)}{|a|}.$$

Then Eq. (A44), when  $S = S'$ , can be written as

$$C_0 \int d^2\mathbf{x} \Psi_{j,S}^\dagger(\mathbf{x}) \frac{1}{2} \{\mathbf{p}, \mathcal{A}(\mathbf{x})\} \Psi_{j,S}(\mathbf{x}). \quad (\text{A47})$$

Now, for  $S \neq S'$ , the summation in Eq. (A45) gives

$$\begin{aligned} & \frac{v_F}{2} [\delta_{\mu\rho}(\bar{\sigma}_{SS'})_v + \delta_{v\rho}(\bar{\sigma}_{SS'})_\mu] + D_0 [\delta_{\mu\rho}(\bar{\sigma}_{SS'})_v + \delta_{v\rho}(\bar{\sigma}_{SS'})_\mu + \delta_{\mu v}(\bar{\sigma}_{SS'})_\rho] \\ & = \left(\frac{v_F}{2} + D_0\right) \delta_{\mu\rho} \bar{\sigma}_{v,SS'} + \left(\frac{v_F}{2} + D_0\right) \delta_{v\rho} \bar{\sigma}_{\mu,SS'} + D_0 \delta_{\mu v} \bar{\sigma}_{\rho,SS'}, \end{aligned} \quad (\text{A48})$$

$$\text{with } D_0 = -\frac{i}{6} \sum_{\mathbf{a}} e^{i\mathbf{K}\cdot(\mathbf{a}+\delta\boldsymbol{\tau}_{AB})} (\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_x^3 \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|}. \quad (\text{A49})$$

This leads to the term

$$\pm \Psi_{j,S}^\dagger(\mathbf{x}) \frac{1}{2} \left( p_\mu \xi_{\mu,SS'}(\mathbf{x}) + \xi_{\mu,SS'}(\mathbf{x}) p_\mu + \left(\frac{v_F}{2} + D_0\right) \frac{1}{2} \{\nabla \cdot \delta\mathbf{U}, \bar{\sigma}_{SS'} \cdot \mathbf{p}\} \right) \Psi_{j,S'}(\mathbf{x}), \quad (\text{A50})$$

$$\text{with } \xi_{x,SS'}(\mathbf{x}) = \left[ \left(\frac{v_F}{2} + 2D_0\right) \partial_x \delta U_x \sigma_1 - \left[\left(\frac{v_F}{2} + D_0\right) \partial_y \delta U_x + D_0 \partial_x \delta U_y\right] \sigma_2 \right]_{SS'},$$

$$\xi_{y,SS'}(\mathbf{x}) = \left[ \left[\left(\frac{v_F}{2} + D_0\right) \partial_x \delta U_y + D_0 \partial_y \delta U_x\right] \sigma_1 - \left(\frac{v_F}{2} + 2D_0\right) (\partial_y \delta U_y) \sigma_2 \right]_{SS'}. \quad (\text{A51})$$

### c. $(\nabla\delta\mathbf{U})^2$ terms

Lastly, we consider the terms containing the square of the gradient of  $\delta\mathbf{U}$ :

$$A_{m\ell g}^{-1} \frac{1}{2} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\delta\boldsymbol{\tau}_{SS'}} \int d^2\mathbf{y} e^{-i(\mathbf{G}+\mathbf{K})\cdot\mathbf{y}} t(|\mathbf{y}|) \frac{i}{2} y_\mu \frac{i}{2} y_\nu \partial_\mu \delta U_\rho(\mathbf{G} + \mathbf{K})_\rho \partial_\nu \delta U_\sigma(\mathbf{G} + \mathbf{K})_\sigma \Psi_{j,S}(\mathbf{x})^\dagger \Psi_{j,S'}(\mathbf{x}). \quad (\text{A52})$$

Again, we consider the coefficient

$$\begin{aligned} & -\frac{1}{8} A_{m\ell g}^{-1} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\delta\boldsymbol{\tau}_{SS'}} \int d^2\mathbf{y} e^{-i(\mathbf{G}+\mathbf{K})\cdot\mathbf{y}} t(|\mathbf{y}|) y_\mu y_\nu (\mathbf{G} + \mathbf{K})_\rho (\mathbf{G} + \mathbf{K})_\sigma \\ & = \frac{1}{8} \sum_{\mathbf{a}} e^{-i\mathbf{K}\cdot(\mathbf{a}+\delta\boldsymbol{\tau}_{SS'})} \frac{\partial^2}{\partial x_\rho \partial x_\sigma} ((\mathbf{x} + \delta\boldsymbol{\tau}_{SS'})_\mu (\mathbf{x} + \delta\boldsymbol{\tau}_{SS'})_\nu t(|\mathbf{x} + \delta\boldsymbol{\tau}_{SS'}|)) \Big|_{\mathbf{x}=\mathbf{a}} \\ & = \frac{1}{8} \sum_{\mathbf{a}} e^{-i\mathbf{K}\cdot(\mathbf{a}+\delta\boldsymbol{\tau}_{SS'})} \left[ (\delta_{\mu\sigma} \delta_{\rho\nu} + \delta_{\nu\sigma} \delta_{\rho\mu}) t(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|) + (\delta_{\nu\sigma} (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\mu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\rho \right. \\ & \quad + \delta_{\mu\sigma} (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\rho + \delta_{\rho\sigma} (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\mu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu \\ & \quad + \delta_{\nu\rho} (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\mu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\sigma + \delta_{\rho\mu} (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\sigma \left. \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|} \right. \\ & \quad \left. + (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\mu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\rho (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\sigma \left( \frac{t''(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|^2} - \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|^3} \right) \right]. \end{aligned} \quad (\text{A53})$$

For  $S = S'$ , the above summation can be simplified as

$$\begin{aligned} & \frac{1}{8} \sum_{\mathbf{a}} e^{-i\mathbf{K}\cdot\mathbf{a}} \left[ (\delta_{\mu\sigma} \delta_{\rho\nu} + \delta_{\nu\sigma} \delta_{\rho\mu}) \left( t(|\mathbf{a}|) + \frac{7}{8} |\mathbf{a}| t'(|\mathbf{a}|) + \frac{1}{8} |\mathbf{a}|^2 t''(|\mathbf{a}|) \right) + \delta_{\rho\sigma} \delta_{\mu\nu} \left( \frac{3}{8} |\mathbf{a}| t'(|\mathbf{a}|) + \frac{1}{8} |\mathbf{a}|^2 t''(|\mathbf{a}|) \right) \right] \\ & = \left( \frac{\mu}{8} + C_1 \right) (\delta_{\mu\sigma} \delta_{\rho\nu} + \delta_{\nu\sigma} \delta_{\rho\mu}) + C_2 \delta_{\rho\sigma} \delta_{\mu\nu}, \end{aligned} \quad (\text{A54})$$

with the expression for the two coefficients  $C_1$  and  $C_2$  listed below:

$$C_1 = \frac{1}{8} \sum_{\mathbf{a}} e^{-i\mathbf{K}\cdot\mathbf{a}} \left( \frac{7}{8} |\mathbf{a}| t'(|\mathbf{a}|) + \frac{1}{8} |\mathbf{a}|^2 t''(|\mathbf{a}|) \right), \quad (\text{A55})$$

$$C_2 = \frac{1}{8} \sum_{\mathbf{a}} e^{-i\mathbf{K}\cdot\mathbf{a}} \left( \frac{3}{8} |\mathbf{a}| t'(|\mathbf{a}|) + \frac{1}{8} |\mathbf{a}|^2 t''(|\mathbf{a}|) \right). \quad (\text{A56})$$

TABLE IV. Numerical values of the parameters in  $\delta H_{\text{intra}}$  in Eqs. (32) and (A5) for two different microscopic tight-binding models.

Intralayer	$\alpha$ (eV)	$C_1$ (eV)	$C_2$ (eV)	$D_1$ (eV)	$D_2$ (eV)
Ref. [65]	-1.0542	-0.1165	0.1448	-0.9829	3.5837
Ref. [66]	2.5472	0.4822	-0.1546	-1.0862	-2.1330

This leads to the term

$$\left[ \left( \frac{\mu}{8} + C_1 \right) (\nabla \cdot \delta U)^2 + (\partial_\mu \delta U_\nu) (\partial_\nu \delta U_\mu) + C_2 (\partial_\mu \delta U_\nu) (\partial_\mu \delta U_\nu) \right] \Psi_{j,S}^\dagger(\mathbf{x}) \Psi_{j,S}(\mathbf{x}). \quad (\text{A57})$$

For  $S \neq S'$ , the first term in the square bracket of Eq. (A53) vanishes by  $C_3$  symmetry. Introducing the notation

$$R_{\mu\nu\rho\sigma,SS'} = \frac{1}{4} v_F \gamma [\delta_{\mu\sigma} ((\tau_3)_{\rho\nu} \sigma_1 + (\tau_1)_{\rho\nu} \sigma_2) + (\mu \leftrightarrow \rho) + (\mu\sigma \leftrightarrow \rho\nu)]_{SS'}, \quad (\text{A58})$$

$$\begin{aligned} V_{\mu\nu\rho\sigma,SS'} &= \frac{1}{8} \sum_a e^{-i\mathbf{K} \cdot (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})} (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\mu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\nu (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\rho (\mathbf{a} + \delta\boldsymbol{\tau}_{SS'})_\sigma \left( \frac{t''(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|^2} - \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{SS'}|^3} \right) \\ &= \frac{D_1}{6} [\delta_{\nu\mu} ((\tau_3)_{\rho\sigma} \sigma_1 + (\tau_1)_{\rho\sigma} \sigma_2) + (\mu \leftrightarrow \rho) + (\mu \leftrightarrow \sigma) + (\rho \leftrightarrow \nu) + (\nu \leftrightarrow \sigma) + (\rho\mu \leftrightarrow \nu\sigma)]_{SS'} \\ &\quad + D_2 \left( \cos\left(\frac{\pi}{2} n_2\right) \sigma_1 - \sin\left(\frac{\pi}{2} n_2\right) \sigma_2 \right)_{SS'}, \end{aligned} \quad (\text{A59})$$

$$\zeta_{SS'}(\mathbf{x}) = (R_{\mu\nu\rho\sigma,SS'} + V_{\mu\nu\rho\sigma,SS'}) \partial_\mu \delta U_\rho \partial_\nu \delta U_\sigma, \quad (\text{A60})$$

where  $n_2$  in Eq. (A59) is the number of times that the index  $y$  appears in the subscripts  $\mu, \nu, \rho$ , and  $\sigma$ , i.e.,

$$n_2 = \delta_{\mu y} + \delta_{\nu y} + \delta_{\rho y} + \delta_{\sigma y}.$$

Equation (A59) is derived by  $C_3$  symmetry, and the constants  $D_1$  and  $D_2$  must be real as constrained by  $m_y$  symmetry. To be more specific,

$$D_1 = \frac{1}{16} \sum_a e^{-i\mathbf{K} \cdot (\mathbf{a} + \delta\boldsymbol{\tau}_{AB})} \left( \frac{t''(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|^2} - \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|^3} \right) ([(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_x]^4 + [(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_y]^4)$$

$$D_2 = \frac{1}{16} \sum_a e^{-i\mathbf{K} \cdot (\mathbf{a} + \delta\boldsymbol{\tau}_{AB})} \left( \frac{t''(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|^2} - \frac{t'(|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|)}{|\mathbf{a} + \delta\boldsymbol{\tau}_{AB}|^3} \right) ([(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_x]^4 - [(\mathbf{a} + \delta\boldsymbol{\tau}_{AB})_y]^4).$$

Thus, we have the terms

$$\sum_j \sum_{SS'} \Psi_{j,S}^\dagger \left( \zeta_{SS'} + \frac{1}{2} v_F \gamma (\nabla \cdot \delta \mathbf{U}) \bar{\boldsymbol{\sigma}}_{SS'} \cdot \mathcal{A} \right) \Psi_{j,S'}. \quad (\text{A61})$$

with the numerical values of the parameters listed in Table IV.

## 6. Deformation potential

In this subsection, we consider how the strain can couple with the charge density of the electrons on a single layer. In the previous subsections, we have calculated the effective continuum Hamiltonian  $H_{\text{eff}}^{\mathbf{K}}$  that contains the term

$$H_{dp,1} = \alpha \int d^2\mathbf{x} \sum_{j,S} \pm (\nabla \cdot \delta \mathbf{U}) \Psi_{j,S}^\dagger \Psi_{j,S} = 2\alpha \int d^2\mathbf{x} \sum_{j,S} (\nabla \cdot \mathbf{U}_j^\parallel) \Psi_{j,S}^\dagger \Psi_{j,S}, \quad (\text{A62})$$

with  $\alpha = \frac{1}{4} \sum_a e^{-i\mathbf{K} \cdot \mathbf{a}} |a| t'(|a|)$ . In the above formula, + and - signs are from the top and bottom layers, respectively, and we have used the fact that  $\nabla \cdot \mathbf{U}_t^\parallel = -\nabla \cdot \mathbf{U}_b^\parallel = \frac{1}{2} \nabla \cdot \delta \mathbf{U}$ . However, we have assumed that the all the hopping constants depend only on the displacement, and neglect the possible dependence of the on-site energy on the nearby atomic configurations. If the on-site energy  $\epsilon$  depends on the lengths of three nearest bonds, it is expected that this dependence should also induce the coupling between the strain and the fermion density. In this case,  $\epsilon$  is a function of  $\{|\mathbf{n}_{j,S}^{(i)}(\mathbf{x})|\}$  ( $i = 1, \dots, 3$ ) where  $\mathbf{n}_{j,S}^{(i)}(\mathbf{x})$  is the nearest-neighbor bond on layer  $j$ , sublattice  $S$ , and at the position of  $\mathbf{x}$  [61]. As a result, the on-site energy can be expanded to the linear order of the strain [61]:

$$\epsilon(\{|\mathbf{n}_{j,S}^{(\alpha)}(\mathbf{x})|\}) \approx \epsilon_0 + \kappa \nabla \cdot \mathbf{U}_j^\parallel, \quad \text{with} \quad \kappa = \frac{\sqrt{3}}{2} a \frac{\partial \epsilon}{\partial |\delta_S^\alpha|}. \quad (\text{A63})$$

This induces another term that couples the strain with the fermion density:

$$H_{dp,2}^K = \sum_{j,S} \int d^2\mathbf{x} (\epsilon_0 + \kappa \nabla \cdot \mathbf{U}_j^\parallel) \Psi_{j,S}^\dagger \Psi_{j,S}. \quad (\text{A64})$$

Now, combining Eqs. (A62) and (A64), we obtain

$$H_{dp}^K = \sum_{j,S} \int d^2\mathbf{x} (\epsilon_0 + \alpha_{dp} \nabla \cdot \mathbf{U}_j^\parallel) \Psi_{j,S}^\dagger \Psi_{j,S}, \quad (\text{A65})$$

with

$$\alpha_{dp} = \kappa + 2\alpha = \frac{\sqrt{3}}{2} a \frac{\partial \epsilon}{\partial |\delta_S^\alpha|} + \frac{1}{2} \sum_a e^{-i\mathbf{K} \cdot \mathbf{a}} |\mathbf{a}| t'(|\mathbf{a}|). \quad (\text{A66})$$

Since Refs. [65,66] do not provide enough information on how the on-site energy depends on the nearby atomic configuration, we can calculate only the value of  $\alpha$ , but not  $\kappa$ . The numerical value of  $\kappa$  can be obtained either by *ab initio* method, such as DFT, or calibrated by the experimental measurements on  $\alpha_{dp}$  with the formula  $\kappa = \alpha_{dp} - 2\alpha$ .

## APPENDIX B: INTERLAYER TUNNELING

In this Appendix, we will present the detailed derivation and the symmetry analysis of the interlayer tunneling terms. Starting from the master formula in Eq. (28), and keeping only the first-order gradient of the fermion fields, we obtain

$$H_{\text{inter}} = A_{mlg}^{-1} \sum_G \sum_{SS'} e^{i\mathbf{G} \cdot \tau_{SS'}} e^{i(\mathbf{G}+\mathbf{K}) \cdot 2U_{i,S}^\parallel(\mathbf{x})} \int d^2\mathbf{y} e^{-i(\mathbf{G}+\mathbf{K}) \cdot \mathbf{y}_t} (\mathbf{y} + d_0 \hat{z}, \{\mathbf{n}_{i,S}^{(\alpha)}(\mathbf{x} + \mathbf{y}/2)\}, \{\mathbf{n}_{b,S'}^{(\alpha)}(\mathbf{x} - \mathbf{y}/2)\}) \\ \times \left[ \Psi_{i,S}^\dagger(\mathbf{x}) \Psi_{b,S'}(\mathbf{x}) + \frac{\mathbf{y}}{2} \cdot (\nabla \Psi_{i,S}^\dagger(\mathbf{x}) \Psi_{b,S'}(\mathbf{x}) - \Psi_{i,S}^\dagger(\mathbf{x}) \nabla \Psi_{b,S'}(\mathbf{x})) \right] + \text{H.c.}, \quad (\text{B1})$$

where we have applied the formula that the lattice distortion  $U_i^\parallel = -U_b^\parallel$  and is independent of the sublattice. In addition, the microscopic interlayer hopping functions are assumed to depend not only on the hopping displacement  $\mathbf{y} + d_0 \hat{z}$  but also on the direction of the nearest-neighbor vectors  $\mathbf{n}_{i,S}^{(\alpha)}$  and  $\mathbf{n}_{b,S'}^{(\alpha)}$  on each layer [66]. Note that the terms in Eq. (28) are expanded only to the first order of the gradients of fermion fields, and therefore, the interlayer tunneling can be written as

$$H_{\text{inter}} = \int d^2\mathbf{x} \Psi_{i,S}(\mathbf{x}) \left[ T_{SS'}(\mathbf{x}) + \frac{1}{2} \{\mathbf{p}, \mathbf{\Lambda}_{SS'}(\mathbf{x})\} \right] \Psi_{b,S'}(\mathbf{x}) + \text{H.c.} \quad (\text{B2})$$

In the rest of this Appendix, we will study the properties of these two tunneling fields  $T_{SS'}(\mathbf{x})$  and  $\mathbf{\Lambda}_{SS'}(\mathbf{x})$  and express them in terms of the microscopic hopping functions.

Different from the intralayer terms that depends only on the gradient of  $\mathbf{U}^\parallel$ , the inter-layer tunnelings are the function of  $\mathbf{U}^\parallel$  that is not tiny compared with  $|\mathbf{a}|$ . As a consequence, we do not expand the interlayer tunneling terms in the powers of  $\delta U$ .

As mentioned in the main text, in general,  $U_i^\parallel(\mathbf{x}) = \frac{\theta}{2} \hat{z} \times \mathbf{x} + \frac{1}{2} \delta U(\mathbf{x})$ , leading to

$$e^{i(\mathbf{G}+\mathbf{K}) \cdot 2U^\parallel(\mathbf{x})} = e^{i\theta(\mathbf{G}+\mathbf{K}) \cdot (\hat{z} \times \mathbf{x})} e^{i(\mathbf{G}+\mathbf{K}) \cdot \delta U} = e^{i\mathbf{x} \cdot (\mathbf{q}_{1,1} + \mathbf{g}_G)} e^{i(\mathbf{G}+\mathbf{K}) \cdot \delta U(\mathbf{x})},$$

where  $\mathbf{q}_{1,1} = -\theta \hat{z} \times \mathbf{K}$  and  $\mathbf{g}_G = -\theta \hat{z} \times \mathbf{G}$ . For small twist angle  $\theta \ll 1$ , it can be shown that  $\mathbf{g}_G$  is a reciprocal lattice vector of the moire superlattice. For  $\mathbf{G}_i$  and  $\mathbf{g}_i$  defined in Fig. 2, it is easy to see that

$$\mathbf{g}_{G_1} = -(\mathbf{g}_1 + \mathbf{g}_2) \quad \text{and} \quad \mathbf{g}_{G_2} = \mathbf{g}_1 \implies \mathbf{g}_{m_1 G_1 + m_2 G_2} = (m_2 - m_1) \mathbf{g}_1 - m_2 \mathbf{g}_2.$$

In addition, the lattice relaxation  $\delta U(\mathbf{x}) = \delta U(\mathbf{x} + \mathbf{L}_i)$  ( $i = 1, 2$ ) is a periodic function, and so is  $e^{i(\mathbf{G}+\mathbf{K}) \cdot \delta U}$ . This suggests that

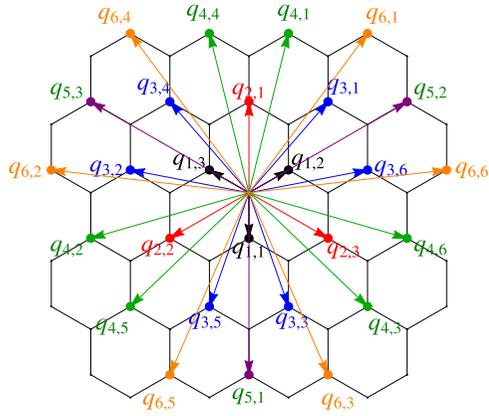
$$e^{i(\mathbf{G}+\mathbf{K}) \cdot \delta U} = \sum_{\mathbf{g}} u_{\mathbf{g}}(\mathbf{G}) e^{i\mathbf{g} \cdot \mathbf{x}} \implies e^{i(\mathbf{G}+\mathbf{K}) \cdot 2U_i^\parallel(\mathbf{x})} = e^{i\mathbf{x} \cdot \mathbf{q}_{(1,1)}} \sum_{\mathbf{g}} e^{i(\mathbf{g}_G + \mathbf{g}) \cdot \mathbf{x}} u_{\mathbf{g}}(\mathbf{G}),$$

where  $\sum_{\mathbf{g}}$  sums over all the reciprocal lattice vector of the moire superlattice. This gives the Fourier transformation of  $e^{i(\mathbf{G}+\mathbf{K}) \cdot 2U_i^\parallel(\mathbf{x})}$ .

Now, we obtain the expression of the interlayer tunneling terms:

$$T_{SS'}(\mathbf{x}) = \sum_G e^{i\mathbf{G} \cdot \tau_{SS'}} e^{i\mathbf{q}_{(1,1)} \cdot \mathbf{x}} \sum_{\mathbf{g}} e^{i(\mathbf{g}_G + \mathbf{g}) \cdot \mathbf{x}} u_{\mathbf{g}}(\mathbf{G}) \int d^2\mathbf{y} e^{-i(\mathbf{G}+\mathbf{K}) \cdot \mathbf{y}_t} (\mathbf{y} + d_0 \hat{z}, \{\mathbf{n}_{i,S}^{(\alpha)}(\mathbf{x} + \mathbf{y}/2)\}, \{\mathbf{n}_{b,S'}^{(\alpha)}(\mathbf{x} - \mathbf{y}/2)\}), \quad (\text{B3})$$

$$\mathbf{\Lambda}_{SS'}(\mathbf{x}) = \sum_G e^{i\mathbf{G} \cdot \tau_{SS'}} e^{i\mathbf{q}_{(1,1)} \cdot \mathbf{x}} \sum_{\mathbf{g}} e^{i(\mathbf{g}_G + \mathbf{g}) \cdot \mathbf{x}} u_{\mathbf{g}}(\mathbf{G}) \int d^2\mathbf{y} (-i\mathbf{y}) e^{-i(\mathbf{G}+\mathbf{K}) \cdot \mathbf{y}_t} (\mathbf{y} + d_0 \hat{z}, \{\mathbf{n}_{i,S}^{(\alpha)}(\mathbf{x} + \mathbf{y}/2)\}, \{\mathbf{n}_{b,S'}^{(\alpha)}(\mathbf{x} - \mathbf{y}/2)\}) \quad (\text{B4})$$


 FIG. 14. The vectors  $\mathbf{q}_{\mu,l}$  in the first six shells.

As shown in the next Appendix, the Fourier transformation of the interlayer tunneling in both microscopic tight-binding models decay fast as a function of the momentum. Therefore, it is convenient to express and calculate the Fourier transformation of the tunneling terms. Since the direction of the nearest bond  $\mathbf{n}_{j,S}$  is also periodic with the period of  $\mathbf{L}_i$ , its Fourier transformation can be written as

$$T_{SS'}(\mathbf{x}) = \sum_{\mu,l} e^{i\mathbf{q}_{\mu,l}\cdot\mathbf{x}} T_{SS'}^{(\mu,l)}, \quad \Lambda_{SS'}(\mathbf{x}) = \sum_{\mu,l} e^{i\mathbf{q}_{\mu,l}\cdot\mathbf{x}} \Lambda_{SS'}^{(\mu,l)}, \quad (\text{B5})$$

where the vectors  $\mathbf{q}_{(\mu,l)}$  are plotted in Fig. 14 for the first six  $\mathbf{q}$  shells ( $\mu \leq 6$ ). The subscript  $\mu$  refers to the shell of the vectors, and  $l$  is used to distinguish different vectors in the same shell. It is easy to see that  $\mathbf{q}_{\mu,l} - \mathbf{q}_{(1,1)}$  is a reciprocal lattice vector of the moire superlattice.

Here, we consider the microscopic tight-binding models with the interlayer hopping depending only on the hopping displacement, as in Ref. [65]. We introduce the notation for the Fourier transformation of the hopping:

$$\begin{aligned} \tilde{t}_{d_0}(\mathbf{q}) &= A_{mlg}^{-1} \int d^2\mathbf{y} e^{-i\mathbf{q}\cdot\mathbf{y}t} (\mathbf{y} + d_0\hat{z}) \\ &\implies A_{mlg}^{-1} \int d^2\mathbf{y} e^{-i\mathbf{q}\cdot\mathbf{y}t} \mathbf{y}t (\mathbf{y} + d_0\hat{z}) \\ &= i\nabla_{\mathbf{q}} \tilde{t}_{d_0}(\mathbf{q}). \end{aligned}$$

Therefore,

$$T_{SS'}(\mathbf{x}) = e^{i\mathbf{q}_{(1,1)}\cdot\mathbf{x}} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\tau_{SS'}} \tilde{t}_{d_0}(\mathbf{G} + \mathbf{K}) \sum_{\mathbf{g}} e^{i(\mathbf{g}_G + \mathbf{g})\cdot\mathbf{x}} u_{\mathbf{g}}(\mathbf{G}), \quad (\text{B6})$$

$$\begin{aligned} \Lambda_{SS'}(\mathbf{x}) &= e^{i\mathbf{q}_{(1,1)}\cdot\mathbf{x}} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\tau_{SS'}} \nabla_{\mathbf{q}} \tilde{t}_{d_0}(\mathbf{q}) \Big|_{\mathbf{q}=\mathbf{G}+\mathbf{K}} \\ &\quad \times \sum_{\mathbf{g}} e^{i(\mathbf{g}_G + \mathbf{g})\cdot\mathbf{x}} u_{\mathbf{g}}(\mathbf{G}). \end{aligned} \quad (\text{B7})$$

With the rigid twist only,  $\delta U = 0$ , and thus  $u_{\mathbf{g}}(\mathbf{G}) = \delta_{\mathbf{g},0}$ . If we only focus on the first  $\mathbf{q}$  shell ( $\mu = 1$ ), we see

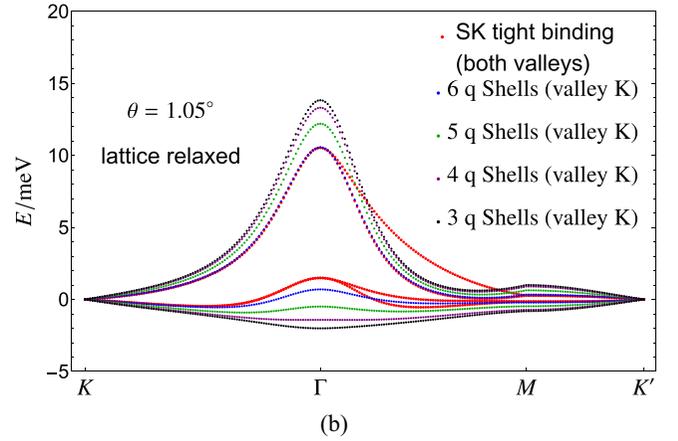
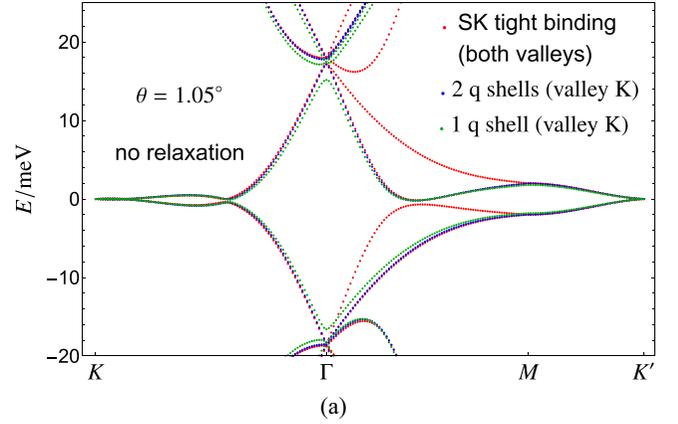


FIG. 15. Spectrum after truncating to a different number of  $\mathbf{q}$  shells in the interlayer tunneling terms. We have considered the effective continuum Hamiltonian  $H_{\text{eff}}^{\mathbf{K}}$  constructed for the microscopic models in Ref. [65]. With the rigid twist only (above), the inclusion of two  $\mathbf{q}$  shells leads to an almost perfect agreement between the spectrum produced by  $H_{\text{eff}}^{\mathbf{K}}$  and the microscopic model. In the presence of the lattice relaxation (below), however, more  $\mathbf{q}$  shells are needed to achieve comparable accuracy. Note that while the tight-binding spectra automatically contain both valleys, for the continuum model we show the spectra only for one valley to avoid clutter.

that

$$\begin{aligned} T_{SS'}(\mathbf{x}) &= e^{i\mathbf{q}_{(1,1)}\cdot\mathbf{x}} \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\tau_{SS'}} \tilde{t}_{d_0}(\mathbf{G} + \mathbf{K}) e^{i\mathbf{g}_G\cdot\mathbf{x}} \\ &\approx e^{i\mathbf{q}_{(1,1)}\cdot\mathbf{x}} \tilde{t}(\mathbf{K}) + e^{i\mathbf{q}_{1,2}\cdot\mathbf{x}} e^{-i\mathbf{G}_1\cdot\tau_{SS'}} \tilde{t}_{d_0}(-\mathbf{G}_1 + \mathbf{K}) \\ &\quad + e^{i\mathbf{q}_{1,2}\cdot\mathbf{x}} e^{-i(\mathbf{G}_1 + \mathbf{G}_2)\cdot\tau_{SS'}} \tilde{t}_{d_0}(-\mathbf{G}_1 - \mathbf{G}_2 + \mathbf{K}). \end{aligned} \quad (\text{B8})$$

Due to  $C_3$  symmetry,  $\tilde{t}_{d_0}(\mathbf{K}) = \tilde{t}_{d_0}(-\mathbf{G}_1 + \mathbf{K}) = \tilde{t}_{d_0}(-\mathbf{G}_1 - \mathbf{G}_2 + \mathbf{K})$  and are real by  $C_{2x}$  symmetry. Thus, by calculating the component of  $w_i^{(1,l)}$ , we see that

$$\begin{aligned} w_0^{(1,1)} &= w_0, \quad w_2^{(1,1)} = 0, \quad w_3^{(1,1)} = 0, \\ w_1^{(1,2)} &= w_1^{(1,3)} = -\frac{1}{2}w_1^{(1,1)}, \quad w_2^{(1,2)} = -w_2^{(1,3)} = -\frac{\sqrt{3}}{2}w_0, \end{aligned}$$

where  $w_0 = \tilde{t}_{d_0}(\mathbf{K})$ . Thus, if we further neglect the gradient coupling fields  $\Lambda_{SS'}$ , we fully recover the BM model [60] in which  $w_0 = w_1$ .

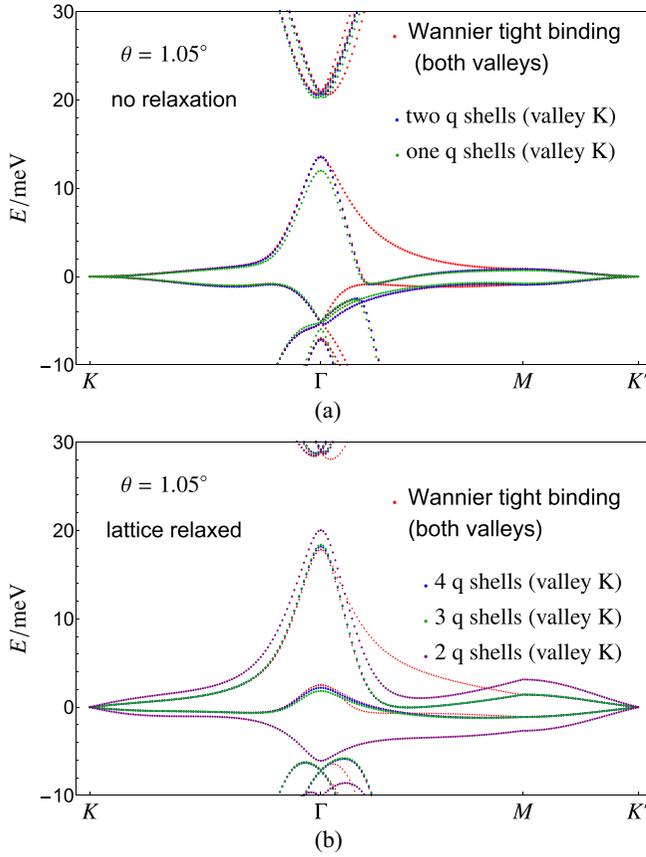


FIG. 16. Same as Fig. 15 but for another microscopic model in Ref. [66].

The values of  $w_i^{(\mu,l)}$  and  $\lambda_i^{(\mu,l)}$  have been listed in Tables III and V for the two microscopic tight-binding models proposed in Refs. [65,66]. Note that we only listed the values of  $w$ s and  $\lambda$ s for the first several  $\mathbf{g}$  shells. The contributions from other shells are negligible.

In Appendix D 2, we will discuss the interlayer tunneling terms for another more complicated microscopic model in which the interlayer hopping depends not only on the hopping displacement but also on the direction of the nearest-neighbor bonds on the two layers.

### APPENDIX C: SYMMETRY ANALYSIS

In this Appendix, we consider the constraints on them by various symmetries. Here, we focus on three different symmetry transformations,  $C_2\mathcal{T}$ ,  $C_3$ , and  $C_{2x}$ , and the fermion fields transform as

$$C_2\mathcal{T} : \Psi_{j,S}(\mathbf{x}) \longrightarrow \mathcal{K}\Psi_{j,\bar{S}}(-\mathbf{x}), \quad (C1)$$

$$C_3 : \Psi_{j,S}(\mathbf{x}) \longrightarrow e^{-i\frac{2\pi}{3}(\sigma_z)_{SS}}\Psi_{j,S}\left(R\left(-\frac{2\pi}{3}\right)\mathbf{x}\right), \quad (C2)$$

$$C_{2x} : \Psi_{j,S}(\mathbf{x}) \longrightarrow \Psi_{\bar{j},\bar{S}}(m_y\mathbf{x}), \quad (C3)$$

where  $\bar{j}$  and  $\bar{S}$  are the layer and sublattice index different from  $j$  and  $S$ , respectively.  $R(-\frac{2\pi}{3})\mathbf{x}$  is the vector  $\mathbf{x}$  rotated clockwise by the angle of  $2\pi/3$ , i.e.,

$$R\left(-\frac{2\pi}{3}\right)(x,y)^T = \left(-\frac{x}{2} + \frac{\sqrt{3}}{2}y, -\frac{y}{2} - \frac{\sqrt{3}}{2}x\right)^T,$$

and  $m_y$  is the reflection symmetry through  $xz$  plane, i.e.,  $m_y(x,y)^T = (x,-y)^T$ .

The interlayer tunneling terms are invariant under  $C_2\mathcal{T}$  transformation. It leads to the constraints

$$(\sigma_1 T(\mathbf{x})\sigma_1)_{SS'} = (T_{SS'}(-\mathbf{x}))^* \quad (\sigma_1 \Lambda(\mathbf{x})\sigma_1)_{SS'} = (\Lambda_{SS'}(-\mathbf{x}))^* \quad (C4)$$

$$\implies (\sigma_1 T^{(\mu,l)}\sigma_1)_{SS'} = (T_{SS'}^{(\mu,l)})^*, \quad (\sigma_1 \Lambda^{(\mu,l)}\sigma_1)_{SS'} = (\Lambda_{SS'}^{(\mu,l)})^*. \quad (C5)$$

This suggests that we can write

$$T_{SS'}^{(\mu,l)} = (w_0^{(\mu,l)}\sigma_0 + w_1^{(\mu,l)}\sigma_1 + w_2^{(\mu,l)}\sigma_2 + iw_3^{(\mu,l)}\sigma_3)_{SS'}, \quad (C6)$$

$$\Lambda_{SS'}^{(\mu,l)} = (\lambda_0^{(\mu,l)}\sigma_0 + \lambda_1^{(\mu,l)}\sigma_1 + \lambda_2^{(\mu,l)}\sigma_2 + i\lambda_3^{(\mu,l)}\sigma_3)_{SS'}, \quad (C7)$$

where all  $w_i^{(\mu,l)}$  ( $i = 0, \dots, 4$ ) are real numbers and  $\lambda_i^{(\mu,l)}$  are two-component real vectors.

The invariance under  $C_3$  transformation leads to the following constraints on  $T_{SS'}$  and  $\Lambda_{SS'}$ :

$$T_{SS'}(\mathbf{x}) = \left( e^{i\frac{2\pi}{3}\sigma_z} T\left(R\left(\frac{2\pi}{3}\right)\mathbf{x}\right) e^{-i\frac{2\pi}{3}\sigma_z} \right)_{SS'},$$

$$\Lambda_{\alpha,SS'}(\mathbf{x}) = \left( e^{i\frac{2\pi}{3}\sigma_z} R_{\alpha\beta}\left(-\frac{2\pi}{3}\right) \Lambda_{\beta}\left(R\left(\frac{2\pi}{3}\right)\mathbf{x}\right) e^{-i\frac{2\pi}{3}\sigma_z} \right)_{SS'}. \quad (C8)$$

For notational convenience, we introduce  $\mathbf{q}_{(\mu,l')} = R(2\pi/3)\mathbf{q}_{(\mu,l)}$ . Then,

$$T_{SS'}^{(\mu,l)} = \left( e^{i\frac{2\pi}{3}\sigma_z} T^{(\mu,l')} e^{-i\frac{2\pi}{3}\sigma_z} \right)_{SS'}, \quad \Lambda_{\alpha,SS'}^{(\mu,l)} = R_{\alpha\beta}\left(-\frac{2\pi}{3}\right) \left( e^{i\frac{2\pi}{3}\sigma_z} \Lambda_{\beta}^{(\mu,l')} e^{-i\frac{2\pi}{3}\sigma_z} \right)_{SS'}. \quad (C9)$$

Eventually, the contact and gradient coupling constants should satisfy the following relations:

$$\begin{aligned}
 w_0^{(\mu,l')} &= w_0^{(\mu,l)}, & w_3^{(\mu,l')} &= w_3^{(\mu,l)}, & w_1^{(\mu,l')} - iw_2^{(\mu,l')} &= e^{i\frac{2\pi}{3}}(w_1^{(\mu,l)} - iw_2^{(\mu,l)}) \\
 (\lambda_0^{(\mu,l')})_\alpha &= \left(R\left(\frac{2\pi}{3}\right)\right)_{\alpha\beta} (\lambda_0^{(\mu,l)})_\beta, & (\lambda_3^{(\mu,l')})_\alpha &= \left(R\left(\frac{2\pi}{3}\right)\right)_{\alpha\beta} (\lambda_3^{(\mu,l)})_\beta \\
 (\lambda_1^{(\mu,l')} - i\lambda_2^{(\mu,l')})_\alpha &= e^{i\frac{2\pi}{3}} \left(R\left(\frac{2\pi}{3}\right)\right)_{\alpha\beta} (\lambda_1^{(\mu,l)} - i\lambda_2^{(\mu,l)})_\beta.
 \end{aligned} \tag{C10}$$

Lastly, we consider  $C_{2x}$  symmetry. It imposes the constraints

$$T_{SS'}(\mathbf{x}) = (\sigma_x T(m_y \mathbf{x}) \sigma_x)_{SS'}^*, \quad \Lambda_{\alpha,SS'}(\mathbf{x}) = (\tau_3)_{\alpha\beta} (\sigma_x \Lambda_\beta(m_y \mathbf{x}) \sigma_x)_{SS'}^*. \tag{C11}$$

Again, we introduce the notation  $\mathbf{q}_{(\mu,n)} = -m_y \mathbf{q}_{(\mu,l)}$ . It leads to

$$\begin{aligned}
 w_0^{(\mu,n)} &= w_0^{(\mu,l)}, & w_1^{(\mu,n)} &= w_1^{(\mu,l)}, & w_2^{(\mu,n)} &= -w_2^{(\mu,l)}, & w_3^{(\mu,n)} &= w_3^{(\mu,l)} \\
 (\lambda_0^{(\mu,n)})_\alpha &= (\tau_3)_{\alpha\beta} (\lambda_0^{(\mu,l)})_\beta, & (\lambda_1^{(\mu,n)})_\alpha &= (\tau_3)_{\alpha\beta} (\lambda_1^{(\mu,l)})_\beta, & (\lambda_2^{(\mu,n)})_\alpha &= -(\tau_3)_{\alpha\beta} (\lambda_2^{(\mu,l)})_\beta \\
 (\lambda_3^{(\mu,n)})_\alpha &= (\tau_3)_{\alpha\beta} (\lambda_3^{(\mu,l)})_\beta.
 \end{aligned} \tag{C12}$$

The formulas above are the general constraints for the interlayer tunneling fields  $T$  and  $\Lambda$ . As an example, here we explicitly write the formula for  $\mathbf{q}$ s in the innermost shell  $\mu = 1$ . It is obvious that  $n = 1$  for  $l = 1$ . Combining Eqs. (C10) and (C12):

$$w_0^{(1,l)} = w_0, \quad w_3^{(1,l)} = w_3, \quad w_2^{(1,1)} = 0, \tag{C13}$$

$$w_1 = w_1^{(1,1)}, \quad w_1^{(1,2)} = w_1^{(1,3)} = -\frac{w_1}{2}, \quad w_2^{(1,2)} = -w_2^{(1,3)} = -\frac{\sqrt{3}}{2}w_1, \tag{C14}$$

$$\lambda_0^{(1,1)} = (\lambda_{0,x}, 0), \quad \lambda_1^{(1,1)} = (\lambda_{1,x}, 0), \quad \lambda_3^{(1,1)} = (\lambda_{3,x}, 0), \quad \lambda_2^{(1,1)} = (0, \lambda_{2,y}). \tag{C15}$$

## APPENDIX D: MICROSCOPIC HOPPING FUNCTION

### 1. Slater-Koster-like hopping parameterization

First, we consider the microscopic tight-binding model proposed in Ref. [65], in which

$$t(\mathbf{d}) = V_{pp\pi}^0 e^{-\frac{|\mathbf{d}|-a_0}{\delta}} \left[ 1 - \left( \frac{\mathbf{d} \cdot \hat{\mathbf{z}}}{|\mathbf{d}|} \right)^2 \right] + V_{pp\sigma}^0 e^{-\frac{|\mathbf{d}|-d_0}{\delta}} \left( \frac{\mathbf{d} \cdot \hat{\mathbf{z}}}{|\mathbf{d}|} \right)^2, \tag{D1}$$

where  $V_{pp\pi}^0 = -2.7\text{eV}$ ,  $V_{pp\sigma}^0 = 0.48\text{eV}$ .  $a_0 = 0.142\text{nm}$  is the distance between the two nearest-neighbor carbon atoms on the same layer;  $d_0 = 0.335\text{nm}$  is the interlayer distance. The decay length for the hopping is  $\delta = 0.319a_0$ . The intralayer hopping thus can be expressed as

$$t_{\text{intra}}(\mathbf{d}) = V_{pp\pi}^0 e^{-\frac{|\mathbf{d}|-a_0}{\delta}}, \tag{D2}$$

where  $\mathbf{d}$  is the in-plane hopping displacement. It is easy to show that the Fourier transformation of the intralayer hopping is

$$\tilde{t}_{\text{intra}}(\mathbf{q}) = A_{mlg}^{-1} \int d^2\mathbf{y} e^{-iy\cdot\mathbf{q}} t(\mathbf{y}) = V_{pp\pi}^0 \frac{2\pi\delta^2}{A_{mlg}} e^{a_0/\delta} [1 + (q\delta)^2]^{-\frac{3}{2}}. \tag{D3}$$

It is now clear that  $\tilde{t}_{\text{intra}}$  decays in  $q^{-3}$  in the momentum space, but the intralayer hopping  $t$  exponentially decays in real space. The Fourier transformations  $\tilde{t}$  of the interlayer hoppings are

$$\begin{aligned}
 \tilde{t}_{d_0}(|\mathbf{q}|) &= \frac{1}{A_{mlg}} \int d^2\mathbf{y} e^{-iy\cdot\mathbf{q}} t(\mathbf{y} + d_0\hat{\mathbf{z}}) \\
 &= \frac{2\pi d_0^2}{A_{mlg}} \left[ V_{pp\pi}^0 \int_0^\infty dy y J_0(qd_0y) e^{-\frac{d_0}{\delta}(\sqrt{y^2+1}-a_0/d_0)} \frac{y^2}{y^2+1} + V_{pp\sigma}^0 \int_0^\infty dy y J_0(qd_0y) e^{-\frac{d_0}{\delta}(\sqrt{y^2+1}-1)} \frac{1}{y^2+1} \right].
 \end{aligned} \tag{D4}$$

With the Fourier transformation in Eq. (D4), we are able to calculate all the components  $w_i^{(\mu,l)}$  and  $\lambda_i^{(\mu,l)}$  of the interlayer tunneling fields  $T_{SS'}(\mathbf{x})$  and  $\Lambda_{SS'}(\mathbf{x})$ , with their values listed in Table V.

TABLE V. Parameters of the interlayer tunneling terms for two microscopic tight-binding models in the presence of the lattice relaxation.  $a$  is the magnitude of the primitive lattice vector and all numbers are in units of meV.

	$w_0^{(4,1)}$	$w_1^{(4,1)}$	$w_2^{(4,1)}$	$w_3^{(4,1)}$	$w_0^{(5,1)}$	$w_1^{(5,1)}$	$w_2^{(5,1)}$	$w_3^{(5,1)}$	$w_0^{(6,1)}$	$w_1^{(6,1)}$	$w_2^{(6,1)}$	$w_3^{(6,1)}$
Ref. [65], relaxed	-0.39	1.13	-0.20	0	5.79	6.36	0	0	5.18	-2.78	-4.98	0
Ref. [66], relaxed	0.16	0.34	0.10	-0.07	negligible							
	$\lambda_0^{(4,1)}/a$	$\lambda_1^{(4,1)}/a$	$\lambda_2^{(4,1)}/a$	$\lambda_3^{(4,1)}/a$	$\lambda_0^{(5,1)}/a$	$\lambda_1^{(5,1)}/a$	$\lambda_2^{(5,1)}/a$	$\lambda_3^{(5,1)}/a$	$\lambda_0^{(6,1)}/a$	$\lambda_1^{(6,1)}/a$	$\lambda_2^{(6,1)}/a$	$\lambda_3^{(6,1)}/a$
Ref. [65], relaxed	(-0.1, -0.2)	(0.5, 0.1)	(-0.1, -0.7)	(0, 0)	(-5.2, 0)	(-5.1, 0)	(0, -0.2)	(0, 0)	(2.2, -4.1)	(-1.1, 2.1)	(-2.1, 3.3)	(0, 0)
Ref. [66], relaxed	(0.3, 0.1)	(0.3, -0.1)	(-0.1, 0)	(0, 0)	negligible							

## 2. Wannier-based hopping parameterization

In this subsection, we consider the model proposed in Ref. [66]. Note that Ref. [66] does not provide a general formula for the intralayer hoppings but lists its magnitude for a set of discrete hopping distance. Here, we fit the intralayer hopping with the following formula:

$$t_{\text{intra}}(\mathbf{r}) = t_0 e^{-\alpha_0 \bar{r}^2} \cos(\beta_0 \bar{r}) + t_1 \bar{r}^2 e^{-\alpha_1 (\bar{r} - r_1)^2}. \quad (\text{D5})$$

The values of the fitted parameters are listed in Table VI. It is obvious that the intralayer hopping decays exponentially as a function of  $|\mathbf{r}|$ .

According to Ref. [66], the general form of the interlayer hopping can be written as

$$t_{\text{inter}}(\mathbf{r}) = V_0(r) + V_3(r)(\cos(3\theta_{12}) + \cos(3\theta_{21})) + V_6(r)(\cos(6\theta_{12}) + \cos(6\theta_{21})), \quad (\text{D6})$$

$$V(r) = \lambda_0 e^{-\xi_0 \bar{r}^2} \cos(\kappa_0 \bar{r}), \quad (\text{D7})$$

$$V_3(r) = \lambda_3 \bar{r}^2 e^{-\xi_3 (\bar{r} - x_3)^2}, \quad (\text{D8})$$

$$V_6(r) = \lambda_6 e^{-\xi_6 (\bar{r} - x_3)^2} \sin(\kappa_6 \bar{r}), \quad (\text{D9})$$

where the vector  $\mathbf{r}$  is the in-plane projected vector of the hopping displacement and  $\bar{r} = r/a$ . The variables  $\theta_{12}$  and  $\theta_{21}$  are the angles between  $\mathbf{r}$  and the nearest-neighbor bond vectors  $\mathbf{n}_{j,S}$  on two layers, i.e.,

$$\theta_{12} = \cos^{-1} \left( -\frac{\mathbf{r} \cdot \mathbf{n}_{j,S}}{r |\mathbf{n}_{j,S}|} \right) = \theta_r - \theta_{j,S} + \pi, \quad (\text{D10})$$

$$\theta_{21} = \cos^{-1} \left( \frac{\mathbf{r} \cdot \mathbf{n}_{j',S'}}{r |\mathbf{n}_{j',S'}|} \right) = \theta_r - \theta_{j',S'}. \quad (\text{D11})$$

In the above formula, we define  $\theta_r$  to be the angle between the vector  $\mathbf{r}$  and the  $x$  axis, and  $\theta_{j,S}$  ( $\theta_{j',S'}$ ) to be the angle between the bond vector  $\mathbf{n}_{j,S}$  ( $\mathbf{n}_{j',S'}$ ) and the  $x$  axis. Note that each carbon atom has three different bond vectors. We label the angles of these three bonds by  $\theta_{j,S}^{(\alpha)}$  ( $\theta_{j',S'}^{(\alpha)}$ ), where the superscript  $\alpha$  is the index of the bond vectors. Without the lattice distortion (e.g., as for a rigid twist), the three in-plane nearest neighbors of a carbon atom are  $C_3$  symmetric about the carbon atom, and  $\theta_{j,S}^{(\alpha)} = \theta_{j,S}^{(1)} + 2\pi(\alpha - 1)/3$ . Therefore, the angles  $\theta_{12}$  and  $\theta_{21}$  could differ by  $2\pi/3$  if choosing a different nearest-neighbor bond, leading to the same  $\cos(3m\theta_{12})$  and  $\cos(3m\theta_{21})$ , with  $m$  being an integer. As a consequence, Eq. (D6) is independent of the choice of the bond vectors. The values of all microscopic hopping parameters are listed in Table VI.

TABLE VI. Parameters for the interlayer and intralayer hoppings in the model proposed in Ref. [66].

Inter	$\lambda_i$ (eV)	$\xi_i$	$x_i$	$\kappa_i$		
$V_0$	0.3155	1.7543		2.0010		
$V_3$	-0.0688	3.4692	0.5212			
$V_6$	-0.0083	2.8764	1.5206	1.5731		
Intra	$t_0$ (eV)	$\alpha_0$	$\beta_0$	$t_1$ (eV)	$\alpha_1$	$r_1$
	-18.4295	1.2771	2.3934	-3.7183	6.2194	0.9071

In the presence of the lattice distortion, however, the local  $C_3$  symmetry is broken, and thus Eq. (D6) depends on the choice of the bond vectors. In this case, we set the interlayer hopping as

$$t_{\text{inter}}(\mathbf{r}) = V_0(r) + V_3(r) \left( \frac{1}{3} \sum_{\alpha=1}^3 \cos(3\theta_{12}^{(\alpha)}) + \frac{1}{3} \sum_{\alpha=1}^3 \cos(3\theta_{21}^{(\alpha)}) \right) + V_6(r) \left( \frac{1}{3} \sum_{\alpha=1}^3 \cos(6\theta_{12}^{(\alpha)}) + \frac{1}{3} \sum_{\alpha=1}^3 \cos(6\theta_{21}^{(\alpha)}) \right), \quad (\text{D12})$$

where  $\theta_{12}^{(\alpha)} = \theta_r - \theta_{j,S}^{(\alpha)} + \pi$  and  $\theta_{21}^{(\alpha)} = \theta_r - \theta_{j',S'}^{(\alpha)}$ .

To obtain the Fourier transformation of the hopping function, consider

$$\begin{aligned} \int d^2\mathbf{y} e^{-i\mathbf{q}\cdot\mathbf{y}} t_n(|\mathbf{y}|) e^{in(\theta_y - \theta_j)} &= \int_0^\infty dy y t_n(y) \int_0^{2\pi} d\theta e^{-iqy \cos(\theta - \theta_j)} e^{in(\theta - \theta_j)} \\ &= \int_0^\infty dy y t_n(y) \int_0^{2\pi} d\theta \sum_{m=-\infty}^\infty (-ie^{i(\theta - \theta_j)})^m J_m(qy) e^{in(\theta - \theta_j)} = 2\pi \int_0^\infty dy y t_n(y) (-i)^{-n} J_{-n}(qy) e^{in(\theta_q - \theta_j)} \\ &= 2\pi (-i)^n e^{in(\theta_q - \theta_j)} \int_0^\infty dy y t_n(y) J_n(qy), \end{aligned} \quad (\text{D13})$$

where  $\theta_q$  is the angle between the vector  $\mathbf{q}$  and the  $\hat{x}$  axis. In the derivation above, we have applied the formula  $e^{\frac{z}{2}(t-t^{-1})} = \sum_{m=-\infty}^\infty t^m J_m(z)$  and  $J_{-n}(z) = (-1)^n J_n(z)$ . Starting from Eq. (D13), it is easy to obtain

$$\int d^2\mathbf{y} e^{-i\mathbf{q}\cdot\mathbf{y}} t_n(|\mathbf{y}|) \cos(n(\theta_y - \theta_j)) = 2\pi (-i)^n \cos(n(\theta_q - \theta_j)) \int_0^\infty dy y J_n(qy) t_n(y). \quad (\text{D14})$$

For notation convenience, we introduce

$$\tilde{V}_i(q) = \frac{2\pi}{A_{\text{mlg}}} \int_0^\infty dy y V_i(y) J_i(qy), \quad (\text{D15})$$

where  $i = 0, 3$ , and  $6$ .

We first consider the Fourier transformation when the lattice is locally  $C_3$  symmetric, i.e., the lattice relaxation is absent. In this case,

$$\begin{aligned} \tilde{t}_{C_3}(\mathbf{q}, \theta_{j,S}, \theta_{j',S'}) &= \tilde{t}(\mathbf{q}, \theta_{j,S}, \theta_{j',S'}) = A_{\text{mlg}}^{-1} \int d^2\mathbf{y} e^{-i\mathbf{q}\cdot\mathbf{y}} t(\mathbf{y} + d_0\hat{z}, \theta_{j,S}, \theta_{j',S'}) \\ &= \tilde{V}_0(q) + i(-\cos(3(\theta_q - \theta_{j,S})) + \cos(3(\theta_q - \theta_{j',S'}))) \tilde{V}_3(q) - (\cos(6(\theta_q - \theta_{j,S})) + \cos(6(\theta_q - \theta_{j',S'}))) \tilde{V}_6(q), \end{aligned} \quad (\text{D16})$$

where  $\tilde{t}_{C_3}$  is introduced to refer to the Fourier transformation of the interlayer hopping function when the lattice is locally  $C_3$  symmetric.

In the presence of the lattice distortion that varies slowly in the real space, the angle  $\theta_{j,S}^{(\alpha)}$  can be approximated as

$$\theta_{j,S}^{(\alpha)} = \theta_{\delta_S}^{(\alpha)} + \delta\theta_{j,S}^{(\alpha)}, \quad (\text{D17})$$

$$\delta\theta_{j,S}^{(\alpha)} = \frac{(\hat{z} \times \boldsymbol{\delta}_S^{(\alpha)}) \cdot \delta\mathbf{n}_{j,S}^{(\alpha)}}{|\boldsymbol{\delta}_S^{(\alpha)}|^2} = \frac{\epsilon_{\mu\nu}}{|\boldsymbol{\delta}_S^{(\alpha)}|^2} \delta_{S,\mu}^{(\alpha)} \frac{\partial U_{j,v}^{\parallel(\alpha)}}{\partial x_\rho} \delta_{S,\rho}^{(\alpha)}, \quad (\text{D18})$$

where  $\theta_{\delta_B}^{(\alpha)} = \theta_{\delta_A}^{(\alpha)} + \pi$ , and for our choice of the coordinate system,  $\theta_{\delta_A}^{(1)} = \pi/6$ ,  $\theta_{\delta_A}^{(2)} = \pi/6 + 2\pi/3$ ,  $\theta_{\delta_A}^{(3)} = \pi/6 - 2\pi/3$ . Since  $\mathbf{U}_{t,S}^{\parallel} = -\mathbf{U}_{b,S}^{\parallel} = \frac{1}{2}(\theta\hat{z} \times \mathbf{y} + \delta\mathbf{U})$ , we obtain

$$\delta\theta_{t,S}^{(\alpha)} = -\delta\theta_{b,S}^{(\alpha)} = \frac{\theta}{2} + \frac{\epsilon_{\mu\nu}}{2|\boldsymbol{\delta}_S^{(\alpha)}|^2} \delta_{S,\mu}^{(\alpha)} \frac{\partial \delta\mathbf{U}_v}{\partial x_\rho} \delta_{S,\rho}^{(\alpha)}.$$

Consequently, the hopping function is approximated as

$$\begin{aligned} t(\mathbf{y} + d_0\hat{z}, \{\mathbf{n}_{j,S}^{(\alpha)}\}, \{\mathbf{n}_{j',S'}^{(\alpha)}\}) &\approx t(\mathbf{y} + d_0\hat{z}, \{\theta_{\delta_S}^{(\alpha)}\}, \{\theta_{\delta_{S'}}^{(\alpha)}\}) + \sum_{\alpha} \frac{\partial t}{\partial \theta_{j,S}^{(\alpha)}} \delta\theta_{j,S}^{(\alpha)} + \sum_{\alpha} \frac{\partial t}{\partial \theta_{j',S'}^{(\alpha)}} \delta\theta_{j',S'}^{(\alpha)} \\ &= V_0(y) + V_3(y) (-\cos 3(\theta_y - \theta_{\delta_S}) + \cos 3(\theta_y - \theta_{\delta_{S'}})) + V_6(y) (\cos 6(\theta_y - \theta_{\delta_S}) + \cos 6(\theta_y - \theta_{\delta_{S'}})) \\ &\quad + V_3(y) \left( \left( \sum_{\alpha} \delta\theta_{j,S}^{(\alpha)} \right) \sin 3(\theta_y - \theta_{\delta_S}) - \left( \sum_{\alpha} \delta\theta_{j',S'}^{(\alpha)} \right) \sin 3(\theta_y - \theta_{\delta_{S'}}) \right) \end{aligned}$$

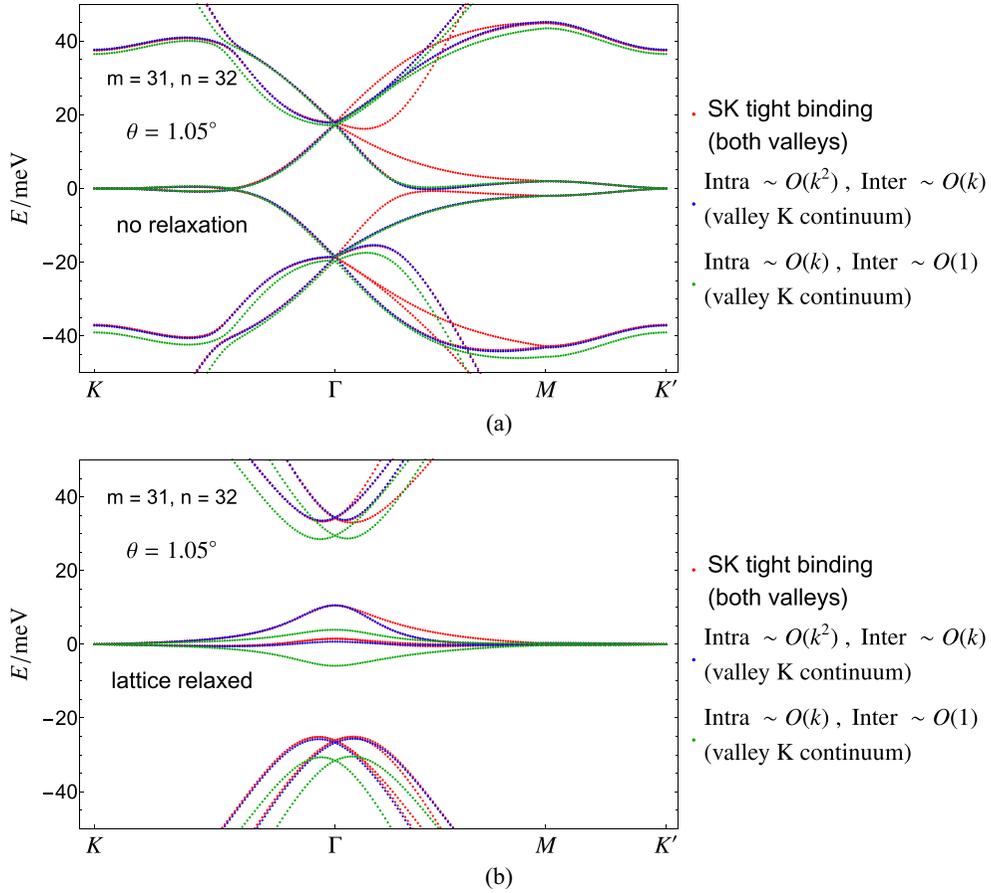


FIG. 17. Impact of the higher order terms in the effective Hamiltonian on the energy spectrum near the CNP. The spectra of  $H_{\text{eff}}^{\text{K}}$  constructed for the microscopic models in Ref. [65] produced by (blue) keeping or (green) dropping the second-order gradient terms in the intralayer continuum Hamiltonian and gradient couplings in the interlayer tunneling terms. For comparison, the spectra of the microscopic tight-binding model are also plotted and marked as red. Dropping the higher order derivative terms leads to a mismatch of  $\sim 5$ – $10$  meV that is consistent with the estimate of this energy scale in the main text. Note that while the tight-binding spectra automatically contain both valleys, for the continuum model we show the spectra only for one valley to avoid clutter.

$$-2V_6(y) \left( \left( \sum_{\alpha} \delta\theta_{j,S}^{(\alpha)} \right) \sin 6(\theta_y - \theta_{\delta_s}) + \left( \sum_{\alpha} \delta\theta_{j',S'}^{(\alpha)} \right) \sin 6(\theta_y - \theta_{\delta_{s'}}) \right). \quad (\text{D19})$$

Correspondingly, its Fourier transformation can be written as

$$\tilde{t}(\mathbf{q}, \{\theta_{j,S}^{(\alpha)}\}, \{\theta_{j',S'}^{(\alpha)}\}) \approx \tilde{t}_{C_3}(\mathbf{q}, \theta_{\delta_s}, \theta_{\delta_{s'}}) + \frac{1}{3} \left( \sum_{\alpha} \delta\theta_{j,S}^{(\alpha)} \right) \frac{\partial \tilde{t}_{C_3}(\mathbf{q}, \theta_{\delta_s}, \theta_{\delta_{s'}})}{\partial \theta_{\delta_s}} + \frac{1}{3} \left( \sum_{\alpha} \delta\theta_{j',S'}^{(\alpha)} \right) \frac{\partial \tilde{t}_{C_3}(\mathbf{q}, \theta_{\delta_s}, \theta_{\delta_{s'}})}{\partial \theta_{\delta_{s'}}}, \quad (\text{D20})$$

where  $\tilde{t}_{C_3}$ , defined in Eq. (D16), is the Fourier transformation of the interlayer hopping for a locally  $C_3$  symmetric lattice.

### 3. Impact of sub-leading terms

We also consider the impact of the subleading terms in the constructed continuum model. Fig. 15 and 16 demonstrates the spectrum after truncating to a different number of  $\mathbf{q}$  shells in the inter-layer tunneling terms, for two microscopic models in Ref. [65] and in Ref. [66]. The impact of the higher order terms on the energy spectrum is illustrated in Fig. 17 and 18. Fig. 19 demonstrates the p-h asymmetry induced by  $w_3$  and momentum dependent  $\Lambda$  in the interlayer tunneling.

## APPENDIX E: APPROXIMATE FORMULA OF $v_{\text{Dirac}}$

In this Appendix, we derive the approximate formula of  $v_{\text{Dirac}}$  in the presence of the pseudovector field that is induced by the lattice relaxation proposed in Ref. [64]. Motivated by Fig. 5, we neglect  $\nabla\varphi^A$ , the irrotational part of the pseudovector field  $\mathcal{A}$ , and thus  $\mathcal{A} \approx \nabla \times (\hat{z}\varepsilon^A)$ . Furthermore, as demonstrated in Table I, the lattice relaxation  $\delta U$  in Ref. [64] is dominated by the lowest harmonics. Because  $\varepsilon_{\mathcal{A}}$  is even and real, its Fourier components must be pure imaginary and also odd. Considering  $C_3$

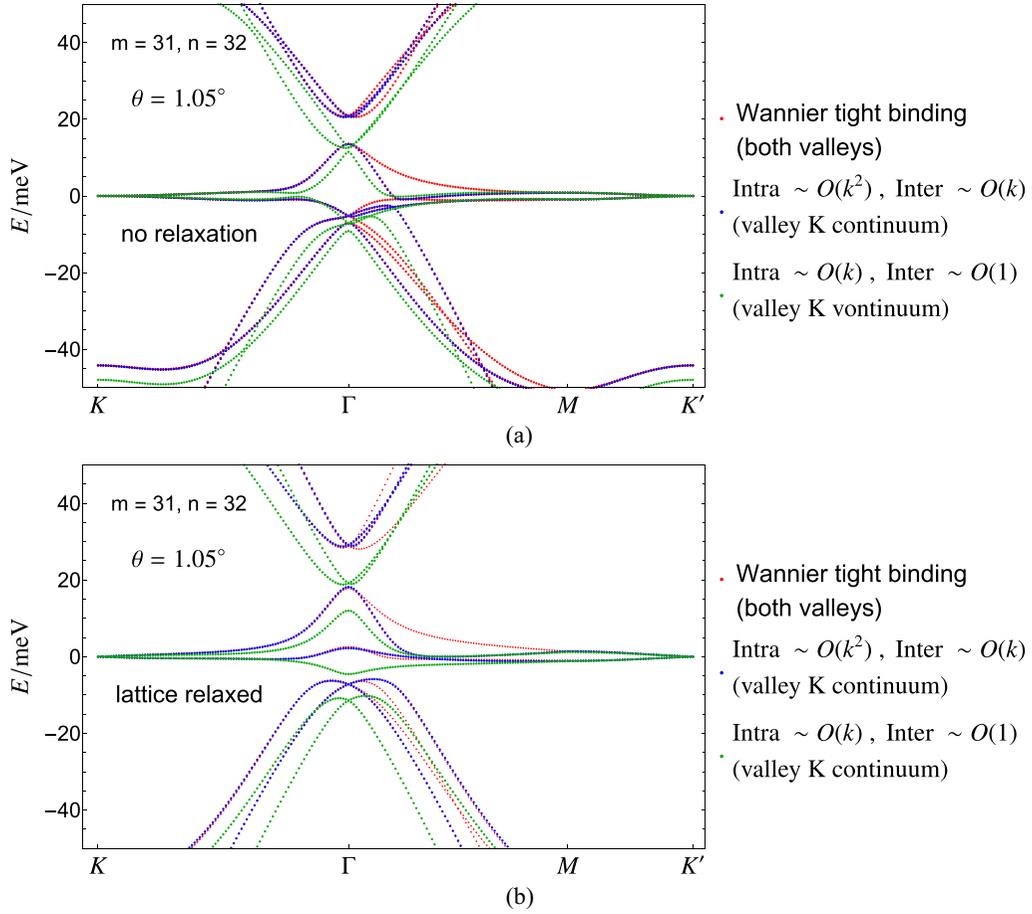


FIG. 18. Same as Fig. 17 but for another microscopic model in Ref. [66].

symmetry, we can assume that

$$\varepsilon^{\mathcal{A}}(\mathbf{x}) \approx i\tilde{\varepsilon}_1^{\mathcal{A}} \sum_{j=1}^3 (e^{ig_j \cdot \mathbf{x}} - e^{-ig_j \cdot \mathbf{x}}), \quad (\text{E1})$$

where  $\mathbf{g}_3 = -(\mathbf{g}_1 + \mathbf{g}_2)$ . Under these approximation, we can obtain

$$\mathcal{A}_+ = \frac{1}{\tilde{\varepsilon}_1^{\mathcal{A}}} (\mathcal{A}_1 + i\mathcal{A}_2) \approx i|\mathbf{g}_1| \sum_{j=1}^3 \omega^{j-1} (e^{ig_j \cdot \mathbf{x}} + e^{-ig_j \cdot \mathbf{x}}), \quad (\text{E2})$$

where  $\mathbf{g}_3 = -(\mathbf{g}_1 + \mathbf{g}_2)$  and  $\omega = e^{i2\pi/3}$ . Now, for the equation

$$v_F \begin{pmatrix} p_+ + \gamma\tilde{\varepsilon}_1^{\mathcal{A}}\mathcal{A}_+ & \alpha U(\mathbf{x}) \\ \alpha U(-\mathbf{x}) & p_+ - \gamma\tilde{\varepsilon}_1^{\mathcal{A}}\mathcal{A}_+ \end{pmatrix} \begin{pmatrix} \Phi_1(\mathbf{x}) \\ \Phi_2(\mathbf{x}) \end{pmatrix} = 0, \quad (\text{E3})$$

where  $\alpha = w_1/(v_F k_\theta)$ , and  $U(\mathbf{x}) = \sum_j \omega^{j-1} e^{iq_j \cdot \mathbf{x}}$ . We further assume that  $\alpha$  and  $\gamma\tilde{\varepsilon}_1^{\mathcal{A}}$  are small and expand the wave function in terms of the powers of  $\alpha$  and  $\gamma\tilde{\varepsilon}_1^{\mathcal{A}}$ :

$$\Phi_1(\mathbf{x}) = 1 + \gamma\tilde{\varepsilon}_1^{\mathcal{A}}\Phi_1^{(0,1)}(\mathbf{x}) + \alpha^2\Phi_1^{(2,0)}(\mathbf{x}) + (\gamma\tilde{\varepsilon}_1^{\mathcal{A}})^2\Phi_1^{(0,2)}(\mathbf{x}) + \dots, \quad (\text{E4})$$

$$\Phi_2(\mathbf{x}) = \alpha\Phi_2^{(1,0)}(\mathbf{x}) + \alpha\gamma\tilde{\varepsilon}_1^{\mathcal{A}}\Phi_2^{(1,1)}(\mathbf{x}) + \dots, \quad (\text{E5})$$

where we have used the fact that  $\Phi_1(\mathbf{x})$  is even in  $\alpha$  and  $\Phi_2(\mathbf{x})$  is odd in  $\alpha$ . Substituting Eqs. (E4) and (E5) into Eq. (E3) and comparing the powers of  $\alpha$  and  $\gamma\tilde{\varepsilon}_1^{\mathcal{A}}$ , we obtain the equations

$$p_+\Phi_1^{(0,1)} + \mathcal{A}_+ = 0, \quad p_+\Phi_1^{(2,0)} + U(\mathbf{x})\Phi_2^{(1,0)} = 0, \quad p_+\Phi_1^{(0,2)} + \mathcal{A}_+\Phi_1^{(0,1)} = 0 \quad (\text{E6})$$

$$p_+\Phi_2^{(1,0)} + U(-\mathbf{x}) = 0, \quad p_+\Phi_2^{(1,1)} - \mathcal{A}_+\Phi_2^{(1,0)} + U(-\mathbf{x})\Phi_1^{(0,1)} = 0. \quad (\text{E7})$$

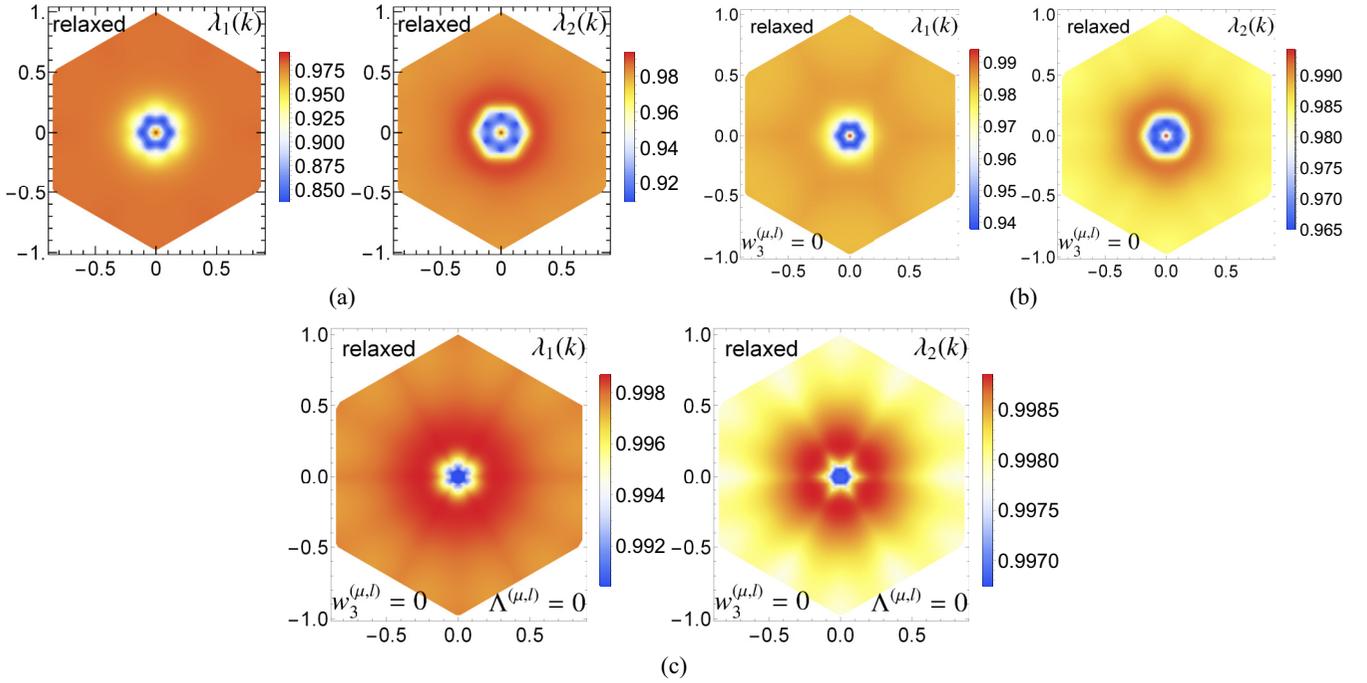


FIG. 19. The two singular values,  $\lambda_1(k)$  (left column) and  $\lambda_2(k)$  (right column), of the projected p-h operator  $\hat{\mathcal{P}}$  defined in Sec. IV by turning off different interlayer terms in  $H_{\text{eff}}^{\mathbf{K}}$  for the microscopic tight-binding model in Ref. [66], and keeping all terms in the intralayer part of  $H_{\text{eff}}^{\mathbf{K}}$  in Eq. (33). (a) All terms in Eq. (40) are kept in the interlayer tunnelings, (b) the contact couplings  $w_3^{(\mu,l)}$  for all the shells are set to 0, and (c) both the contact  $w_3^{(\mu,l)}$  and the gradient couplings  $\Lambda_{Ss'}^{(\mu,l)}$  are set to 0. It is clear that the p-h asymmetry is dominated by the contribution from  $w_3^{(\mu,l)}$ .

We consider the correction of  $v_{\text{Dirac}}$  by the inclusion of  $\mathcal{A}$  by expanding it to the powers of  $O(\alpha^2 \gamma \tilde{\epsilon}_1^A)$ . This allows us to focus only on the lowest harmonics of the Fourier expansion of  $\Phi_1$  and  $\Phi_2$ . After some calculations, we found

$$\Phi_1^{(0,1)} = -i \sum_{j=1}^3 (e^{ig_j \cdot x} - e^{-ig_j \cdot x}), \quad \Phi_2^{(1,0)} = i \sum_j e^{-iq_j \cdot x}, \quad (\text{E8})$$

$$\Phi_1^{(2,0)} = \frac{-i}{\sqrt{3}} \sum_j (\omega e^{ig_j \cdot x} - \omega^* e^{-ig_j \cdot x}), \quad \Phi_2^{(1,1)} \approx -2i\sqrt{3} \sum_j e^{-iq_j \cdot x}, \quad (\text{E9})$$

$$\Phi_1^{(0,2)} = - \sum_j (e^{ig_j \cdot x} + e^{-ig_j \cdot x}). \quad (\text{E10})$$

Now, Eq. (68) gives

$$v_{\text{Dirac}} \approx \frac{1 - 6(\gamma \tilde{\epsilon}_1^A)^2 - 3\alpha^2(1 - \frac{14}{3}\sqrt{3}\gamma \tilde{\epsilon}_1^A)}{1 + 3\alpha^2 + 6(\gamma \tilde{\epsilon}_1^A)^2}. \quad (\text{E11})$$

Based on Tables I and II,  $\gamma \tilde{\epsilon}_1^A \approx 0.06$ . Equation (E11) gives  $\alpha \approx 0.79$  when  $v_{\text{Dirac}} = 0$ . This value is very close to  $\alpha = 0.7857$ , the numerical result obtained in Sec. V.

[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, R. C. Ashoori, and P. Jarillo-Herrero, Correlated insulator behaviour at half-filling in magic-angle graphene superlattices, *Nature* **556**, 43 (2018).

[2] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Unconventional superconductivity in magic-angle graphene superlattices, *Nature* **556**, 80 (2018).

[3] A. L. Sharpe, E. J. Fox, A. W. Barnard, J. Finney, K. Watanabe, T. Taniguchi, M. A. Kastner, and D. Goldhaber-Gordon, Emergent ferromagnetism near three-quarters filling in twisted bilayer graphene, *Science* **365**, 605 (2019).

[4] M. Serlin, C. L. Tschirhart, H. Polshyn, Y. Zhang, J. Zhu, K. Watanabe, T. Taniguchi, L. Balents, and A. F. Young, Intrinsic quantized anomalous hall effect in a moire heterostructure, *Science* **367**, 900 (2019).

- [5] M. Yankowitz, S. Chen, H. Polshyn, Y. Zhang, K. Watanabe, T. Taniguchi, D. Graf, A. F. Young, and C. R. Dean, Tuning superconductivity in twisted bilayer graphene, *Science* **363**, 1059 (2019).
- [6] A. Kerelsky, L. J. McGilly, D. M. Kennes, L. Xian, M. Yankowitz, S. Chen, K. Watanabe, T. Taniguchi, J. Hone, C. Dean, A. Rubio, and A. N. Pasupathy, Maximized electron interactions at the magic angle in twisted bilayer graphene, *Nature* **572**, 95 (2019).
- [7] Y. Xie, B. Lian, B. Jack, X. Liu, C.-L. Chiu, K. Watanabe, T. Taniguchi, B. A. Bernevig, and A. Yazdani, Spectroscopic signatures of many body correlations in magic-angle twisted bilayer graphene, *Nature* **572**, 101 (2019).
- [8] S. L. Tomarken, Y. Cao, A. Demir, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, and R. C. Ashoori, Electronic compressibility of magic-angle graphene superlattices, *Phys. Rev. Lett.* **123**, 046601 (2019).
- [9] Y. Jiang, X. Lai, K. Watanabe, T. Taniguchi, K. Haule, J. Mao, and E. Y. Andrei, Charge order and broken rotational symmetry in magic-angle twisted bilayer graphene, *Nature* **573**, 91 (2019).
- [10] X. Lu, P. Stepanov, W. Yang, M. Xie, M. A. Aamir, I. Das, C. Urgell, K. Watanabe, T. Taniguchi, G. Zhang, A. Bachtold, A. H. MacDonald, and D. K. Efetov, Superconductors, orbital magnets and correlated states in magic-angle bilayer graphene, *Nature* **574**, 653 (2019).
- [11] Y. Choi, J. Kemmer, Y. Peng, A. Thomson, H. Arora, R. Polski, Y. Zhang, H. Ren, J. Alicea, G. Refael, F. von Oppen, K. Watanabe, T. Taniguchi, and S. Nadj-Perge, Electronic correlations in twisted bilayer graphene near the magic angle, *Nat. Phys.* **15**, 1174 (2019).
- [12] A. Uri, S. Grover, Y. Cao, J. A. Crosse, K. Bagani, D. Rodan-Legrain, Y. Myasoedov, K. Watanabe, T. Taniguchi, P. Moon, M. Koshino, P. Jarillo-Herrero, and E. Zeldov, Mapping the twist-angle disorder and Landau levels in magic-angle graphene, *Nature* **581**, 47 (2020).
- [13] D. Wong, K. P. Nuckolls, M. Oh, B. Lian, Y. Xie, S. Jeon, K. Watanabe, T. Taniguchi, B. A. Bernevig, and A. Yazdani, Cascade of electronic transitions in magic-angle twisted bilayer graphene, *Nature* **582**, 198 (2020).
- [14] U. Zondiner, A. Rozen, D. Rodan-Legrain, Y. Cao, R. Queiroz, T. Taniguchi, K. Watanabe, Y. Oreg, F. von Oppen, A. Stern, E. Berg, P. Jarillo-Herrero, and S. Ilani, Cascade of phase transitions and Dirac revivals in magic angle graphene, *Nature* **582**, 203 (2020).
- [15] P. Stepanov, I. Das, X. Lu, A. Fahimniya, K. Watanabe, T. Taniguchi, F. H. L. Koppens, J. Lischner, L. Levitov, and D. K. Efetov, Untying the insulating and superconducting orders in magic-angle graphene, *Nature* **583**, 375 (2020).
- [16] Y. Saito, J. Ge, K. Watanabe, T. Taniguchi, and A. F. Young, Independent superconductors and correlated insulators in twisted bilayer graphene, *Nat. Phys.* **16**, 926 (2020).
- [17] X. Liu, Z. Wang, K. Watanabe, T. Taniguchi, O. Vafeek, and J. I. A. Li, Tuning electron correlation in magic-angle twisted bilayer graphene using Coulomb screening, *Science* **371**, 1261 (2021).
- [18] Y. Cao, D. Rodan-Legrain, J. M. Park, F. N. Yuan, K. Watanabe, T. Taniguchi, R. M. Fernandes, L. Fu, and P. Jarillo-Herrero, Nematicity and competing orders in superconducting magic-angle graphene, *Science* **372**, 264 (2021).
- [19] C. L. Tschirhart, M. Serlin, H. Polshyn, A. Shragai, Z. Xia, J. Zhu, Y. Zhang, K. Watanabe, T. Taniguchi, M. E. Huber, A. F. Young, Imaging orbital ferromagnetism in a moire Chern insulator, *Science* **372**, 1323 (2021).
- [20] J. M. Park, Y. Cao, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, Flavour Hund's coupling, correlated Chern gaps, and diffusivity in Moiré flat bands, *Nature* **592**, 43 (2021).
- [21] A. Rozen, J. M. Park, U. Zondiner, Y. Cao, D. Rodan-Legrain, T. Taniguchi, K. Watanabe, Y. Oreg, A. Stern, E. Berg, P. Jarillo-Herrero, and S. Ilani, Entropic evidence for a Pomeranchuk effect in magic angle graphene, *Nature* **592**, 214 (2021).
- [22] Y. Saito, F. Yang, J. Ge, X. Liu, K. Watanabe, T. Taniguchi, J. I. A. Li, E. Berg, and A. F. Young, Isospin Pomeranchuk effect and the entropy of collective excitations in twisted bilayer graphene, *Nature* **592**, 220 (2021).
- [23] M. Oh, K. P. Nuckolls, D. Wong, R. L. Lee, X. Liu, K. Watanabe, T. Taniguchi, and A. Yazdani, Evidence for unconventional superconductivity in twisted bilayer graphene, *Nature* **600**, 240 (2021).
- [24] Y. Xie, A. T. Pierce, J. M. Park, D. E. Parker, E. Khalaf, P. Ledwith, Y. Cao, S. H. Lee, S. Chen, P. R. Forrester, K. Watanabe, T. Taniguchi, A. Vishwanath, P. Jarillo-Herrero, and A. Yacoby, Fractional Chern insulators in magic-angle twisted bilayer graphene, *Nature* **600**, 439 (2021).
- [25] A. T. Pierce, Y. Xie, J. M. Park, E. Khalaf, S. Hwan Lee, Y. Cao, D. E. Parker, P. R. Forrester, S. Chen, K. Watanabe, T. Taniguchi, A. Vishwanath, P. Jarillo-Herrero, and A. Yacoby, Unconventional sequence of correlated Chern insulators in magic-angle twisted bilayer graphene, *Nat. Phys.* **17**, 1210 (2021).
- [26] J.-X. Lin, Y.-H. Zhang, E. Morissette, Z. Wang, S. Liu, D. Rhodes, K. Watanabe, T. Taniguchi, J. Hone, and J. I. A. Li, Spin-orbit driven ferromagnetism at half moiré filling in magic-angle twisted bilayer graphene, *Science* **375**, 437 (2022).
- [27] H. Polshyn, Y. Zhang, M. A. Kumar, T. Soejima, P. Ledwith, K. Watanabe, T. Taniguchi, A. Vishwanath, M. P. Zaletel, and A. F. Young, Topological charge density waves at half-integer filling of a moiré superlattice, *Nat. Phys.* **18**, 42 (2022).
- [28] M. Koshino, N. F. Q. Yuan, T. Koretsune, M. Ochi, K. Kuroki, and L. Fu, Maximally Localized Wannier Orbitals and the Extended Hubbard Model for Twisted Bilayer Graphene, *Phys. Rev. X* **8**, 031087 (2018).
- [29] J. Kang and O. Vafeek, Symmetry, Maximally Localized Wannier States, and a Low-Energy Model for Twisted Bilayer Graphene Narrow Bands, *Phys. Rev. X* **8**, 031088 (2018).
- [30] H. C. Po, L. Zou, A. Vishwanath, and T. Senthil, Origin of Mott Insulating Behavior and Superconductivity in Twisted Bilayer Graphene, *Phys. Rev. X* **8**, 031089 (2018).
- [31] F. Wu, A. H. MacDonald, and I. Martin, Theory of Phonon-Mediated Superconductivity in Twisted Bilayer Graphene, *Phys. Rev. Lett.* **121**, 257001 (2018).
- [32] F. Guinea and N. R. Walet, Electrostatic effects, band distortions, and superconductivity in twisted graphene bilayers, *Proc. Natl. Acad. Sci. USA* **115**, 13174 (2018).
- [33] L. Balents, General continuum model for twisted bilayer graphene and arbitrary smooth deformations, *SciPost Phys.* **7**, 048 (2019).

- [34] J. Ahn, S. Park, and B.-J. Yang, Failure of Nielsen-Ninomiya Theorem and Fragile Topology in Two-Dimensional Systems with Space-Time Inversion Symmetry: Application to Twisted Bilayer Graphene at Magic Angle, *Phys. Rev. X* **9**, 021013 (2019).
- [35] Z. Song, Z. Wang, W. Shi, G. Li, C. Fang, and B. A. Bernevig, All Magic Angles in Twisted Bilayer Graphene are Topological, *Phys. Rev. Lett.* **123**, 036401 (2019).
- [36] K. Hejazi, C. Liu, H. Shapourian, X. Chen, and L. Balents, Multiple topological transitions in twisted bilayer graphene near the first magic angle, *Phys. Rev. B* **99**, 035111 (2019).
- [37] J. Liu, J. Liu, and X. Dai, The pseudo-Landau-level representation of twisted bilayer graphene: band topology and the implications on the correlated insulating phase, *Phys. Rev. B* **99**, 155415 (2019).
- [38] G. Tarnopolsky, A. J. Kruchkov, and A. Vishwanath, Origin of Magic Angles in Twisted Bilayer Graphene, *Phys. Rev. Lett.* **122**, 106405 (2019).
- [39] J. Kang and O. Vafek, Strong Coupling Phases of Partially Filled Twisted Bilayer Graphene Narrow Bands, *Phys. Rev. Lett.* **122**, 246401 (2019).
- [40] Y.-H. Zhang, D. Mao, Y. Cao, P. Jarillo-Herrero, and T. Senthil, Nearly flat Chern bands in moire superlattices, *Phys. Rev. B* **99**, 075127 (2019).
- [41] M. Xie, and A. H. MacDonald, Nature of the Correlated Insulator States in Twisted Bilayer Graphene, *Phys. Rev. Lett.* **124**, 097601 (2020).
- [42] N. Bultinck, S. Chatterjee, and M. P. Zaletel, Mechanism for Anomalous Hall Ferromagnetism in Twisted Bilayer Graphene, *Phys. Rev. Lett.* **124**, 166601 (2020).
- [43] N. Bultinck, E. Khalaf, S. Liu, S. Chatterjee, A. Vishwanath, and M. P. Zaletel, Ground State and Hidden Symmetry of Magic Angle Graphene at Even Integer Filling, *Phys. Rev. X* **10**, 031034 (2020).
- [44] J. Kang and O. Vafek, Non-Abelian Dirac node braiding and near-degeneracy of correlated phases at odd integer filling in magic angle twisted bilayer graphene, *Phys. Rev. B* **102**, 035161 (2020).
- [45] T. Soejima, D. E. Parker, N. Bultinck, J. Hauschild, and M. P. Zaletel, Efficient simulation of moire materials using the density matrix renormalization group, *Phys. Rev. B* **102**, 205111 (2020).
- [46] F. Wu and S. Das Sarma, Collective Excitations of Quantum Anomalous Hall Ferromagnets in Twisted Bilayer Graphene, *Phys. Rev. Lett.* **124**, 046403 (2020).
- [47] S. Liu, E. Khalaf, J. Y. Lee, and A. Vishwanath, Nematic topological semimetal and insulator in magic angle bilayer graphene at charge neutrality, *Phys. Rev. Res.* **3**, 013033 (2021).
- [48] Y. Alavirad and J. D. Sau, Ferromagnetism and its stability from the one-magnon spectrum in twisted bilayer graphene, *Phys. Rev. B* **102**, 235123 (2020).
- [49] J. Liu and X. Dai, Theories for the correlated insulating states and quantum anomalous Hall effect phenomena in twisted bilayer graphene, *Phys. Rev. B* **103**, 035427 (2021).
- [50] Y. H. Kwan, G. Wagner, T. Soejima, M. P. Zaletel, S. H. Simon, S. A. Parameswaran, and Nick Bultinck, Kekulé Spiral Order at All Nonzero Integer Fillings in Twisted Bilayer Graphene, *Phys. Rev. X* **11**, 041063 (2021).
- [51] Y. Zhang, K. Jiang, Z. Wang, and F. C. Zhang, Correlated insulating phases of twisted bilayer graphene at commensurate filling fractions: A Hartree-Fock study, *Phys. Rev. B* **102**, 035136 (2020).
- [52] T. Cea and F. Guinea, Band structure and insulating states driven by Coulomb interaction in twisted bilayer graphene, *Phys. Rev. B* **102**, 045107 (2020).
- [53] E. Khalaf, N. Bultinck, A. Vishwanath, and M. P. Zaletel, Soft modes in magic angle twisted bilayer graphene, [arXiv:2009.14827](https://arxiv.org/abs/2009.14827).
- [54] D. E. Parker, T. Soejima, J. Hauschild, M. P. Zaletel, and N. Bultinck, Strain-Induced Quantum Phase Transitions in Magic Angle Graphene, *Phys. Rev. Lett.* **127**, 027601 (2021).
- [55] R. M. Fernandes and J. W. F. Venderbos, Nematicity with a twist: rotational symmetry breaking in a moiré superlattice, *Sci. Adv.* **6**, eaba8834 (2020).
- [56] X. Wang and O. Vafek, Narrow bands in magnetic field and strong-coupling Hofstadter spectra, *Phys. Rev. B* **106**, L121111 (2022).
- [57] D. Parker, P. Ledwith, E. Khalaf, T. Soejima, J. Hauschild, Y. Xie, A. Pierce, M. P. Zaletel, A. Yacoby, and A. Vishwanath, Field-tuned and zero-field fractional Chern insulators in magic angle graphene, [arXiv:2112.13837](https://arxiv.org/abs/2112.13837).
- [58] B. A. Bernevig, Z. D. Song, N. Regnault, and B. Lian, Twisted bilayer graphene. I. Matrix elements, approximations, perturbation theory, and a two-band model, *Phys. Rev. B* **103**, 205411 (2021); Z. D. Song, B. Lian, N. Regnault, and B. A. Bernevig, Twisted bilayer graphene. II. Stable symmetry anomaly, *ibid.* **103**, 205412 (2021); B. A. Bernevig, Z. D. Song, N. Regnault, and B. Lian, Twisted bilayer graphene. III. Interacting Hamiltonian and exact symmetries, *ibid.* **103**, 205413 (2021); B. Lian, Z. D. Song, N. Regnault, D. K. Efetov, A. Yazdani, and B. A. Bernevig, Twisted bilayer graphene. IV. Exact insulator ground states and phase diagram, *ibid.* **103**, 205414 (2021); B. A. Bernevig, B. Lian, A. Cowsik, F. Xie, N. Regnault, and Z. D. Song, Twisted bilayer graphene. V. Exact analytic many-body excitations in Coulomb Hamiltonians: Charge gap, Goldstone modes, and absence of Cooper pairing, *ibid.* **103**, 205415 (2021); F. Xie, A. Cowsik, Z. D. Song, B. Lian, B. A. Bernevig, and N. Regnault, Twisted bilayer graphene. VI. An exact diagonalization study at nonzero integer filling, *ibid.* **103**, 205416 (2021).
- [59] Z.-D. Song and B. A. Bernevig, Magic-Angle Twisted Bilayer Graphene as a Topological Heavy Fermion Problem, *Phys. Rev. Lett.* **129**, 047601 (2022).
- [60] R. Bistritzer and A. H. MacDonald, Moire bands in twisted double-layer graphene, *Proc. Natl. Acad. Sci. USA* **108**, 12233 (2011).
- [61] O. Vafek and J. Kang, Continuum effective Hamiltonian for graphene bilayers for an arbitrary smooth lattice deformation from microscopic theories, *Phys. Rev. B* **107**, 075123 (2023).
- [62] O. Vafek and J. Kang, Renormalization group study of hidden symmetry in twisted bilayer graphene with coulomb interactions, *Phys. Rev. Lett.* **125**, 257602 (2020).
- [63] N. N. T. Nam and M. Koshino, Lattice relaxation and energy band modulation in twisted bilayer graphene, *Phys. Rev. B* **96**, 075311 (2017); Erratum: Lattice relaxation and energy band modulation in twisted bilayer graphene [Phys. Rev. B **96**, 075311 (2017)], **101**, 099901(E) (2020).

- [64] S. Carr, D. Massatt, S. B. Torrisi, P. Cazeaux, M. Luskin, and E. Kaxiras, Relaxation and domain formation in incommensurate two-dimensional heterostructures, *Phys. Rev. B* **98**, 224102 (2018).
- [65] P. Moon and M. Koshino, Energy spectrum and quantum Hall effect in twisted bilayer graphene, *Phys. Rev. B* **85**, 195458 (2012).
- [66] S. Fang and E. Kaxiras, Electronic structure theory of weakly interacting bilayers, *Phys. Rev. B* **93**, 235153 (2016).
- [67] S. S. Sunku, G. X. Ni, B.-Y. Jiang, H. Yoo, A. Sternbach, A. S. McLeod, T. Stauber, L. Xiong, T. Taniguchi, K. Watanabe, P. Kim, M. M. Fogler, and D. N. Basov, Photonic crystals for nano-light in moiré graphene superlattices, *Science* **362**, 1153 (2018).
- [68] H. Ochoa and R. M. Fernandes, Degradation of Phonons in Disordered Moiré Superlattices, *Phys. Rev. Lett.* **128**, 065901 (2022).
- [69] N. P. Kazmierczak, M. Van Winkle, C. Ophus, K. C. Bustillo, S. Carr, H. G. Brown, J. Ciston, T. Taniguchi, K. Watanabe, and D. K. Bediako, Strain fields in twisted bilayer graphene, *Nat. Mater.* **20**, 956 (2021).
- [70] F. de Juan, M. Sturla, and M. A. H. Vozmediano, Space Dependent Fermi Velocity in Strained Graphene, *Phys. Rev. Lett.* **108**, 227205 (2012).
- [71] See Appendices for more detailed derivations of the interlayer Hamiltonian  $H_{\text{intra}}$ , the interlayer tunneling, and the approximate formula of  $v_{\text{Dirac}}$  when the relaxation-induced pseudomagnetic field is included. Note that the numerical values of the coupling constants of  $\delta H_{\text{intra}}$ , the contact and gradient couplings of the interlayer tunnelings with  $qs$  in the fourth and fifth shells are also listed in the Appendices.
- [72] F. Guinea and N. R. Walet, Continuum models for twisted bilayer graphene: Effect of lattice deformation and hopping parameters, *Phys. Rev. B* **99**, 205134 (2019).
- [73] M. Koshino and N. N. T. Nam, Continuum model for relaxed twisted bilayer graphenes and moiré electron-phonon interaction, *Phys. Rev. B* **101**, 195425 (2020).
- [74] Z. Bi, N. F. Q. Yuan, and L. Fu, Designing flat bands by strain, *Phys. Rev. B* **100**, 035448 (2019).
- [75] S. Fang, S. Carr, Z. Zhu, D. Massatt, and E. Kaxiras, Angle-dependent *ab initio* low-energy Hamiltonians for a relaxed twisted bilayer graphene heterostructure, [arXiv:1908.00058](https://arxiv.org/abs/1908.00058).