Editors' Suggestion

Magic-angle twisted symmetric trilayer graphene as a topological heavy-fermion problem

Jiabin Yu,[1,2] Ming Xie,[1] B. Andrei Bernevig,[2,3,4] and Sankar Das Sarma[1]
[1]*Condensed Matter Theory Center and Joint Quantum Institute, Department of Physics, University of Maryland, College Park, Maryland 20742, USA*
[2]*Department of Physics, Princeton University, Princeton, New Jersey 08544, USA*
[3]*Donostia International Physics Center, P. Manuel de Lardizabal 4, 20018 Donostia-San Sebastian, Spain*
[4]*IKERBASQUE, Basque Foundation for Science, Bilbao, Spain*

(Received 6 March 2023; revised 5 June 2023; accepted 5 June 2023; published 17 July 2023)

Recently, Song and Bernevig [Phys. Rev. Lett. **129**, 047601 (2022)] reformulated magic-angle twisted bilayer graphene as a topological heavy fermion problem, and used this reformulation to provide a deeper understanding for the correlated phases at integer fillings. In this work, we generalize this heavy-fermion paradigm to magic-angle twisted symmetric trilayer graphene, and propose a low-energy f-c-d model that reformulates magic-angle twisted symmetric trilayer graphene as heavy localized f modes coupled to itinerant topological semimetalic c modes and itinerant Dirac d modes. Our f-c-d model well reproduces the single-particle band structure of magic-angle twisted symmetric trilayer graphene at low energies for displacement field $\mathcal{E} \in [0, 300]$ meV. By performing Hartree-Fock calculations with the f-c-d model for $\nu = 0, -1, -2$ electrons per Moiré unit cell, we reproduce all the correlated ground states obtained from the previous numerical Hartree-Fock calculations with the Bistritzer-MacDonald-type model, and we find additional new correlated ground states at high displacement field. Based on the numerical results, we propose a simple rule for the ground states at high displacement fields by using the f-c-d model, and provide analytical derivation for the rule at charge neutrality. We also provide analytical symmetry arguments for the (nearly) degenerate energies of the high-$\mathcal{E}$ ground states at all the integer fillings of interest, and make experimental predictions of which charge-neutral states are stabilized in magnetic fields. Our f-c-d model provides a new perspective for understanding the correlated phenomena in magic-angle twisted symmetric trilayer graphene, suggesting that the heavy fermion paradigm of Song and Bernevig [Phys. Rev. Lett. **129**, 047601 (2022)] should be the generic underpinning of correlated physics in multilayer moire graphene structures.

DOI: 10.1103/PhysRevB.108.035129

I. INTRODUCTION

Magic-angle twisted bilayer graphene (MATBG) [1] hosts superconductivity [2–12] and various other interaction-induced phenomena [13–34]. In the last several years, models have been constructed (in the real space [35–43], in the momentum space [44–51], or phenomenologically [52–59]) to understand the physics observed in MATBG, among other research efforts [60–124]. Recently a physically relevant and symmetry-preserving model that separates the correct energy scales and is convenient for studying the correlated phenomena was proposed in Ref. [125]. It is called the topological heavy-fermion model. At the single-particle level, the model proposed in Ref. [125] consists of localized heavy f modes (of $p_x \pm ip_y$ symmetry) and itinerant c modes, where the nearly flat bands in MATBG are given by coupling f and c modes (mainly around Γ_M). The model is topological because the c modes are anomalous in one valley (when the normal-state particle-hole symmetry is imposed exactly) and have a double-vortex dispersion akin to that in one of the valleys of *untwisted* bilayer graphene, but at the Γ_M point. Using the topological heavy-fermion model, Ref. [125] finds that the filling of the system is governed by the heavy fermions, which in a Hartree-Fock calculation polarize. The Hartree-Fock calculation can be done efficiently for the correlated states at integer fillings, and a simple rule for the stability of the correlated ground-states can be derived analytically for those correlated states. Furthermore, the hope is that, using the differentiation of degrees of freedom in local and itinerant, progress can be made in the hard physics at non-integer filling, as well as at nonzero temperature. Recently, the heavy-fermion picture has been used to construct Kondo lattice model in MATBG [126–129], and has been generalized to twisted $(M + N)$-layer graphenes [130] and to a variant of the kagome lattice [131].

Motivated by Ref. [125], in this work, we generalize the topological heavy-fermion picture to magic-angle twisted symmetric trilayer graphene (MATSTG) [132–166], which has also been experimentally confirmed to host correlated insulating states and superconductivity [167–175]. Specifically, we first follow Ref. [125] to construct the heavy f and itinerant c modes, and then generalize the framework to include the nonzero displacement field $\mathcal{E}$, which couples f and c electrons to the relativistic Dirac (d) modes. The resultant single-particle f-c-d model can reproduce almost identically the band structure of the Bistritzer-MacDonald-type (BM-type) model [142] in the energy window $[-50$ meV, 50 meV$]$ and for displacement field $\mathcal{E} \in [0, 300]$ meV. We find that the f modes dominate the low-energy single-particle physics for $\mathcal{E} \in [0, 300]$ meV.

The interaction in the f-c-d model is obtained by projecting the Coulomb interaction to the f-c-d basis. Using this model, we perform the self-consistent Hartree-Fock calculation for the correlated states at fillings $\nu = 0, -1, -2$ per Moiré unit cell. The numerical results of our Hartree-Fock calculation are generally consistent with the previous numerical results in Refs. [146,149,151,152,156,175], where a phase transition to states with zero intervalley coherence at all $\nu = 0, -1, -2$ fillings exists when increasing the displacement field. Nevertheless, we find *more* additional correlated ground states than found in the previous literature at high displacement fields. We further perform analytical one-shot Hartree-Fock analysis at the considered integer fillings. At $\nu = 0$, we provide analytical understanding of the loss of intervalley coherence for high displacement field, and derive a simple rule for the ground states at high fields. The same rule is also derived for $\nu = -1, -2$ under an unrealistic approximation, but the rule turns out to be consistent with the self-consistent calculation for $\nu = -1, -2$. We also find a symmetry reason for the similar energies of the ground states at high displacement fields at all $\nu = 0, -1, -2$. Finally, we discuss the experimental implication of our results.

The rest of the paper is organized as follows. In Sec. II, we review the BM-type model for MATSTG. In Sec. III, we build the heavy fermion f-c-d model for MATSTG. In Sec. IV, we perform numerical Hartree-Fock calculations with the f-c-d model for $\nu = 0, -1, -2$. In Sec. V, we perform analytical one-shot Hartree-Fock analysis for the correlated states with $\nu = 0, -1, -2$. In Sec. VI, we conclude the paper and discuss the experimental predictions. A series of appendices provide all the technical details of our theory.

II. REVIEW: INTERACTING BM-TYPE MODEL FOR MATSTG

In this section, we review the interacting BM-type model of MATSTG, which has been theoretically studied in Refs. [132–161]. Here the interaction is the Coulomb interaction screened by a top gate and a bottom gate, where the sample is placed in the middle of the two gates. We will only review the contents that are essential for our later discussions and are specific to our theory presented in this work; a more complete and detailed discussion can be found in Ref. [142].

A. Single-Particle BM-type model

In this part, we review the BM-type model for MATSTG following Refs. [132,142].

MATSTG is constructed from a AAA-stacking trilayer graphene by rotating the graphene layers alternatively, i.e., rotating the top ($l = 3$) and bottom ($l = 1$) layers by $-\theta/2$ and rotating middle ($l = 2$) layer by $\theta/2$, where $\theta > 0$ corresponds to the counterclockwise rotation and $l = 1,\ 2,\ 3$ is the layer index. We label the lattice constant and the Fermi velocity of the monolayer graphene as $a_G = 2.46$ Å and $v_0 = 5944$ meV Å, respectively. We refer to the unit system in which Å is the length unit and meV is the energy unit as the experimental unit system (EUS), since this unit system is convenient for the comparison to the experiments. However, EUS is not the most convenient unit system for the theoretical study of MATSTG. The most convenient unit system is the following simplified unit system in which

$$\hbar = 1, \quad \epsilon_0 = 1, \quad k_\theta = 1, \quad v_0 = 1, \tag{1}$$

where $k_\theta = \frac{4\pi}{3a_G} 2\sin(\frac{\theta}{2})$ and ϵ_0 is the vacuum permittivity. Throughout the entire work, we will use Eq. (1) unless otherwise (e.g., EUS) is specified.

With the unit system specified by Eq. (1), the single-particle BM-type model for MATSTG reads

$$H_0 = H_{0,+} + H_{0,-}\,. \tag{2}$$

Here "+" and "−" label two graphene valleys, which are related by time-reversal (TR) symmetry as

$$H_{0,-} = \mathcal{T} H_{0,+} \mathcal{T}^{-1}\,. \tag{3}$$

Specifically, $H_{0,+}$ reads

$$H_{0,+} = \int d^2r\, (\psi^\dagger_{+,\boldsymbol{r},1} \quad \psi^\dagger_{+,\boldsymbol{r},2} \quad \psi^\dagger_{+,\boldsymbol{r},3}) \begin{pmatrix} -i\boldsymbol{\sigma}\cdot\boldsymbol{\nabla} - \frac{\mathcal{E}}{2} & T(\boldsymbol{r}) & \\ T^\dagger(\boldsymbol{r}) & -i\boldsymbol{\sigma}\cdot\boldsymbol{\nabla} & T^\dagger(\boldsymbol{r}) \\ & T(\boldsymbol{r}) & -i\boldsymbol{\sigma}\cdot\boldsymbol{\nabla} + \frac{\mathcal{E}}{2} \end{pmatrix} \otimes s_0 \begin{pmatrix} \psi_{+,\boldsymbol{r},1} \\ \psi_{+,\boldsymbol{r},2} \\ \psi_{+,\boldsymbol{r},3} \end{pmatrix}, \tag{4}$$

where $\psi^\dagger_{+,\boldsymbol{r},l} = (\psi^\dagger_{+,\boldsymbol{r},l,A,\uparrow}, \psi^\dagger_{+,\boldsymbol{r},l,A,\downarrow}, \psi^\dagger_{+,\boldsymbol{r},l,B,\uparrow}, \psi^\dagger_{+,\boldsymbol{r},l,B,\downarrow})$ is the vector of creation operators for the + valley and the lth layer, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y)$, and $\sigma_{0,x,y,z}$ and $s_{0,x,y,z}$ label the Pauli matrices for the sublattice index $\sigma = A/B$ and the spin index $s = \uparrow/\downarrow$, respectively. The expression of $H_{0,-}$ can be obtained via

$$\mathcal{T}\psi^\dagger_{+,\boldsymbol{r},l}\mathcal{T}^{-1} = \psi^\dagger_{-,\boldsymbol{r},l}\sigma_0 i s_y. \tag{5}$$

In Eq. (4), we assume that the twist angle θ is small enough such that the kinetic terms of order $O(\theta)$ can be safely neglected. Moreover, $T(\boldsymbol{r}) = \sum_{j=1,2,3} T_j e^{i\boldsymbol{r}\cdot\boldsymbol{q}_j}$ stands for the interlayer hopping between neighboring layers with

$$\begin{aligned} \boldsymbol{q}_1 &= (0, 1)^T, \\ \boldsymbol{q}_2 &= \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)^T, \\ \boldsymbol{q}_3 &= \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)^T, \end{aligned} \tag{6}$$

and

$$T_j = w_0\sigma_0 + w_1\left[\cos\left(\frac{2\pi}{3}(j-1)\right)\sigma_x + \sin\left(\frac{2\pi}{3}(j-1)\right)\sigma_y\right]. \tag{7}$$

Here w_0 and w_1 are the AA and AB interlayer tunnellings, respectively, with $w_0 = 88$ meV and $w_1 = 110$ meV in EUS. The values of $w_{0,1}$ in the unit system specified by Eq. (1) depend on θ.

In Eq. (4), $\mathcal{E}$ is the energy difference generated by the displacement field, i.e., the external electric field perpendicular to MATSTG. When the displacement field is zero ($\mathcal{E} = 0$), H_0 has a mirror symmetry m_z with mirror plane lying in the middle layer, which is represented as

$$m_z \psi^\dagger_{\eta,\boldsymbol{r},l} m_z^{-1} = -\psi^\dagger_{\eta,\boldsymbol{r},4-l} \tag{8}$$

with $\eta = \pm$ the graphene valley index. Here the extra minus sign comes from the fact that $\psi^\dagger_{\eta,\boldsymbol{r},l}$ are constructed from the p_z orbital of graphene. In fact, MATSTG is called symmetric owing to the presence of m_z symmetry for $\mathcal{E} = 0$. m_z allows us to recombine $\psi^\dagger_{\eta,\boldsymbol{r},l}$ into a m_z-odd sector

$$\begin{aligned}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},t} &= \frac{1}{\sqrt{2}}(\psi^\dagger_{\eta,\boldsymbol{r},3} + \psi^\dagger_{\eta,\boldsymbol{r},1}) \\ \widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},b} &= \psi^\dagger_{\eta,\boldsymbol{r},2}\end{aligned} \tag{9}$$

and a m_z-even sector

$$d^\dagger_{\eta,\boldsymbol{r}} = \frac{1}{\sqrt{2}}(\psi^\dagger_{\eta,\boldsymbol{r},3} - \psi^\dagger_{\eta,\boldsymbol{r},1})\,. \tag{10}$$

With the recombination, $H_{0,\eta}$ can be split into three parts

$$H_{0,\eta} = H_{0,\mathrm{TBG},\eta} + H_{0,D,\eta} + H_{0,\mathcal{E},\eta}\,, \tag{11}$$

where

$$H_{0,\mathrm{TBG},+} = \int d^2r\ \widetilde{\psi}^\dagger_{+,\boldsymbol{r}} \begin{pmatrix} -\mathrm{i}\boldsymbol{\sigma}\cdot\nabla & \sqrt{2}T(\boldsymbol{r}) \\ \sqrt{2}T^\dagger(\boldsymbol{r}) & -\mathrm{i}\boldsymbol{\sigma}\cdot\nabla \end{pmatrix} \otimes s_0\ \widetilde{\psi}_{+,\boldsymbol{r}} \tag{12}$$

is equivalent to the (valley +) BM model of the ordinary TBG with $w_0 \to \sqrt{2}w_0$ and $w_1 \to \sqrt{2}w_1$,

$$H_{0,D,+} = \int d^2r\ d^\dagger_{+,\boldsymbol{r}}(-\mathrm{i})\boldsymbol{\sigma}\cdot\nabla d_{+,\boldsymbol{r}} \tag{13}$$

is just a Dirac cone, the displacement field term

$$H_{0,\mathcal{E},+} = \int d^2r\ \frac{\mathcal{E}}{2}\widetilde{\psi}^\dagger_{+,\boldsymbol{r},t} d_{+,\boldsymbol{r}} + \mathrm{H.c.} \tag{14}$$

becomes the coupling between the TBG modes and the Dirac modes, and

$$\begin{aligned}H_{0,\mathrm{TBG},-} &= \mathcal{T} H_{0,\mathrm{TBG},+}\mathcal{T}^{-1}, \\ H_{0,D,-} &= \mathcal{T} H_{0,D,+}\mathcal{T}^{-1}, \\ H_{0,\mathcal{E},-} &= \mathcal{T} H_{0,\mathcal{E},+}\mathcal{T}^{-1}\,.\end{aligned} \tag{15}$$

It is clear that $H_{0,\mathrm{TBG},\eta}$ and $H_{0,D,\eta}$ commute with m_z, while $H_{0,\mathcal{E},\eta}$ anticommutes with m_z.

H_0 has Moiré lattice translation symmetry, which is represented as

$$\begin{aligned}T_{\boldsymbol{R}}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\widetilde{l}}T^{-1}_{\boldsymbol{R}} &= \widetilde{\psi}^\dagger_{\eta,\boldsymbol{r}+\boldsymbol{R},\widetilde{l}}\, e^{-\eta\mathrm{i}\boldsymbol{K}_{\widetilde{l}}\cdot\boldsymbol{R}}, \\ T_{\boldsymbol{R}} d^\dagger_{\eta,\boldsymbol{r}} T^{-1}_{\boldsymbol{R}} &= d^\dagger_{\eta,\boldsymbol{r}+\boldsymbol{R}}\, e^{-\eta\mathrm{i}\boldsymbol{K}_t\cdot\boldsymbol{R}}\,,\end{aligned} \tag{16}$$

where $\widetilde{l} = t, b$ labels the "layer" of the TBG part, and $\boldsymbol{R}$ is the Moiré lattice vector with two primitive Moiré vectors being $\boldsymbol{a}_{\mathrm{M},1} = \frac{4\pi}{3}(0,-1)^T$ and $\boldsymbol{a}_{\mathrm{M},2} = \frac{4\pi}{3}(\frac{\sqrt{3}}{2}, \frac{1}{2})^T$. In particular, $\boldsymbol{K}_t$ and $\boldsymbol{K}_b$ in Eq. (16) arise from the graphene valley as shown in Appendix A, which read

$$\boldsymbol{K}_t = \tfrac{1}{2}(\cot(\theta/2), -1)^T\,, \quad \boldsymbol{K}_b = \tfrac{1}{2}(\cot(\theta/2), 1)^T\,. \tag{17}$$

To exploit the Moiré lattice translational symmetry of H_0, it is better to transform the Hamiltonian to the momentum space. To do so, we first transform the basis to the momentum space as

$$\begin{aligned}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{p},\widetilde{l}} &= \frac{1}{\sqrt{\mathcal{A}}}\int d^2r e^{\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\widetilde{l}}, \\ d^\dagger_{\eta,\boldsymbol{p}} &= \frac{1}{\sqrt{\mathcal{A}}}\int d^2r e^{\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}} d^\dagger_{\eta,\boldsymbol{r}}\,,\end{aligned} \tag{18}$$

where $\boldsymbol{p} \in \mathbb{R}^2$ and $\mathcal{A}$ is the area of MATSTG. Then, we define

$$\mathcal{Q} = \mathcal{Q}_+ \cup \mathcal{Q}_-\,, \quad \mathcal{Q}_\pm = \boldsymbol{b}_{\mathrm{M},1}\mathbb{Z} + \boldsymbol{b}_{\mathrm{M},2}\mathbb{Z} \pm \boldsymbol{q}_1 \tag{19}$$

with

$$\begin{aligned}\boldsymbol{b}_{\mathrm{M},1} &= \boldsymbol{q}_3 - \boldsymbol{q}_1 = \left(\frac{\sqrt{3}}{2}, -\frac{3}{2}\right)^T, \\ \boldsymbol{b}_{\mathrm{M},2} &= \boldsymbol{q}_3 - \boldsymbol{q}_2 = (\sqrt{3}, 0)^T,\end{aligned} \tag{20}$$

forming the basis of the Moiré reciprocal lattice. Finally, we define

$$\begin{aligned}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q}} &= \widetilde{\psi}^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{Q},\widetilde{l}^\eta_{\boldsymbol{Q}}} \quad \text{with} \quad \boldsymbol{Q} \in \mathcal{Q}, \\ d^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q}} &= d^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{Q}} \quad \text{with} \quad \boldsymbol{Q} \in \mathcal{Q}_\eta\end{aligned} \tag{21}$$

with $\widetilde{l}^\eta_{\boldsymbol{Q}} = t$ for $\boldsymbol{Q} \in \mathcal{Q}_\eta$ and $\widetilde{l}^\eta_{\boldsymbol{Q}} = b$ for $\boldsymbol{Q} \in \mathcal{Q}_{-\eta}$. With Eq. (21), the Hamiltonian becomes

$$\begin{aligned}H_{0,\mathrm{TBG},\eta} = &\sum_{\boldsymbol{k}\in\mathrm{MBZ}}\sum_{\boldsymbol{Q},\boldsymbol{Q}'\in\mathcal{Q}} \widetilde{\psi}^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q}} \\ &\times \left[h^D_{\eta,\boldsymbol{Q}}(\boldsymbol{k})\delta_{\boldsymbol{Q}\boldsymbol{Q}'} + \sqrt{2}h^I_{\eta,\boldsymbol{Q}\boldsymbol{Q}'}\right]s_0\widetilde{\psi}_{\eta,\boldsymbol{k},\boldsymbol{Q}'},\end{aligned} \tag{22}$$

$$H_{0,D,\eta} = \sum_{\boldsymbol{k}\in\mathrm{MBZ}}\sum_{\boldsymbol{Q}\in\mathcal{Q}_\eta} d^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q}} h^D_{\eta,\boldsymbol{Q}}(\boldsymbol{k}) s_0 d_{\eta,\boldsymbol{k},\boldsymbol{Q}}, \tag{23}$$

and

$$H_{0,\mathcal{E},\eta} = \sum_{\boldsymbol{k}\in\mathrm{MBZ}}\sum_{\boldsymbol{Q}\in\mathcal{Q}_\eta} \frac{\mathcal{E}}{2}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q}} d_{\eta,\boldsymbol{k},\boldsymbol{Q}} + \mathrm{H.c.} \tag{24}$$

Here $h^D_{+,\boldsymbol{Q}}(\boldsymbol{k}) = (\boldsymbol{k}-\boldsymbol{Q})\cdot\boldsymbol{\sigma}$, $h^I_{+,\boldsymbol{Q}\boldsymbol{Q}'} = \sum_j T_j(\delta_{\boldsymbol{Q},\boldsymbol{Q}'+\boldsymbol{q}_j} + \delta_{\boldsymbol{Q}',\boldsymbol{Q}+\boldsymbol{q}_j})$, $h^D_{-,\boldsymbol{Q}}(\boldsymbol{k}) = [h^D_{+,-\boldsymbol{Q}}(-\boldsymbol{k})]^*$, $h^I_{-,\boldsymbol{Q}\boldsymbol{Q}'} = [h^I_{+,(-\boldsymbol{Q})(-\boldsymbol{Q}')}]^*$, and MBZ is short for the Moiré Brillouin zone. In this work, all numerical calculations with H_0 are done in the momentum space by using Eqs. (22)–(24). The numerical band structure of H_0 in the + valley is shown in Figs. 1(a)–1(d) as red lines. The definitions of various high-symmetry points in MBZ are illustrated in Fig. 2.

At the end of this part, we list the symmetries of H_0 for generic $\mathcal{E}$. We have discussed TR and Moiré lattice translations, which are symmetries of H_0 for any values of $\mathcal{E}$. Beside these two, H_0 has spin-charge U(2) symmetry in each valley, the spinless threefold rotation symmetry C_3 along z, $C_2\mathcal{T}$ symmetry (the combination of the spinless twofold rotation C_2 along z and the TR operation), an effective unitary

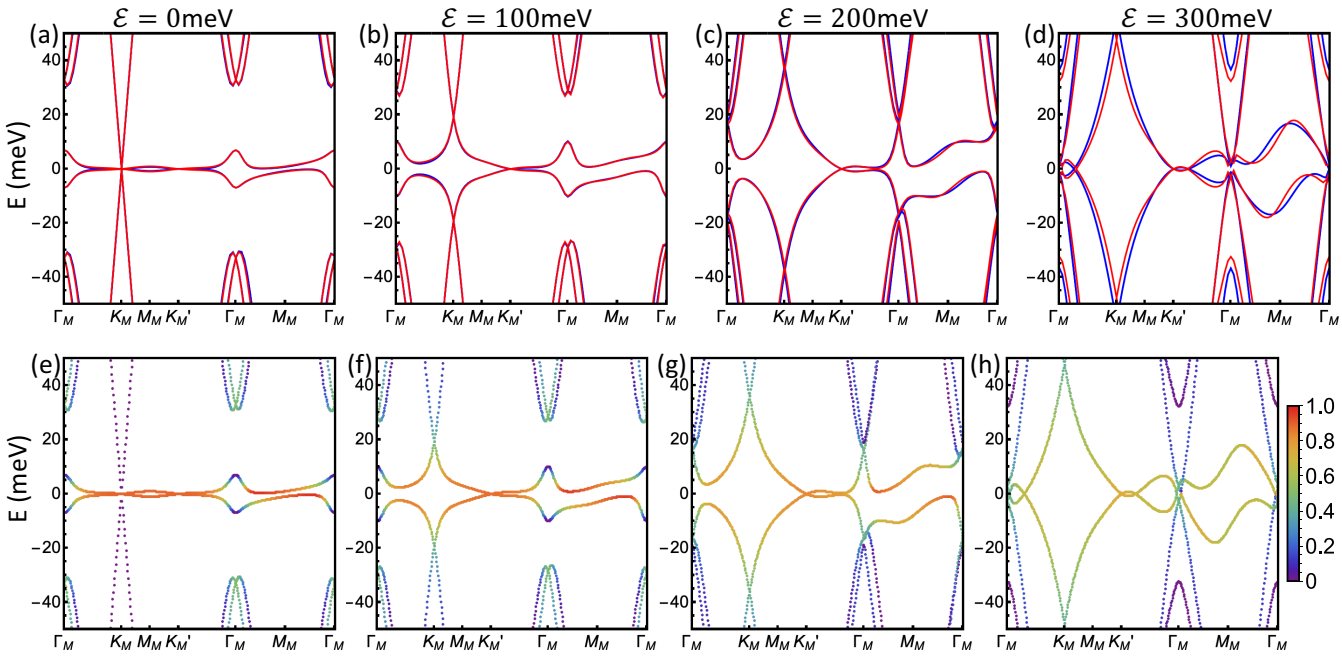

FIG. 1. The single-particle band structures of MATSTG in the + valley for the parameter values in Eq. (44) and Table I. In this figure, we use EUS. The momentum cutoffs of the BM-type model and the f-c-d model are $2\sqrt{7}$ and $2\sqrt{3}$, respectively. (See Appendix B for details on the f-c-d model.) In (a)–(d), we plot the band structure of the single-particle BM-type model Eq. (2) in red, and the band structure of the single-particle f-c-d model Eq. (54) in blue. $\mathcal{E}$ is the energy difference between the top and bottom layer generated by the displacement field. In (e)–(h), we replot the band structure of the single-particle BM-type model Eq. (2) in (a)–(d), respectively. The colors of the points show the (square of the absolute values of) overlaps between the Bloch states and the trial Wannier functions according to the color bar on the right of (h).

antisymmetry $C_{2x}P$, and the charge conjugate antisymmetry $\mathcal{C}$. Here antisymmetry means that the symmmetry operation anticommutes with the Hamiltonian, i.e., $C_{2x}PH_0(C_{2x}P)^{-1} = -H_0$ and $\mathcal{C}H_0\mathcal{C}^{-1} = -H_0$. (See the symmetry representations in Appendix A.)

B. Coulumb interaction

In this part, we review the Coulomb interaction in the BM-type model for MATSTG following Refs. [142,149].

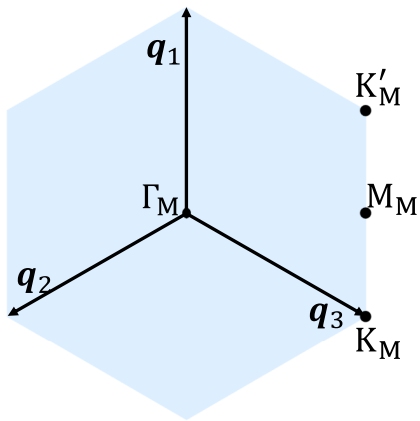

FIG. 2. This figure shows MBZ, as well as the $\boldsymbol{q}_{1,2,3}$ and various high-symmetry points. Note that K'_M is equivalent to $-K_M$.

The Coulomb interaction in MATSTG is screened by the top and bottom gates, which are parallel to the MATSTG sample. For simplicity, we assume that MATSTG lies in the middle of two gates, and then the Coulomb interaction between two electrons separated by $\boldsymbol{r}$ has the following form:

$$V(\boldsymbol{r}) = \frac{1}{\mathcal{A}} \sum_{\boldsymbol{p}} e^{-i\boldsymbol{p}\cdot\boldsymbol{r}} V(\boldsymbol{p}) \,, \tag{25}$$

where

$$V(\boldsymbol{p}) = \pi \xi^2 V_\xi \frac{\tanh(\xi|\boldsymbol{p}|/2)}{\xi|\boldsymbol{p}|/2} \,, \tag{26}$$

ξ is the distance between two gates, and $V_\xi = \frac{e^2}{4\pi\epsilon\xi}$ with e the elementary charge and ϵ the dielectric constant. Throughout this work, we choose

$$\xi = 100 \text{ Å} \quad \text{and} \quad V_\xi = 24 \text{ meV} \tag{27}$$

in EUS for all numerical calculations, unless specified otherwise. In Eq. (25), we have included the screening due to the two gates. It is clear that

$$V^*(\boldsymbol{r}) = V(\boldsymbol{r}) \quad \text{and} \quad V(g\boldsymbol{r}) = V(\boldsymbol{r}) \; \forall g \in \mathrm{O}(2) \,. \tag{28}$$

With the form of the Coulomb interaction [Eq. (25)], the Hamiltonian for the interaction reads

$$H_{\text{int}} = \frac{1}{2} \int d^2r d^2r' V(\boldsymbol{r} - \boldsymbol{r}') : \rho(\boldsymbol{r}) :: \rho(\boldsymbol{r}') : \,, \tag{29}$$

where $\rho(\boldsymbol{r}) = \sum_{\eta,l} \psi^\dagger_{\eta,\boldsymbol{r},l}\psi_{\eta,\boldsymbol{r},l}$ is the electron number density operator. The normal-ordering is defined as $:O:= O - \langle G_0|O|G_0\rangle$ with $|G_0\rangle$ chosen such that

$$\langle G_0|\psi^\dagger_{\eta,\boldsymbol{r},l,\sigma,s}\psi_{\eta',\boldsymbol{r}',l',\sigma',s'}|G_0\rangle = \tfrac{1}{2}\delta_{\eta\eta'}\delta(\boldsymbol{r}-\boldsymbol{r}')\delta_{ll'}\delta_{\sigma\sigma'}\delta_{ss'} \,. \tag{30}$$

The usage of the normal ordering is just to include a uniform positive charge background that makes half filling charge-neutral, as discussed in the following. Based on the form of the interaction [Eq. (29)], $(-e:\rho(\boldsymbol{r}):)$ should be the total charge density at $\boldsymbol{r}$. Since we know $-e:\rho(\boldsymbol{r}):= -e\rho(\boldsymbol{r}) + e\langle G_0|\rho(\boldsymbol{r})|G_0\rangle$ and $-e\rho(\boldsymbol{r})$ is the electron charge density, $e\langle G_0|\rho(\boldsymbol{r})|G_0\rangle$ should be the background charge density. Note that $e\langle G_0|\rho(\boldsymbol{r})|G_0\rangle = 12e\ \delta(\boldsymbol{r}=0) = \frac{e\sum_p 2\times3\times2\times2}{2A}$ is nothing but the charge density of a uniform positive charge background that corresponds to half filling, justifying the meaning of the normal ordering.

H_{int} is invariant under TR, C_3, $C_2\mathcal{T}$, m_z, Moiré lattice translations, $C_{2x}P$, $\mathcal{C}$, and U(2) × U(2). (See more details in Appendix A.)

C. Interacting BM-type Hamiltonian and filling

In this part, we review some general properties of the interacting BM-type model for MATSTG following Refs. [142,149].

The interacting BM-type Hamiltonian for MATSTG is the sum of the single-particle BM-type Hamiltonian and the Coulomb interaction as

$$H = H_0 + H_{\text{int}} \,. \tag{31}$$

The total Hamiltonian H has U(2) × U(2), $\mathcal{T}$, C_3, $C_2\mathcal{T}$ and $T_{\boldsymbol{R}}$ symmetries, as well as m_z if combined with the action $\mathcal{E} \to -\mathcal{E}$ on the electric field. However, due to the opposite behaviors of H_{int} and H_0 under $C_{2x}P$ and $\mathcal{C}$, H does not preserve $C_{2x}P$ or $\mathcal{C}$, but it preserves the combination of them, i.e., $\mathcal{C}C_{2x}P$. Therefore the symmetry properties of the total Hamiltonian are

$$\begin{aligned}[\mathcal{T}, H] &= [C_3, H] = [C_2\mathcal{T}, H] = [T_{\boldsymbol{R}}, H] \\ &= [\mathcal{C}C_{2x}P, H] = [\mathrm{U}(2)\times\mathrm{U}(2), H] = 0 \\ m_z H m_z^{-1} &= H|_{\mathcal{E}\to-\mathcal{E}} \,.\end{aligned} \tag{32}$$

Based on the symmetry properties of H [Eq. (32)], we know that we only need to study $\mathcal{E} \geqslant 0$ since the negative $\mathcal{E}$ are related by m_z. Furthermore, we also only need to study the non-positive fillings, owing to $\mathcal{C}C_{2x}P$. To see this, we first define the filling operator

$$\hat{\nu} = \frac{1}{N}\int d^2r :\rho(\boldsymbol{r}):, \tag{33}$$

where N is the number of Moiré unit cells. The eigenvalue ν of $\hat{\nu}$ is the filling, i.e., the averaged number of electrons per Moiré unit cell counted from the charge neutrality. Owing to $[\hat{\nu}, H] = 0$ derived from the charge-U(1) invariance of H, we can label the energy eigenstates with definite filling ν.

As the filling operator anticommutes with $\mathcal{C}C_{2x}P$ as $\{\hat{\nu}, \mathcal{C}C_{2x}P\} = 0$ (Appendix A), we only need to study $\nu \leqslant 0$. To be more specific, for any many-body energy eigenstate $|\psi_{\nu,E}\rangle$ of H with filling ν and energy E, $\mathcal{C}C_{2x}P|\psi_{\nu,E}\rangle$ is an energy eigenstate with the same energy E and opposite filling $-\nu$. Then, if we obtain the set of all othornormal energy eigenstates $\{|i, \nu, E_i\rangle\}$ with filling ν, $\{\mathcal{C}C_{2x}P|i, \nu, E_i\rangle\}$ is the set of all orthonormal energy eigenstates with opposite filling $-\nu$, and the two energy eiegnstates have the same energy if they are related by $CC_{2x}P$. Therefore we only need to diagonalize H for $\nu \leqslant 0$, and we will adopt this simplification in later calculations.

III. *f-c-d* MODEL

In this section, we construct the heavy fermion f-c-d model for MATSTG. We start with the single-particle f-c-d model, and then project the Coulomb interaction to the heavy fermion basis to obtain the interacting Hamiltonian.

A. Single-particle *f-c-d* model

We start with the single-particle f-c-d model.

1. Review: f and c modes

We first discuss the construction of the f and c modes in the TBG part of the Hamiltonian. As shown in Eq. (12), the TBG part $H_{0,\text{TBG}}$ is just ordinary TBG with a $\sqrt{2}$ scaling of the interlayer tunneling [142]. Such rescaling can be canceled by the same rescaling of the energy unit. Thus, given any statement about the ordinary TBG with twist angle θ_{TBG}, the same statement holds for $H_{0,\text{TBG}}$ with θ satisfying $\sin(\theta/2) = \sqrt{2}\sin(\theta_{\text{TBG}}/2)$ [142]. When θ is very small (e.g., around the first magic angle), the condition can be approximated by $\theta \approx \sqrt{2}\theta_{\text{TBG}}$.

According to Ref. [125], in the ordinary TBG, localized heavy f modes and itinerant c modes can be constructed in each valley for each spin by mixing the nearly flat bands with the four lowest (two above and two below the flat bands) remote bands around Γ_M. Owing to the correspondence between the TBG part $H_{0,\text{TBG}}$ of MATSTG and the ordinary TBG, we are also able to construct such f and c modes from $H_{0,\text{TBG}}$. In the rest of this part, we follow Ref. [125] to show such construction. The discussion in this part is the same as that in Ref. [125], and thus can be viewed as a review of Ref. [125].

First, the f and c modes have the following expressions:

$$f^\dagger_{\eta,\boldsymbol{k},\alpha,s} = \sum_{\boldsymbol{Q}\sigma} \widetilde{\psi}^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q},\sigma,s}[\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]_{\boldsymbol{Q}\sigma} \quad \text{for } \boldsymbol{k}\in\text{MBZ} \tag{34}$$

and

$$c^\dagger_{\eta,\boldsymbol{k},\beta,s} = \sum_{\boldsymbol{Q}\sigma} \widetilde{\psi}^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q},\sigma,s}[\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})]_{\boldsymbol{Q}\sigma} \quad \text{for } |\boldsymbol{k}| \leqslant \Lambda_c \,, \tag{35}$$

where $\alpha = 1,\ 2$; $\beta = 1,\ 2,\ 3,\ 4$; and Λ_c is a small momentum cutoff for the c modes (small compared to the length of the primitive Moiré reciprocal vectors). $\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})$ is a smooth function of $\boldsymbol{k}\in\mathbb{R}^2$ while keeping $f^\dagger_{\eta,\boldsymbol{k}+\boldsymbol{G},\alpha,s} = f^\dagger_{\eta,\boldsymbol{k},\alpha,s}$ with $\boldsymbol{G}$ the Moiré lattice vector, and $\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})$ is a smooth function of $\boldsymbol{k}$ for $|\boldsymbol{k}| \leqslant \Lambda_c$. Here $\widetilde{v}_{\eta,f}$ and $\widetilde{u}_{\eta,c}$ are all in the eigen-subspace of the lowest six spinless bands in valley η; $\widetilde{v}_{\eta,f}$ belongs to the subspace for the nearly flat bands (the remote bands) far away from Γ_M (at Γ_M).

f modes are exponentially localized functions with physical symmetry representations (reps). Specifically, we can define

$$f^\dagger_{\eta,\boldsymbol{R}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}\in \mathrm{MBZ}} e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}} f^\dagger_{\eta,\boldsymbol{k}} , \tag{36}$$

where

$$f^\dagger_{\eta,\boldsymbol{k}} = (f^\dagger_{\eta,\boldsymbol{k},1,\uparrow}, f^\dagger_{\eta,\boldsymbol{k},1,\downarrow}, f^\dagger_{\eta,\boldsymbol{k},2,\uparrow}, f^\dagger_{\eta,\boldsymbol{k},2,\downarrow}) . \tag{37}$$

The smoothness of $\widetilde{v}_{\eta,f}(\boldsymbol{k})$ guarantees the exponential localization, and the symmetry reps in Appendix A suggest that $f_{\eta,\boldsymbol{R}}$ creates two spinful p-like orbitals localized at $\boldsymbol{R}$. This is why the f modes are localized.

Now we construct the low-energy Hamiltonain of $H_{0,\mathrm{TBG}}$ based on the f modes and c modes. First, note that

$$\widetilde{\psi}^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q},\sigma,s} = \sum_{\alpha=1,2} f^\dagger_{\eta,\boldsymbol{k},\alpha,s} [\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]^*_{\boldsymbol{Q}\sigma} + \sum_{\beta=1}^{4} c^\dagger_{\eta,\boldsymbol{k},\beta,s} \times [\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})]^*_{\boldsymbol{Q}\sigma} \theta(\Lambda_c - |\boldsymbol{k}|) + \dots , \tag{38}$$

where "..." labels the high-energy modes in the subspace spanned by $\widetilde{\psi}$. Then, we can separate out the low-energy part of $H_{0,\mathrm{TBG},\eta}$ (in the f and c basis) as

$$H_{0,\mathrm{TBG},\eta} = H_{0,\eta,f} + H_{0,\eta,c} + H_{0,\eta,fc} + \dots , \tag{39}$$

where $H_{0,\eta,fc}$ involves both f and c modes, $H_{0,\eta,f}$ only involves f modes, and $H_{0,\eta,c}$ only involves c modes. Based on the symmetry, the expressions of the low-energy terms are

$$H_{0,\eta,f} = 0 , \tag{40}$$

$$H_{0,\eta,c} = \sum_{\boldsymbol{k}}^{|\boldsymbol{k}|\leqslant\Lambda_c} c^\dagger_{\eta,\boldsymbol{k}} \begin{pmatrix} 0_{2\times 2} & v_\star(\eta k_x\tau_0 + \mathrm{i}k_y\tau_z) \\ v_\star(\eta k_x\tau_0 - \mathrm{i}k_y\tau_z) & M\tau_x \end{pmatrix} \otimes s_0 c_{\eta,\boldsymbol{k}} , \tag{41}$$

and

$$H_{0,\eta,fc} = \sum_{\boldsymbol{k}}^{|\boldsymbol{k}|\leqslant\Lambda_c} f^\dagger_{\eta,\boldsymbol{k}} e^{-\frac{|\boldsymbol{k}|^2\lambda^2}{2}} (\gamma\tau_0 + v'_\star(\eta k_x\tau_x + k_y\tau_y) \quad v''_\star(\eta k_x\tau_x - k_y\tau_y)) \otimes s_0 c_{\eta,\boldsymbol{k}} + \mathrm{H.c.}, \tag{42}$$

where

$$\begin{aligned} c^\dagger_{\eta,\boldsymbol{k}} &= (c^\dagger_{\eta,\boldsymbol{k},\Gamma_3}, c^\dagger_{\eta,\boldsymbol{k},\Gamma_1\Gamma_2}), \\ c^\dagger_{\eta,\boldsymbol{k},\Gamma_3} &= (c^\dagger_{\eta,\boldsymbol{k},1,\uparrow}, c^\dagger_{\eta,\boldsymbol{k},1,\downarrow}, c^\dagger_{\eta,\boldsymbol{k},2,\uparrow}, c^\dagger_{\eta,\boldsymbol{k},2,\downarrow}), \\ c^\dagger_{\eta,\boldsymbol{k},\Gamma_1\Gamma_2} &= (c^\dagger_{\eta,\boldsymbol{k},3,\uparrow}, c^\dagger_{\eta,\boldsymbol{k},3,\downarrow}, c^\dagger_{\eta,\boldsymbol{k},4,\uparrow}, c^\dagger_{\eta,\boldsymbol{k},4,\downarrow}) , \end{aligned} \tag{43}$$

and λ^2 is the Wannier spread of the f modes. Here we choose $H_{0,\eta,f} = 0$ because the hopping among f modes is very small ($\sim$0.1 meV), and we only keep terms up to $O(\boldsymbol{k}^2)$ for $H_{0,\eta,c}$ and $H_{0,\eta,fc}$. Owing to the zero kinetic energy of f modes, they are heavy. Here the factor $e^{-\frac{|\boldsymbol{k}|^2\lambda^2}{2}}$ is added, since it is the coupling between a wave packet with spread λ^2 and an itinerant electron with momentum $\boldsymbol{k}$. This factor can be neglected for small $\boldsymbol{k}$, but adding it allows us to choose a larger Λ_c in later numerical calculations.

To determine the values of the parameters, we need to specify θ. Specifically, we choose

$$\theta = 1.4703^\circ , \tag{44}$$

which is close to the first magic angle of MATSTG. In the rest of this work, we choose Eq. (44) for all numerical calculations unless specified otherwise. [We note that the framework discussed here is not limited to the value of θ in Eq. (44), as discussed in Appendix A.] Then, by projecting $H_{0,\mathrm{TBG}}$ to the f and c modes, we can get the numerical values for the parameters in Eqs. (40)–(42), as shown in Table I.

Before moving to other parts of the f-c-d model, we note that approximate analytic expressions exist for the f modes. Explicitly, the f modes have general expressions as

$$f^\dagger_{\eta,\boldsymbol{R},\alpha,s} = \int d^2r \sum_{\widetilde{l},\sigma} e^{\mathrm{i}\boldsymbol{R}\cdot\Delta K_{\widetilde{l}}} w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R}) \widetilde{\psi}^\dagger_{\eta,\widetilde{l},\boldsymbol{r},\sigma,s} , \tag{45}$$

where

$$\Delta K_t = -\boldsymbol{q}_3 , \quad \Delta K_b = \boldsymbol{q}_2 , \tag{46}$$

and

$$w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}) = \frac{1}{N\sqrt{\Omega}} \sum_{\boldsymbol{k}}^{\mathrm{MBZ}} \sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}} e^{\mathrm{i}(\boldsymbol{k}-\boldsymbol{Q})\cdot\boldsymbol{r}} [\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]_{\boldsymbol{Q}\sigma} \tag{47}$$

with Ω being the area of the Moiré unit cell, $\mathcal{Q}_{\eta,t} = \mathcal{Q}_\eta$ and $\mathcal{Q}_{\eta,b} = \mathcal{Q}_{-\eta}$. Symmetry properties of $w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r})$ are listed in Appendix A. The approximate expressions of $w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})$

TABLE I. Numerical values of the parameters in the single-particle f-c-d model [Eq. (54)] for the value of θ in Eq. (44).

M	γ	$v_\star$	$v'_\star$	$v''_\star$
−0.02678	0.1265	0.7176	0.2711	0.005768
M_1	B_γ	B_M	$B_{v''}$	λ
−0.1394	−0.09818	−0.08583	0.08760	1.407

are

$$w_{+1tA}^{\text{approx}}(\boldsymbol{r}) = \frac{N_0}{\sqrt{2}} \frac{1}{\sqrt{\pi\lambda_1^2}} e^{-\frac{|\boldsymbol{r}|^2}{2\lambda_1^2}} e^{-\mathrm{i}\frac{\pi}{4}},$$
$$w_{+1tB}^{\text{approx}}(\boldsymbol{r}) = \frac{N_1}{\sqrt{2}} \frac{1}{\sqrt{\pi\lambda_2^4}} e^{-\frac{|\boldsymbol{r}|^2}{2\lambda_2^2}} (x+\mathrm{i}y) e^{-\mathrm{i}\frac{5\pi}{4}}, \tag{48}$$

and other expressions can be obtained by acting with the symmetries on the basis. With $N_0 = -0.8193$, $N_1 = -0.5734$, $\lambda_1 = 0.7502$, and $\lambda_2 = 0.8001$, we find the overlapping probability between the numerical f modes and the analytical f modes is at least 86% as varying momentum, meaning that the analytical f modes are good approximations. By using the analytical expressions, we find that the low-energy bands of MATSTG are dominated by the f modes, as shown in Figs. 1(e)–1(h).

2. d Modes

The low-energy Dirac modes are just $d_{\eta,\boldsymbol{p}}^\dagger$ with small momentum $|\boldsymbol{p}| < \Lambda_d$, where Λ_d is the small momentum cutoff for the d modes. Then, the corresponding low-energy model of d modes is

$$H_{0,d,\eta} = \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_d} d_{\eta,\boldsymbol{p}}^\dagger (\eta p_x \sigma_x + p_y \sigma_y) \otimes s_0 \, d_{\eta,\boldsymbol{p}} \,, \tag{49}$$

where

$$d_{\eta,\boldsymbol{p}}^\dagger = (d_{\eta,\boldsymbol{p},A,\uparrow}^\dagger, d_{\eta,\boldsymbol{p},A,\downarrow}^\dagger, d_{\eta,\boldsymbol{p},B,\uparrow}^\dagger, d_{\eta,\boldsymbol{p},B,\downarrow}^\dagger) \,. \tag{50}$$

3. f-d coupling around ηK_M

The displacement field would couple f/c modes to d modes. The leading-order coupling should happen between f and d modes around ηK_M. This is because, at $\mathcal{E} = 0$, d modes cross with the nearly flat bands around ηK_M in valley η, and the nearly flat bands around ηK_M are purely given by f modes. [See Figs. 1(a) and 1(e).]

Based on the symmetry reps in Eq. (A24) in Sec. II A, the leading-order coupling reads

$$H_{0,\eta,fd} = \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_d} e^{-\frac{|\boldsymbol{p}|^2\lambda^2}{2}} f_{\eta,\eta K_M+\boldsymbol{p}}^\dagger \, M_1 \mathcal{E}(\tau_0 + \eta\mathrm{i}\tau_z) s_0 \, d_{\eta,\boldsymbol{p}} + \text{H.c.}, \tag{51}$$

where the $\boldsymbol{p}$-dependent terms are small (and are neglected) since the rep of $H_{0,\mathcal{E}}$ in the original basis is momentum independent. (See Appendix B 1 for details.) Again, we add the factor $e^{-\frac{|\boldsymbol{k}|^2\lambda^2}{2}}$ to allow a larger Λ_d in later numerical calculations, in the same spirit of the factor for the c modes [125]. By projecting $H_{0,\mathcal{E}}$ to f and d at ηK_M, we can get the numerical value of M_1 for the θ value in Eq. (44), as shown in Table I. Interestingly, the value of M_1 can also be estimated by the approximate analytical expressions of f modes in Eq. (48), resulting in

$$M_1 \approx \frac{N_0}{2}\sqrt{\frac{\pi\lambda_1^2}{\Omega}} = -0.1397, \tag{52}$$

which is quite close to the numerical value, suggesting the good quality of the analytical approximation.

4. f-d and c-d couplings around Γ_M

We did not yet include a c-d coupling, since we focused on the ηK_M, where c does not appear at low energies. To make our model more precise, we add the c-d (as well as f-d) coupling around Γ_M. The forms of those couplings are tedious, and we find that a more convenient way is to include them as corrections to the low-energy TBG part [i.e., Eqs. (40)–(42)]. Such corrections can be obtained by using the perturbation theory, since the f-d and c-d couplings are small compared to the gaps between f/c modes and d modes around Γ, as elaborated in Appendix B 2. As $H_{0,\mathcal{E}}$ has lower symmetries than $H_{0,\text{TBG}}$, the correction would bring in terms that break the extra symmetries of $H_{0,\text{TBG}}$. Nevertheless, we numerically find that those terms that break extra symmetries can be neglected without affecting the precision too much. As a result, the correction due to terms that preserve the extra symmetries can be incorporated by performing the following replacement in Eqs. (40)–(42)

$$\gamma \to \gamma + B_\gamma \mathcal{E}^2, \quad v_\star'' \to v_\star'' + B_{v''} \mathcal{E}^2, \quad M \to M + B_M \mathcal{E}^2. \tag{53}$$

We can directly obtain the values of B_γ, $B_{v''}$, and B_M for the θ value in Eq. (44) from the perturbation methods and show the results in Table I.

5. Single-particle f-c-d model

Combining Sec. III A 1–III A 4, we arrive at the single-particle f-c-d model as

$$H_{0,\eta}^{\text{eff}} = H_{0,\eta,f} + H_{0,\eta,c} + H_{0,\eta,fc} + H_{0,\eta,d} + H_{0,\eta,fd} \,, \tag{54}$$

where $H_{0,\eta,f}$, $H_{0,\eta,c}$, and $H_{0,\eta,fc}$ are Eq. (40)-(42) with the replacement in Eq. (53), $H_{0,\eta,d}$ is in Eq. (49), and $H_{0,\eta,fd}$ is in Eq. (51). With the parameter values in Table I, we plot the band structure of Eq. (54) in valley + in Figs. 1(a)–1(d). We find that the bands of Eq. (54) match those of the single-particle BM-type model H_0 in Eq. (2) very well for $0 \leqslant \mathcal{E} \leqslant 300$ meV and for the energy window $[-50\text{ meV}, 50\text{ meV}]$ (in EUS). The details on the numerical calculation can be found in Appendix B.

Before moving to the interacting part of the f-c-d model, we comment on the difference between our heavy localied f modes and the heavy modes mentioned in previous works [135,176] on MATSTG.

First, we emphasize that our heavy localized f modes are not the heavy modes mentioned in Ref. [135]. Reference [135] directly refers to the nearly flat bands in TBG part as the ultraheavy quasiparticles: however, these cannot be localized if physical symmetry reps are required due to the nontrivial topology of the bands. On the other hand, our f modes are localized, since the Wannier obstruction has been broken by mixing the nearly flat bands and the remote bands in the construction.

Second, although the heavy-fermion physics in MATSTG was also discussed in Ref. [176], the heavy modes in Ref. [176] are different from our heavy localized f modes. In Ref. [176], the dispersionless localized modes are phenomenologically constructed by coupling the TBG nearly flat

bands to the Dirac modes at a relatively large displacement field. It is not clear whether their construction can be applied to small displacement fields, since at zero displacement field, the TBG nearly flat bands are decoupled from the Dirac cones, and cannot be directly treated as localized modes due to their nontrivial topology. Our dispersionless localized f modes are constructed by combining the TBG flat bands with the remote bands around Γ_M, which does not rely on the displacement field. One manifestation of such differences is that the heavy modes in Ref. [176] have in total four flavors per Moiré unit cell and couple to dispersive modes around $\mathrm{M_M}$, while our heavy f modes have eight flavors per Moiré unit cell and couple to dispersive modes around Γ_M (and around ηK_M via the displacement field). Nevertheless, despite the difference, it is interesting to study (as future works) whether the model in Ref. [176] and our f-c-d model give qualitatively consistent phases after including the interaction.

B. Interaction among f, c, and d modes

We now discuss the interaction among f, c, and d modes, which is derived by projecting the Coulomb interaction to the f, c, and d modes.

1. Review: interaction among f and c modes

Both f and c modes are constructed solely from the TBG part of the model. Therefore the interaction among f and c modes should have the same form as those in Ref. [125], which we will review in this part. More details can be found in Appendix C.

First, for the interaction among f modes, the leading-order term is the density-density interaction, which reads

$$H_{\mathrm{int},U} = \frac{U_1}{2}\sum_{\boldsymbol{R}} :\rho_f(\boldsymbol{R})::\rho_f(\boldsymbol{R}): + \frac{U_2}{2}\sum_{\boldsymbol{R},\boldsymbol{R}'}^{|\boldsymbol{R}-\boldsymbol{R}'|=|\boldsymbol{a}_{M,1}|} :\rho_f(\boldsymbol{R})::\rho_f(\boldsymbol{R}'):, \tag{55}$$

where $\rho_f(\boldsymbol{R}) = \sum_{\eta,\alpha,s} f^\dagger_{\eta,\boldsymbol{R},\alpha,s} f_{\eta,\boldsymbol{R},\alpha,s}$. The expressions of U_1 and U_2 can be found in Appendix C. In Eq. (55), we neglect the density-density interactions of further ranges, as they are exponentially lower owing to the localized nature of the f modes.

Second, the interaction among c modes turns out to have the Coulomb form to the leading order as

$$H_{\mathrm{int},V,c} = \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_c(\boldsymbol{r})::\rho_c(\boldsymbol{r}'):, \tag{56}$$

where $\rho_c(\boldsymbol{r}) = \sum_\beta \rho_{c,\beta}(\boldsymbol{r})$, $\rho_{c,\beta}(\boldsymbol{r}) = \sum_{\eta,s} c^\dagger_{\eta,\boldsymbol{r},\beta,s} c_{\eta,\boldsymbol{r},\beta,s}$, and

$$c^\dagger_{\eta,\boldsymbol{r},\beta,s} = \frac{1}{\sqrt{\mathcal{A}}}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} e^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}} c^\dagger_{\eta,\boldsymbol{p},\beta,s}. \tag{57}$$

Third, the interaction between f and c modes has two non-negligible terms. One term is the channel-resolved density-density interaction as

$$H_{\mathrm{int},W,fc} = \Omega\sum_{\boldsymbol{R},\beta} W_\beta :\rho_f(\boldsymbol{R})::\rho_{c,\beta}(\boldsymbol{R}): \tag{58}$$

TABLE II. Numerical values of the parameters in Eq. (63). Values in the second line of the table is in the unit system specified in Eq. (1), while those in the third line are in EUS. More details can be found in Appendix C.

Unit	U_1	U_2	W_1	W_3	J	W_{fd}
Eq. (1)	0.3523	0.02388	0.3409	0.3761	0.09337	0.3647
EUS (meV)	91.50	6.203	88.54	97.67	24.25	94.71

with $W_1 = W_2$ and $W_3 = W_4$. The last term is

$$H_{\mathrm{int},J} = -\frac{J\Omega}{2}\sum_{\boldsymbol{R}}\sum_{\eta\alpha s}\sum_{\eta'\alpha's'}(\eta\eta' + (-)^{\alpha+\alpha'}) :f^\dagger_{\eta,\boldsymbol{R},\alpha,s} f_{\eta',\boldsymbol{R},\alpha',s'}::c^\dagger_{\eta',\boldsymbol{R},\alpha'+2,s'} c_{\eta,\boldsymbol{R},\alpha+2,s}:. \tag{59}$$

The interaction only occurs at the Moiré lattice positions, which is consistent with the fact that f modes are localized at Moiré lattice positions. The expressions of W_β and J can be found in Appendix C.

2. Interaction among d modes

Inherited from the total Coulomb interaction, the interaction among d modes is given by the Coulomb form

$$H_{\mathrm{int},V,d} = \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_d(\boldsymbol{r})::\rho_d(\boldsymbol{r}'):, \tag{60}$$

where $\rho_d(\boldsymbol{r}) = \sum_{\eta,\sigma,s}\widetilde{d}^\dagger_{\eta,\boldsymbol{r},\sigma,s}\widetilde{d}_{\eta,\boldsymbol{r},\sigma,s}$ and $\widetilde{d}^\dagger_{\eta,\boldsymbol{r},\sigma,s} = \frac{1}{\sqrt{\mathcal{A}}}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_d} e^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}} d^\dagger_{\eta,\boldsymbol{p},\sigma,s}$, which becomes $d_{\eta,\boldsymbol{r},\sigma,s}$ in the limit of $\Lambda_d\to\infty$.

3. f-d and c-d interaction

We find that the interaction between f and d modes and the interaction between c and d modes are both in the form of density-density interaction in the leading order, as discussed in details in Appendix C. Specifically, we find that the leading-order interaction between f and d modes reads

$$H_{\mathrm{int},W,fd} = \Omega W_{fd}\sum_{\boldsymbol{R}} :\rho_f(\boldsymbol{R})::\rho_d(\boldsymbol{R}):, \tag{61}$$

and the leading-order interaction between c and d modes has the Coulomb form as

$$H_{\mathrm{int},V,cd} = \int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_c(\boldsymbol{r})::\rho_d(\boldsymbol{r}'):. \tag{62}$$

The expression of W_{fd} can be found in Appendix C.

4. Total interaction

The total interaction among the f, c and d modes is the sum of Eqs. (55), (56), (58)–(62), which reads

$$H^{\mathrm{eff}}_{\mathrm{int}} = H_{\mathrm{int},U} + H_{\mathrm{int},V,c} + H_{\mathrm{int},W,fc} + H_{\mathrm{int},J} + H_{\mathrm{int},V,d} + H_{\mathrm{int},V,cd} + H_{\mathrm{int},W,fd}. \tag{63}$$

We numerically evaluate the values of the interaction parameters, and the results are listed in Table II. Among the interaction strengths, we can see that the largest energy scale is 90–100 meV in EUS. We have W_1, W_3, W_{fd}, and U_1 at this scale. Unlike Ref. [125], W_3 is slightly larger than

U_1 here, since the gate distance is not scaled by $1/\sqrt{2}$ for MATSTG compared to that in Ref. [125]. (See more details in Appendix C.) Moreover, W_{fd} is also slightly larger than U_1 here. Nevertheless, we would expect that the onsite repulsive interaction U_1 among f modes is the dominant interaction channel at low energies, since U_1 only involves the f modes (which dominate in the low energy), while W_1, W_3, and W_{fd} involve the c and d modes with relatively higher energies.

C. *f-c-d* model for MATSTG

The f-c-d model for MATSTG is just the sum of the single-particle part Eq. (54) and the interaction Eq. (63) as

$$H_{fcd} = \sum_{\eta} H^{\text{eff}}_{0,\eta} + H^{\text{eff}}_{\text{int}} . \tag{64}$$

This is the low-energy model that we propose for MATSTG with only Coulomb interaction. The single-particle band structure (Fig. 1) already shows that the f-c-d model well captures the single-particle physics of MATSTG for $\mathcal{E} \in [-300, 300]$ meV and for the energy window $[-50, 50]$ meV in EUS. Since the largest energy scale of the interaction is $U_1 \sim 100$ meV, the energy window corresponds to $[-U_1/2, U_1/2]$, covering the main low-energy modes affected by the interaction. Therefore we expect the f-c-d model Eq. (64) to work for the specified $\mathcal{E}$ range and energy window even at the many-body level. We will perform Hartree-Fock calculations with the model in the following section.

IV. NUMERICAL HARTREE-FOCK CALCULATIONS

With our model [Eq. (64)], we perform numerical Hartree-Fock calculations for $\nu = 0, -1, -2$. We will not study the positive fillings since they are related to the negative fillings by $\mathcal{C}C_{2x}\mathcal{P}$ as discussed in Sec. II C.

Similar to the TBG case [125], the initial states that we choose for the Hartree-Fock calculation have the following general form

$$|\Psi_{\text{initial}}\rangle = \prod_{\boldsymbol{R}} f^\dagger_{\boldsymbol{R}}\zeta_1 f^\dagger_{\boldsymbol{R}}\zeta_2 \cdots f^\dagger_{\boldsymbol{R}}\zeta_{4+\nu}|\text{Fermi Sea}\rangle , \tag{65}$$

where

$$f^\dagger_{\boldsymbol{R}} = (f^\dagger_{+,\boldsymbol{R},1,\uparrow}, f^\dagger_{+,\boldsymbol{R},1,\downarrow}, f^\dagger_{+,\boldsymbol{R},2,\uparrow}, f^\dagger_{+,\boldsymbol{R},2,\downarrow}, f^\dagger_{-,\boldsymbol{R},1,\uparrow}, f^\dagger_{-,\boldsymbol{R},1,\downarrow}, f^\dagger_{-,\boldsymbol{R},2,\uparrow}, f^\dagger_{-,\boldsymbol{R},2,\downarrow}) , \tag{66}$$

each of $\zeta_1, \ldots, \zeta_{4+\nu}$ has eight components, e.g.,

$$\zeta_1 = \begin{pmatrix} (\zeta_1)_{+1\uparrow} \\ (\zeta_1)_{+1\downarrow} \\ (\zeta_1)_{+2\uparrow} \\ (\zeta_1)_{+2\downarrow} \\ (\zeta_1)_{-1\uparrow} \\ (\zeta_1)_{-1\downarrow} \\ (\zeta_1)_{-2\uparrow} \\ (\zeta_1)_{-2\downarrow} \end{pmatrix}, \tag{67}$$

$$f^\dagger_{\boldsymbol{R}}\zeta_1 = \sum_{\eta,\alpha,s} f^\dagger_{\eta,\boldsymbol{R},\alpha,s}(\zeta_1)_{\eta\alpha s} , \tag{68}$$

and $|\text{Fermi Sea}\rangle$ is the half-filled Fermi sea of the free c and d modes. (See the choice of the initial states in Appendix D 2.) Eq. (65) means that we specify different initial states by specifying different combinations of the f modes, i.e., specifying

$$\zeta = (\zeta_1 \quad \zeta_2 \quad \ldots \quad \zeta_{4+\nu}). \tag{69}$$

We can do so because the f modes and its onsite interaction dominate the low-energy physics as discussed in the last section. By using Eq. (65), we perform self-consistent Hartree-Fock calculations for $\nu = 0, -1, -2$, and the results are summarized below and shown in Fig. 3. (See details in Appendix D.)

As shown in Figs. 3(a), 3(d) and 3(g), for all the considered fillings, increasing the displacement field $\mathcal{E}$ would lead to a phase transition, at which the ground states lose intervalley coherence.

For $\nu = 0$, the low-$\mathcal{E}$ ground states are the Kramers-intervalley-coherent (K-IVC) states, while there are four types of competing ground states at high $\mathcal{E}$, namely, Chern states ($\text{Ch} = \pm 2$), half-Chern states ($\text{Ch} = \pm 1$), valley-Hall (VH) states and $C_2\mathcal{T}$-invariant states, where "competing" means that the differences in their ground-state energies are beyond our numerical resolution, VH refers to the state with nonzero valley Chern numbers but zero total Chern number, and Ch stands for the Chern number. The low-$\mathcal{E}$ states are metallic, while the high-$\mathcal{E}$ states are insulating. [See Figs. 3(b) and 3(c).]

For $\nu = -1$, the low-$\mathcal{E}$ ground states are a combination of valley-polarized (VP) and intervalley-coherent (IVC) states, while there are three types of competing partially valley-polarized (PVP) ground states at high $\mathcal{E}$, where PVP means that one valley has one more electron than the other valley per Moiré unit cell and the state has no intervalley coherence. PVP is "partial" because the VP state for $\nu = -1$ should have three more electrons in one valley than in the other. Both the low-$\mathcal{E}$ and high-$\mathcal{E}$ states are metallic. [See Figs. 3(e) and 3(f).]

For $\nu = -2$, the low-$\mathcal{E}$ ground states are K-IVC states, while there are four types of competing ground states at high $\mathcal{E}$—two types of VP states and two types of valley unpolarized states. Both the low-$\mathcal{E}$ and high-$\mathcal{E}$ states are metallic. [See Figs. 3(h) and 3(i).]

All these self-consistent Hartree-Fock results obtained from our f-c-d model [Eq. (64)] are generally consistent with previous numerical results in Refs. [149,152,156], verifying the validity of our f-c-d model. Moreover, our calculation finds some high-$\mathcal{E}$ ground states (like the half-Chern states for $\nu = 0$) that are missed in Refs. [149,152,156], meaning that our calculation actually refines the previous results [149,152,156].

In particular, we find that the phase transitions characterized by the loss of intervalley coherence [Figs. 3(a), 3(d) and 3(g)] can be qualitatively captured by the one-shot Hartree-Fock calculation, where "one-shot" means only performing the first step of the iteration, which is numerically simple to do compared with the full self-consistent calculation and can even be done analytically as discussed in the next section. Furthermore, we find that the competing energies of the high-$\mathcal{E}$ ground states can be precisely captured in the one-shot Hartree-Fock calculation. Therefore our choice of the initial states in Eq. (65) are considerably close to the final

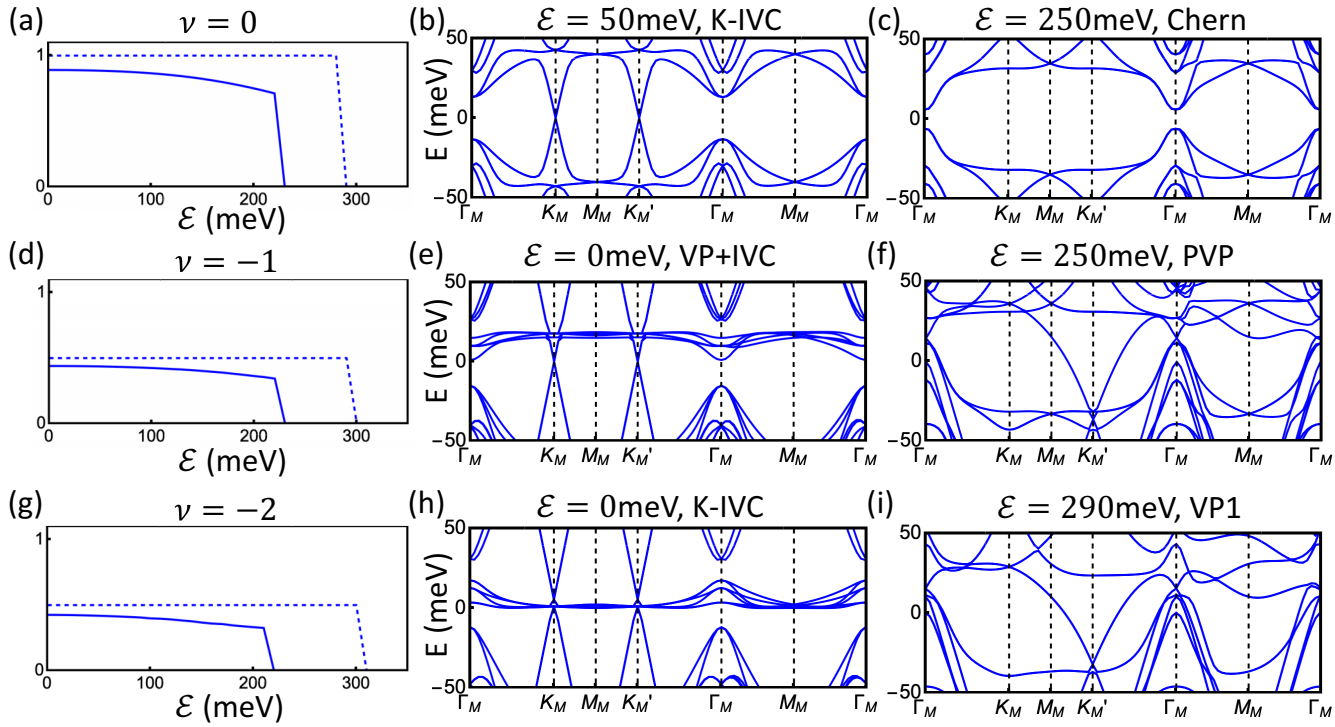

FIG. 3. This figure shows the numerical Hartree-Fock results for MATSTG based on Eq. (64), where [(a)–(c)] are for $\nu = 0$, [(d)–(f)] are for $\nu = -1$, and [(g)–(i)] are for $\nu = -2$. (a), (d), and (g) shows the intervalley coherence of the Hartree-Fock ground state as $\mathcal{E}$ varies, where the zero (nonzero) value corresponds to the absence (presence) of the intervalley coherence. The solid line is given by the self-consistent Hartree-Fock calculation, while the dashed is the one-shot result. (b), (c), (e), (f), (h), and (i) are the Hartree-Fock band structures of the ground state (or one of the competing ground states) at the corresponding filling and $\mathcal{E}$, plotted with the density matrices given by the self-consistent Hartree-Fock calculation.

Hartree-Fock ground states given by the self-consistent Hartree-Fock calculation, verifying the fact that the f modes and their onsite repulsive interaction dominate the low-energy physics.

We note that our Hartree-Fock calculation is done only for the translationally invariant initial states listed in Appendix D 2. It is possible that the true ground state is beyond our chosen initial states in Appendix D 2 (e.g., beyond the translationally invairant subspace). We leave a complete Hartree-Fock study as a future work.

V. ANALYTICAL UNDERSTANDING

In this section, we provide an analytical understanding for the key numerical results in Sec. IV. As discussed at the end of Sec. IV, the one-shot Hartree-Fock calculation (i) can qualitatively capture the phase transition between states with and without intervalley-coherence [Figs. 3(a), 3(d) and 3(h)] and (ii) can precisely capture the competing energies of the several found high-$\mathcal{E}$ states. Therefore we will use the analytical one-shot Hartree-Fock Hamiltonian of the f-c-d model [Eq. (64)] derived from the expression of the initial states [Eq. (65)] to answer two questions: (i) why the states without intervalley coherence are favored at high $\mathcal{E}$, and (ii) why those high-$\mathcal{E}$ ground states have nearly the same energies.

Let us start with the first question: why the states without intervalley coherence are favored at high $\mathcal{E}$. Since we care about high $\mathcal{E}$, let us consider the limit where $\mathcal{E}$ is infinitely large. The validity of this assumption will be discussed right beneath proposition 1. The low-energy itinerant modes are mainly around Γ_M and $\pm K_M$. In the following, we will look at $\pm K_M$ first and then look at Γ_M.

We want to minimize the total energy of all the occupied states at K_M and $-K_M$, which is labeled by $E_{\pm K_M}$. To do so, let us define ζ_η. We know ζ_l (with $l = 1, 2, \ldots, 4+\nu$) in Eq. (65) has eight components as $(\zeta_l)_{\eta\alpha s}$, where η, α and s are indices of the f modes. We define ζ_η as a $4 \times (4+\nu)$ matrix such that $(\zeta_\eta)_{\alpha s,l} = (\zeta_l)_{\eta\alpha s}$, which means that

$$\zeta = \begin{pmatrix} \zeta_+ \\ \zeta_- \end{pmatrix}. \tag{70}$$

Then, as elaborated in Appendix E 2, in the high-$\mathcal{E}$ limit, the one-shot Hartree-Fock Hamiltonian at ηK_M of MBZ to the first order of $|U_1/\mathcal{E}|$ (up to unitary transformation and total energy shift) reads

$$\begin{pmatrix} \epsilon_0 \mathbb{1}_{4\times 4} & & \\ & \epsilon_1 \mathbb{1}_{4\times 4} & \\ & & \nu(U_1+6U_2)\mathbb{1}_{4\times 4} \end{pmatrix} - U_1 \left(\begin{array}{cc|c} \begin{pmatrix} |\chi_{0,1}|^2 & \\ & |\chi_{1,1}|^2 \end{pmatrix} \otimes \left(\zeta_\eta \zeta_\eta^\dagger - \frac{1}{2}\right) & & \\ & & \zeta_{-\eta}\zeta_{-\eta}^\dagger - \frac{1}{2} \end{array} \right), \tag{71}$$

where

$$\begin{pmatrix} \nu(U_1+6U_2) & \sqrt{2}M_1\mathcal{E} \\ \sqrt{2}M_1\mathcal{E} & W_{fd}\nu \end{pmatrix}\chi_\gamma = \epsilon_\gamma \chi_\gamma, \tag{72}$$

$\gamma = 0, 1$, χ_γ is real, and

$$\epsilon_\gamma = \frac{\nu(U_1+6U_2+W_{fd})}{2} + (-)^\gamma\sqrt{\left[\frac{\nu(U_1+6U_2-W_{fd})}{2}\right]^2 + 2M_1^2\mathcal{E}^2}\,. \tag{73}$$

(See Appendix E 2 for details.) Since the chemical potential can be estimated as $\mu \approx \nu(U_1+6U_2)$ (f modes give the filling) as discussed in Appendix E 2 (also in Ref. [125]), the occupied states of the approximated Hamiltonian in Eq. (71) are all eigenstates of

$$[\epsilon_1 - \nu(U_1+6U_2)]\mathbb{1}_{4\times4} - U_1|\chi_{1,1}|^2\left(\zeta_\eta\zeta_\eta^\dagger - \tfrac{1}{2}\right) \tag{74}$$

and all negative-energy eigensstates of

$$-U_1\left(\zeta_\eta\zeta_\eta^\dagger - \tfrac{1}{2}\right) \tag{75}$$

for both $\eta = \pm$, where we have subtracted the chemical potential. The total energy of these occupied states give $E_{\pm K_M}$ to the first order of $|U_1/\mathcal{E}|$.

Now let us minimize $E_{\pm K_M}$. To express $E_{\pm K_M}$, we use λ_i ($i = 1, 2, \ldots, 8$) to label the eigenvalues of

$$\begin{pmatrix} \zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger \end{pmatrix}. \tag{76}$$

Then, we can choose $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 1/2 \geqslant \lambda_{n+1} \geqslant \cdots \geqslant \lambda_8$ without loss of generality, resulting in

$$E_{\pm K_M} = 8[\epsilon_1 - \nu(U_1+6U_2)] - U_1|\chi_{1,1}|^2\nu - U_1\sum_{i=1}^{n}\left(\lambda_i - \frac{1}{2}\right) + O\left(U_1^2/\mathcal{E}\right), \tag{77}$$

where we have used

$$\sum_\eta \mathrm{Tr}[\zeta_\eta\zeta_\eta^\dagger] = \mathrm{Tr}[\zeta\zeta^\dagger] = 4+\nu\,. \tag{78}$$

To proceed, we note that $\lambda_i \in [0, 1]$ and $\sum_{i=1}^8 \lambda_i = 4+\nu$. Then, we know

$$E_{\pm K_M} \geqslant 8[\epsilon_1 - \nu(U_1+6U_2)] - U_1|\chi_{1,1}|^2\nu - U_1\frac{4+\nu}{2} + O\left(U_1^2/\mathcal{E}\right). \tag{79}$$

As elaborated in Appendix E 2, it turns out that the equality happens if and only if

$$\begin{pmatrix} \zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger \end{pmatrix} \cong \mathrm{diag}(\underbrace{1, 1, \ldots, 1}_{4+\nu}, \underbrace{0, 0, \ldots, 0}_{4-\nu}), \tag{80}$$

which is equivalent to $\zeta_+\zeta_-^\dagger = 0$ (i.e., zero intervalley coherence). Here $\cong$ means being equal up to any unitary transformations. Therefore we know $E_{\pm K_M}$ is minimized if and only if the intervalley coherence of the state vanishes.

Now we turn to the Γ_M point. As discussed in Appendix E 2 (and also in Ref. [125]), the main origin of the symmetry breaking is the J interaction term, which appears in the diagonal block of the one-shot Hartree-Fock Hamiltonian for $c^\dagger_{\Gamma_1\Gamma_2}$, expressed as $\nu W_3 + h_{\Gamma_1\Gamma_2}$. In our case, $h_{\Gamma_1\Gamma_2}$ reads

$$h_{\Gamma_1\Gamma_2} = \widetilde{M}\eta_0\sigma_x s_0 - \frac{J}{2}(\eta_z\sigma_0 s_0\zeta\zeta^\dagger\eta_z\sigma_0 s_0 + \eta_0\sigma_z s_0\zeta\zeta^\dagger\eta_0\sigma_z s_0 - \mathbb{1}_{8\times8})\,, \tag{81}$$

where $\widetilde{M} = M + B_M\mathcal{E}^2$. Since we consider the high-$\mathcal{E}$ limit, we have $|\widetilde{M}| \gg J$. Then, the energy difference between different states given by $h_{\Gamma_1\Gamma_2}$ should be of order J, which is generally much smaller than the energy difference at $\pm K_M$ which is of the order U_1. Therefore we should only focus on the states with lowest $E_{\pm K_M}$, i.e., states with zero intervalley coherence. In other words, the discussion at $\pm K_M$ already suggests that only states without intervalley coherence should be favored at large $\mathcal{E}$.

To further pick out the high-$\mathcal{E}$ ground states among all states without IVC, let us minimize the energy at Γ_M. Since we are considering the high-$\mathcal{E}$ limit, we have $|\widetilde{M}| \gg |\nu(U_1 + 6U_2 - W_3)|$. Then, by minimizing the total energy of all the occupied states of $\nu W_3 + h_{\Gamma_1\Gamma_2}$ [i.e., states of $h_{\Gamma_1\Gamma_2}$ that are energetically lower than $\nu(U_1+6U_2-W_3)$] while keeping the intervalley coherence zero, we find that the energetically favored states are (and only are) the states whose $\zeta\zeta^\dagger$ [up to U(2) × U(2)] are also spin-diagonal with $4+\nu$ diagonal blocks (labelled by valley and spin) being $(\sigma_0 \pm \sigma_z)/2$ and $4-\nu$ diagonal blocks being zero. (See details in Appendix E 2.) Eventually, we arrive at the following rule for the high-$\mathcal{E}$ ground states, which resolves the first question raised at the beginning of this section.

Proposition 1. For $\nu = 0, -1, -2$, at the one-shot Hartree-Fock level, a state is energetically favored at high $\mathcal{E}$ if and only if its $\zeta\zeta^\dagger$, up to U(2) × U(2), is spin-valley diagonal with $4+\nu$ diagonal blocks (labelled by valley and spin) being $(\sigma_0 \pm \sigma_z)/2$ and $4-\nu$ diagonal blocks being zero.

Now let us discuss the validity of the derivation that leads to proposition 1. We know that the derivation is done in the limit that $\mathcal{E}$ is infinitely large, which seems to contradict the fact that the f-c-d model is valid within $\mathcal{E} = 300$ meV (EUS), since $|\sqrt{2}M_1\mathcal{E}| \sim U_1$ for $\mathcal{E} = 300$ meV. However, we show in Appendix E 3 that the derivation should still be valid for $\nu = 0$ at $\mathcal{E} = 300$ meV, since the quantities required to be small in the derivation are still small for $\nu = 0$ at $\mathcal{E} = 300$ meV. Although the derivation is not entirely reasonable for $\nu = -1, -2$, we find that proposition 1 is consistent with the self-consistent Hartree-Fock calculation for $\nu = -1, -2$. Specifically, we numerate all initial states that satisfy proposition 1 for $\nu = 0, -1, -2$, and we find that they all become high-$\mathcal{E}$ ground states in the self-consistent Hartree-Fock calculation discussed in Sec. IV.

Before proceeding to the second question raised at the beginning of this section, let us provide an understanding of the appearance of the phase transition with gradually increasing $\mathcal{E}$. In the earlier part of this section, we have shown that the Hartree-Fock Hamiltonian at $\pm K_M$ should favor states without intervalley coherence at high $\mathcal{E}$; on the other hand, Ref. [125] suggests the TBG part around Γ_M should favor states with nonzero intervalley coherence. Then, the transition should be

a result of the competition between Γ_M and $\pm K_M$. To make it concrete, let us consider $\nu = 0$ and treat $\mathcal{E}$ perturbatively, to consider the case where $\mathcal{E}$ is gradually increased. We focus on the competence between K-IVC and Chern states. By using second-order perturbation, we derive the effective energies for the two states at $\pm K_M$ as

$$E^{\text{K-IVC},\nu=0}_{\pm K_M} = -4 \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|<\Lambda_d} |\boldsymbol{p}|,$$
$$E^{\text{Ch},\nu=0}_{\pm K_M} = -4 \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|<\Lambda_d} \sqrt{|\boldsymbol{p}|^2 + \frac{16 M_1^4 \mathcal{E}^4}{U_1^2}} \tag{82}$$

and at Γ_M as

$$E^{\text{K-IVC},\nu=0}_{\Gamma_M} = -\sum_{\boldsymbol{k}}^{|\boldsymbol{k}|<\Lambda_c} \Big[\sqrt{U_1^2 + 16(|v'_\star \boldsymbol{p}| - |\widetilde{\gamma}|)^2} + \sqrt{U_1^2 + 16(|v'_\star \boldsymbol{p}| + |\widetilde{\gamma}|)^2}\Big],$$
$$E^{\text{Ch},\nu=0}_{\Gamma_M} = -\sum_{\boldsymbol{k}}^{|\boldsymbol{k}|<\Lambda_c} \sum_{z=\pm} \Big[\sqrt{U_1^2 + 16|v'_\star \boldsymbol{p}|^2} + z\sqrt{U_1^2 + 16\widetilde{\gamma}^2}\Big], \tag{83}$$

where $\mathcal{E}$ is treated perturbatively. (See Appendix E 5 for details.) Then, the total effective energies are

$$E^{\text{K-IVC},\nu=0}_{\text{eff}} = E^{\text{K-IVC},\nu=0}_{\Gamma_M} + E^{\text{K-IVC},\nu=0}_{\pm K_M},$$
$$E^{\text{Ch},\nu=0}_{\text{eff}} = E^{\text{Ch},\nu=0}_{\Gamma_M} + E^{\text{Ch},\nu=0}_{\pm K_M}. \tag{84}$$

As elaborated in Appendix E 5, at $\mathcal{E} = 0$, we have $E^{\text{K-IVC},\nu=0}_{\text{eff}} < E^{\text{Ch},\nu=0}_{\text{eff}}$ since $E^{\text{K-IVC},\nu=0}_{\pm K_M} = E^{\text{Ch},\nu=0}_{\pm K_M}$ and $E^{\text{K-IVC},\nu=0}_{\Gamma_M} < E^{\text{Ch},\nu=0}_{\Gamma_M}$. Moreover, at $\mathcal{E} = \mathcal{E}_c$ ($\approx$294.816 meV in EUS) that satisfies $\gamma + B_\gamma \mathcal{E}_c^2 = 0$, we have $E^{\text{K-IVC},\nu=0}_{\text{eff}} > E^{\text{Ch},\nu=0}_{\text{eff}}$ since $E^{\text{K-IVC},\nu=0}_{\pm K_M} > E^{\text{Ch},\nu=0}_{\pm K_M}$ and $E^{\text{K-IVC},\nu=0}_{\Gamma_M} = E^{\text{Ch},\nu=0}_{\Gamma_M}$, demonstrating the existence of the transition (as increasing $\mathcal{E}$ from $\mathcal{E} = 0$ to $\mathcal{E} = \mathcal{E}_c$).

Combining the low-$\mathcal{E}$ with the high-$\mathcal{E}$ picture, we arrive at the following picture. At low $\mathcal{E}$, Γ_M dominates and favors nonzero intervalley coherence. At high $\mathcal{E}$, $\pm K_M$ dominate and favor zero intervalley coherence, and the secondary Γ_M effect picks out specific states among all states without intervalley coherence.

Now we turn to the second question: why the numerically found high-$\mathcal{E}$ low-energy states have competing energies. We answer this question by showing that those high-$\mathcal{E}$ states have exactly the same Hartree-Fock energies at the one-shot level. At the one-shot level, we find (Appendix E 4) that the Hartree-Fock Hamiltonian is block diagonalzied in spin and valley for all the high-$\mathcal{E}$ ground states for $\nu = 0, -1, -2$. Interestingly, the one-shot Hartree-Fock Hamiltnoians for different types of states are related by performing spinless version of $C_2\mathcal{T}$, noted as $C_2\overline{\mathcal{T}}$, or spinless TR symmetries on certain blocks. Taking VH and Chern states at $\nu = 0$ as an example, we have

$$H^{A,\text{OS}} = H^{A,\text{OS}}_{+,\uparrow} + H^{A,\text{OS}}_{+,\downarrow} + H^{A,\text{OS}}_{-,\uparrow} + H^{A,\text{OS}}_{-,\downarrow} - E^{\text{OS}}_0, \tag{85}$$

where $A =$ VH and Chern, and OS is short for one-shot, E^{OS}_0 depends on the initial states only through the filling, and $H^{\text{Chern,OS}}$ is related to $H^{\text{VH,OS}}$ as

$$H^{\text{Chern,OS}} = H^{\text{VH,OS}}_{+,\uparrow} + H^{\text{VH,OS}}_{+,\downarrow} + C_2\overline{\mathcal{T}} H^{\text{VH,OS}}_{-,\uparrow} (C_2\overline{\mathcal{T}})^{-1} + C_2\overline{\mathcal{T}} H^{\text{VH,OS}}_{-,\downarrow} (C_2\overline{\mathcal{T}})^{-1}. \tag{86}$$

Therefore the one-shot Hartree-Fock energies are exactly the same for the high-$\mathcal{E}$ ground states with the same filling. (See Appendix E 4 for more details.)

Before concluding the paper, we compare and contrast our analytic discussion to those in Refs. [152,156,175] Instead of choosing the f-c-d basis in our work, Refs. [152,156,175] chose the TBG nearly flat bands and the Dirac cones as basis for the analytical discussions. As a result, Refs. [152,156,175] did not give a general simple analytic rule for high-$\mathcal{E}$ states as proposition 1 or simple symmetry argument for competing energies as ours, indicating the great simplification brought by our f-c-d model. Furthermore, Refs. [152,156,175] do not provide an understanding of the appearance of the transition; the simple picture of the heavy fermion model explains the transition based on the competition between the energies at Γ_M and $\pm K_M$ points.

VI. CONCLUSION AND DISCUSSION

In conclusion, we construct an effective heavy fermionic f-c-d model for MATSTG with localized heavy f modes and itinerant c and d modes. Our f-c-d model can reproduce the previously obtained single-particle band structure of MATSTG in the energy window $[-50\text{ meV}, 50\text{ meV}]$ and for displacement field $\mathcal{E} \in [0, 300]$ meV in EUS. Our f-c-d model can also reproduce and refine the previous numerical Hartree-Fock results for $\nu = 0, -1, -2$. Remarkably, based on our f-c-d model at $\nu = 0, -1, -2$, we propose a simple analytical rule for the high-$\mathcal{E}$ ground states, which explains the general loss of intervalley coherence observed in numerical results, and we find analytical symmetry arguments that explain the completing energies of the nearly degenerate high-$\mathcal{E}$ ground states.

For experiments, we predict that at charge neutrality and high displacement fields, Chern gaps for $\text{Ch} = \pm 1, \pm 2$ can be observed by scanning tunneling microscope in the presence of an out-of-plane magnetic field. In particular, we predict that $\text{Ch} = \pm 2$ gaps should be most pronounced, since our arguments in Appendix F show that the orbital effect of the magnetic field can lower the energy of the Chern states. More specifically, the projection of the orbital effect of the magnetic field to the two TR-realted Chern states is proportional to σ_z, which always lowers the energy of one Chern state regardless of the sign of the coefficient. We leave a more detailed study of such prediction for the future. Our work both generalizes and puts on firmer footing through analytical reasoning the applicability and the importance of the topological heavy fermion model in naturally explaining the emergence of Coulomb interaction-driven correlated phases in Moiré multilayer graphene systems.

ACKNOWLEDGMENTS

J.Y. thanks Yang-Zhi Chou, Seth Davis, Biao Lian, Jay D. Sau, Zhi-Da Song, and Fang Xie for helpful discussions. This

work is supported by the Laboratory of Physical Sciences at the University of Maryland. J.Y. and B.A.B. were supported by DOE Grant No. DE-SC0016239. Additional support was provided by Gordon and Betty Moore Foundation through the EPiQS Initiative, Grant No. GBMF11070 and Grant No. GBMF8685 towards the Princeton theory program.

APPENDIX A: MORE DETAILS ON THE BASIS OF THE HAMILTONIAN

In this section, we provide more details on the basis of the Hamiltonian and the furnished symmetry reps.

Let us use $\pm K$ to label the two graphene valleys with $K = \frac{4\pi}{3a_G}(1,0)$ in EUS. (Recall that EUS is the unit system in which Å is the length unit and meV is the energy unit, as discussed at the beginning of Sec. II.) Given a single graphene, its electron basis near $\pm K$ reads

$$c^\dagger_{\pm K+\boldsymbol{p},\sigma,s} = \frac{1}{\sqrt{N_G}} \sum_{\boldsymbol{R}_G} e^{\mathrm{i}(\pm K+\boldsymbol{p})\cdot(\boldsymbol{R}_G+\boldsymbol{\tau}_\sigma)} c^\dagger_{\boldsymbol{R}_G+\boldsymbol{\tau}_\sigma,s}\,, \tag{A1}$$

where N_G is the number of lattice points for graphene, $\boldsymbol{R}_G$ labels the Bravais lattice points of graphene, $\boldsymbol{\tau}_\sigma$ labels the vector for the sublattice, and $c^\dagger_{\boldsymbol{R}_G+\boldsymbol{\tau}_\sigma,s}$ creates an electron with p_z orbital and spin s at $\boldsymbol{R}_G+\boldsymbol{\tau}_\sigma$. We note that c here is for the electron basis of the original graphene following the notation in Ref. [125], not to be confused with the c modes in Eq. (35).

Now we rotate the graphene by a generic angle θ counterclockwisely about the out-of-plane axis (denoted as C_θ) and shift the graphene along the out-of-plane axis by d_z, we have

$$T_{d_z} C_\theta c^\dagger_{\boldsymbol{R}_G+\boldsymbol{\tau}_\sigma} C_\theta^{-1} T_{d_z}^{-1} = c^\dagger_{d_z,C_\theta(\boldsymbol{R}_G+\boldsymbol{\tau}_\sigma)} e^{-\mathrm{i}\frac{s_z}{2}\theta} \tag{A2}$$

and

$$T_{d_z} C_\theta c^\dagger_{\pm K+\boldsymbol{p},\sigma} C_\theta^{-1} T_{d_z}^{-1} = \frac{1}{\sqrt{N_G}} \sum_{C_\theta \boldsymbol{R}_G} e^{\mathrm{i}(\pm C_\theta K + C_\theta \boldsymbol{p})\cdot(C_\theta \boldsymbol{R}_G + C_\theta \boldsymbol{\tau}_\sigma)} c^\dagger_{d_z, C_\theta \boldsymbol{R}_G + C_\theta \boldsymbol{\tau}_\sigma} e^{-\mathrm{i}\frac{s_z}{2}\theta}\,, \tag{A3}$$

where we define

$$\begin{aligned} c^\dagger_{\boldsymbol{R}_G+\boldsymbol{\tau}_\sigma} &= (c^\dagger_{\boldsymbol{R}_G+\boldsymbol{\tau}_\sigma,\uparrow}, c^\dagger_{\boldsymbol{R}_G+\boldsymbol{\tau}_\sigma,\downarrow}), \\ c^\dagger_{\pm K+\boldsymbol{p},\sigma} &= (c^\dagger_{\pm K+\boldsymbol{p},\sigma,\uparrow}, c^\dagger_{\pm K+\boldsymbol{p},\sigma,\downarrow})\,. \end{aligned} \tag{A4}$$

We can then define

$$\begin{aligned} c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},\sigma,s} &= \frac{1}{\sqrt{N_G}} \sum_{C_\theta \boldsymbol{R}_G} e^{\mathrm{i}(\pm C_\theta K+\boldsymbol{p})\cdot(C_\theta \boldsymbol{R}_G + C_\theta \boldsymbol{\tau}_\sigma)} c^\dagger_{d_z, C_\theta \boldsymbol{R}_G + C_\theta \boldsymbol{\tau}_\sigma, s}, \\ c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p}} &= (c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},A,\uparrow}, c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},A,\downarrow}, c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},B,\uparrow}, c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},B,\downarrow}), \end{aligned} \tag{A5}$$

which gives

$$c^\dagger_{d_z,\theta,\pm C_\theta K+C_\theta \boldsymbol{p}} = T_{d_z} C_\theta c^\dagger_{\pm K+\boldsymbol{p}} C_\theta^{-1} T_{d_z}^{-1} \sigma_0 e^{\mathrm{i}\frac{s_z}{2}\theta}\,. \tag{A6}$$

Based on Eq. (A6), we clearly see that the symmetry reps of C_3, C_2, and $\mathcal{T}$ symmetries are the same for $c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},\sigma,s}$ and $c^\dagger_{0,0,\pm K+\boldsymbol{p},\sigma,s}$, since C_3, C_2, and $\mathcal{T}$ commutes with C_θ and T_{d_z}. The symmetry reps of m_z are also the same, except that d_z is flipped by m_z in $c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},\sigma,s}$. Specifically, we have

$$\begin{aligned} C_3 c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p}} C_3^{-1} &= c^\dagger_{d_z,\theta,\pm C_\theta K+C_3\boldsymbol{p}} e^{\pm\mathrm{i}\frac{2\pi}{3}\sigma_z} s_0, \\ C_2 c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p}} C_2^{-1} &= c^\dagger_{d_z,\theta,\mp C_\theta K-\boldsymbol{p}} \sigma_x s_0, \\ \mathcal{T} c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p}} \mathcal{T}^{-1} &= c^\dagger_{d_z,\theta,\mp C_\theta K-\boldsymbol{p}} \sigma_0 \mathrm{i} s_y, \\ m_z c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p}} m_z^{-1} &= c^\dagger_{-d_z,\theta,\pm C_\theta K+\boldsymbol{p}} (-\sigma_0 s_0)\,. \end{aligned} \tag{A7}$$

(Recall that C_3, C_2, and m_z are defined to be spinless operators.) The lattice translations for $c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},\sigma}$ now becomes

$$T_{C_\theta \boldsymbol{R}_G} c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p}} T^{-1}_{C_\theta \boldsymbol{R}_G} = c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p}} e^{-\mathrm{i}(\pm C_\theta K+\boldsymbol{p})\cdot C_\theta \boldsymbol{R}_G}\,. \tag{A8}$$

Now we take the continuum limit, i.e., treating the graphene lattice as a continuous media. Then, $C_\theta \boldsymbol{R}_G \to \boldsymbol{r}$ with $\boldsymbol{r}$ taking continuous values in $\mathbb{R}^2$, $\pm C_\theta K$ and σ become internal degrees of freedom, and $\boldsymbol{p}$ now also takes values in $\mathbb{R}^2$. Specifically, we have

$$c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},\sigma,s} \to c^\dagger_{d_z,\theta,\pm C_\theta K,\boldsymbol{p},\sigma,s}\,. \tag{A9}$$

Symmetric reps of $c^\dagger_{d_z,\theta,\pm C_\theta K,\boldsymbol{p},\sigma,s}$ and $c^\dagger_{d_z,\theta,\pm C_\theta K+\boldsymbol{p},\sigma,s}$ are exactly the same as Eq. (A7) for C_3, C_2, $\mathcal{T}$, and m_z. The translation operation of $c^\dagger_{d_z,\theta,\pm C_\theta K,\boldsymbol{p},\sigma,s}$ becomes continuous as

$$T_{\boldsymbol{r}} c^\dagger_{d_z,\theta,\pm C_\theta K,\boldsymbol{p},\sigma,s} T_{\boldsymbol{r}}^{-1} = c^\dagger_{d_z,\theta,\pm C_\theta K,\boldsymbol{p},\sigma,s} e^{-\mathrm{i}(\pm C_\theta K+\boldsymbol{p})\cdot\boldsymbol{r}}\,. \tag{A10}$$

$\psi^\dagger_{\eta,\boldsymbol{r},l,\sigma,s}$ in Eq. (4) is defined as

$$\psi^\dagger_{\eta,\boldsymbol{r},l,\sigma,s} = c^\dagger_{d_{z,l},\theta_l,\eta C_{\theta_l} K,\boldsymbol{r},\sigma,s}\,, \tag{A11}$$

where θ_l is the twist angle for the lth layer, $d_{z,l}$ is the position of the lth layer along the out-of-plane axis, and

$$c^\dagger_{d_z,\theta,\pm C_\theta K,\boldsymbol{r},\sigma,s} = \frac{1}{\sqrt{\mathcal{A}}} \sum_{\boldsymbol{p}} e^{-\mathrm{i}\boldsymbol{r}\cdot\boldsymbol{p}} c^\dagger_{d_z,\theta,\pm C_\theta K,\boldsymbol{p},\sigma,s}\,. \tag{A12}$$

Then, we have

$$T_{\boldsymbol{r}_0} \psi^\dagger_{\eta,\boldsymbol{r},l,\sigma,s} T^{-1}_{\boldsymbol{r}_0} = \psi^\dagger_{\eta,\boldsymbol{r}+\boldsymbol{r}_0,l,\sigma,s} e^{-\mathrm{i}\eta(C_{\theta_l} K)\cdot\boldsymbol{r}_0}\,. \tag{A13}$$

Eventually, based on Eq. (9), we know that $\boldsymbol{K}_{t/b}$ in Eq. (16) are determined by $\boldsymbol{K}_t = C_{-\theta/2} K$ and $\boldsymbol{K}_b = C_{\theta/2} K$, since we choose $\theta_1 = \theta_3 = -\theta/2$ and $\theta_2 = \theta/2$.

We define

$$\psi^\dagger_{\eta,\boldsymbol{r},l} = (\psi^\dagger_{\eta,\boldsymbol{r},l,A,\uparrow}, \psi^\dagger_{\eta,\boldsymbol{r},l,A,\downarrow}, \psi^\dagger_{\eta,\boldsymbol{r},l,B,\uparrow}, \psi^\dagger_{\eta,\boldsymbol{r},l,B,\downarrow})\,, \tag{A14}$$

and define $\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\tilde{l}}$ and $d^\dagger_{\eta,\boldsymbol{r}}$ by Eqs. (9) and (10). Then, C_3 is represented as

$$\begin{aligned} C_3\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\tilde{l}}C_3^{-1} &= \widetilde{\psi}^\dagger_{\eta,C_3\boldsymbol{r},\tilde{l}}\, e^{\eta i\sigma_z\frac{2\pi}{3}} s_0, \\ C_3 d^\dagger_{\eta,\boldsymbol{r}}C_3^{-1} &= d^\dagger_{\eta,C_3\boldsymbol{r}}\, e^{\eta i\sigma_z\frac{2\pi}{3}} s_0\ ; \end{aligned} \tag{A15}$$

$C_2\mathcal{T}$ is represented as

$$\begin{aligned} C_2\mathcal{T}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\tilde{l}}(C_2\mathcal{T})^{-1} &= \widetilde{\psi}^\dagger_{\eta,-\boldsymbol{r},\tilde{l}}\,\sigma_x i s_y, \\ C_2\mathcal{T} d^\dagger_{\eta,\boldsymbol{r}}(C_2\mathcal{T})^{-1} &= d^\dagger_{\eta,-\boldsymbol{r}}\,\sigma_x i s_y\ ; \end{aligned} \tag{A16}$$

we can define an effective C_{2x} as

$$\begin{aligned} C_{2x}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\tilde{l}}(C_{2x})^{-1} &= \sum_{\tilde{l}'}\widetilde{\psi}^\dagger_{\eta,C_{2x}\boldsymbol{r},\tilde{l}'}(-\sigma_x)s_0\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}_{\tilde{l}'\tilde{l}}, \\ C_{2x}d^\dagger_{\eta,\boldsymbol{r}}(C_{2x})^{-1} &= d^\dagger_{-\eta,C_{2x}\boldsymbol{r}}\,\sigma_x s_0\ , \end{aligned} \tag{A17}$$

we can also define an effective P as

$$\begin{aligned} P\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\tilde{l}}P^{-1} &= \sum_{\tilde{l}'}\widetilde{\psi}^\dagger_{\eta,-\boldsymbol{r},\tilde{l}'}\eta\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}_{\tilde{l}'\tilde{l}}, \\ Pd^\dagger_{\eta,\boldsymbol{r}}P^{-1} &= \eta d^\dagger_{-\eta,-\boldsymbol{r}}\ , \end{aligned} \tag{A18}$$

then $C_{2x}P$ is represented as

$$\begin{aligned} C_{2x}P\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\tilde{l}}(C_{2x}P)^{-1} &= \widetilde{\psi}^\dagger_{\eta,-C_{2x}\boldsymbol{r},\tilde{l}}\,(-1)^{\tilde{l}}\eta(-\sigma_x)s_0, \\ C_{2x}Pd^\dagger_{\eta,\boldsymbol{r}}(C_{2x}P)^{-1} &= d^\dagger_{\eta,-C_{2x}\boldsymbol{r}}\,\eta\sigma_x s_0 \end{aligned} \tag{A19}$$

with $(-1)^t = -(-1)^b = 1$ and $C_{2x}\boldsymbol{r} = (x,-y)^T$; the rep of $T_{\boldsymbol{R}}$ is in Eq. (16); $\mathcal{T}$ is represented as

$$\begin{aligned} \mathcal{T}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\tilde{l}}\mathcal{T}^{-1} &= \widetilde{\psi}^\dagger_{-\eta,\boldsymbol{r},\tilde{l}}\,\sigma_0 i s_y, \\ \mathcal{T}d^\dagger_{\eta,\boldsymbol{r}}\mathcal{T}^{-1} &= d^\dagger_{-\eta,\boldsymbol{r}}\,\sigma_0 i s_y\ ; \end{aligned} \tag{A20}$$

$\mathcal{C}$ is represented as

$$\mathcal{C}\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\tilde{l}}\mathcal{C}^{-1} = \widetilde{\psi}^T_{\eta,\boldsymbol{r},\tilde{l}}, \quad \mathcal{C}d^\dagger_{\eta,\boldsymbol{r}}\mathcal{C}^{-1} = d^T_{\eta,\boldsymbol{r}} \tag{A21}$$

with $\mathcal{C}^2 = 1$. The symmetry reps in the momentum space can be naturally obtained by using Eq. (21).

The symmetry properties of $:\rho(\boldsymbol{r}):$ are

$$\begin{aligned} \mathcal{T} : \rho(\boldsymbol{r}) : \mathcal{T}^{-1} &=: \rho(\boldsymbol{r}) :\\ C_3 : \rho(\boldsymbol{r}) : C_3^{-1} &=: \rho(C_3\boldsymbol{r}) :\\ C_2\mathcal{T} : \rho(\boldsymbol{r}) : (C_2\mathcal{T})^{-1} &=: \rho(-\boldsymbol{r}) :\\ m_z : \rho(\boldsymbol{r}) : m_z^{-1} &=: \rho(\boldsymbol{r}) :\\ T_{\boldsymbol{R}} : \rho(\boldsymbol{r}) : T_{\boldsymbol{R}}^{-1} &=: \rho(\boldsymbol{r}+\boldsymbol{R}) :\\ C_{2x}P : \rho(\boldsymbol{r}) : (C_{2x}P)^{-1} &=: \rho(-C_{2x}\boldsymbol{r}) :\\ \mathcal{C} : \rho(\boldsymbol{r}) : (\mathcal{C})^{-1} &= - : \rho(\boldsymbol{r}) : . \end{aligned} \tag{A22}$$

Combined with the fact that $\rho(\boldsymbol{r})$ is invariant under the spin charge U(2) in each valley, we have

$$\begin{aligned} [\mathcal{T}, H_{\text{int}}] = [C_3, H_{\text{int}}] &= [C_2\mathcal{T}, H_{\text{int}}] = [m_z, H_{\text{int}}] \\ = [T_{\boldsymbol{R}}, H_{\text{int}}] &= [C_{2x}P, H_{\text{int}}] = [\mathcal{C}, H_{\text{int}}] \\ = [\text{U}(2)\times\text{U}(2), H_{\text{int}}] &= 0\ . \end{aligned} \tag{A23}$$

The symmetry reps of f and c are particularly important for deriving the low-energy effective model. The relevant high-symmetry points in MBZ for $H_{0,\text{TBG}}$ are Γ_M, K_M, and M_M (shown in Fig. 2). Based on the origin of the f modes, we know that the symmetry reps of f should carry the symmetry reps of the nearly flat bands at K_M and M_M, and carry one 2D irreducible rep (irrep) of the remote bands at Γ_M. According to Eq. (34), the symmetry rep of f is determined by the form of $\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})$. Then, we require $\widetilde{v}_{\eta,f}$ to guarantee the following the rep of f

$$\begin{aligned} C_2\mathcal{T}f^\dagger_{+,\boldsymbol{k}}(C_2\mathcal{T})^{-1} &= f^\dagger_{+,\boldsymbol{k}}\tau_x i s_y, \\ C_3 f^\dagger_{+,\boldsymbol{k}}(C_3)^{-1} &= f^\dagger_{+,C_3\boldsymbol{k}}e^{i\tau_z\frac{2\pi}{3}}s_0, \\ C_{2x}f^\dagger_{+,\boldsymbol{k}}C_{2x}^{-1} &= f^\dagger_{+,C_{2x}\boldsymbol{k}}\tau_x s_0, \\ Pf^\dagger_{+,\boldsymbol{k}}P^{-1} &= f^\dagger_{+,-\boldsymbol{k}}i\tau_z s_0, \\ T_{\boldsymbol{R}}f^\dagger_{+,\boldsymbol{k}}T_{\boldsymbol{R}}^{-1} &= f^\dagger_{+,\boldsymbol{k}}e^{-i(\boldsymbol{K}_b+\boldsymbol{q}_2+\boldsymbol{k})\cdot\boldsymbol{R}}, \\ f^\dagger_{-,\boldsymbol{k}} &= \mathcal{T}f^\dagger_{+,-\boldsymbol{k}}\mathcal{T}^{-1}\tau_0(-is_y)\ . \end{aligned} \tag{A24}$$

The spinlesss parts of the reps here are the same as those of f in Ref. [125], except the extra $e^{-i(\boldsymbol{K}_b+\boldsymbol{q}_2)\cdot\boldsymbol{R}}$ factor in the rep of translation, which will be discussed carefully below. Furthermore, to guarantee the exponential decay of the Wannier functions of f modes, we have to require $\widetilde{v}_{\eta,f}$ to be smooth while keeping

$$[\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k}+\boldsymbol{G})]_{Q\sigma} = [\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]_{Q-G\sigma}\ . \tag{A25}$$

The existence of such smooth $\widetilde{v}_{\eta,f}$ is numerically verified in Ref. [125]. Note that the 2D irrep carried by f_{+,Γ_M} is just the spinless Γ_3 if we only consider D_3 (spanned by C_3 and C_{2x}) [177].

Now we show that the extra $e^{-i(\boldsymbol{K}_b+\boldsymbol{q}_2)\cdot\boldsymbol{R}}$ factor of f modes under Moiré lattice translations [shown in Eq. (A24)] can be safely neglected for any values of angle, similar to Ref. [125]. First, we show under certain special values of the angles, we can make $\boldsymbol{K}_b+\boldsymbol{q}_2$ a Moiré reciprocal lattice vector, and thus $e^{-i(\boldsymbol{K}_b+\boldsymbol{q}_2)\cdot\boldsymbol{R}}$ becomes 1. Combined with Eqs. (6), (17), and (20), we have

$$\begin{aligned} \boldsymbol{K}_b+\boldsymbol{q}_2 &\in \boldsymbol{b}_{M,1}\mathbb{Z}+\boldsymbol{b}_{M,2}\mathbb{Z} \\ &\Leftrightarrow \frac{1}{2}(\cot(\theta/2)-\sqrt{3},0) \\ &\in \left\{\left(\frac{\sqrt{3}}{2}n_1+\sqrt{3}n_2, -\frac{3}{2}n_1\right)\middle| n_1,n_2\in\mathbb{Z}\right\} \\ &\Leftrightarrow \frac{1}{2}\cot(\theta/2)-\frac{1}{2}\sqrt{3}\in\sqrt{3}\mathbb{Z}\ . \end{aligned} \tag{A26}$$

Therefore we can choose θ to satisfy

$$\frac{1}{2}\cot\left(\frac{\theta}{2}\right)-\frac{\sqrt{3}}{2} = 0 \bmod \sqrt{3} \tag{A27}$$

such that $e^{-i(\boldsymbol{K}_b+\boldsymbol{q}_2)\cdot\boldsymbol{R}}$ becomes 1. Second, even if θ does not satisfy Eq. (A27), we can define an operation as

$$Y_{\boldsymbol{R}}f^\dagger_{\eta,\boldsymbol{k}}Y_{\boldsymbol{R}}^{-1} = f^\dagger_{\eta,\boldsymbol{k}}e^{i\eta(\boldsymbol{K}_b+\boldsymbol{q}_2)\cdot\boldsymbol{R}}\ , \tag{A28}$$

where $Y_{\boldsymbol{R}}$ belongs to the valley U(1), which is obeyed by the system. Then, we can redefine $Y_{\boldsymbol{R}}T_{\boldsymbol{R}}$ as the new

lattice translation, which does not has the $e^{-\mathrm{i}(\boldsymbol{K}_b+\boldsymbol{q}_2)\cdot\boldsymbol{R}}$ factor in Eq. (A24). This is what was done in Ref. [125]. For the convenience of the derivation in this study, we simply choose θ to have the value in Eq. (44), which approximately satisfies Eq. (A27). Nevertheless, Eq. (44) is not required for omitting the $e^{-\mathrm{i}(\boldsymbol{K}_b+\boldsymbol{q}_2)\cdot\boldsymbol{R}}$ factor of the lattice translation in Eq. (A24).

At Γ_M, the remote bands have one remaining 2D irrep (also corresponding to Γ_3 of D_3) of the remote bands at Γ_M, and the near-flat bands have two 1D irreps (corresponding to Γ_1 and Γ_2 of D_3). They should be carried by the c modes. As a result, the reps furnished by the c modes are

$$\begin{aligned}
C_2\mathcal{T}c^\dagger_{+,\boldsymbol{k}}(C_2\mathcal{T})^{-1} &= c^\dagger_{+,\boldsymbol{k}}\begin{pmatrix}\tau_x & \\ & \tau_x\end{pmatrix}\mathrm{i}s_y,\\
C_3c^\dagger_{+,\boldsymbol{k}}(C_3)^{-1} &= c^\dagger_{+,C_3\boldsymbol{k}}\begin{pmatrix}e^{\mathrm{i}\tau_z\frac{2\pi}{3}} & \\ & \tau_0\end{pmatrix}s_0,\\
C_{2x}c^\dagger_{+,\boldsymbol{k}}C_{2x}^{-1} &= c^\dagger_{+,C_{2x}\boldsymbol{k}}\begin{pmatrix}\tau_x & \\ & \tau_x\end{pmatrix}s_0,\\
Pc^\dagger_{+,\boldsymbol{k}}P^{-1} &= c^\dagger_{+,-\boldsymbol{k}}\begin{pmatrix}-\mathrm{i}\tau_z & \\ & -\mathrm{i}\tau_z\end{pmatrix}s_0,\\
T_{\boldsymbol{R}}c^\dagger_{+,\boldsymbol{k}}T_{\boldsymbol{R}}^{-1} &= c^\dagger_{+,\boldsymbol{k}}e^{-\mathrm{i}(\boldsymbol{K}_b+\boldsymbol{q}_2+\boldsymbol{k})\cdot\boldsymbol{R}},\\
c^\dagger_{-,\boldsymbol{k}} &= \mathcal{T}c^\dagger_{+,-\boldsymbol{k}}\mathcal{T}^{-1}\mathbb{1}_4(-\mathrm{i}s_y)\,,
\end{aligned}\tag{A29}$$

where $c^\dagger_{\eta,\boldsymbol{k}}=(c^\dagger_{\eta,\boldsymbol{k},\Gamma_3},c^\dagger_{\eta,\boldsymbol{k},\Gamma_1\Gamma_2})$, $c^\dagger_{\eta,\boldsymbol{k},\Gamma_3}=(c^\dagger_{\eta,\boldsymbol{k},1},c^\dagger_{\eta,\boldsymbol{k},2})$, and $c^\dagger_{\eta,\boldsymbol{k},\Gamma_1\Gamma_2}=(c^\dagger_{\eta,\boldsymbol{k},3},c^\dagger_{\eta,\boldsymbol{k},4})$. Note that $\tau_{0,x,y,z}$ carries the index α for $f^\dagger_{\eta,\boldsymbol{k}}$ and carries the index β for $c^\dagger_{\eta,\boldsymbol{k},\Gamma_3}$ and $c^\dagger_{\eta,\boldsymbol{k},\Gamma_1\Gamma_2}$. According to Eq. (35), Eq. (A29) is guaranteed by choosing a special $\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})$ with $|\boldsymbol{k}|<\Lambda_c$. Furthermore, in order to guarantee the resultant effective Hamiltonian to have a smooth matrix rep, we need to require $\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})$ to be smooth. Such required $\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})$ always exists for $|\boldsymbol{k}|<\Lambda_c$. The reason is that $\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})$ is effectively defined on an open manifold instead of a torus, as we do not impose any relation between $\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})$ and $\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k}+\boldsymbol{G})$ if both $\boldsymbol{k}$ and $\boldsymbol{k}+\boldsymbol{G}$ have magnitudes smaller than Λ_c.

We would like the compare the lattice translations of the f, c and d modes after considering Eq. (A27), which read

$$\begin{aligned}
T_{\boldsymbol{R}}f^\dagger_{\eta,\boldsymbol{k}}T_{\boldsymbol{R}}^{-1} &= f^\dagger_{\eta,\boldsymbol{k}}e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}},\\
T_{\boldsymbol{R}}c^\dagger_{\eta,\boldsymbol{k}}T_{\boldsymbol{R}}^{-1} &= c^\dagger_{\eta,\boldsymbol{k}}e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}},\\
T_{\boldsymbol{R}}d^\dagger_{\eta,\boldsymbol{p}}T_{\boldsymbol{R}}^{-1} &= d^\dagger_{\eta,\boldsymbol{p}}e^{-\mathrm{i}(\eta\boldsymbol{K}_t+\boldsymbol{p})\cdot\boldsymbol{R}}\\
&= d^\dagger_{\eta,\boldsymbol{p}}e^{-\mathrm{i}(\eta\boldsymbol{K}_b+\eta\boldsymbol{q}_2+(\boldsymbol{p}+\eta K_M))\cdot\boldsymbol{R}}\\
&= d^\dagger_{\eta,\boldsymbol{p}}e^{-\mathrm{i}(\boldsymbol{p}+\eta K_M)\cdot\boldsymbol{R}}\,,
\end{aligned}\tag{A30}$$

where we used Eqs. (A24), (A29), (6), (16)–(18) and Fig. 2. According to Eq. (A30), $c^\dagger_{\eta,\boldsymbol{k}}$ and $f^\dagger_{\eta,\boldsymbol{k}}$ transforms in the same way under the Moiré lattice translations. It means that $c^\dagger_{\eta,\boldsymbol{k}}$ is around the Γ_M point of the $f^\dagger_{\eta,\boldsymbol{k}}$ modes for small $\boldsymbol{k}$. On the other hand, according to Eq. (A30), $d^\dagger_{\eta,\boldsymbol{k}-\eta K_M}$ and $f^\dagger_{\eta,\boldsymbol{k}}$ transforms in the same way under the Moiré lattice translations, i.e., $\boldsymbol{k}$ in $d^\dagger_{\eta,\boldsymbol{k}-\eta K_M}$ is the same as $\boldsymbol{k}$ in $f^\dagger_{\eta,\boldsymbol{k}}$. Thus $d^\dagger_{\eta,\boldsymbol{p}}$ with small $\boldsymbol{p}$ are around the ηK_M point of the f modes.

Symmetry properties of $w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r})$ [defined in Eq. (47)] are listed below.

$$\begin{aligned}
&w_{-\alpha\widetilde{l}\sigma}(\boldsymbol{r}) = w^*_{+\alpha\widetilde{l}\sigma}(\boldsymbol{r}),\\
&w_{+1tA}\big(C_3^{-1}\boldsymbol{r}\big) = w_{+1tA}(\boldsymbol{r}),\\
&w_{+1tB}\big(C_3^{-1}\boldsymbol{r}\big) = w_{+1tB}(\boldsymbol{r})e^{-\mathrm{i}2\pi/3},\\
&w_{+1t\sigma}(\boldsymbol{r}) = -\mathrm{i}w^*_{+1t\sigma}\big(C_{2x}^{-1}\boldsymbol{r}\big),\\
&w_{+1b\sigma}(\boldsymbol{r}) = -w^*_{+1t\sigma}\big(-C_{2x}^{-1}\boldsymbol{r}\big),\\
&w_{+2\widetilde{l}\sigma}(\boldsymbol{r}) = w^*_{+1\widetilde{l}\overline{\sigma}}(-\boldsymbol{r}),
\end{aligned}\tag{A31}$$

where $\overline{\sigma}=A/B$ for $\sigma=B/A$.

APPENDIX B: MORE DETAILS ON THE SINGLE-PARTICLE *f-c-d* MODEL

1. More details on the *f-d* coupling around ηK_M

In general, the f-d coupling reads

$$H_{0,\eta,fd}=\sum_{\boldsymbol{k}\in\mathrm{MBZ}}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|<\Lambda_d}f^\dagger_{\eta,\boldsymbol{k}}\widetilde{h}_{\eta,fd}(\boldsymbol{k},\boldsymbol{p})\otimes s_0d_{\eta,\boldsymbol{p}}+\mathrm{H.c.},\tag{B1}$$

where we have used $\mathrm{U}(2)\times\mathrm{U}(2)$ to rule out the intervalley coupling and the spin-orbit coupling, and $\widetilde{h}_{\eta,fd}(\boldsymbol{k}+\boldsymbol{G},\boldsymbol{p})=\widetilde{h}_{\eta,fd}(\boldsymbol{k},\boldsymbol{p})$ for any Moiré reciprocal lattice vector $\boldsymbol{G}$. $H_{0,\eta,fd}$ preserves the Moiré lattice translation $T_{\boldsymbol{R}}$, $C_2\mathcal{T}$, C_3, $C_{2x}P$, TR, and the combination of m_z and $\mathcal{E}\to-\mathcal{E}$. Then, according to Eqs. (A24) and (16), $T_{\boldsymbol{R}}$ gives

$$\begin{aligned}
&T_{\boldsymbol{R}}H_{0,\eta,fd}T_{\boldsymbol{R}}^{-1}=H_{0,\eta,fd}\ \forall\boldsymbol{R}\\
\Leftrightarrow\ &\widetilde{h}_{\eta,fd}(\boldsymbol{k},\boldsymbol{p})e^{-\mathrm{i}(\eta\boldsymbol{q}_2+\eta\boldsymbol{q}_3+\boldsymbol{k}-\boldsymbol{p})\cdot\boldsymbol{R}}=\widetilde{h}_{\eta,fd}(\boldsymbol{k},\boldsymbol{p})\ \forall\boldsymbol{R}\\
\Leftrightarrow\ &\widetilde{h}_{\eta,fd}(\boldsymbol{k},\boldsymbol{p})=\sum_{\boldsymbol{G}}\delta_{\boldsymbol{p},\boldsymbol{k}-\eta K_M+\boldsymbol{G}}\widetilde{h}_{\eta,fd}(\boldsymbol{k},\boldsymbol{k}-\eta K_M+\boldsymbol{G})\\
\Leftrightarrow\ &H_{0,\eta,fd}=\sum_{\boldsymbol{k}\in\mathrm{MBZ}}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|<\Lambda_d}f^\dagger_{\eta,\boldsymbol{k}}\sum_{\boldsymbol{G}}\delta_{\boldsymbol{p},\boldsymbol{k}-\eta K_M+\boldsymbol{G}}\widetilde{h}_{\eta,fd}(\boldsymbol{k},\boldsymbol{k}-\eta K_M+\boldsymbol{G})\otimes s_0d_{\eta,\boldsymbol{p}}+\mathrm{H.c.}
\end{aligned}$$

$$\Leftrightarrow H_{0,\eta,fd} = \sum_{\boldsymbol{p}'} \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|<\Lambda_d} f^\dagger_{\eta,\boldsymbol{p}'} \delta_{\boldsymbol{p},\boldsymbol{p}'-\eta K_M} \widetilde{h}_{\eta,fd}(\boldsymbol{p}', \boldsymbol{p}' - \eta K_M) \otimes s_0 d_{\eta,\boldsymbol{p}} + \text{H.c.}$$

$$\Leftrightarrow H_{0,\eta,fd} = \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|<\Lambda_d} f^\dagger_{\eta,\boldsymbol{p}+\eta K_M} h_{\eta,fd}(\boldsymbol{p}) \otimes s_0 d_{\eta,\boldsymbol{p}} + \text{H.c.}, \tag{B2}$$

where $h_{\eta,fd}(\boldsymbol{p}) = \widetilde{h}_{\eta,fd}(\boldsymbol{p} + \eta K_M, \boldsymbol{p})$. Since we are considering the coupling around ηK_M, we only consider $\boldsymbol{p}$ to the first order. Then, $C_2\mathcal{T}$ and C_3 give

$$\begin{aligned} \{C_2\mathcal{T}: \ & \tau_x h^*_{+,fd}(\boldsymbol{p})\tau_x = h_{+,fd}(\boldsymbol{p}) \\ C_3: \ & e^{i\tau_z 2\pi/3} h_{+,fd}(\boldsymbol{p}) e^{-i\tau_z 2\pi/3} = h_{+,fd}(C_3\boldsymbol{p}) \\ & \Leftrightarrow h_{+,fd}(\boldsymbol{p}) = \widetilde{M}_1\tau_0 + i\widetilde{M}'_1\tau_z + v_1(p_x + ip_y)(\tau_x - i\tau_y) + v'_1(p_x - ip_y)(\tau_x + i\tau_y) + O(p^2)\,, \end{aligned} \tag{B3}$$

where $\widetilde{M}_1$, $\widetilde{M}'_1$ are real and v_1, v'_1 are complex. Furthermore, $C_{2x}P$ gives

$$\tau_y h_{+,fd}(\boldsymbol{p})\tau_x = -h_{+,fd}(C_{2y}\boldsymbol{p}) \Leftrightarrow \widetilde{M}_1 = \widetilde{M}'_1 \ \& \ v_1 = v'^*_1 = |v_1| e^{i\frac{\pi}{4}}\,. \tag{B4}$$

Therefore we have

$$h_{+,fd}(\boldsymbol{p}) = \widetilde{M}_1(\tau_0 + i\tau_z) + v_{fd,1}[e^{i\pi/4}(p_x + ip_y)(\tau_x - i\tau_y) + e^{-i\pi/4}(p_x - ip_y)(\tau_x + i\tau_y)] + O(p^2)\,, \tag{B5}$$

where the combination of m_z and $\mathcal{E} \to -\mathcal{E}$ requires that $\widetilde{M}_1$ and $v_{fd,1}$ are odd in $\mathcal{E}$. To further simplify $h_{+,fd}(\boldsymbol{p})$, we project $H_{0,\mathcal{E},+}$ in Eq. (24) to the f and d basis at K_M. Explicitly, we have

$$\begin{aligned} H_{0,\mathcal{E},+} &= \sum_{\boldsymbol{k}\in\text{MBZ}} \sum_{\boldsymbol{Q}\in\mathcal{Q}_+} \sum_{\sigma,s} \frac{\mathcal{E}}{2} \widetilde{\psi}^\dagger_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} d_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} + \text{H.c.} \\ &= \sum_{\boldsymbol{k}\in\text{MBZ}} \sum_{\boldsymbol{Q}\in\mathcal{Q}_+} \sum_{\sigma,s} \frac{\mathcal{E}}{2} \sum_{\alpha} f^\dagger_{+,\boldsymbol{k},\alpha,s} [\widetilde{v}_{+,f,\alpha}(\boldsymbol{k})]^*_{\boldsymbol{Q}\sigma} d_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} + \text{H.c.} + \ldots \\ &= \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|<\Lambda_d} \sum_{\sigma,s} \frac{\mathcal{E}}{2} \sum_{\alpha} f^\dagger_{+,\boldsymbol{p}+K_M,\alpha,s} \sum_{\boldsymbol{Q}\in\mathcal{Q}_+} [\widetilde{v}_{+,f,\alpha}(\boldsymbol{p} + K_M)]^*_{\boldsymbol{Q}\sigma} d_{+,\boldsymbol{p}+K_M-\boldsymbol{Q},\sigma,s} + \text{H.c.} + \ldots \\ &= \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|<\Lambda_d} \sum_{\sigma,s} \sum_{\alpha} f^\dagger_{+,\boldsymbol{p}+K_M,\alpha,s} \frac{\mathcal{E}}{2} [\widetilde{v}_{+,f,\alpha}(\boldsymbol{p} + K_M)]^*_{\boldsymbol{Q}=K_M,\sigma} d_{+,\boldsymbol{p},\sigma,s} + \text{H.c.} + \ldots, \end{aligned} \tag{B6}$$

where we have used Eq. (38) for the second equality, and $\ldots$ contains high-energy modes. By comparing Eq. (B6) to Eq. (B2), we arrive at

$$[h_{+,fd}(\boldsymbol{p})]_{\alpha\sigma} = \frac{\mathcal{E}}{2} [\widetilde{v}_{+,f,\alpha}(\boldsymbol{p} + K_M)]^*_{\boldsymbol{Q}=K_M,\sigma} \approx \frac{\mathcal{E}}{2} [\widetilde{v}_{+,f,\alpha}(K_M)]^*_{\boldsymbol{Q}=K_M,\sigma}\,, \tag{B7}$$

where we neglect the momentum dependence of $\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})$ in the last step since Ref. [125] shows the momentum dependence of $\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})$ should be small as the f modes have small Wannier spread and have Wannier center at 1a position. Owing to Eqs. (B7) and (B5), we get

$$\widetilde{M}_1 = M_1\mathcal{E}, \quad v_{fd,1} = 0 \tag{B8}$$

with the value of M_1 in Table I. Therefore combined with the $\mathcal{T}$ symmetry and the extra exponential decay factor, we have Eq. (51) as the leading-order term of the f-d coupling around ηK_M.

2. More details on f-d and c-d couplings around Γ_M

In this part, we provide more details on the how we project out the f-d and c-d couplings around Γ_M. We will focus on the $+$ valley, since the Hamiltonian at the $-$ valley can be obtained by the TR symmtry.

According to Eq. (23), the d modes with lowest energies at Γ_M in the $+$ valley are $d^\dagger_{+,0,\boldsymbol{q}_1}$, $d^\dagger_{+,0,\boldsymbol{q}_2}$ and $d^\dagger_{+,0,\boldsymbol{q}_3}$, which gives energies ± 1 owing to Eq. (6). Then, we consider the following f-d and c-d couplings around Γ:

$$(c^\dagger_{+,\boldsymbol{k},\Gamma_3} \quad c^\dagger_{+,\boldsymbol{k},\Gamma_1\Gamma 2} \quad f^\dagger_{+,\boldsymbol{k}}) \begin{pmatrix} h_{+,c-d,\Gamma} \\ h_{+,f-d,\Gamma} \end{pmatrix} \otimes s_0 \begin{pmatrix} d_{+,\boldsymbol{k},\boldsymbol{q}_1} \\ d_{+,\boldsymbol{k},\boldsymbol{q}_2} \\ d_{+,\boldsymbol{k},\boldsymbol{q}_3} \end{pmatrix}, \tag{B9}$$

where f and c are defined in Eqs. (34) and (35), respectively, $h_{+,f-d,\Gamma}$ is a 2×6 matrix, and $h_{+,c-d,\Gamma}$ is a 4×6 matrix. Here we neglect the momentum dependence of the f-d and c-d coupling away from Γ_M since the matrix rep of Eq. (24) is momentum independent. To obtain the forms of $h_{+,f-d,\Gamma}$ and $h_{+,c-d,\Gamma}$, we use Eq. (38) to project Eq. (24) to the low-energy modes around Γ_M:

$$\begin{aligned}
H_{0,\mathcal{E},+} &= \sum_{\boldsymbol{k}\in \mathrm{MBZ}} \sum_{\boldsymbol{Q}\in\mathcal{Q}_+} \sum_{\sigma,s} \frac{\mathcal{E}}{2} \widetilde{\psi}^\dagger_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} d_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} + \mathrm{H.c.} \\
&= \sum_{\boldsymbol{k}}^{|\boldsymbol{k}|<\Lambda_c} \sum_{\boldsymbol{Q}\in\mathcal{Q}_+} \sum_{\sigma,s} \frac{\mathcal{E}}{2} \left(\sum_\alpha f^\dagger_{+,\boldsymbol{k},\alpha,s} [\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]^*_{\boldsymbol{Q}\sigma} d_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} + \sum_\beta c^\dagger_{+,\boldsymbol{k},\beta,s} [\widetilde{u}_{\eta,f,\beta}(\boldsymbol{k})]^*_{\boldsymbol{Q}\sigma} d_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} \right) + \mathrm{H.c.} + \ldots \\
&= \sum_{\boldsymbol{k}}^{|\boldsymbol{k}|<\Lambda_c} \sum_{\boldsymbol{Q}=\boldsymbol{q}_1,\boldsymbol{q}_3,\boldsymbol{q}_3} \sum_{\sigma,s} \left(\sum_\alpha f^\dagger_{+,\boldsymbol{k},\alpha,s} \frac{\mathcal{E}}{2} [\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]^*_{\boldsymbol{Q}\sigma} d_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} + \sum_\beta c^\dagger_{+,\boldsymbol{k},\beta,s} \frac{\mathcal{E}}{2} [\widetilde{u}_{\eta,f,\beta}(\boldsymbol{k})]^*_{\boldsymbol{Q}\sigma} d_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} \right) + \mathrm{H.c.} + \ldots \\
&= \sum_{\boldsymbol{k}}^{|\boldsymbol{k}|<\Lambda_c} \sum_{\boldsymbol{Q}=\boldsymbol{q}_1,\boldsymbol{q}_3,\boldsymbol{q}_3} \sum_{\sigma,s} \left(\sum_\alpha f^\dagger_{+,\boldsymbol{k},\alpha,s} \frac{\mathcal{E}}{2} [\widetilde{v}_{\eta,f,\alpha}(0)]^*_{\boldsymbol{Q}\sigma} d_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} + \sum_\beta c^\dagger_{+,\boldsymbol{k},\beta,s} \frac{\mathcal{E}}{2} [\widetilde{u}_{\eta,f,\beta}(0)]^*_{\boldsymbol{Q}\sigma} d_{+,\boldsymbol{k},\boldsymbol{Q},\sigma,s} \right) + \mathrm{H.c.} + \ldots ,
\end{aligned} \tag{B10}$$

where we neglect the momentum dependence of $\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})$ and $\widetilde{u}_{\eta,f,\beta}(\boldsymbol{k})$ again, and … contains high-energy modes. By comparing Eq. (B10) to Eq. (B9), we can get the forms of $h_{+,f-d,\Gamma}$ and $h_{+,c-d,\Gamma}$, which read

$$h_{+,f-d,\Gamma} = \frac{\mathcal{E}}{2} \begin{pmatrix} [\widetilde{v}_{+,f,1}(0)]_{\boldsymbol{q}_1,A} & [\widetilde{v}_{+,f,1}(0)]_{\boldsymbol{q}_1,B} & [\widetilde{v}_{+,f,1}(0)]_{\boldsymbol{q}_2,A} & [\widetilde{v}_{+,f,1}(0)]_{\boldsymbol{q}_2,B} & [\widetilde{v}_{+,f,1}(0)]_{\boldsymbol{q}_3,A} & [\widetilde{v}_{+,f,1}(0)]_{\boldsymbol{q}_3,B} \\ [\widetilde{v}_{+,f,2}(0)]_{\boldsymbol{q}_1,A} & [\widetilde{v}_{+,f,2}(0)]_{\boldsymbol{q}_1,B} & [\widetilde{v}_{+,f,2}(0)]_{\boldsymbol{q}_2,A} & [\widetilde{v}_{+,f,2}(0)]_{\boldsymbol{q}_2,B} & [\widetilde{v}_{+,f,2}(0)]_{\boldsymbol{q}_3,A} & [\widetilde{v}_{+,f,2}(0)]_{\boldsymbol{q}_3,B} \end{pmatrix}^* \tag{B11}$$

and

$$h_{+,c-d,\Gamma} = \frac{\mathcal{E}}{2} \begin{pmatrix} [\widetilde{u}_{+,c,1}(0)]_{\boldsymbol{q}_1,A} & [\widetilde{u}_{+,c,1}(0)]_{\boldsymbol{q}_1,B} & [\widetilde{u}_{+,c,1}(0)]_{\boldsymbol{q}_2,A} & [\widetilde{u}_{+,c,1}(0)]_{\boldsymbol{q}_2,B} & [\widetilde{u}_{+,c,1}(0)]_{\boldsymbol{q}_3,A} & [\widetilde{u}_{+,c,1}(0)]_{\boldsymbol{q}_3,B} \\ [\widetilde{u}_{+,c,2}(0)]_{\boldsymbol{q}_1,A} & [\widetilde{u}_{+,c,2}(0)]_{\boldsymbol{q}_1,B} & [\widetilde{u}_{+,c,2}(0)]_{\boldsymbol{q}_2,A} & [\widetilde{u}_{+,c,2}(0)]_{\boldsymbol{q}_2,B} & [\widetilde{u}_{+,c,2}(0)]_{\boldsymbol{q}_3,A} & [\widetilde{u}_{+,c,2}(0)]_{\boldsymbol{q}_3,B} \\ [\widetilde{u}_{+,c,3}(0)]_{\boldsymbol{q}_1,A} & [\widetilde{u}_{+,c,3}(0)]_{\boldsymbol{q}_1,B} & [\widetilde{u}_{+,c,3}(0)]_{\boldsymbol{q}_2,A} & [\widetilde{u}_{+,c,3}(0)]_{\boldsymbol{q}_2,B} & [\widetilde{u}_{+,c,3}(0)]_{\boldsymbol{q}_3,A} & [\widetilde{u}_{+,c,3}(0)]_{\boldsymbol{q}_3,B} \\ [\widetilde{u}_{+,c,4}(0)]_{\boldsymbol{q}_1,A} & [\widetilde{u}_{+,c,4}(0)]_{\boldsymbol{q}_1,B} & [\widetilde{u}_{+,c,4}(0)]_{\boldsymbol{q}_2,A} & [\widetilde{u}_{+,c,4}(0)]_{\boldsymbol{q}_2,B} & [\widetilde{u}_{+,c,4}(0)]_{\boldsymbol{q}_3,A} & [\widetilde{u}_{+,c,4}(0)]_{\boldsymbol{q}_3,B} \end{pmatrix}^* . \tag{B12}$$

On the other hand, according to Eqs. (40)–(42), the f and c block without the f-d and c-d corrections [i.e., the f and c block that comes from $H_{0,\mathrm{TBG}}$ in Eq. (22)] reads

$$\begin{pmatrix} c^\dagger_{+,\boldsymbol{k},\Gamma_3} & c^\dagger_{+,\boldsymbol{k},\Gamma_1\Gamma 2} & f^\dagger_{+,\boldsymbol{k}} \end{pmatrix} h_{+,0}(\boldsymbol{k}) \otimes s_0 \begin{pmatrix} c_{+,\boldsymbol{k},\Gamma_3} \\ c_{+,\boldsymbol{k},\Gamma_1\Gamma 2} \\ f_{+,\boldsymbol{k}} \end{pmatrix} \tag{B13}$$

with

$$h_{+,0}(\boldsymbol{k}) = \begin{pmatrix} 0_{2\times 2} & v_\star(k_x\tau_0 + \mathrm{i}k_y\tau_z) & \gamma\tau_0 + v'_\star(k_x\tau_x + k_y\tau_y) \\ v_\star(k_x\tau_0 - \mathrm{i}k_y\tau_z) & M\tau_x & v''_\star(k_x\tau_x - k_y\tau_y) \\ \gamma\tau_0 + v'_\star(k_x\tau_x + k_y\tau_y) & v''_\star(k_x\tau_x - k_y\tau_y) & 0_{2\times 2} \end{pmatrix}. \tag{B14}$$

Based on Eq. (23), the Hamiltonian for the low-energy d modes around Γ reads

$$\begin{pmatrix} d^\dagger_{+,\boldsymbol{k},\boldsymbol{q}_1} & d^\dagger_{\boldsymbol{k},\boldsymbol{q}_2} & d^\dagger_{+,\boldsymbol{k},\boldsymbol{q}_3} \end{pmatrix} h_{+,1}(\boldsymbol{k}) \otimes s_0 \begin{pmatrix} d_{+,\boldsymbol{k},\boldsymbol{q}_1} \\ d_{+,\boldsymbol{k},\boldsymbol{q}_2} \\ d_{+,\boldsymbol{k},\boldsymbol{q}_3} \end{pmatrix} \tag{B15}$$

with

$$h_{+,1}(\boldsymbol{k}) = \begin{pmatrix} (\boldsymbol{k}-\boldsymbol{q}_1)\cdot\boldsymbol{\sigma} & 0_{2\times 2} & 0_{2\times 2} \\ 0_{2\times 2} & (\boldsymbol{k}-\boldsymbol{q}_2)\cdot\boldsymbol{\sigma} & 0_{2\times 2} \\ 0_{2\times 2} & 0_{2\times 2} & (\boldsymbol{k}-\boldsymbol{q}_3)\cdot\boldsymbol{\sigma} \end{pmatrix}. \tag{B16}$$

Now we show it is reasonable for us to treat $h_{+,f-d,\Gamma}$ and $h_{+,c-d,\Gamma}$ as perturbations, and then we will project out the d modes around Γ. Since $h_{+,f-d,\Gamma}$ and $h_{+,c-d,\Gamma}$ depend linearly on $\mathcal{E}$, we choose $\mathcal{E} = 300$ meV (EUS) and find that the absolute values of the matrix elements of $h_{+,f-d,\Gamma}$ and $h_{+,c-d,\Gamma}$ are no larger than 0.22, while the gaps between the levels of $h_{+,0}(0)$ and the levels of $h_{+,1}(0)$ are no smaller than 0.87. Therefore we can treat $h_{+,f-d,\Gamma}$ and $h_{+,c-d,\Gamma}$ as perturbations. Then, according to the

second-order perturbation theory, we may project out the d modes. Explicitly, we have

$$\begin{pmatrix} h_{+,0}(\boldsymbol{k}) & \begin{pmatrix} h_{+,c-d,\Gamma} \\ h_{+,f-d,\Gamma} \end{pmatrix} \\ \begin{pmatrix} h_{+,c-d,\Gamma} \\ h_{+,f-d,\Gamma} \end{pmatrix}^{\dagger} & h_{+,1}(\boldsymbol{k}) \end{pmatrix} \otimes s_0 \tag{B17}$$

as the effective Hamiltonian around Γ_M, and by treating $h_{+,f-d,\Gamma}$ and $h_{+,c-d,\Gamma}$ as perturbations, $h_{+,0}(\boldsymbol{k})$ gains the following correction according to the second-order perturbation theory [178]:

$$\begin{pmatrix} h_{+,c-d,\Gamma} \\ h_{+,f-d,\Gamma} \end{pmatrix} \frac{1}{h_{+,0}(\boldsymbol{k}) - h_{+,1}(\boldsymbol{k})} \begin{pmatrix} h_{+,c-d,\Gamma} \\ h_{+,f-d,\Gamma} \end{pmatrix}^{\dagger}. \tag{B18}$$

To the first order of $\boldsymbol{k}$, Eq. (B18) brings in many different corrections. By comparing the resultant band structure of the f-c-d model to that of the BM model, we find the most important three corrections are those in Eq. (53). Specifically, the correction to M accounts for the increasing gap of the nearly flat bands at Γ_M when increasing $\mathcal{E}$, the correction to γ accounts for the decreasing gap of the remote bands at Γ_M when increasing $\mathcal{E}$, and the correction to $v_\star''$ accounts for the change of the band structure along Γ_M-M_M due to $\mathcal{E}$. The values of B parameters of Eq. (53), which are shown in Table I are also directly given by Eq. (B18).

3. More details on the band structure calculation

In this part, we present more details on how the numerical calculations for Fig. 1 are carried out. Owing to the exponentially decayed factor in $H_{0,\eta,fd}$ and $H_{0,\eta,fc}$, we are allowed to extend Λ_c and Λ_d to outside MBZ [125]. Then, we can reexpress $c^\dagger_{\eta,\boldsymbol{p}}$ and $d^\dagger_{\eta,\boldsymbol{p}}$ in Eq. (54) as $c^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{G}}$ and $d^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{Q}}$, respectively, where $\boldsymbol{k} \in$ MBZ, $\boldsymbol{G}$ is the reciprocal lattice vector, and $\boldsymbol{Q} \in \mathcal{Q}_\eta$. In this case, the original definitions of Λ_c and Λ_d require $|\boldsymbol{k}-\boldsymbol{G}| \leqslant \Lambda_c$ and $|\boldsymbol{k}-\boldsymbol{Q}| \leqslant \Lambda_d$. For the convenience of numerical calculation, we alter the definitions, and instead require $|\boldsymbol{G}| \leqslant \Lambda_c$ and $|\boldsymbol{Q}| \leqslant \Lambda_d$, while keeping $\boldsymbol{k}$ running over the entire MBZ. As a result, the terms in the single-particle f-c-d model in Eq. (54) become

$$H_{0,\eta,f} = 0\,, \tag{B19}$$

$$H_{0,\eta,c} = \sum_{\boldsymbol{k}\in\text{MBZ}} \sum_{|\boldsymbol{G}|\leqslant\Lambda_c} c^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{G}} \begin{pmatrix} 0_{2\times 2} & v_\star(\eta(k_x-G_x)\tau_0 + \mathrm{i}(k_y-G_y)\tau_z) \\ v_\star(\eta(k_x-G_x)\tau_0 - \mathrm{i}(k_y-G_y)\tau_z) & (M + B_M\mathcal{E}^2)\tau_x \end{pmatrix} s_0 c_{\eta,\boldsymbol{k}-\boldsymbol{G}}\,, \tag{B20}$$

$$\begin{aligned} H_{0,\eta,fc} = \sum_{\boldsymbol{k}\in\text{MBZ}} \sum_{|\boldsymbol{G}|\leqslant\Lambda_c} & e^{-\frac{|\boldsymbol{k}-\boldsymbol{G}|^2\lambda^2}{2}} [f^\dagger_{\eta,\boldsymbol{k}}[(\gamma + B_\gamma\mathcal{E}^2)\tau_0 + v_\star'(\boldsymbol{k}-\boldsymbol{G})\cdot(\eta\tau_x,\tau_y)] s_0 c_{\eta,\boldsymbol{k}-\boldsymbol{G},\Gamma_3} \\ & + f^\dagger_{\eta,\boldsymbol{k}}(v_\star'' + B_{v''}\mathcal{E}^2)(\boldsymbol{k}-\boldsymbol{G})\cdot(\eta\tau_x,-\tau_y) s_0 c_{\eta,\boldsymbol{k}-\boldsymbol{G},\Gamma_1\Gamma_2}] + \text{H.c.} \end{aligned} \tag{B21}$$

with $c^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{G},\Gamma_3} = (c^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{G},1}, c^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{G},2})$ and $c^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{G},\Gamma_1\Gamma_2} = (c^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{G},3}, c^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{G},4})$,

$$H_{0,\eta,d} = \sum_{\boldsymbol{k}\in\text{MBZ}} \sum_{\boldsymbol{Q}\in\mathcal{Q}_\eta}^{|\boldsymbol{Q}|\leqslant\Lambda_d} d^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{Q}}(\eta(k_x-Q_x)\sigma_x + (k_y-Q_y)\sigma_y) s_0\, d_{\eta,\boldsymbol{k}-\boldsymbol{Q}}\,, \tag{B22}$$

and

$$H_{0,\eta,fd} = \sum_{\boldsymbol{k}\in\text{MBZ}} \sum_{\boldsymbol{Q}\in\mathcal{Q}_\eta}^{|\boldsymbol{Q}|\leqslant\Lambda_d} e^{-\frac{|\boldsymbol{k}-\boldsymbol{Q}|^2\lambda^2}{2}} f^\dagger_{\eta,\boldsymbol{k}}\, M_1\mathcal{E}(\tau_0 + \eta\mathrm{i}\tau_z) s_0\, d_{\eta,\boldsymbol{k}-\boldsymbol{Q}} + \text{H.c.} \tag{B23}$$

Figure 1 is plotted by choosing $\Lambda_c = \Lambda_d = 2\sqrt{3}$, Eq. (44) and Table I.

APPENDIX C: MORE DETAILS ON THE INTERACTION AMONG f, c AND d MODES

In this section, we provide more details on the interaction among f, c and d modes, which is derived by projecting the gate-screened Coulomb interaction Eq. (29) to the f-c-d basis. Throughout this section, we assume Λ_c and Λ_d to be small, i.e., $\Lambda_c, \Lambda_d \ll 1$. Owing to the assumption that Λ_c is small, we, just for the convenience, formally define

$$[\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k}+\boldsymbol{G})]_{\boldsymbol{Q}\sigma} = [\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})]_{\boldsymbol{Q}-\boldsymbol{G}\sigma} \quad \text{for } \boldsymbol{k} < \Lambda_c, \text{only when } \Lambda_c \ll 1\,, \tag{C1}$$

where $\widetilde{u}_{\eta,c,\beta}$ is defined in Eq. (35), and $\boldsymbol{G}$ is any Moiré reciprocal lattice vector. Note that if Λ_c becomes large such that there exists $\boldsymbol{k}$ and $\boldsymbol{k}+\boldsymbol{G}$ with $|\boldsymbol{k}|, |\boldsymbol{k}+\boldsymbol{G}| < \Lambda_c$, then Eq. (C1) cannot be used anymore, since we want $c^\dagger_{\boldsymbol{k}}$ to be independent from $c^\dagger_{\boldsymbol{k}+\boldsymbol{G}}$.

As the density operator $\rho(\boldsymbol{r})$ can be split into two parts as

$$\rho(\boldsymbol{r}) = \widetilde{\rho}(\boldsymbol{r}) + \rho_D(\boldsymbol{r}) \tag{C2}$$

with

$$\widetilde{\rho}(\boldsymbol{r}) = \sum_{\eta,\widetilde{l}} \widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\widetilde{l}} \widetilde{\psi}_{\eta,\boldsymbol{r},\widetilde{l}} \text{ for the TBG block and } \rho_D(\boldsymbol{r}) = \sum_{\eta} d^\dagger_{\eta,\boldsymbol{r}} d_{\eta,\boldsymbol{r}} \,, \tag{C3}$$

the interaction Eq. (29) can be split into three parts as

$$H_{\text{int}} = H_{\text{int}}^{\text{TBG}} + H_{\text{int}}^{\text{TBG-}D} + H_{\text{int}}^{D} \,, \tag{C4}$$

where

$$\begin{aligned} H_{\text{int}}^{\text{TBG}} &= \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \widetilde{\rho}(\boldsymbol{r}) :: \widetilde{\rho}(\boldsymbol{r}') :, \\ H_{\text{int}}^{\text{TBG-}D} &= \int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \widetilde{\rho}(\boldsymbol{r}) :: \rho_D(\boldsymbol{r}') :, \\ H_{\text{int}}^{D} &= \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_D(\boldsymbol{r}) :: \rho_D(\boldsymbol{r}') : \,. \end{aligned} \tag{C5}$$

Before discussing each part in Eq. (C5), we derive the following expressions for the convenience of the latter evaluation of the normal ordering. Based on the choice of $|G_0\rangle$ in Eq. (30), we have

$$\langle G_0 | \psi^\dagger_{\eta,\boldsymbol{p},l,\sigma,s} \psi_{\eta',\boldsymbol{p}',l',\sigma',s'} | G_0 \rangle = \tfrac{1}{2}\delta_{\eta\eta'}\delta_{\boldsymbol{p}\boldsymbol{p}'}\delta_{ll'}\delta_{\sigma\sigma'}\delta_{ss'} \,. \tag{C6}$$

Then, combined with Eqs. (9), (10), (18), and (21), we have

$$\begin{aligned} \langle G_0 | \widetilde{\psi}^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q},\sigma,s} \widetilde{\psi}_{\eta',\boldsymbol{k}',\boldsymbol{Q}',\sigma',s'} | G_0 \rangle &= \tfrac{1}{2}\delta_{\eta\eta'}\delta_{\boldsymbol{k}\boldsymbol{k}'}\delta_{\boldsymbol{Q}\boldsymbol{Q}'}\delta_{\sigma\sigma'}\delta_{ss'}, \\ \langle G_0 | \widetilde{\psi}^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q},\sigma,s} d_{\eta',\boldsymbol{p},\sigma',s'} | G_0 \rangle &= 0, \\ \langle G_0 | d^\dagger_{\eta,\boldsymbol{p},\sigma,s} d_{\eta',\boldsymbol{p}',\sigma',s'} | G_0 \rangle &= \tfrac{1}{2}\delta_{\eta\eta'}\delta_{\boldsymbol{p}\boldsymbol{p}'}\delta_{\sigma\sigma'}\delta_{ss'} \,. \end{aligned} \tag{C7}$$

Then, combined with Eqs. (34) and (35), we have

$$\begin{aligned} \langle G_0 | f^\dagger_{\eta,\boldsymbol{k},\alpha,s} f_{\eta',\boldsymbol{k}',\alpha',s'} | G_0 \rangle &= \tfrac{1}{2}\delta_{\eta\eta'}\delta_{\boldsymbol{k}\boldsymbol{k}'}\delta_{\alpha\alpha'}\delta ss', \\ \langle G_0 | f^\dagger_{\eta,\boldsymbol{k},\alpha,s} c_{\eta',\boldsymbol{k}',\beta',s'} | G_0 \rangle &= 0, \\ \langle G_0 | c^\dagger_{\eta,\boldsymbol{k},\beta,s} c_{\eta',\boldsymbol{k}',\beta',s'} | G_0 \rangle &= \tfrac{1}{2}\delta_{\eta\eta'}\delta_{\boldsymbol{k}\boldsymbol{k}'}\delta_{\beta\beta'}\delta ss', \\ \langle G_0 | f^\dagger_{\eta,\boldsymbol{k},\alpha,s} d_{\eta',\boldsymbol{p}',\sigma',s'} | G_0 \rangle &= 0, \\ \langle G_0 | c^\dagger_{\eta,\boldsymbol{k},\beta,s} d_{\eta',\boldsymbol{p}',\sigma',s'} | G_0 \rangle &= 0 \,. \end{aligned} \tag{C8}$$

In the following, we will discuss how we project the three parts in Eq. (C5) onto the f, c and d modes. All the numerical evaluations throughout this section are done with the parameter values in Table I and Eq. (27).

1. Review on $H_{\text{int}}^{\text{TBG}}$

We discuss $H_{\text{int}}^{\text{TBG}}$ in Eq. (C5) first. Since $H_{\text{int}}^{\text{TBG}}$ only depends on $\widetilde{\psi}$, its projection onto the f and c modes should have the same form as the interaction terms in the ordinary MATBG as discussed Ref. [125]. This subsection is a review of the interaction in Ref. [125], except that we use the parameter values for MATSTG.

To do the projection, we first need to figure out the projection of $\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\widetilde{l},\sigma,s}$ to $f^\dagger$ and $c^\dagger$. Combining Eq. (38) with

$$\widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\widetilde{l}} = \frac{1}{\sqrt{\mathcal{A}}}\sum_{\boldsymbol{p}} e^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}} \widetilde{\psi}^\dagger_{\eta,\boldsymbol{p},\widetilde{l}} = \frac{1}{\sqrt{\mathcal{A}}}\sum_{\boldsymbol{k}}^{\text{MBZ}} \sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}} e^{-\mathrm{i}(\boldsymbol{k}-\boldsymbol{Q})\cdot\boldsymbol{r}} \widetilde{\psi}^\dagger_{\eta,\boldsymbol{k}-\boldsymbol{Q},\widetilde{l}^\eta_{\boldsymbol{Q}}} = \frac{1}{\sqrt{\mathcal{A}}}\sum_{\boldsymbol{k}}^{\text{MBZ}} \sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}} e^{-\mathrm{i}(\boldsymbol{k}-\boldsymbol{Q})\cdot\boldsymbol{r}} \widetilde{\psi}^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q}} \,, \tag{C9}$$

we have

$$\begin{aligned} \widetilde{\psi}^\dagger_{\eta,\boldsymbol{r},\widetilde{l},\sigma,s} = &\frac{1}{\sqrt{\mathcal{A}}}\sum_{\boldsymbol{k}}^{\text{MBZ}} \sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}} e^{-\mathrm{i}(\boldsymbol{k}-\boldsymbol{Q})\cdot\boldsymbol{r}} \sum_{\alpha=1,2} f^\dagger_{\eta,\boldsymbol{k},\alpha,s} [\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]^*_{\boldsymbol{Q}\sigma} \\ &+ \frac{1}{\sqrt{\mathcal{A}}}\sum_{\boldsymbol{k}}^{|\boldsymbol{k}|<\Lambda_c} \sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}} e^{-\mathrm{i}(\boldsymbol{k}-\boldsymbol{Q})\cdot\boldsymbol{r}} \sum_{\beta=1,\ldots,4} c^\dagger_{\eta,\boldsymbol{k},\beta,s} [\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})]^*_{\boldsymbol{Q}\sigma} + \ldots \,, \end{aligned} \tag{C10}$$

where "..." means the high-energy modes. Combined with Eqs. (36) and (57), we get

$$\widetilde{\psi}^{\dagger}_{\eta,\boldsymbol{r},\widetilde{l},\sigma,s}=\frac{1}{\sqrt{N\mathcal{A}}}\sum_{\alpha=1,2}\sum_{\boldsymbol{R}}f^{\dagger}_{\eta,\boldsymbol{R},\alpha,s}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}}\sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}}e^{-\mathrm{i}(\boldsymbol{k}-\boldsymbol{Q})\cdot(\boldsymbol{r}-\boldsymbol{R})}e^{\mathrm{i}\boldsymbol{Q}\cdot\boldsymbol{R}}[\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]^{*}_{\boldsymbol{Q}\sigma}+\sum_{\beta=1,\ldots,4}c^{\dagger}_{\eta,\boldsymbol{r},\beta,s}\sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}}e^{\mathrm{i}\boldsymbol{Q}\cdot\boldsymbol{r}}[\widetilde{u}_{\eta,c,\beta}(\boldsymbol{k})]^{*}_{\boldsymbol{Q}\sigma}+\ldots\,, \tag{C11}$$

where $f^{\dagger}_{\eta,\boldsymbol{R},\alpha,s}$ is defined in Eq. (36), and $c^{\dagger}_{\eta,\boldsymbol{r},\beta,s}$ is defined in Eq. (57). By defining

$$g_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})=\sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}}e^{\mathrm{i}\boldsymbol{Q}\cdot\boldsymbol{r}}[\widetilde{u}_{\eta,c,\beta}(0)]_{\boldsymbol{Q}\sigma} \tag{C12}$$

and using $\Delta K_{\widetilde{l}}$ defined in Eq. (46), we eventually get

$$\widetilde{\psi}^{\dagger}_{\eta,\boldsymbol{r},\widetilde{l},\sigma,s}\approx\sum_{\alpha=1,2}\sum_{\boldsymbol{R}}f^{\dagger}_{\eta,\boldsymbol{R},\alpha,s}e^{-\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})^{*}+\sum_{\beta=1,\ldots,4}c^{\dagger}_{\eta,\boldsymbol{r},\beta,s}g^{*}_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})\,, \tag{C13}$$

where we use Eq. (47) and

$$e^{-\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}=e^{\mathrm{i}\eta(-)^{\widetilde{l}}\boldsymbol{q}_1\cdot\boldsymbol{R}}=e^{\mathrm{i}\boldsymbol{Q}\cdot\boldsymbol{R}}\quad\text{for }\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}\,, \tag{C14}$$

and $(-)^{t}=-(-)^{b}=1$. With Eq. (C13), we can derive the projection of $\widetilde{\rho}(\boldsymbol{r})$ to f and c modes, resulting in

$$\begin{aligned}\widetilde{\rho}(\boldsymbol{r})&=\sum_{\eta,\widetilde{l},\sigma,s}\widetilde{\psi}^{\dagger}_{\eta,\boldsymbol{r},\widetilde{l},\sigma,s}\widetilde{\psi}_{\eta,\boldsymbol{r},\widetilde{l},\sigma,s}\\&\approx\sum_{\eta,\widetilde{l},\sigma,s}\left[\sum_{\alpha=1,2}\sum_{\boldsymbol{R}}f^{\dagger}_{\eta,\boldsymbol{R},\alpha,s}e^{-\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}w^{*}_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})+\sum_{\beta=1,\ldots,4}c^{\dagger}_{\eta,\boldsymbol{r},\beta,s}g^{*}_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})\right]\\&\quad\times\left[\sum_{\alpha'=1,2}\sum_{\boldsymbol{R}'}f_{\eta,\boldsymbol{R}',\alpha',s}e^{\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}'}w_{\eta\alpha'\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R}')+\sum_{\beta'=1,\ldots,4}c_{\eta,\boldsymbol{r},\beta',s}g_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r})\right]\\&=\sum_{\eta,\widetilde{l},\sigma,s}\left\{\sum_{\alpha=1,2}\sum_{\boldsymbol{R}}\sum_{\alpha'=1,2}\sum_{\boldsymbol{R}'}f^{\dagger}_{\eta,\boldsymbol{R},\alpha,s}f_{\eta,\boldsymbol{R}',\alpha',s}e^{-\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}+\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}'}w^{*}_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})w_{\eta\alpha'\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R}')\right.\\&\quad+\left[\sum_{\beta=1,\ldots,4}\sum_{\alpha=1,2}\sum_{\boldsymbol{R}}c^{\dagger}_{\eta,\boldsymbol{r},\beta,s}g^{*}_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})f_{\eta,\boldsymbol{R},\alpha,s}e^{\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})+\mathrm{H.c.}\right]\\&\quad\left.+\sum_{\beta=1,\ldots,4}\sum_{\beta'=1,\ldots,4}c^{\dagger}_{\eta,\boldsymbol{r},\beta,s}c_{\eta,\boldsymbol{r},\beta',s}g^{*}_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})g_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r})\right\}.\end{aligned} \tag{C15}$$

At the single particle level, we mentioned that it is legitimate to neglect the hopping among f modes due to the extreme localization of the Wannier functions, meaning that we can adopt the following approximation:

$$w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})w^{*}_{\eta'\alpha'\widetilde{l}'\sigma'}(\boldsymbol{r}-\boldsymbol{R}')\approx 0\quad\text{if}\quad\boldsymbol{R}\neq\boldsymbol{R}'\,. \tag{C16}$$

With this approximation, we have

$$\begin{aligned}\widetilde{\rho}(\boldsymbol{r})\approx\sum_{\eta,\widetilde{l},\sigma,s}\Bigg\{&\sum_{\alpha=1,2}\sum_{\boldsymbol{R}}\sum_{\alpha'=1,2}f^{\dagger}_{\eta,\boldsymbol{R},\alpha,s}f_{\eta,\boldsymbol{R},\alpha',s}w^{*}_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})w_{\eta\alpha'\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})\\&+\left[\sum_{\beta=1,\ldots,4}\sum_{\alpha=1,2}\sum_{\boldsymbol{R}}c^{\dagger}_{\eta,\boldsymbol{r},\beta,s}g^{*}_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})f_{\eta,\boldsymbol{R},\alpha,s}e^{\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})+\mathrm{H.c.}\right]\\&+\sum_{\beta=1,\ldots,4}\sum_{\beta'=1,\ldots,4}c^{\dagger}_{\eta,\boldsymbol{r},\beta,s}c_{\eta,\boldsymbol{r},\beta',s}g^{*}_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})g_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r})\Bigg\}.\end{aligned} \tag{C17}$$

Furthermore, according to the symmetry properties of the Wannier functions Eq. (A31), we have

$$w_{+2\widetilde{l}\sigma}(\boldsymbol{r}) = w^*_{+1\widetilde{l}\sigma}(-\boldsymbol{r}) = [(-)^{\widetilde{l}}\mathrm{i}w_{+1\widetilde{\underline{l}}\overline{\sigma}}(\boldsymbol{r})]^* = (-)^{\widetilde{\underline{l}}}\mathrm{i}w^*_{+1\widetilde{\underline{l}}\overline{\sigma}}(\boldsymbol{r}),$$
$$w_{-\alpha\widetilde{l}\sigma}(\boldsymbol{r}) = w^*_{+\alpha\widetilde{l}\sigma}(\boldsymbol{r}) \tag{C18}$$

with $\widetilde{\underline{l}} = t/b$ for $\widetilde{l} = b/t$ and $\overline{\sigma} = A/B$ for $\sigma = B/A$, just as Ref. [125]. Then, we have

$$\sum_{\widetilde{l}\sigma} w_{+1\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})w^*_{+2\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R}) = -\mathrm{i}\sum_{\widetilde{l}\sigma} w_{+1\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})(-)^{\widetilde{\underline{l}}}w_{+1\widetilde{\underline{l}}\overline{\sigma}}(\boldsymbol{r}-\boldsymbol{R}) = 0\,, \tag{C19}$$

resulting in

$$\sum_{\widetilde{l}\sigma} w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})w^*_{\eta\alpha'\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R}) = 0 \quad \text{for } \alpha \neq \alpha'. \tag{C20}$$

This expression brings simplification to the projection of $\widetilde{\rho}(\boldsymbol{r})$ as

$$\begin{aligned}\widetilde{\rho}(\boldsymbol{r}) \approx \sum_{\eta,\widetilde{l},\sigma,s}\Bigg\{&\sum_{\alpha=1,2}\sum_{\boldsymbol{R}} f^\dagger_{\eta,\boldsymbol{R},\alpha,s}f_{\eta,\boldsymbol{R},\alpha,s}w^*_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})\\ &+\Bigg[\sum_{\beta=1,\ldots,4}\sum_{\alpha=1,2}\sum_{\boldsymbol{R}} c^\dagger_{\eta,\boldsymbol{r},\beta,s}g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})f_{\eta,\boldsymbol{R},\alpha,s}e^{\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R}) + \mathrm{H.c.}\Bigg]\\ &+\sum_{\beta=1,\ldots,4}\sum_{\beta'=1,\ldots,4} c^\dagger_{\eta,\boldsymbol{r},\beta,s}c_{\eta,\boldsymbol{r},\beta',s}g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})g_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r})\Bigg\}\,.\end{aligned} \tag{C21}$$

Furthermore, Eq. (C18) shows that $\sum_{\widetilde{l}\sigma}|w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r})|^2$ is independent of η and α, and then we can define

$$n_f(\boldsymbol{r}) = \sum_{\widetilde{l}\sigma}|w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r})|^2. \tag{C22}$$

Then, we have

$$\begin{aligned}\widetilde{\rho}(\boldsymbol{r}) \approx \sum_{\boldsymbol{R}}\rho_f(\boldsymbol{R})n_f(\boldsymbol{r}-\boldsymbol{R}) + \sum_{\eta,s}\Bigg[\sum_{\beta=1,\ldots,4}\sum_{\alpha=1,2}\sum_{\boldsymbol{R}} c^\dagger_{\eta,\boldsymbol{r},\beta,s}f_{\eta,\boldsymbol{R},\alpha,s}\sum_{\widetilde{l}\sigma} g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})e^{\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R}) + \mathrm{H.c.}\Bigg]\\ +\sum_{\eta,s}\sum_{\beta=1,\ldots,4}\sum_{\beta'=1,\ldots,4} c^\dagger_{\eta,\boldsymbol{r},\beta,s}c_{\eta,\boldsymbol{r},\beta',s}\sum_{\widetilde{l}\sigma} g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})g_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r})\,,\end{aligned} \tag{C23}$$

where $\rho_f(\boldsymbol{R})$ is defined under Eq. (55). By further defining

$$\begin{aligned}\rho_{ff}(\boldsymbol{r}) &= \sum_{\boldsymbol{R}}\rho_f(\boldsymbol{R})n_f(\boldsymbol{r}-\boldsymbol{R}),\\ \rho_{cc}(\boldsymbol{r}) &= \sum_{\eta,s}\sum_{\beta=1,\ldots,4}\sum_{\beta'=1,\ldots,4} c^\dagger_{\eta,\boldsymbol{r},\beta,s}c_{\eta,\boldsymbol{r},\beta',s}\sum_{\widetilde{l}\sigma} g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})g_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r}),\\ \rho_{cf}(\boldsymbol{r}) &= \sum_{\eta,s}\sum_{\beta=1,\ldots,4}\sum_{\alpha=1,2}\sum_{\boldsymbol{R}} c^\dagger_{\eta,\boldsymbol{r},\beta,s}f_{\eta,\boldsymbol{R},\alpha,s}\sum_{\widetilde{l}\sigma} g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})e^{\mathrm{i}\eta\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R}),\\ \rho_{fc}(\boldsymbol{r}) &= \rho^\dagger_{cf}(\boldsymbol{r})\,,\end{aligned} \tag{C24}$$

we eventually arrive at

$$\widetilde{\rho}(\boldsymbol{r}) \approx \rho_{ff}(\boldsymbol{r}) + \rho_{cc}(\boldsymbol{r}) + \rho_{cf}(\boldsymbol{r}) + \rho_{fc}(\boldsymbol{r})\,. \tag{C25}$$

With Eq. (C25), the expanded $H^{\mathrm{TBG}}_{\mathrm{int}}$ becomes

$$\begin{aligned}H^{\mathrm{TBG}}_{\mathrm{int}} \approx \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{ff}(\boldsymbol{r})::\rho_{ff}(\boldsymbol{r}'):\\ +\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{cc}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}'):\end{aligned}$$

$$
\begin{aligned}
&+\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_{cc}(\boldsymbol{r}') : \\
&+\left[\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cf}(\boldsymbol{r}) :: \rho_{cf}(\boldsymbol{r}') : +\text{H.c.}\right] \\
&+\left[\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_{cf}(\boldsymbol{r}') : +\text{H.c.}\right] \\
&+\left[\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_{fc}(\boldsymbol{r}') : +\text{H.c.}\right] \\
&+\left[\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cc}(\boldsymbol{r}) :: \rho_{cf}(\boldsymbol{r}') : +\text{H.c.}\right] \\
&+\left[\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cc}(\boldsymbol{r}) :: \rho_{fc}(\boldsymbol{r}') : +\text{H.c.}\right] \\
&+\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}')[: \rho_{cf}(\boldsymbol{r}) :: \rho_{fc}(\boldsymbol{r}') : + : \rho_{fc}(\boldsymbol{r}) :: \rho_{cf}(\boldsymbol{r}') :] \,. \qquad \text{(C26)}
\end{aligned}
$$

In the following, we will discuss each term in Eq. (C26), as were discussed in Ref. [125].

a. $\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_{ff}(\boldsymbol{r}') :$

To simplify this term, we first evaluate the Fourier transformation of $n_f(\boldsymbol{r})$ as

$$
n_f(\boldsymbol{p}) = \int d^2r\, n_f(\boldsymbol{r}) e^{i\boldsymbol{p}\cdot\boldsymbol{r}} = \int d^2r \sum_{\widetilde{l}\sigma} |w_{+1\widetilde{l}\sigma}(\boldsymbol{r})|^2 e^{i\boldsymbol{p}\cdot\boldsymbol{r}} = \frac{1}{N}\sum_{\boldsymbol{k}\in\text{MBZ}} U^{\dagger}_{+,f,1}(\boldsymbol{k}+\boldsymbol{p}) U_{+,f,1}(\boldsymbol{k}) \,, \qquad \text{(C27)}
$$

where Eq. (47) is used. Based on Eq. (A31), we can derive useful symmetry properties of $n_f(\boldsymbol{r})$ and $n_f(\boldsymbol{p})$ as

$$
\begin{cases} n_f(\boldsymbol{r}) = n_f^*(-\boldsymbol{r}) \\ n_f(\boldsymbol{r}) = n_f(C_3\boldsymbol{r}) \\ n_f(\boldsymbol{r}) = n_f(C_{2x}\boldsymbol{r}) \\ n_f(\boldsymbol{r}) = n_f(-\boldsymbol{r}) \end{cases} \quad \text{and} \quad \begin{cases} n_f(\boldsymbol{p}) = n_f^*(\boldsymbol{p}) \\ n_f(\boldsymbol{p}) = n_f(C_3\boldsymbol{p}) \\ n_f(\boldsymbol{p}) = n_f(C_{2x}\boldsymbol{p}) \\ n_f(\boldsymbol{p}) = n_f(-\boldsymbol{p}) \end{cases} . \qquad \text{(C28)}
$$

With the definition of $n_f(\boldsymbol{p})$, we have

$$
\begin{aligned}
\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_{ff}(\boldsymbol{r}') : &= \frac{1}{2}\sum_{\boldsymbol{R},\boldsymbol{R}'}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') n_f(\boldsymbol{r}-\boldsymbol{R}) n_f(\boldsymbol{r}'-\boldsymbol{R}') : \rho_f(\boldsymbol{R}) :: \rho_f(\boldsymbol{R}') : \\
&= \frac{1}{2}\sum_{\boldsymbol{R},\boldsymbol{R}'} : \rho_f(\boldsymbol{R}) :: \rho_f(\boldsymbol{R}') : U(\boldsymbol{R}-\boldsymbol{R}') \,, \qquad \text{(C29)}
\end{aligned}
$$

where

$$
\begin{aligned}
U(\boldsymbol{R}-\boldsymbol{R}') &= \int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') n_f(\boldsymbol{r}-\boldsymbol{R}) n_f(\boldsymbol{r}'-\boldsymbol{R}') \\
&= \int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') n_f(\boldsymbol{r}) n_f(\boldsymbol{r}'-\boldsymbol{R}'+\boldsymbol{R}) \\
&= \int d^2r d^2r' \frac{1}{\mathcal{A}^3}\sum_{\boldsymbol{p}_1} n_f(\boldsymbol{p}_1) e^{-i\boldsymbol{p}_1\cdot\boldsymbol{r}} \sum_{\boldsymbol{p}} V(\boldsymbol{p}) e^{-i\boldsymbol{p}\cdot(\boldsymbol{r}-\boldsymbol{r}')} \sum_{\boldsymbol{p}_2} n_f(\boldsymbol{p}_2) e^{-i\boldsymbol{p}_2\cdot(\boldsymbol{r}-\boldsymbol{R}'+\boldsymbol{R})} \\
&= \frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}} n_f^*(\boldsymbol{p}) V(\boldsymbol{p}) n_f(\boldsymbol{p}) e^{-i\boldsymbol{p}\cdot(\boldsymbol{R}-\boldsymbol{R}')} \,. \qquad \text{(C30)}
\end{aligned}
$$

Numerically, we find

$$
U(0) = 91.50\ \text{meV}, \quad U(\boldsymbol{a}_1) = 5.387\ \text{meV}, \quad U(\boldsymbol{a}_1-\boldsymbol{a}_2) = 0.5079\ \text{meV} \qquad \text{(C31)}
$$

in EUS, which shows that $U(\boldsymbol{R})$ decays very as $|\boldsymbol{R}|$ increases. Therefore we only keep the terms up to the nearest-neighboring terms and get

$$
\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_{ff}(\boldsymbol{r}') := H_{\text{int},U} \,, \qquad \text{(C32)}
$$

where $H_{\text{int},U}$ is defined in Eq. (55). In Eq. (55), we have

$$U_1 = U(0) \tag{C33}$$

and

$$U_2 = \frac{1}{6}\sum_{\boldsymbol{R}\neq 0} U(\boldsymbol{R})\,. \tag{C34}$$

Since $U(\boldsymbol{R})$ decays very fast as $|\boldsymbol{R}|$ increases, U_2 is dominated by the $|\boldsymbol{R}| = |\boldsymbol{a}_1|$ contribution. The reason for choosing an expression of U_2 that is not equal to $U(\boldsymbol{a}_1)$ is that such choice can reduce the error in calculating the interaction-induced chemical potential shift, as discussed in Ref. [125]. The numerical values of U_1 and U_2 are in Table II.

b. $\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cc}(\boldsymbol{r}) :: \rho_{cc}(\boldsymbol{r}') :$

To simplify this term, we first derive the expression of $\sum_{\widetilde{l}\sigma} g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r}) g_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r})$ as

$$\begin{aligned}
\sum_{\widetilde{l}\sigma} g_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r}) g^*_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r}) &= \sum_{\widetilde{l}\sigma}\sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}} e^{-i\boldsymbol{Q}\cdot\boldsymbol{r}}[\widetilde{u}_{\eta,c,\beta}(0)]_{\boldsymbol{Q}\sigma}\sum_{\boldsymbol{Q}'\in\mathcal{Q}_{\eta,\widetilde{l}}} e^{i\boldsymbol{Q}'\cdot\boldsymbol{r}}[\widetilde{u}_{\eta,c,\beta'}(0)]^*_{\boldsymbol{Q}'\sigma} \\
&= \sum_{\widetilde{l}\sigma}\sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}}\sum_{\boldsymbol{G}} e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}[\widetilde{u}_{\eta,c,\beta'}(0)]^*_{\boldsymbol{Q}-\boldsymbol{G}\sigma}[\widetilde{u}_{\eta,c,\beta}(0)]_{\boldsymbol{Q}\sigma} \\
&= \sum_{\boldsymbol{Q}\in\mathcal{Q}}\sum_{\sigma}\sum_{\boldsymbol{G}} e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}[\widetilde{u}_{\eta,c,\beta'}(\boldsymbol{G})]^*_{\boldsymbol{Q}\sigma}[\widetilde{u}_{\eta,c,\beta}(0)]_{\boldsymbol{Q}\sigma} \\
&= \sum_{\boldsymbol{G}} e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}\widetilde{u}^\dagger_{\eta,c,\beta'}(\boldsymbol{G})\widetilde{u}_{\eta,c,\beta}(0)
\end{aligned} \tag{C35}$$

where Eqs. (C1) and (C12) are used. Then, we have

$$\sum_{\widetilde{l}\sigma} g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r}) g_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}\widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G})\widetilde{u}_{\eta,c,\beta'}(0)\,. \tag{C36}$$

Then,

$$\begin{aligned}
&\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cc}(\boldsymbol{r}) :: \rho_{cc}(\boldsymbol{r}') : \\
&= \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}')\sum_{\eta,s}\sum_{\beta}\sum_{\beta'} : c^\dagger_{\eta,\boldsymbol{r},\beta,s} c_{\eta,\boldsymbol{r},\beta',s} : \sum_{\widetilde{l}\sigma} g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r}) g_{\eta\beta'\widetilde{l}\sigma}(\boldsymbol{r}) \\
&\quad\times\sum_{\eta_1,s_1}\sum_{\beta_1}\sum_{\beta_1'} : c^\dagger_{\eta_1,\boldsymbol{r}',\beta_1,s} c_{\eta_1,\boldsymbol{r}',\beta_1',s_1} : \sum_{\widetilde{l}_1\sigma_1} g^*_{\eta_1\beta_1\widetilde{l}_1\sigma_1}(\boldsymbol{r}') g_{\eta_1\beta_1'\widetilde{l}_1\sigma_1}(\boldsymbol{r}') \\
&= \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}')\sum_{\eta,s}\sum_{\beta,\beta'}\sum_{\eta_1,s_1}\sum_{\beta_1,\beta_1'} : c^\dagger_{\eta,\boldsymbol{r},\beta,s} c_{\eta,\boldsymbol{r},\beta',s} :: c^\dagger_{\eta_1,\boldsymbol{r}',\beta_1,s_1} c_{\eta_1,\boldsymbol{r}',\beta_1',s_1} : \\
&\quad\times\sum_{\boldsymbol{G}} e^{-i\boldsymbol{G}\cdot\boldsymbol{r}}\widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G})\widetilde{u}_{\eta,c,\beta'}(0)\sum_{\boldsymbol{G}'} e^{-i\boldsymbol{G}'\cdot\boldsymbol{r}'}\widetilde{u}^\dagger_{\eta,c,\beta_1}(\boldsymbol{G}')\widetilde{u}_{\eta,c,\beta_1'}(0) \\
&= \frac{1}{2}\frac{1}{\mathcal{A}}\sum_{\boldsymbol{k}}^{\text{MBZ}}\sum_{\boldsymbol{G},\boldsymbol{G}',\boldsymbol{G}''} V(\boldsymbol{k}+\boldsymbol{G}'')\sum_{\eta,s}\sum_{\beta,\beta'}\sum_{\eta_1,s_1}\sum_{\beta_1,\beta_1'}\int d^2r : c^\dagger_{\eta,\boldsymbol{r},\beta,s} c_{\eta,\boldsymbol{r},\beta',s} : e^{-i(\boldsymbol{k}+\boldsymbol{G}''+\boldsymbol{G})\cdot\boldsymbol{r}} \\
&\quad\times\int d^2r' : c^\dagger_{\eta_1,\boldsymbol{r}',\beta_1,s} c_{\eta_1,\boldsymbol{r}',\beta_1',s_1} : e^{-i(-\boldsymbol{k}-\boldsymbol{G}''+\boldsymbol{G}')\cdot\boldsymbol{r}'}\widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G})\widetilde{u}_{\eta,c,\beta'}(0)U^\dagger_{\eta_1,c,\beta_1}(\boldsymbol{G}')U_{\eta_1,c,\beta_1'}(0)\,.
\end{aligned} \tag{C37}$$

Owing to small $\Lambda_c \ll |\boldsymbol{q}_1| = 1$, we have

$$\int d^2r : c^\dagger_{\eta,\boldsymbol{r},\beta,s} c_{\eta,\boldsymbol{r},\beta',s} : e^{-i(\boldsymbol{k}+\boldsymbol{G}''+\boldsymbol{G})\cdot\boldsymbol{r}} = 0 \text{ if } \boldsymbol{G}''+\boldsymbol{G}\neq 0\,. \tag{C38}$$

Then, we have

$$\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{cc}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}'): = \frac{1}{2}\frac{1}{\mathcal{A}}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}}\sum_{\boldsymbol{G}} V(\boldsymbol{k}+\boldsymbol{G})\sum_{\eta,s}\sum_{\beta,\beta'}\sum_{\eta_1,s_1}\sum_{\beta_1,\beta_1'}\int d^2r : c^\dagger_{\eta,\boldsymbol{r},\beta,s}c_{\eta,\boldsymbol{r},\beta',s} : e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \times \int d^2r' : c^\dagger_{\eta_1,\boldsymbol{r}',\beta_1,s_1}c_{\eta_1,\boldsymbol{r}',\beta_1',s_1} : e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}'}\widetilde{u}^\dagger_{\eta,c,\beta}(-\boldsymbol{G})\widetilde{u}_{\eta,c,\beta'}(0)U^\dagger_{\eta_1,c,\beta_1}(\boldsymbol{G})U_{\eta_1,c,\beta_1'}(0)\,. \tag{C39}$$

Furthermore, we numerically find that

$$\sum_{\boldsymbol{G}} V(\boldsymbol{k}+\boldsymbol{G})\widetilde{u}^\dagger_{\eta,c,\beta}(-\boldsymbol{G})U_{\eta_1,c,\beta'}(0)\widetilde{u}^\dagger_{\eta,c,\beta_1}(\boldsymbol{G})U_{\eta_1,c,\beta_1'}(0) \approx V(\boldsymbol{k})\delta_{\beta\beta'}\delta_{\beta_1\beta_1'} \tag{C40}$$

with only 8% error. Then, we have

$$\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{cc}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}'): \approx \frac{1}{2}\int d^2r \int d^2r' \frac{1}{\mathcal{A}}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}} V(\boldsymbol{k})e^{-\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')}\sum_{\eta,s}\sum_{\beta}\sum_{\eta_1,s_1}\sum_{\beta_1} : c^\dagger_{\eta,\boldsymbol{r},\beta,s}c_{\eta,\boldsymbol{r},\beta,s} :: c^\dagger_{\eta_1,\boldsymbol{r}',\beta_1,s}c_{\eta_1,\boldsymbol{r}',\beta_1,s_1} : \,. \tag{C41}$$

Again owing to small Λ_c, we can extend the summation of $\boldsymbol{k}$ from MBZ to $\mathbb{R}^2$, leading to

$$\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{cc}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}'): \approx \frac{1}{2}\int d^2r \int d^2r' \frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}} V(\boldsymbol{p})e^{-\mathrm{i}\boldsymbol{p}\cdot(\boldsymbol{r}-\boldsymbol{r}')}\sum_{\eta,s}\sum_{\beta}\sum_{\eta_1,s_1}\sum_{\beta_1} : c^\dagger_{\eta,\boldsymbol{r},\beta,s}c_{\eta,\boldsymbol{r},\beta,s} :: c^\dagger_{\eta_1,\boldsymbol{r}',\beta_1,s}c_{\eta_1,\boldsymbol{r}',\beta_1,s_1} :\,, \tag{C42}$$

resulting in

$$\frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{cc}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}'): \approx H_{\mathrm{int},V,c}\,, \tag{C43}$$

where $H_{\mathrm{int},V,c}$ is defined in Eq. (56).

c. $\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{ff}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}'):$

First, by using Eqs. (C36) and (C28), we have

$$\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{ff}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}') := \frac{1}{\mathcal{A}}\sum_{\boldsymbol{R}} :\rho_f(\boldsymbol{R}): \sum_{\eta,s}\sum_{\beta\beta'}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}} e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}} \times \sum_{\boldsymbol{G},\boldsymbol{G}'}\int d^2r' e^{\mathrm{i}(\boldsymbol{k}+\boldsymbol{G}'-\boldsymbol{G})\cdot\boldsymbol{r}'}\widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G})\widetilde{u}_{\eta,c,\beta'}(0)V(\boldsymbol{k}+\boldsymbol{G}')n_f(\boldsymbol{k}+\boldsymbol{G}') : c^\dagger_{\eta,\boldsymbol{r}',\beta,s}c_{\eta,\boldsymbol{r}',\beta',s} :\,. \tag{C44}$$

Then, by using Eq. (C38) derived from the small Λ_c, we get

$$\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{ff}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}') := \frac{1}{\mathcal{A}}\sum_{\boldsymbol{R}} :\rho_f(\boldsymbol{R}): \sum_{\eta,s}\sum_{\beta\beta'}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}} e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}} \times \int d^2r' e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}'} : c^\dagger_{\eta,\boldsymbol{r}',\beta,s}c_{\eta,\boldsymbol{r}',\beta',s} : \sum_{\boldsymbol{G}}\widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G})\widetilde{u}_{\eta,c,\beta'}(0)V(\boldsymbol{k}+\boldsymbol{G})n_f(\boldsymbol{k}+\boldsymbol{G})\,. \tag{C45}$$

Again owing to small Λ_c, we can choose $\boldsymbol{k}=0$ in $V(\boldsymbol{k}+\boldsymbol{G})n_f(\boldsymbol{k}+\boldsymbol{G})$ as a good approximation, resulting in

$$\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') :\rho_{ff}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}'): \approx \frac{1}{\mathcal{A}}\sum_{\boldsymbol{R}} :\rho_f(\boldsymbol{R}): \sum_{\eta,s}\sum_{\beta\beta'}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}} e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}}\int d^2r' e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}'} : c^\dagger_{\eta,\boldsymbol{r}',\beta,s}c_{\eta,\boldsymbol{r}',\beta',s} : [X_\eta]_{\beta\beta'}\,, \tag{C46}$$

where

$$[X_\eta]_{\beta\beta'} = \sum_{\boldsymbol{G}}\widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G})\widetilde{u}_{\eta,c,\beta'}(0)V(\boldsymbol{G})n_f(\boldsymbol{G})\,. \tag{C47}$$

Based on with Eqs. (A29) and (C28), we have

$$X_+=\begin{pmatrix} e^{-\mathrm{i}\frac{2\pi}{3}\sigma_z} & \\ & \sigma_0\end{pmatrix}X_+\begin{pmatrix} e^{\mathrm{i}\frac{2\pi}{3}\sigma_z} & \\ & \sigma_0\end{pmatrix}=\begin{pmatrix}\sigma_x & \\ & \sigma_x\end{pmatrix}X_+\begin{pmatrix}\sigma_x & \\ & \sigma_x\end{pmatrix}=\begin{pmatrix}\sigma_x & \\ & \sigma_x\end{pmatrix}X_\eta^*\begin{pmatrix}\sigma_x & \\ & \sigma_x\end{pmatrix}=\begin{pmatrix}\sigma_z & \\ & \sigma_z\end{pmatrix}X_+\begin{pmatrix}\sigma_z & \\ & \sigma_z\end{pmatrix}=X_-^* \tag{C48}$$

resulting in

$$X_\eta=\Omega\begin{pmatrix}W_1\sigma_0 & \\ & W_3\sigma_0\end{pmatrix} \tag{C49}$$

with $W_1, W_3 \in \mathbb{R}$. Then, combined with the fact that small Λ_c allows us to extend the summation of $\boldsymbol{k}$ to $\mathbb{R}^2$, we arrive at

$$\begin{aligned}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{ff}(\boldsymbol{r})::\rho_{cc}(\boldsymbol{r}'):&\approx\frac{1}{\mathcal{A}}\sum_{\boldsymbol{R}}:\rho_f(\boldsymbol{R}):\sum_{\eta,s}\sum_{\beta\beta'}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}}e^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}}\int d^2r'e^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}'}:c^\dagger_{\eta,\boldsymbol{r}',\beta,s}c_{\eta,\boldsymbol{r}',\beta',s}:\Omega W_\beta\delta_{\beta\beta'}\\&=\Omega\sum_{\boldsymbol{R}}:\rho_f(\boldsymbol{R}):\sum_{\eta,s}\sum_{\beta}\frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}}e^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{R}}\int d^2r'e^{\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}'}:c^\dagger_{\eta,\boldsymbol{r}',\beta,s}c_{\eta,\boldsymbol{r}',\beta,s}:W_\beta\\&=H_{\mathrm{int},W,fc}\ ,\end{aligned} \tag{C50}$$

where $H_{\mathrm{int},W,fc}$ is defined in Eq. (58).

d. $\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}')[:\rho_{cf}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):+:\rho_{fc}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):]$

To simplify this term, first note that

$$\begin{aligned}\sum_{\widetilde{l}\sigma}g^*_{\eta\beta\widetilde{l}\sigma}(\boldsymbol{r})e^{\eta\mathrm{i}\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}w_{\eta\alpha\widetilde{l}\sigma}(\boldsymbol{r}-\boldsymbol{R})&=\sum_{\widetilde{l}\sigma}\sum_{\boldsymbol{Q}\in\mathcal{Q}_{\eta,\widetilde{l}}}e^{-\mathrm{i}\boldsymbol{Q}\cdot\boldsymbol{r}}[\widetilde{u}_{\eta,c,\beta}(0)]^*_{\boldsymbol{Q}\sigma}e^{\eta\mathrm{i}\Delta K_{\widetilde{l}}\cdot\boldsymbol{R}}\frac{1}{N\sqrt{\Omega}}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}}\sum_{\boldsymbol{Q}'\in\mathcal{Q}_{\eta,\widetilde{l}}}e^{\mathrm{i}(\boldsymbol{k}-\boldsymbol{Q}')\cdot\boldsymbol{r}}[\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]_{\boldsymbol{Q}'\sigma}\\&=\frac{1}{N\sqrt{\Omega}}\sum_{\widetilde{l}\sigma}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}}e^{\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{R})}\sum_{\boldsymbol{Q},\boldsymbol{Q}'\in\mathcal{Q}_{\eta,\widetilde{l}}}e^{\mathrm{i}(\boldsymbol{Q}-\boldsymbol{Q}')\cdot\boldsymbol{r}}[\widetilde{u}_{\eta,c,\beta}(0)]^*_{\boldsymbol{Q}\sigma}[\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]_{\boldsymbol{Q}'\sigma}\\&=\frac{1}{N\sqrt{\Omega}}\sum_{\boldsymbol{k}}^{\mathrm{MBZ}}\sum_{\boldsymbol{G}}e^{\mathrm{i}(\boldsymbol{k}-\boldsymbol{G})\cdot(\boldsymbol{r}-\boldsymbol{R})}\sum_{\boldsymbol{Q},\sigma}[\widetilde{u}_{\eta,c,\beta}(0)]^*_{\boldsymbol{Q}\sigma}[\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{k})]_{\boldsymbol{Q}+\boldsymbol{G}\sigma}\\&=\frac{1}{N\sqrt{\Omega}}\sum_{\boldsymbol{p}}e^{\mathrm{i}\boldsymbol{p}\cdot(\boldsymbol{r}-\boldsymbol{R})}\widetilde{u}_{\eta,c,\beta}(0)^\dagger\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p})\ ,\end{aligned} \tag{C51}$$

where we use Eqs. (47), (C12), (A25), and (C14). With Eqs. (C51) and (C16), we have

$$\begin{aligned}&\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cf}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):\\&\approx\frac{1}{2}\frac{1}{\mathcal{A}^2N}\int d^2rd^2r'\sum_{\eta,s}\sum_{\beta\alpha\boldsymbol{R}}:c^\dagger_{\eta,\boldsymbol{r},\beta,s}f_{\eta,\boldsymbol{R},\beta,s}:\sum_{\eta',s'}\sum_{\beta'\alpha'}:f^\dagger_{\eta',\boldsymbol{R},\alpha',s'}c_{\eta',\boldsymbol{r}',\beta',s'}:\\&\times\sum_{\boldsymbol{p},\boldsymbol{p}_1,\boldsymbol{p}_2}e^{\mathrm{i}\boldsymbol{p}_1\cdot\boldsymbol{r}-\mathrm{i}\boldsymbol{r}'\cdot\boldsymbol{p}_2-\mathrm{i}\boldsymbol{p}_1\cdot\boldsymbol{R}+\mathrm{i}\boldsymbol{p}_2\cdot\boldsymbol{R}}V(\boldsymbol{p})\widetilde{u}^\dagger_{\eta,c,\beta}(0)\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p}_1+\boldsymbol{p})\widetilde{v}^\dagger_{\eta',f,\alpha'}(\boldsymbol{p}_2+\boldsymbol{p})\widetilde{u}_{\eta',c,\beta'}(0)\ .\end{aligned} \tag{C52}$$

Here $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are carried by the c modes, and thus we can set them to be zero in $\widetilde{u}^\dagger_{\eta,c,\beta}(0)\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p}_1+\boldsymbol{p})\widetilde{v}^\dagger_{\eta',f,\alpha'}(\boldsymbol{p}_2+\boldsymbol{p})\widetilde{u}_{\eta',c,\beta'}(0)$ as a good approximation, resulting in

$$\begin{aligned}&\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cf}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):\\&\approx\frac{1}{2}\frac{1}{\mathcal{A}^2N}\int d^2rd^2r'\sum_{\eta,s}\sum_{\beta\alpha\boldsymbol{R}}:c^\dagger_{\eta,\boldsymbol{r},\beta,s}f_{\eta,\boldsymbol{R},\beta,s}:\sum_{\eta',s'}\sum_{\beta'\alpha'}:f^\dagger_{\eta',\boldsymbol{R},\alpha',s'}c_{\eta',\boldsymbol{r}',\beta',s'}:\\&\times\sum_{\boldsymbol{p},\boldsymbol{p}_1,\boldsymbol{p}_2}e^{\mathrm{i}\boldsymbol{p}_1\cdot\boldsymbol{r}-\mathrm{i}\boldsymbol{r}'\cdot\boldsymbol{p}_2-\mathrm{i}\boldsymbol{p}_1\cdot\boldsymbol{R}+\mathrm{i}\boldsymbol{p}_2\cdot\boldsymbol{R}}V(\boldsymbol{p})\widetilde{u}^\dagger_{\eta,c,\beta}(0)\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p})\widetilde{v}^\dagger_{\eta',f,\alpha'}(\boldsymbol{p})\widetilde{u}_{\eta',c,\beta'}(0)\\&=\frac{\Omega}{2}\sum_{\eta,s}\sum_{\alpha\beta\boldsymbol{R}}:c^\dagger_{\eta,\boldsymbol{R},\beta,s}f_{\eta,\boldsymbol{R},\alpha,s}:\sum_{\eta',s'}\sum_{\alpha'\beta'}:f^\dagger_{\eta,'\boldsymbol{R},\alpha',s'}c_{\eta',\boldsymbol{R},\beta',s'}:J^*_{\eta\alpha\beta,\eta'\alpha'\beta'}\ ,\end{aligned} \tag{C53}$$

where

$$J_{\eta\alpha\beta,\eta'\alpha'\beta'} = \frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}} V(\boldsymbol{p})\widetilde{v}^{\dagger}_{\eta,f,\alpha}(\boldsymbol{p})\widetilde{u}_{\eta,c,\beta}(0)\widetilde{u}_{\eta',c,\beta'}(0)\widetilde{v}^{\dagger}_{\eta',f,\alpha'}(\boldsymbol{p}) \tag{C54}$$

which satisfies

$$J_{\eta\alpha\beta,\eta'\alpha'\beta'} = J^{*}_{\eta'\alpha'\beta',\eta\alpha\beta}\ . \tag{C55}$$

Similarly, we have

$$\begin{aligned}\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{fc}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'): &= \frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{fc}(\boldsymbol{r}')::\rho_{cf}(\boldsymbol{r}):\\ &\approx \frac{\Omega}{2}\sum_{\boldsymbol{R}}\sum_{\eta,s,\alpha,\beta}\sum_{\eta',s',\alpha'\beta'}: f^{\dagger}_{\eta',\boldsymbol{R},\alpha',s'}c_{\eta',\boldsymbol{R},\beta',s'}::c^{\dagger}_{\eta,\boldsymbol{R},\beta,s}f_{\eta,\boldsymbol{R},\alpha,s}: J^{*}_{\eta\alpha\beta,\eta'\alpha'\beta'}\ .\end{aligned} \tag{C56}$$

As a result, we have

$$\begin{aligned}&\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}')[:\rho_{cf}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):+:\rho_{fc}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):]\\ &\quad\approx -\Omega\sum_{\boldsymbol{R}}\sum_{\eta,s,\alpha,\beta}\sum_{\eta',s',\alpha'\beta'}J_{\eta\alpha\beta,\eta'\alpha'\beta'}: f^{\dagger}_{\eta,\boldsymbol{R},\alpha,s}f_{\eta',\boldsymbol{R},\alpha',s'}::c^{\dagger}_{\eta',\boldsymbol{R},\beta',s'}c_{\eta,\boldsymbol{R},\beta,s}: +\text{const.}\end{aligned} \tag{C57}$$

Numerically, we find the biggest components of $J_{\eta\alpha\beta,\eta'\alpha'\beta'}$ are equal to $J_{\eta 13,\eta 13}$, $J_{\eta 24,\eta 24}$, $J_{\eta 24,-\eta 13}$, and $J_{\eta 13,-\eta 24}$, whose magnitudes are 24.25 meV in EUS. The next biggest components of $J_{\eta\alpha\beta,\eta'\alpha'\beta'}$ have magnitudes being 6.478 meV in EUS, which are roughly a quarter of those of the biggest components. Therefore we only keep the biggest components of $J_{\eta\alpha\beta,\eta'\alpha'\beta'}$. Furthermore, based on Eqs. (A24) and (A29), we find [125] that

$$J_{\eta 13,\eta 13} = J_{\eta 24,\eta 24} = -J_{\eta 24,-\eta 13} = -J_{\eta 13,-\eta 24}, \tag{C58}$$

which is independent of η. Then, we define

$$J = J_{+13,+13}\ , \tag{C59}$$

leading to

$$\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}')[:\rho_{cf}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):+:\rho_{fc}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):]\approx H_{\text{int},J}+\text{const.}\ , \tag{C60}$$

where $H_{\text{int},J}$ is defined in Eq. (59).

e. $\frac{1}{2}\int \boldsymbol{d^2rd^2r'V(r-r'):\rho_{cf}(r)::\rho_{cf}(r'):+\text{H.c.}}$

By using Eq. (C51) and Eq. (C16), the term can be simplfied to

$$\begin{aligned}&\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cf}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):+\text{H.c.}\\ &\approx\frac{1}{2}\sum_{\boldsymbol{k_1}}^{|\boldsymbol{k_1}|\leqslant\Lambda_c}\sum_{\boldsymbol{k_2}}^{|\boldsymbol{k_2}|\leqslant\Lambda_c}\sum_{\eta,s}\sum_{\alpha\beta\boldsymbol{R}}c^{\dagger}_{\eta,\boldsymbol{k}_1,\beta,s}f_{\eta,\boldsymbol{R},\alpha,s}\sum_{\eta',s'}\sum_{\alpha'\beta'}c^{\dagger}_{\eta',\boldsymbol{k}_2,\beta',s'}f_{\eta',\boldsymbol{R},\alpha',s'}\\ &\quad\times\frac{1}{N}e^{-i\boldsymbol{k}_1\cdot\boldsymbol{R}}e^{-i\boldsymbol{k}_2\cdot\boldsymbol{R}}\frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}}V(\boldsymbol{p})(\widetilde{u}^{\dagger}_{\eta,c,\beta}(0)\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p}+\boldsymbol{k}_1))(\widetilde{u}^{\dagger}_{\eta',c,\beta'}(0)\widetilde{v}_{\eta',f,\alpha'}(-\boldsymbol{p}+\boldsymbol{k}_2))+\text{H.c.}\end{aligned} \tag{C61}$$

Then, due to the small Λ_c, we can set $\boldsymbol{k}_1=0$ and $\boldsymbol{k}_2=0$ in $(\widetilde{u}^{\dagger}_{\eta,c,\beta}(0)\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p}+\boldsymbol{k}_1))(\widetilde{u}^{\dagger}_{\eta',c,\beta'}(0)\widetilde{v}_{\eta',f,\alpha'}(-\boldsymbol{p}+\boldsymbol{k}_2))$, resulting in

$$\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cf}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):+\text{H.c.}\approx H_{\text{int},\widetilde{J}}\ , \tag{C62}$$

where

$$H_{\text{int},\widetilde{J}} = \frac{\Omega}{2}\sum_{\boldsymbol{R}}\sum_{\eta,s,\alpha,\beta}\sum_{\eta',s',\alpha',\beta'}J_{-\eta'\beta'\alpha',\eta\beta\alpha}c^{\dagger}_{\eta,\boldsymbol{R},\beta,s}c^{\dagger}_{\eta',\boldsymbol{R},\beta',s'}f_{\eta',\boldsymbol{R},\alpha',s'}f_{\eta,\boldsymbol{R},\alpha,s}+\text{H.c.} \tag{C63}$$

f. $[\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{ff}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):+\text{H.c.}]$ **&** $[\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{ff}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):+\text{H.c.}]$

With Eqs. (25), (C16), (C24), (C27), and (C51), we can get

$$\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{ff}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):+\text{H.c.}\approx\frac{1}{2}\sum_{\boldsymbol{R}}:\rho_f(\boldsymbol{R}):\int d^2r'\sum_{\eta',s'}\sum_{\beta',\alpha'}c^\dagger_{\eta',\boldsymbol{r}',\beta',s'}f_{\eta',\boldsymbol{R},\beta',s'}\frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}}V(\boldsymbol{p})n_f(-\boldsymbol{p})$$
$$\times\frac{1}{\sqrt{\Omega}N}\sum_{\boldsymbol{p}_2}e^{i\boldsymbol{p}_2\cdot\boldsymbol{r}'}e^{-i\boldsymbol{p}_2\cdot\boldsymbol{R}}\widetilde{u}^\dagger_{\eta',c,\beta'}(0)\widetilde{v}_{\eta',f,\alpha'}(\boldsymbol{p}_2-\boldsymbol{p})+\text{H.c.}\tag{C64}$$

Since $\boldsymbol{p}_2$ is carried by c modes and thus is small due to the small Λ_c, we can set $\boldsymbol{p}_2=0$ in $\widetilde{u}^\dagger_{\eta',c,\beta'}(0)\widetilde{v}_{\eta',f,\alpha'}(\boldsymbol{p}_2-\boldsymbol{p})$ as a good approximation, resulting in

$$\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{ff}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):+\text{H.c.}\approx\frac{1}{2}\sum_{\boldsymbol{R}}:\rho_f(\boldsymbol{R}):\sum_{\eta',s'}\sum_{\beta',\alpha'}c^\dagger_{\eta',\boldsymbol{R},\beta',s'}f_{\eta',\boldsymbol{R},\beta',s'}[\widetilde{Y}_{\eta'}]_{\beta'\alpha'}+\text{H.c.}\,,\tag{C65}$$

where

$$\widetilde{Y}_{\eta'}=\frac{1}{\sqrt{\Omega}N}\sum_{\boldsymbol{p}}V(\boldsymbol{p})n_f(\boldsymbol{p})\widetilde{u}^\dagger_{\eta',c}(0)\widetilde{v}_{\eta',f}(\boldsymbol{p})\,.\tag{C66}$$

Based on Eqs. (A24), (A29), and (C28), we find

$$\widetilde{Y}_+=\begin{pmatrix}e^{-i\frac{2\pi}{3}\sigma_z} & \\ & \sigma_0\end{pmatrix}\widetilde{Y}_+e^{i\frac{2\pi}{3}\sigma_z}=\begin{pmatrix}\sigma_x & \\ & \sigma_x\end{pmatrix}\widetilde{Y}_+\sigma_x=-\begin{pmatrix}\sigma_z & \\ & \sigma_z\end{pmatrix}\widetilde{Y}_+\sigma_z=\widetilde{Y}^*_-\,,\tag{C67}$$

leading to

$$\widetilde{Y}_\eta=0\Rightarrow\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{ff}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):+\text{H.c.}\approx0\,.\tag{C68}$$

Similarly, we have

$$\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{ff}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):+\text{H.c.}\approx0\,.\tag{C69}$$

g. $[\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cc}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):+\text{H.c.}]+[\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cc}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):+\text{H.c.}]$

By using Eqs. (25), (C24), (C36), and (C51), we can get

$$\left[\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cc}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):+\text{H.c.}\right]+\left[\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cc}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):+\text{H.c.}\right]$$
$$=\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}')\{:\rho_{cc}(\boldsymbol{r}):,:\rho_{cf}(\boldsymbol{r}'):\}+\text{H.c.}$$
$$=\frac{1}{2}\int d^2rd^2r'\sum_{\eta,s}\sum_{\beta,\beta'}\sum_{\eta_1,s_1,\beta_1,\alpha_1}\sum_{\boldsymbol{R}}\{:c^\dagger_{\eta,\boldsymbol{r},\beta,s}c_{\eta,\boldsymbol{r},\beta',s}:,\ c^\dagger_{\eta_1,\boldsymbol{r}',\beta_1,s_1}f_{\eta_1,\boldsymbol{R},\alpha_1,s_1}\}$$
$$\times\frac{\sqrt{\Omega}}{\mathcal{A}^2}\sum_{\boldsymbol{p}}\sum_{\boldsymbol{G}}\sum_{\boldsymbol{p}_1}e^{-i\boldsymbol{r}\cdot\boldsymbol{p}}V(\boldsymbol{p}-\boldsymbol{G})\widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G})\widetilde{u}_{\eta,c,\beta'}(0)e^{i\boldsymbol{r}'\cdot\boldsymbol{p}}e^{-i(\boldsymbol{p}_1-\boldsymbol{p})\cdot\boldsymbol{R}}U^\dagger_{\eta_1,c,\beta_1}(0)U_{\eta_1,f,\alpha_1}(\boldsymbol{p}_1-\boldsymbol{p}+\boldsymbol{G})\,.\tag{C70}$$

Clearly, both $\boldsymbol{p}$ and $\boldsymbol{p}_1$ are carried by c modes, and are small due to the small Λ_c. Then, we can set $\boldsymbol{p}=\boldsymbol{p}_1=0$ in $U^\dagger_{\eta_1,c,\beta_1}(0)U_{\eta_1,f,\alpha_1}(\boldsymbol{p}_1-\boldsymbol{p}+\boldsymbol{G})$ as a good approximation, resulting in

$$\left[\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cc}(\boldsymbol{r})::\rho_{cf}(\boldsymbol{r}'):+\text{H.c.}\right]+\left[\frac{1}{2}\int d^2rd^2r'V(\boldsymbol{r}-\boldsymbol{r}'):\rho_{cc}(\boldsymbol{r})::\rho_{fc}(\boldsymbol{r}'):+\text{H.c.}\right]\approx H_{\text{int},K}\,,\tag{C71}$$

where

$$H_{\text{int},K}=\frac{1}{2}\Omega^{3/2}\sum_{\eta,s}\sum_{\beta,\beta'}\sum_{\eta_1,s_1,\beta_1,\alpha_1}\sum_{\boldsymbol{R}}\{:c^\dagger_{\eta,\boldsymbol{R},\beta,s}c_{\eta,\boldsymbol{R},\beta',s}:,c^\dagger_{\eta_1,\boldsymbol{R},\beta_1,s_1}f_{\eta_1,\boldsymbol{R},\alpha_1,s_1}\}K_{\eta\beta\beta',\eta_1\beta_1\alpha_1}+\text{H.c.}\tag{C72}$$

and

$$K_{\eta\beta\beta',\eta_1\beta_1\alpha_1}=\frac{1}{\Omega}\sum_{\boldsymbol{G}}V(\boldsymbol{G})\widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G})\widetilde{u}_{\eta,c,\beta'}(0)U^\dagger_{\eta_1,c,\beta_1}(0)\widetilde{v}_{\eta,f,\alpha_1}(\boldsymbol{G})\,.\tag{C73}$$

Numerically, we find the biggest components of $K_{\eta\beta\beta',\eta_1\beta_1\alpha_1}$ have magnitudes being 7.054 meV in EUS.

h. In sum

In sum, we have

$$H_{\text{int}}^{\text{TBG}} \approx H_{\text{int},U} + H_{\text{int},V,c} + H_{\text{int},W,fc} + H_{\text{int},J} + H_{\text{int},\tilde{J}} + H_{\text{int},K} + \text{const.}\,, \tag{C74}$$

where the definitions of $H_{\text{int},U}$, $H_{\text{int},V,c}$, $H_{\text{int},W,fc}$, $H_{\text{int},J}$, $H_{\text{int},\tilde{J}}$, and $H_{\text{int},K}$ can be found in Eqs. (55), (56), (58), (59), (C63), and (C72), respectively. Among all these terms, only $H_{\text{int},\tilde{J}}$ and $H_{\text{int},K}$ do not preserve the number of f modes. Moreover, according to Table II, the strengths of $H_{\text{int},\tilde{J}}$ and $H_{\text{int},K}$ are small compared to the onsite interaction among f modes in $H_{\text{int},U}$, as $|J| \sim U_1/4$ and $|K_{\eta\beta\beta',\eta_1\beta_1\alpha_1}| < U_1/10$. Therefore we neglect $H_{\text{int},\tilde{J}}$ and $H_{\text{int},K}$. We can also neglect the *const*. in $H_{\text{int}}^{\text{TBG}}$, since it is just a shift in the total energy, leading to

$$H_{\text{int}}^{\text{TBG}} \approx H_{\text{int},U} + H_{\text{int},V,c} + H_{\text{int},W,fc} + H_{\text{int},J}\,, \tag{C75}$$

At the end of this part, we address the issue of the $\sqrt{2}$ scaling. As discussed in Sec. III A 1, the parameters values of the single-particle TBG block of MATSTG are $\sqrt{2}$ scaled compared to those of the ordinary MATBG discussed in Ref. [125]. As shown in Table II, the same $\sqrt{2}$ scaling does not necessarily occur to the interaction strengthes in $H_{\text{int}}^{\text{TBG}}$ of the MATSTG compared to those in Ref. [125]. It is because we choose the gate distance [Eq. (27)] for MATSTG to be the same as that for MATBG, since there is no obvious reason for us to decrease the gate distance by a factor of $\sqrt{2}$ when switching MATBG to MATSTG. Therefore the relative ratios among the interaction strenghes in $H_{\text{int}}^{\text{TBG}}$ of the MATSTG are not the same as those in Ref. [125], allowing W_1 and W_3 to be slightly larger than U_1. Nevertheless, we should still expect U_1 dominates the low-energy physics since W_1 and W_3 involve c modes with relatively higher energies, while U_1 only involves the low-energy f modes.

2. Details on $H_{\text{int}}^{\text{TBG-}D}$ and H_{int}^{D}

Now we turn to the other two terms in Eq. (C5), i.e., $H_{\text{int}}^{\text{TBG-}D}$ and H_{int}^{D}, which are not covered in Ref. [125]. First note that

$$d^\dagger_{\eta,\boldsymbol{r},\sigma,s} = \frac{1}{\sqrt{\mathcal{A}}}\sum_{\boldsymbol{p}} e^{-i\boldsymbol{p}\cdot\boldsymbol{r}} d^\dagger_{\eta,\boldsymbol{p},\sigma,s} = \frac{1}{\sqrt{\mathcal{A}}}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_d} e^{-i\boldsymbol{p}\cdot\boldsymbol{r}} d^\dagger_{\eta,\boldsymbol{p},\sigma,s} + \cdots = \widetilde{d}^\dagger_{\eta,\boldsymbol{r},\sigma,s} + \ldots\,, \tag{C76}$$

where $\widetilde{d}^\dagger_{\eta,\boldsymbol{r},\sigma,s}$ is defined under Eq. (60), and "..." represents the higher-energy d modes. Then, we know

$$\rho_D(\boldsymbol{r}) = \rho_d(\boldsymbol{r}) + \ldots\,, \tag{C77}$$

where $\rho_d(\boldsymbol{r})$ is defined under Eq. (60). Combined with Eq. (C25), we have

$$\begin{aligned} H_{\text{int}}^{\text{TBG-}D} = &\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \widetilde{\rho}(\boldsymbol{r}) :: \rho_D(\boldsymbol{r}'): \approx \int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') : \\ &+ \left[\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cf}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') : + \text{H.c.}\right] + \int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cc}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') :\,. \end{aligned} \tag{C78}$$

Furthermore, we have

$$H_{\text{int}}^{D} = \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_D(\boldsymbol{r}) :: \rho_D(\boldsymbol{r}'): \approx \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_d(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') := H_{\text{int},V,d} \tag{C79}$$

with $H_{\text{int},V,d}$ defined in Eq. (60).

In the following, we will discuss each term in Eq. (C26).

a. $\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') :$

With Eqs. (25), (C24), and (C27), we have

$$\begin{aligned} &\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') : \\ &= \int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}') \sum_{\boldsymbol{R}} : \rho_f(\boldsymbol{R}) : n_f(\boldsymbol{r}-\boldsymbol{R}) : \rho_d(\boldsymbol{r}') : \\ &= \int d^2r d^2r' \frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}} e^{-i\boldsymbol{p}\cdot(\boldsymbol{r}-\boldsymbol{r}')} V(\boldsymbol{p}) \sum_{\boldsymbol{R}} : \rho_f(\boldsymbol{R}) : \frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}_1} n_f(\boldsymbol{p}_1) e^{-i\boldsymbol{p}_1\cdot(\boldsymbol{r}-\boldsymbol{R})} : \rho_d(\boldsymbol{r}') : \\ &= \sum_{\boldsymbol{R}} : \rho_f(\boldsymbol{R}) : \int d^2r' \frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}} e^{i\boldsymbol{p}\cdot\boldsymbol{r}'} V(\boldsymbol{p}) n_f(-\boldsymbol{p}) e^{-i\boldsymbol{p}\cdot\boldsymbol{R}} : \rho_d(\boldsymbol{r}') :\,. \end{aligned} \tag{C80}$$

Since $\boldsymbol{p}$ is carried by d modes here and Λ_d is small, we can adopt $V(\boldsymbol{p})n_f(-\boldsymbol{p}) \approx V(\boldsymbol{p}=0)n_f(\boldsymbol{p}=0)$. This approximation is rather good since if we choose $\boldsymbol{p} = \frac{\boldsymbol{q}_3}{6}$ with $\boldsymbol{q}_3$ defined in Eq. (6), the error is less than 6%, i.e.,

$$\frac{V\left(\frac{\boldsymbol{q}_3}{6}\right)n_f\left(-\frac{\boldsymbol{q}_3}{6}\right)}{V(\boldsymbol{p}=0)n_f(\boldsymbol{p}=0)} > 94\% \,. \tag{C81}$$

Then, by defining

$$W_{fd} = \frac{1}{\Omega}V(\boldsymbol{p}=0)n_f(\boldsymbol{p}=0) \,, \tag{C82}$$

we have

$$\int d^2rd^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{ff}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}'): \approx \sum_{\boldsymbol{R}} : \rho_f(\boldsymbol{R}) :: \rho_d(\boldsymbol{R}) : V(\boldsymbol{p}=0)n_f(\boldsymbol{p}=0) = \Omega \sum_{\boldsymbol{R}} : \rho_f(\boldsymbol{R}) :: \rho_d(\boldsymbol{R}) : W_{fd} = H_{\text{int},W,fd} \,, \tag{C83}$$

where $H_{\text{int},W,fd}$ is defined in Eq. (61). The numerical value of W_{fd} is listed in Table II.

b. $\int d^2rd^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cf}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') : +\text{H.c.}$

With Eqs. (25), (C24), and (C51), we can get

$$\begin{aligned}
&\int d^2rd^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{fc}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') : +\text{H.c.} \\
&= \int d^2rd^2r' \frac{1}{\mathcal{A}} \sum_{\boldsymbol{p}} e^{-\mathrm{i}\boldsymbol{p}\cdot(\boldsymbol{r}-\boldsymbol{r}')} V(\boldsymbol{p}) \sum_{\eta,s} \sum_{\beta,\alpha,\boldsymbol{R}} : c^\dagger_{\eta,\boldsymbol{r},\beta,s} f_{\eta,\boldsymbol{R},\alpha,s} : \frac{1}{N\sqrt{\Omega}} \sum_{\boldsymbol{p}_1} e^{\mathrm{i}\boldsymbol{p}_1\cdot(\boldsymbol{r}-\boldsymbol{R})} \widetilde{u}^\dagger_{\eta,c,\beta}(0) \widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p}_1) : \rho_d(\boldsymbol{r}') : +\text{H.c.} \\
&= \int d^2rd^2r' \frac{\sqrt{\Omega}}{\mathcal{A}^2} \sum_{\eta,s} \sum_{\beta,\alpha,\boldsymbol{R}} : c^\dagger_{\eta,\boldsymbol{r},\beta,s} f_{\eta,\boldsymbol{R},\alpha,s} : \sum_{\boldsymbol{p},\boldsymbol{p}_1} e^{\mathrm{i}(\boldsymbol{p}_1-\boldsymbol{p})\cdot\boldsymbol{r}} V(\boldsymbol{p}) e^{-\mathrm{i}\boldsymbol{p}_1\cdot\boldsymbol{R}} \widetilde{u}^\dagger_{\eta,c,\beta}(0) \widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p}_1) e^{\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}'} : \rho_d(\boldsymbol{r}') : +\text{H.c.} \\
&= \int d^2rd^2r' \frac{\sqrt{\Omega}}{\mathcal{A}^2} \sum_{\eta,s} \sum_{\beta,\alpha,\boldsymbol{R}} : c^\dagger_{\eta,\boldsymbol{r},\beta,s} f_{\eta,\boldsymbol{R},\alpha,s} : \sum_{\boldsymbol{p},\boldsymbol{p}_1} e^{\mathrm{i}\boldsymbol{p}_1\cdot\boldsymbol{r}} V(\boldsymbol{p}) e^{-\mathrm{i}(\boldsymbol{p}_1+\boldsymbol{p})\cdot\boldsymbol{R}} \widetilde{u}^\dagger_{\eta,c,\beta}(0) \widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p}_1+\boldsymbol{p}) e^{\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}'} : \rho_d(\boldsymbol{r}') : +\text{H.c.}
\end{aligned} \tag{C84}$$

As $\boldsymbol{p}_1$ is carried by c modes and $\boldsymbol{p}$ is carried by d modes, both of them are small, and we can adopt

$$\widetilde{u}^\dagger_{\eta,c,\beta}(0)\widetilde{v}_{\eta,f,\alpha}(\boldsymbol{p}_1+\boldsymbol{p}) \approx \widetilde{u}^\dagger_{\eta,c,\beta}(0)\widetilde{v}_{\eta,f,\alpha}(0) = 0 \tag{C85}$$

as a good approximation, where the second equality comes from the orthogonality of $\widetilde{u}$ and $\widetilde{v}$ at the same momentum. Then, we know

$$\int d^2rd^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cf}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') : +\text{H.c.} \approx 0 \,. \tag{C86}$$

c. $\int d^2rd^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cc}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') :$

With Eqs. (25), (C24), and (C36), we can get

$$\begin{aligned}
&\int d^2rd^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cc}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') : \\
&= \int d^2rd^2r' \frac{1}{\mathcal{A}} \sum_{\boldsymbol{p}} e^{-\mathrm{i}\boldsymbol{p}\cdot(\boldsymbol{r}-\boldsymbol{r}')} V(\boldsymbol{p}) \sum_{\eta,s,\beta,\beta'} : c^\dagger_{\eta,\boldsymbol{r},\beta,s} c_{\eta,\boldsymbol{r},\beta',s} : \sum_{\boldsymbol{G}} e^{-\mathrm{i}\boldsymbol{G}\cdot\boldsymbol{r}} \widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G}) \widetilde{u}_{\eta,c,\beta'}(0) : \rho_d(\boldsymbol{r}') : \\
&= \int d^2rd^2r' \frac{1}{\mathcal{A}} \sum_{\boldsymbol{G}} \sum_{\boldsymbol{p}} e^{-\mathrm{i}(\boldsymbol{p}+\boldsymbol{G})\cdot\boldsymbol{r}} V(\boldsymbol{p}) \sum_{\eta,s,\beta,\beta'} : c^\dagger_{\eta,\boldsymbol{r},\beta,s} c_{\eta,\boldsymbol{r},\beta',s} : e^{\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}'} \widetilde{u}^\dagger_{\eta,c,\beta}(\boldsymbol{G}) \widetilde{u}_{\eta,c,\beta'}(0) : \rho_d(\boldsymbol{r}') : \,.
\end{aligned} \tag{C87}$$

As $\boldsymbol{p}+\boldsymbol{G}$ is carried by c modes and $\boldsymbol{p}$ is carried by d modes, both of them should be small, and thus we should only keep $\boldsymbol{G}=0$ in summation, resulting in

$$\begin{aligned}
\int d^2rd^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_{cc}(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') : &= \int d^2rd^2r' \frac{1}{\mathcal{A}} \sum_{\boldsymbol{p}} e^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}} V(\boldsymbol{p}) \sum_{\eta,s,\beta,\beta'} : c^\dagger_{\eta,\boldsymbol{r},\beta,s} c_{\eta,\boldsymbol{r},\beta',s} : e^{\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}'} \widetilde{u}^\dagger_{\eta,c,\beta}(0) \widetilde{u}_{\eta,c,\beta'}(0) : \rho_d(\boldsymbol{r}') : \\
&= \int d^2rd^2r' \frac{1}{\mathcal{A}} \sum_{\boldsymbol{p}} e^{-\mathrm{i}\boldsymbol{p}\cdot(\boldsymbol{r}-\boldsymbol{r}')} V(\boldsymbol{p}) \sum_{\eta,s,\beta} : c^\dagger_{\eta,\boldsymbol{r},\beta,s} c_{\eta,\boldsymbol{r},\beta,s} :: \rho_d(\boldsymbol{r}') : \\
&= \int d^2rd^2r' V(\boldsymbol{r}-\boldsymbol{r}') : \rho_c(\boldsymbol{r}) :: \rho_d(\boldsymbol{r}') := H_{\text{int},V,cd} \,,
\end{aligned} \tag{C88}$$

where $H_{\text{int},V,cd}$ is defined in Eq. (62).

d. In sum

In sum, we have

$$H_{\text{int}}^{\text{TBG-}D} \approx H_{\text{int},W,fd} + H_{\text{int},V,cd} \ , \tag{C89}$$

where $H_{\text{int},W,fd}$ is defined in Eq. (61), and $H_{\text{int},V,cd}$ is defined in Eq. (62). Furthermore, we have

$$H_{\text{int}}^{D} \approx H_{\text{int},V,d} \tag{C90}$$

with $H_{\text{int},V,d}$ defined in Eq. (60).

APPENDIX D: MORE DETAILS ON THE NUMERICAL HARTREE-FOCK CALCULATIONS

In this section, we provide more details on the numerical Hartree-Fock Calculations.

1. Hartree-Fock Hamiltonian

We first present more details for the Hartree-Fock Hamiltonian. In general, given an interacting Hamiltonian of the form

$$H = \sum_{i,j} \psi_i^\dagger \psi_j t_{ij} + \frac{1}{2} \sum_{i_1,i_2,i_3,i_4} U_{i_1,i_2,i_3,i_4} \psi_{i_1}^\dagger \psi_{i_2}^\dagger \psi_{i_3} \psi_{i_4} \tag{D1}$$

with some generic fermion annihilation operator ψ_i. The Hartree-Fock approximation is to choose the ground state as a single Slater determinant:

$$|\Psi\rangle = a_1^\dagger a_2^\dagger a_3^\dagger \dots a_N^\dagger |0\rangle \ , \tag{D2}$$

where $a_n^\dagger = \sum_i \psi_i^\dagger (\zeta_n)_i$ and ζ_1, ζ_2,..., ζ_N are orthonormal vectors. Then, the Hartree-Fock Hamiltonian is derived as

$$H_{\text{HF}} = \sum_{i,j} \psi_i^\dagger \psi_j t_{ij} + \frac{1}{2} \sum_{i_1,i_2,i_3,i_4} U_{i_1 i_2 i_3 i_4} \left(\psi_{i_1}^\dagger \psi_{i_4} O_{i_2 i_3} + \psi_{i_2}^\dagger \psi_{i_3} O_{i_1 i_4} - \psi_{i_2}^\dagger \psi_{i_4} O_{i_1 i_3} - \psi_{i_1}^\dagger \psi_{i_3} O_{i_2 i_4} \right) - E_0 \ , \tag{D3}$$

where $O_{i_1 i_2} = \langle \Psi | \psi_{i_1}^\dagger \psi_{i_2} | \Psi \rangle = \sum_{n=1}^N (\zeta_n^* \zeta_n^T)_{ij}$ and

$$E_0 = \frac{1}{2} \sum_{i_1,i_2,i_3,i_4} U_{i_1 i_2 i_3 i_4} \left(O_{i_1 i_4} O_{i_2 i_3} - O_{i_1 i_3} O_{i_2 i_4} \right). \tag{D4}$$

H_{HF} satisfies $\langle \Psi | H_{\text{HF}} | \Psi \rangle = \langle \Psi | H | \Psi \rangle$. Note that H_{HF} has the same form as the mean-field Hamiltonian; in fact, the Hartree-Fock approximation is equivalent to the mean-field approximation. $O_{i_1 i_2}$ is called the order parameter or the density matrix. $\langle \Psi | H_{\text{HF}} | \Psi \rangle$ is also called the Hartree-Fock energy.

Now we come back to MATSTG. We only consider the states that are invariant under the Moiré lattice translations. Moreover, similar to Ref. [125], we only care about the following averaged density matrices for simplicity

$$\begin{aligned}
O^{ff}_{\eta_1 \alpha_1 s_1, \eta_2 \alpha_2 s_2} &= \frac{1}{N} \sum_{\boldsymbol{R}} \left\langle f^\dagger_{\eta_1, \boldsymbol{R}, \alpha_1, s_1} f_{\eta_2, \boldsymbol{R}, \alpha_2, s_2} \right\rangle, \\
O^{cc}_{\eta_1 \beta_1 s_1, \eta_2 \beta_2 s_2} &= \frac{1}{N} \sum_{\boldsymbol{p}}^{|\boldsymbol{p}| \leqslant \Lambda_c} \left\langle c^\dagger_{\eta_1, \boldsymbol{p}, \beta_1, s_1} c_{\eta_2, \boldsymbol{p}, \beta_2, s_2} \right\rangle - \frac{n_{\Lambda_c}}{2} \delta_{\eta_1 \eta_2} \delta_{\beta_1 \beta_2} \delta_{s_1 s_2}, \\
O^{dd}_{\eta \sigma_1 s_1, \eta \sigma_2 s_2} &= \frac{1}{N} \sum_{\boldsymbol{p}}^{|\boldsymbol{p}| \leqslant \Lambda_d} \left\langle d^\dagger_{\eta, \boldsymbol{p}, \sigma_1, s_1} d_{\eta, \boldsymbol{p}, \sigma_2, s_2} \right\rangle - \frac{n_{\Lambda_d}}{2} \delta_{\sigma_1 \sigma_2} \delta_{s_1 s_2}, \\
O^{dd}_{\eta \beta_1 s_1, -\eta \beta_2 s_2} &= \frac{1}{3N} \sum_{\boldsymbol{p}}^{|\boldsymbol{p}| \leqslant \Lambda_d} \sum_{\boldsymbol{p}'}^{|\boldsymbol{p}'| \leqslant \Lambda_d} \sum_{n=0,1,2} \delta_{\boldsymbol{p} - \boldsymbol{p}', C_3^n \eta \boldsymbol{q}_1} \left\langle d^\dagger_{\eta, \boldsymbol{p}, \sigma_1, s_1} d_{-\eta, \boldsymbol{p}', \sigma_2, s_2} \right\rangle, \\
O^{cf}_{\eta_1 \beta_1 s_1, \eta_2 \alpha_2 s_2} &= \frac{1}{N} \sum_{\boldsymbol{p}}^{|\boldsymbol{p}| \leqslant \Lambda_c} \left\langle c^\dagger_{\eta_1, \boldsymbol{p}, \beta_1, s_1} f_{\eta_2, \boldsymbol{p}, \alpha_2, s_2} \right\rangle, \\
O^{fc} &= [O^{cf}]^\dagger,
\end{aligned}$$

$$
\begin{aligned}
O^{df}_{\eta_1\sigma_1 s_1,\eta_2\alpha_2 s_2} &= \frac{1}{N}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c}\left\langle d^\dagger_{\eta_1,\boldsymbol{p},\sigma_1,s_1} f_{\eta_2,\boldsymbol{p}+\eta_1 K_M,\alpha_2,s_2}\right\rangle,\\
O^{fd} &= [O^{df}]^\dagger,\\
O^{cd}_{\eta_1\beta_1 s_1,\eta_2\sigma_2 s_2} &= \frac{1}{3N}\sum_{n=0,1,2}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c\&|\boldsymbol{p}-\eta_2 C_3^n\boldsymbol{q}_1|\leqslant\Lambda_d}\left\langle c^\dagger_{\eta_1,\boldsymbol{p},\beta_1,s_1} d_{\eta_2,\boldsymbol{p}-\eta_2 C_3^n\boldsymbol{q}_1,\sigma_2,s_2}\right\rangle,\\
O^{dc} &= [O^{cd}]^\dagger\,,
\end{aligned}
\tag{D5}
$$

where $\langle\dots\rangle$ is the expectation done with respect to Hartree-Fock ground state, $n_{\Lambda_c}=\frac{1}{N}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c}$, and $n_{\Lambda_d}=\frac{1}{N}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_d}$. We note that the expressions O^{ff}, O^{cc}, O^{fc} and O^{cf} are the same as those in Ref. [125]. We also note that for $d^\dagger_{\eta,\boldsymbol{p},\sigma_1,s_1}$ and $d_{-\eta,\boldsymbol{p}',\sigma_2,s_2}$ in $O^{dd}_{\eta\beta_1 s_1,-\eta\beta_2 s_2}$, $\boldsymbol{p}$ and $\boldsymbol{p}'$ must be different in order to preserve the Moiré lattice translations, owing to the fact that $d^\dagger_{+,\boldsymbol{p}}$ and $d^\dagger_{-,\boldsymbol{p}}$ are around K_M and $-K_M$ points, respectively, as discussed in and below Eq. (A30). Because of the same reason, we choose the d and f modes in $O^{df}_{\eta_1\sigma_1 s_1,\eta_2\alpha_2 s_2}$ as $d^\dagger_{\eta_1,\boldsymbol{p},\sigma_1,s_1}$ and $f_{\eta_2,\boldsymbol{p}+\eta_1 K_M,\alpha_2,s_2}$ to preserve the Moiré lattice translations.

Then, combining Eq. (64) with Eqs. (D5) and (D3), the Hartree-Fock Hamiltonian reads

$$
\begin{aligned}
H_{\rm HF} &= \sum_\eta H^{\rm eff}_{0,\eta} + H_U + H_{V,c} + H_{V,d} + H_{V,cd} + H_{W,fc} + H_{W,fd} + H_J\\
&\quad - (E_U + E_{V,c} + E_{V,d} + E_{V,cd} + E_{W,fc} + E_{W,fd} + E_J) + \text{const.}\,,
\end{aligned}
\tag{D6}
$$

where $H^{\rm eff}_{0,\eta}$ is in Eq. (54), "const." stands for a scalar that is independent of the ground state, and the rest of the terms are discussed in the following. Before going in to details, we define $f^\dagger_{\boldsymbol{k}}=(\dots,f^\dagger_{\eta,\boldsymbol{k},\alpha,s},\dots)$, $f^\dagger_{\boldsymbol{R}}=(\dots,f^\dagger_{\eta,\boldsymbol{R},\alpha,s},\dots)$, $c^\dagger_{\boldsymbol{p}}=(\dots,c^\dagger_{\eta,\boldsymbol{p},\beta,s},\dots)$, and $d^\dagger_{\boldsymbol{p}}=(\dots,d^\dagger_{\eta,\boldsymbol{p},\sigma,s},\dots)$.

First, we go over H_U, $H_{V,c}$, $H_{W,fc}$, H_J, E_U, $E_{V,c}$, $E_{W,fc}$, and E_J, which are the same as the corresponding Hartree-Fock terms in Ref. [125] since they only involve the f and c modes derived from the TBG part. For more details, one can refer to Ref. [125].

$$
H_U = \sum_{\boldsymbol{R}}\left\{\rho_f(\boldsymbol{R})[U_1(\mathrm{Tr}[O^{ff}]-3.5)+6U_2(\mathrm{Tr}[O^{ff}]-4)] - U_1 f^\dagger_{\boldsymbol{R}}[O^{ff}]^T f_{\boldsymbol{R}}\right\}\,,
\tag{D7}
$$

$$
E_U = \frac{N}{2}\mathrm{Tr}[O^{ff}]^2(U_1+6U_2) - U_1\frac{N}{2}\mathrm{Tr}[O^{ff}O^{ff}]\,,
\tag{D8}
$$

$$
H_{V,c} = \frac{1}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} c^\dagger_{\boldsymbol{p}}c_{\boldsymbol{p}}\,\mathrm{Tr}[O^{cc}]\,,
\tag{D9}
$$

and

$$
E_{V,c} = \frac{N}{2\Omega}V(\boldsymbol{p}=0)(\mathrm{Tr}[O^{cc}]^2 + 16n_{\Lambda_c}\,\mathrm{Tr}[O^{cc}])\,.
\tag{D10}
$$

Here we neglect the Fock channels for $H_{V,c}$ and $E_{V,c}$, same to Ref. [125], since otherwise the Hartree-Fock calculations would heavily depend on the cutoff Λ_c due to the simplified density matrices chosen in Eq. (D5).

$$
\begin{aligned}
H_{W,fc} &= \sum_{\boldsymbol{k}\in\mathrm{MBZ}} f^\dagger_{\boldsymbol{k}}f_{\boldsymbol{k}}\,\mathrm{Tr}\left[O^{cc}\begin{pmatrix}W_1\eta_0\tau_0 s_0 & \\ & W_3\eta_0\tau_0 s_0\end{pmatrix}\right] - \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} c^\dagger_{\boldsymbol{p}}\begin{pmatrix}W_1\eta_0\tau_0 s_0 & \\ & W_3\eta_0\tau_0 s_0\end{pmatrix}[O^{fc}]^T f_{\boldsymbol{p}}\\
&\quad - \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} f^\dagger_{\boldsymbol{p}}[O^{cf}]^T\begin{pmatrix}W_1\eta_0\tau_0 s_0 & \\ & W_3\eta_0\tau_0 s_0\end{pmatrix}c_{\boldsymbol{p}} + \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} c^\dagger_{\boldsymbol{p}}\begin{pmatrix}W_1\eta_0\tau_0 s_0 & \\ & W_3\eta_0\tau_0 s_0\end{pmatrix}c_{\boldsymbol{p}}(\mathrm{Tr}[O^{ff}]-4)\,,
\end{aligned}
\tag{D11}
$$

$\eta_{0,x,y,z}$ are Pauli matrices for the valley index, and

$$
\begin{aligned}
E_{W,fc} &= N\,\mathrm{Tr}[O^{ff}]\,\mathrm{Tr}\left[O^{cc}\begin{pmatrix}W_1\eta_0\tau_0 s_0 & \\ & W_3\eta_0\tau_0 s_0\end{pmatrix}\right] + 2Nn_{\Lambda_c}\left(\sum_\beta W_\beta\right)(\mathrm{Tr}[O^{ff}]-4)\\
&\quad - N\,\mathrm{Tr}\left[O^{cf}O^{fc}\begin{pmatrix}W_1\eta_0\tau_0 s_0 & \\ & W_3\eta_0\tau_0 s_0\end{pmatrix}\right].
\end{aligned}
\tag{D12}
$$

$$
\begin{aligned}
H_J = &-\frac{J}{2}\sum_{\boldsymbol{k}\in\mathrm{MBZ}} f_{\boldsymbol{k}}^{\dagger}\left[\eta_z\tau_0 s_0\left(O^{cc}_{\Gamma_1\Gamma_2,\Gamma_1\Gamma_2}\right)^T\eta_z\tau_0 s_0+\eta_0\tau_z s_0\left(O^{cc}_{\Gamma_1\Gamma_2,\Gamma_1\Gamma_2}\right)^T\eta_0\tau_z s_0\right]f_{\boldsymbol{k}}\\
&+\frac{J}{2}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} c^{\dagger}_{\boldsymbol{p},\Gamma_1\Gamma_2}\left[\eta_z\tau_0 s_0\,\mathrm{Tr}\left[O^{fc}\begin{pmatrix}0_{8\times8}\\ \eta_z\tau_0 s_0\end{pmatrix}\right]+\eta_0\tau_z s_0\,\mathrm{Tr}\left[O^{fc}\begin{pmatrix}0_{8\times8}\\ \eta_0\tau_z s_0\end{pmatrix}\right]\right]f_{\boldsymbol{p}}\\
&+\frac{J}{2}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} f^{\dagger}_{\boldsymbol{p}}\left[\eta_z\tau_0 s_0\,\mathrm{Tr}\left[O^{fc}\begin{pmatrix}0_{8\times8}\\ \eta_z\tau_0 s_0\end{pmatrix}\right]^*+\eta_0\tau_z s_0\,\mathrm{Tr}\left[O^{fc}\begin{pmatrix}0_{8\times8}\\ \eta_0\tau_z s_0\end{pmatrix}\right]^*\right]c_{\boldsymbol{p},\Gamma_1\Gamma_2}\\
&-\frac{J}{2}\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} c^{\dagger}_{\boldsymbol{p},\Gamma_1\Gamma_2}[\eta_z\tau_0 s_0(O^{ff})^T\eta_z\tau_0 s_0+\eta_0\tau_z s_0(O^{ff})^T\eta_0\tau_z s_0-\eta_0\tau_0 s_0]c_{\boldsymbol{p},\Gamma_1\Gamma_2}\,,
\end{aligned}
\tag{D13}
$$

and

$$
\begin{aligned}
E_J = &-\frac{JN}{2}\sum_{\eta\eta'}\sum_{\alpha\alpha' ss'} O^{ff}_{\eta\alpha s,\eta'\alpha' s'}O^{cc}_{\eta'(\alpha'+2)s',\eta(\alpha+2)s}(\eta\eta'+(-1)^{\alpha+\alpha'})-JNn_{\Lambda_c}\frac{\mathrm{Tr}[O^{ff}]}{2}\\
&+\frac{JN}{2}\sum_{\eta\eta',ss',\alpha\alpha'} O^{cf}_{\eta(\alpha+2)s,\eta\alpha s}O^{fc}_{\eta'\alpha' s',\eta'(\alpha'+2)s'}(\eta\eta'+(-1)^{\alpha+\alpha'})\,,
\end{aligned}
\tag{D14}
$$

where $O^{cc}_{\Gamma_1\Gamma_2,\Gamma_1\Gamma_2}$ is the 8×8 diagonal block of O^{cc} that correspond to $c^{\dagger}_{\eta,\boldsymbol{k},\beta=3,4,s}$.

Now we move onto the terms that are not covered in Ref. [125]. First, $H_{\mathrm{int},V,d}$ in Eq. (60) can be rewritten as

$$
\begin{aligned}
H_{\mathrm{int},V,d} &= \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}')\,{:}\,\rho_d(\boldsymbol{r})::\rho_d(\boldsymbol{r}')\,{:}\\
&= \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}')\left(\rho_d(\boldsymbol{r})-\frac{4}{\Omega}n_{\Lambda_d}\right)\left(\rho_d(\boldsymbol{r}')-\frac{4}{\Omega}n_{\Lambda_d}\right)\\
&= \frac{1}{2}\int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}')\left[\rho_d(\boldsymbol{r})\rho_d(\boldsymbol{r}')-\frac{4}{\Omega}n_{\Lambda_d}(\rho_d(\boldsymbol{r})+\rho_d(\boldsymbol{r}'))\right]+\mathrm{const.}\\
&= \frac{1}{2}\frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}}\sum_{\boldsymbol{p}_1\boldsymbol{p}_2\boldsymbol{p}_3\boldsymbol{p}_4}^{\Lambda_d}\delta_{\boldsymbol{p}_4,\boldsymbol{p}+\boldsymbol{p}_1}\delta_{\boldsymbol{p}_2,\boldsymbol{p}+\boldsymbol{p}_3}V(\boldsymbol{p})\sum_{\eta\sigma s}\sum_{\eta'\sigma' s'} d^{\dagger}_{\eta,\boldsymbol{p}_1,\sigma,s}d^{\dagger}_{\eta',\boldsymbol{p}_2,\sigma',s'}d_{\eta',\boldsymbol{p}_3,\sigma',s'}d_{\eta,\boldsymbol{p}_4,\sigma,s}\\
&\quad+\frac{1}{2}\frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}_1\boldsymbol{p}}^{\Lambda_d}V(\boldsymbol{p}_1-\boldsymbol{p})\sum_{\eta\sigma s} d^{\dagger}_{\eta,\boldsymbol{p},\sigma,s}d^{\dagger}_{\eta,\boldsymbol{p},\sigma,s}-\frac{4n_{\Lambda_d}}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{\Lambda_d}\sum_{\eta\sigma s} d^{\dagger}_{\eta,\boldsymbol{p},\sigma,s}d^{\dagger}_{\eta,\boldsymbol{p},\sigma,s}\,,
\end{aligned}
\tag{D15}
$$

which leads to the following Hartree-Fock $H_{V,d}$

$$
\begin{aligned}
H_{V,d} = &\frac{1}{2}\frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}}\sum_{\boldsymbol{p}_1\boldsymbol{p}_2\boldsymbol{p}_3\boldsymbol{p}_4}^{\Lambda_d}\delta_{\boldsymbol{p}_4,\boldsymbol{p}+\boldsymbol{p}_1}\delta_{\boldsymbol{p}_2,\boldsymbol{p}+\boldsymbol{p}_3}V(\boldsymbol{p})\sum_{\eta\sigma s}\sum_{\eta'\sigma' s'}\Big[d^{\dagger}_{\eta,\boldsymbol{p}_1,\sigma,s}d_{\eta,\boldsymbol{p}_4,\sigma,s}\langle d^{\dagger}_{\eta',\boldsymbol{p}_2,\sigma',s'}d_{\eta',\boldsymbol{p}_3,\sigma',s'}\rangle\\
&-d^{\dagger}_{\eta,\boldsymbol{p}_1,\sigma,s}d_{\eta',\boldsymbol{p}_3,\sigma',s'}\langle d^{\dagger}_{\eta',\boldsymbol{p}_2,\sigma',s'}d_{\eta,\boldsymbol{p}_4,\sigma,s}\rangle-d^{\dagger}_{\eta',\boldsymbol{p}_2,\sigma',s'}d_{\eta,\boldsymbol{p}_4,\sigma,s}\langle d^{\dagger}_{\eta,\boldsymbol{p}_1,\sigma,s}d_{\eta',\boldsymbol{p}_3,\sigma',s'}\rangle\\
&+d^{\dagger}_{\eta',\boldsymbol{p}_2,\sigma',s'}d_{\eta',\boldsymbol{p}_3,\sigma',s'}\langle d^{\dagger}_{\eta,\boldsymbol{p}_1,\sigma,s}d_{\eta,\boldsymbol{p}_4,\sigma,s}\rangle\Big]+\frac{1}{2}\frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}_1\boldsymbol{p}}^{\Lambda_d}V(\boldsymbol{p}_1-\boldsymbol{p})\sum_{\eta\sigma s} d^{\dagger}_{\eta,\boldsymbol{p},\sigma,s}d^{\dagger}_{\eta,\boldsymbol{p},\sigma,s}\\
&-\frac{4n_{\Lambda_d}}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{\Lambda_d}\sum_{\eta\sigma s} d^{\dagger}_{\eta,\boldsymbol{p},\sigma,s}d^{\dagger}_{\eta,\boldsymbol{p},\sigma,s}\\
&\xrightarrow{\text{neglecting Fock channel}}\frac{1}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{\Lambda_d}d^{\dagger}_{\boldsymbol{p}}d_{\boldsymbol{p}}\,\mathrm{Tr}[O^{dd}]\,,
\end{aligned}
\tag{D16}
$$

where we have used Eq. (D5). Similarly, we get $E_{V,d}$ as

$$
E_{V,d} = \frac{N}{2\Omega}V(\boldsymbol{p}=0)(\mathrm{Tr}[O^{dd}]^2+8n_{\Lambda_d}\,\mathrm{Tr}[O^{dd}])\,.
\tag{D17}
$$

Second, $H_{\text{int},V,cd}$ in Eq. (62) can be rewritten as

$$\begin{aligned}
H_{\text{int},V,cd} &= \int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}')\left(\rho_c(\boldsymbol{r})-\frac{8}{\Omega}n_{\Lambda_c}\right)\left(\rho_d(\boldsymbol{r}')-\frac{4}{\Omega}n_{\Lambda_d}\right)\\
&= \int d^2r d^2r' V(\boldsymbol{r}-\boldsymbol{r}')\left[\rho_c(\boldsymbol{r})\rho_d(\boldsymbol{r}')-\frac{8}{\Omega}n_{\Lambda_c}\rho_d(\boldsymbol{r}')-\frac{4}{\Omega}n_{\Lambda_d}\rho_c(\boldsymbol{r})\right]+\text{const.}\\
&= \frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}}\sum_{\boldsymbol{p}_1\boldsymbol{p}_2}^{\Lambda_c}\sum_{\boldsymbol{p}_3\boldsymbol{p}_4}^{\Lambda_d}\delta_{\boldsymbol{p}_2,\boldsymbol{p}+\boldsymbol{p}_1}\delta_{\boldsymbol{p}_3,\boldsymbol{p}+\boldsymbol{p}_4}V(\boldsymbol{p})\sum_{\eta\sigma s}\sum_{\eta'\sigma's'}c^\dagger_{\eta,\boldsymbol{p}_1,\beta,s}d^\dagger_{\eta',\boldsymbol{p}_3,\sigma',s'}d_{\eta',\boldsymbol{p}_4,\sigma',s'}c_{\eta,\boldsymbol{p}_2,\beta,s}\\
&\quad -\frac{8n_{\Lambda_c}}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{\Lambda_d}d^\dagger_{\boldsymbol{p}}d_{\boldsymbol{p}}-\frac{4n_{\Lambda_d}}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{\Lambda_c}c^\dagger_{\boldsymbol{p}}c_{\boldsymbol{p}}+\text{const.}\,,
\end{aligned}\tag{D18}$$

which leads to the following Hartree-Fock $H_{\text{int},V,cd}$:

$$\begin{aligned}
H_{V,cd} &= \frac{1}{\mathcal{A}}\sum_{\boldsymbol{p}}\sum_{\boldsymbol{p}_1\boldsymbol{p}_2}^{\Lambda_c}\sum_{\boldsymbol{p}_3\boldsymbol{p}_4}^{\Lambda_d}\delta_{\boldsymbol{p}_2,\boldsymbol{p}+\boldsymbol{p}_1}\delta_{\boldsymbol{p}_3,\boldsymbol{p}+\boldsymbol{p}_4}V(\boldsymbol{p})\sum_{\eta\sigma s}\sum_{\eta'\sigma's'}[c^\dagger_{\eta,\boldsymbol{p}_1,\beta,s}c_{\eta,\boldsymbol{p}_2,\beta,s}\langle d^\dagger_{\eta',\boldsymbol{p}_3,\sigma',s'}d_{\eta',\boldsymbol{p}_4,\sigma',s'}\rangle\\
&\quad +\langle c^\dagger_{\eta,\boldsymbol{p}_1,\beta,s}c_{\eta,\boldsymbol{p}_2,\beta,s}\rangle d^\dagger_{\eta',\boldsymbol{p}_3,\sigma',s'}d_{\eta',\boldsymbol{p}_4,\sigma',s'}-\langle c^\dagger_{\eta,\boldsymbol{p}_1,\beta,s}d_{\eta',\boldsymbol{p}_4,\sigma',s'}\rangle d^\dagger_{\eta',\boldsymbol{p}_3,\sigma',s'}c_{\eta,\boldsymbol{p}_2,\beta,s}\\
&\quad -c^\dagger_{\eta,\boldsymbol{p}_1,\beta,s}d_{\eta',\boldsymbol{p}_4,\sigma',s'}\langle d^\dagger_{\eta',\boldsymbol{p}_3,\sigma',s'}c_{\eta,\boldsymbol{p}_2,\beta,s}\rangle]\\
&\quad -\frac{4n_{\Lambda_c}}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{\Lambda_d}d^\dagger_{\boldsymbol{p}}d_{\boldsymbol{p}}-\frac{4n_{\Lambda_d}}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{\Lambda_c}c^\dagger_{\boldsymbol{p}}c_{\boldsymbol{p}}+\text{const.}\\
&\xrightarrow{\text{Neglecting Fock channel}} \frac{1}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c}c^\dagger_{\boldsymbol{p}}c_{\boldsymbol{p}}\,\text{Tr}[O^{dd}]+\frac{1}{\Omega}V(\boldsymbol{p}=0)\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_d}d^\dagger_{\boldsymbol{p}}d_{\boldsymbol{p}}\,\text{Tr}[O^{cc}]\,,
\end{aligned}\tag{D19}$$

Similarly, we get $E_{V,cd}$ as

$$E_{V,cd}=\frac{N}{2\Omega}V(\boldsymbol{p}=0)(2\,\text{Tr}[O^{dd}]\,\text{Tr}[O^{cc}]+8n_{\Lambda_d}\,\text{Tr}[O^{cc}]+16n_{\Lambda_c}\,\text{Tr}[O^{dd}])\tag{D20}$$

Third, $H_{\text{int},W,fd}$ in Eq. (61) can be rewritten as

$$\begin{aligned}
H_{\text{int},W,fd} &= \Omega W_{fd}\sum_{\boldsymbol{R}}(\rho_f(\boldsymbol{R})-4)\left(\rho_d(\boldsymbol{R})-\frac{4}{\Omega}n_{\Lambda_d}\right)\\
&= \Omega W_{fd}\sum_{\boldsymbol{R}}\rho_f(\boldsymbol{R})\rho_d(\boldsymbol{R})-4\Omega W_{fd}\sum_{\boldsymbol{R}}\rho_d(\boldsymbol{R})-4n_{\Lambda_d}W_{fd}\sum_{\boldsymbol{R}}\rho_f(\boldsymbol{R})+\text{const.}\\
&= \frac{W_{fd}}{N}\sum_{\eta\alpha s}\sum_{\eta'\sigma s'}\sum_{\boldsymbol{k}_1\boldsymbol{k}_4}^{\text{MBZ}}\sum_{\boldsymbol{p}_2\boldsymbol{p}_3}^{\Lambda_d}\sum_{\boldsymbol{G}}\delta_{\boldsymbol{k}_1+\boldsymbol{p}_2,\boldsymbol{p}_3+\boldsymbol{k}_4+\boldsymbol{G}}f^\dagger_{\eta,\boldsymbol{k}_1,\alpha,s}d^\dagger_{\eta',\boldsymbol{p}_2,\sigma',s'}d_{\eta',\boldsymbol{p}_3,\sigma',s'}f_{\eta,\boldsymbol{k}_4,\alpha,s}\\
&\quad -4W_{fd}\sum_{\boldsymbol{p}}^{\Lambda_d}d^\dagger_{\boldsymbol{p}}d_{\boldsymbol{p}}-4n_{\Lambda_d}W_{fd}\sum_{\boldsymbol{k}}^{\text{MBZ}}f^\dagger_{\boldsymbol{k}}f_{\boldsymbol{k}}+\text{const.}\,,
\end{aligned}\tag{D21}$$

which leads to the following Hartree-Fock $H_{\text{int},W,fd}$

$$\begin{aligned}
H_{W,fd} &= \frac{W_{fd}}{N}\sum_{\eta\alpha s}\sum_{\eta'\sigma s'}\sum_{\boldsymbol{k}_1\boldsymbol{k}_4}^{\text{MBZ}}\sum_{\boldsymbol{p}_2\boldsymbol{p}_3}^{\Lambda_d}\sum_{\boldsymbol{G}}\delta_{\boldsymbol{k}_1+\boldsymbol{p}_2,\boldsymbol{p}_3+\boldsymbol{k}_4+\boldsymbol{G}}[f^\dagger_{\eta,\boldsymbol{k}_1,\alpha,s}f_{\eta,\boldsymbol{k}_4,\alpha,s}\langle d^\dagger_{\eta',\boldsymbol{p}_2,\sigma',s'}d_{\eta',\boldsymbol{p}_3,\sigma',s'}\rangle\\
&\quad +\langle f^\dagger_{\eta,\boldsymbol{k}_1,\alpha,s}f_{\eta,\boldsymbol{k}_4,\alpha,s}\rangle d^\dagger_{\eta',\boldsymbol{p}_2,\sigma',s'}d_{\eta',\boldsymbol{p}_3,\sigma',s'}-\langle f^\dagger_{\eta,\boldsymbol{k}_1,\alpha,s}d_{\eta',\boldsymbol{p}_3,\sigma',s'}\rangle d^\dagger_{\eta',\boldsymbol{p}_2,\sigma',s'}f_{\eta,\boldsymbol{k}_4,\alpha,s}\\
&\quad -f^\dagger_{\eta,\boldsymbol{k}_1,\alpha,s}d_{\eta',\boldsymbol{p}_3,\sigma',s'}\langle d^\dagger_{\eta',\boldsymbol{p}_2,\sigma',s'}f_{\eta,\boldsymbol{k}_4,\alpha,s}\rangle]-4W_{fd}\sum_{\boldsymbol{p}}^{\Lambda_d}d^\dagger_{\boldsymbol{p}}d_{\boldsymbol{p}}-4n_{\Lambda_d}W_{fd}\sum_{\boldsymbol{k}}^{\text{MBZ}}f^\dagger_{\boldsymbol{k}}f_{\boldsymbol{k}}\\
&= W_{fd}\,\text{Tr}[O^{dd}]\sum_{\boldsymbol{R}}f^\dagger_{\boldsymbol{R}}f_{\boldsymbol{R}}+W_{fd}(\text{Tr}[O^{dd}]-4)\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_d}d^\dagger_{\boldsymbol{p}}d_{\boldsymbol{p}}-W_{fd}\left(\sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_d}\sum_{\eta\eta'}d^\dagger_{\eta',\boldsymbol{p}}[O^{fd}_{\eta\eta'}]^T f_{\eta,\boldsymbol{p}+\eta'K_M}+\text{H.c.}\right),
\end{aligned}\tag{D22}$$

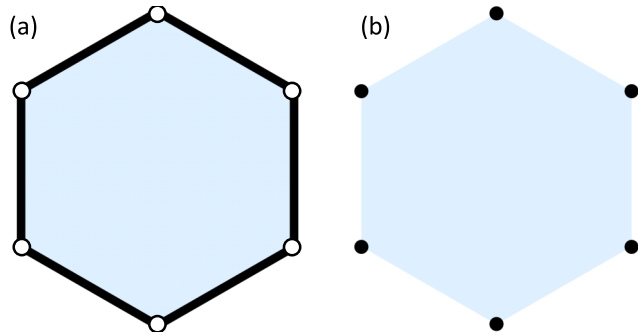

FIG. 4. The edge of the MBZ is shown in (a) as the black solid line, where the cornors are excluded. The corners of the MBZ are shown in (b) as the black dots. Only half of the edge in (a) belongs to MBZ, and only one third of the corners in (b) belongs to MBZ.

where we have used Eq. (D5). Similarly, we get $E_{W,fd}$ as

$$E_{W,fd} = NW_{fd}(\mathrm{Tr}[O^{dd}]\,\mathrm{Tr}[O^{ff}] + 4n_{\Lambda_d}(\mathrm{Tr}[O^{ff}] - 4)) - NW_{fd}\,\mathrm{Tr}[O^{df}O^{fd}]. \tag{D23}$$

Comparing Eq. (D22) to Eq. (D11), we can see $d^\dagger_{\eta,\boldsymbol{p}}$ couples to $f_{\eta',\boldsymbol{p}+\eta K_M}$, while $c^\dagger_{\boldsymbol{p}}$ couples to $f_{\boldsymbol{p}}$, showing that d modes are around the ηK_M points and c modes are around the Γ_M point.

For the calculation of Hartree-Fock density matrices, we choose $\Lambda_c = \Lambda_d = \sqrt{3}$ [$|\boldsymbol{b}_{\mathrm{M},1}| = |\boldsymbol{b}_{\mathrm{M},2}| = \sqrt{3}$ according to Eq. (21) as a comparison], and the iteration for the self-consistent calculation stops when the error of the Hartree-Fock ground state energy is smaller than 10^{-4} meV in EUS. The initial Hartree-Fock density matrices are given by the initial states, which are specified below. To rule out the symmetry breaking induced by the artificial cutoffs, we address the momentum points in MBZ in a symmetric way. Specifically, when we need to sum $\boldsymbol{k}$ over MBZ for determining the density matrices in Eq. (D5), instead of actually summing $\boldsymbol{k}$ over MBZ, we sum $\boldsymbol{k}$ over the completion of MBZ, i.e., the union of MBZ with its all edges and corners (shown in Fig. 4), and include a factor of $1/2$ for terms with $\boldsymbol{k}$ on the edge and $1/3$ for terms with $\boldsymbol{k}$ at the corners.

The Hartree-Fock band structures are plotted for $\Lambda_c = \Lambda_d = 2\sqrt{3}$, in order to compare with the single-particle band structure.

2. Initial states

Now we specify the initial states for the self-consistent Hartree-Fock calculations for $\nu = 0, -1, -2$. The choice of the initial states are inspired by the numerical results in Ref. [149], which show that (i) the ground state at low-$\mathcal{E}$ is similar to TBG, and (ii) the ground states at high-$\mathcal{E}$ have zero intervalley coherence. Therefore, for all the considered fillings, we will include all the initial states that correspond to those used in the study of TBG in Ref. [125] and include representative states without intervalley coherence.

Recall that we choose the initial states to have the form of Eq. (65), where |Fermi Sea⟩ stands for the hall-filled Fermi sea of the free c and d modes. For the initial states without intervalley coherence, the filling in each valley is well-defined and can be evaluated as $\nu_\eta = \mathrm{Tr}[\zeta_\eta\zeta_\eta^\dagger] - 2$ for the η valley, where ζ_η is defined in Eq. (70); we have $\nu_+ + \nu_- = \nu$. Then, we choose certain representative initial states without intervalley coherence for all combinations of (ν_+, ν_-) with $\nu_+ \leqslant \nu_-$, since the $\nu_+ < \nu_-$ subspace is related to the $\nu_+ > \nu_-$ subspace by the TR symmetry. The initial states that we choose for the self-consistent calculations at $\nu = 0$ are

$$|\mathrm{VP}_0^{\nu=0}\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{+,\boldsymbol{R},2,\uparrow} f^\dagger_{+,\boldsymbol{R},2,\downarrow} |\text{Fermi Sea}\rangle\,, \tag{D24}$$

$$|\mathrm{IVC}_0^{\nu=0}\rangle = \prod_{\boldsymbol{R}} \frac{1}{4}(f^\dagger_{+,\boldsymbol{R},1,\uparrow} - \mathrm{i}f^\dagger_{-,\boldsymbol{R},2,\uparrow})(f^\dagger_{+,\boldsymbol{R},1,\downarrow} - \mathrm{i}f^\dagger_{-,\boldsymbol{R},2,\downarrow})(-\mathrm{i}f^\dagger_{-,\boldsymbol{R},1,\uparrow} + f^\dagger_{+,\boldsymbol{R},2,\uparrow})(-\mathrm{i}f^\dagger_{-,\boldsymbol{R},1,\downarrow} + f^\dagger_{+,\boldsymbol{R},2,\downarrow})|\text{Fermi Sea}\rangle\,, \tag{D25}$$

$$|\text{K-IVC}_0^{\nu=0}\rangle = \prod_{\boldsymbol{R}} \frac{1}{4}(f^\dagger_{+,\boldsymbol{R},1,\uparrow} + f^\dagger_{-,\boldsymbol{R},2,\uparrow})(f^\dagger_{+,\boldsymbol{R},1,\downarrow} + f^\dagger_{-,\boldsymbol{R},2,\downarrow})(-f^\dagger_{-,\boldsymbol{R},1,\uparrow} + f^\dagger_{+,\boldsymbol{R},2,\uparrow})(-f^\dagger_{-,\boldsymbol{R},1,\downarrow} + f^\dagger_{+,\boldsymbol{R},2,\downarrow})|\text{Fermi Sea}\rangle\,, \tag{D26}$$

$$|\mathrm{PVP}_0^{1,\nu=0}\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{+,\boldsymbol{R},2,\downarrow} f^\dagger_{-,\boldsymbol{R},1,\downarrow} |\text{Fermi Sea}\rangle\,, \tag{D27}$$

$$|\mathrm{PVP}_0^{2,\nu=0}\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{+,\boldsymbol{R},2,\uparrow} f^\dagger_{-,\boldsymbol{R},2,\downarrow} |\text{Fermi Sea}\rangle\,, \tag{D28}$$

$$|\mathrm{VH}_0^{\nu=0}\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{-,\boldsymbol{R},1,\uparrow} f^\dagger_{-,\boldsymbol{R},1,\downarrow} |\text{Fermi Sea}\rangle\,, \tag{D29}$$

$$|\mathrm{Chern}_0^{\nu=0}\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{-,\boldsymbol{R},2,\uparrow} f^\dagger_{-,\boldsymbol{R},2,\downarrow} |\text{Fermi Sea}\rangle\,, \tag{D30}$$

$$|\text{half-Chern}_0^{\nu=0}\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{-,\boldsymbol{R},1,\downarrow} f^\dagger_{-,\boldsymbol{R},2,\uparrow} |\text{Fermi Sea}\rangle\,, \tag{D31}$$

and

$$|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{+,\boldsymbol{R},2,\uparrow} f^\dagger_{-,\boldsymbol{R},1,\downarrow} f^\dagger_{-,\boldsymbol{R},2,\uparrow} |\text{Fermi Sea}\rangle\,. \tag{D32}$$

Here $|\mathrm{VP}_0^{\nu=0}\rangle$, $|\mathrm{IVC}_0^{\nu=0}\rangle$ and $|\text{K-IVC}_0^{\nu=0}\rangle$ are chosen because the corresponding states are used in Ref. [125] for TBG. $|\mathrm{VP}_0^{\nu=0}\rangle$ is also a representative state without intervalley coherence for $(\nu_+, \nu_-) = (2, -2)$. We choose $|\mathrm{PVP}_0^{1,\nu=0}\rangle$ and $|\mathrm{PVP}_0^{2,\nu=0}\rangle$ as

the representative states without intervalley coherence for $(\nu_+, \nu_-) = (1, -1)$. We choose $|\mathrm{VH}_0^{\nu=0}\rangle$, $|\mathrm{Chern}_0^{\nu=0}\rangle$, $|\text{half-Chern}_0^{\nu=0}\rangle$ and $|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\rangle$ as the representative states without intervalley coherence for $(\nu_+, \nu_-) = (0, 0)$. The initial states that we choose for the self-consistent calculations at $\nu = -1$ are

$$\left|\mathrm{VP}_0^{\nu=-1}\right\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{+,\boldsymbol{R},2,\uparrow} |\text{Fermi Sea}\rangle \,, \tag{D33}$$

$$\left|\mathrm{IVC}_0^{\nu=-1}\right\rangle = \prod_{\boldsymbol{R}} \frac{1}{2\sqrt{2}} (f^\dagger_{+,\boldsymbol{R},1,\uparrow} - \mathrm{i} f^\dagger_{-,\boldsymbol{R},2,\uparrow})(f^\dagger_{+,\boldsymbol{R},1,\downarrow} - \mathrm{i} f^\dagger_{-,\boldsymbol{R},2,\downarrow})(-\mathrm{i} f^\dagger_{-,\boldsymbol{R},1,\uparrow} + f^\dagger_{+,\boldsymbol{R},2,\uparrow}) |\text{Fermi Sea}\rangle \,, \tag{D34}$$

$$\left|\mathrm{VP{+}IVC}_0^{\nu=-1}\right\rangle = \prod_{\boldsymbol{R}} \frac{1}{2} (f^\dagger_{+,\boldsymbol{R},1,\uparrow} + f^\dagger_{-,\boldsymbol{R},2,\uparrow})(-f^\dagger_{-,\boldsymbol{R},1,\uparrow} + f^\dagger_{+,\boldsymbol{R},2,\uparrow}) f^\dagger_{+,\boldsymbol{R},1,\downarrow} |\text{Fermi Sea}\rangle \,, \tag{D35}$$

$$\left|\mathrm{PVP}_0^{1,\nu=-1}\right\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{-,\boldsymbol{R},1,\downarrow} |\text{Fermi Sea}\rangle \,, \tag{D36}$$

$$\left|\mathrm{PVP}_0^{2,\nu=-1}\right\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},2,\uparrow} f^\dagger_{+,\boldsymbol{R},2,\downarrow} f^\dagger_{-,\boldsymbol{R},1,\downarrow} |\text{Fermi Sea}\rangle \,, \tag{D37}$$

and

$$\left|\mathrm{PVP}_0^{3,\nu=-1}\right\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{+,\boldsymbol{R},2,\uparrow} f^\dagger_{-,\boldsymbol{R},1,\downarrow} |\text{Fermi Sea}\rangle \,. \tag{D38}$$

Here $|\mathrm{VP}_0^{\nu=-1}\rangle$, $|\mathrm{IVC}_0^{\nu=-1}\rangle$ and $|\mathrm{VP{+}IVC}_0^{\nu=-1}\rangle$ are chosen because the corresponding states are used in Ref. [125] for TBG. $|\mathrm{VP}_0^{\nu=-1}\rangle$ is also a representative state without intervalley coherence for $(\nu_+, \nu_-) = (1, -2)$. We choose $|\mathrm{PVP}_0^{1,\nu=-1}\rangle$, $|\mathrm{PVP}_0^{2,\nu=-1}\rangle$, and $|\mathrm{PVP}_0^{3,\nu=-1}\rangle$ as the representative states without intervalley coherence for $(\nu_+, \nu_-) = (0, -1)$.

The initial states that we choose for the self-consistent calculations at $\nu = -2$ are

$$\left|\text{K-IVC}_0^{\nu=-2}\right\rangle = \prod_{\boldsymbol{R}} \frac{1}{2} (f^\dagger_{+,\boldsymbol{R},1,\uparrow} + f^\dagger_{-,\boldsymbol{R},2,\uparrow})(f^\dagger_{-,\boldsymbol{R},1,\uparrow} - f^\dagger_{+,\boldsymbol{R},2,\uparrow}) |\text{Fermi Sea}\rangle \,, \tag{D39}$$

$$\left|\mathrm{IVC}_0^{\nu=-2}\right\rangle = \prod_{\boldsymbol{R}} \frac{1}{2} (f^\dagger_{+,\boldsymbol{R},1,\uparrow} - \mathrm{i} f^\dagger_{-,\boldsymbol{R},2,\uparrow})(-\mathrm{i} f^\dagger_{-,\boldsymbol{R},1,\uparrow} + f^\dagger_{+,\boldsymbol{R},2,\uparrow}) |\text{Fermi Sea}\rangle \,, \tag{D40}$$

$$\left|\mathrm{VP}_0^{\nu=-2}\right\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},2,\uparrow} |\text{Fermi Sea}\rangle \,, \tag{D41}$$

$$\left|\mathrm{VP}_0^{1,\nu=-2}\right\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\downarrow} f^\dagger_{+,\boldsymbol{R},2,\uparrow} |\text{Fermi Sea}\rangle \,, \tag{D42}$$

$$\left|\mathrm{VP}_0^{2,\nu=-2}\right\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{+,\boldsymbol{R},1,\downarrow} |\text{Fermi Sea}\rangle \,, \tag{D43}$$

$$\left|\text{valley-unpolarized}_0^{1,\nu=-2}\right\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{-,\boldsymbol{R},1,\uparrow} |\text{Fermi Sea}\rangle \,, \tag{D44}$$

and

$$\left|\text{valley-unpolarized}_0^{2,\nu=-2}\right\rangle = \prod_{\boldsymbol{R}} f^\dagger_{+,\boldsymbol{R},1,\uparrow} f^\dagger_{-,\boldsymbol{R},2,\downarrow} |\text{Fermi Sea}\rangle \,. \tag{D45}$$

Here $|\text{K-IVC}_0^{\nu=-2}\rangle$, $|\mathrm{IVC}_0^{\nu=-2}\rangle$ and $|\mathrm{VP}_0^{\nu=-2}\rangle$ are chosen because the corresponding states are used in Ref. [125] for TBG. We choose $|\mathrm{VP}_0^{1,\nu=-2}\rangle$ and $|\mathrm{VP}_0^{2,\nu=-2}\rangle$ (as well as $|\mathrm{VP}_0^{\nu=-2}\rangle$) as the representative states without intervalley coherence for $(\nu_+, \nu_-) = (0, -2)$. We choose $|\text{valley-unpolarized}_0^{1,\nu=-2}\rangle$ and $|\text{valley-unpolarized}_0^{2,\nu=-2}\rangle$ as the representative states without intervalley coherence for $(\nu_+, \nu_-) = (-1, -1)$.

APPENDIX E: MORE DETAILS ON ANALYTICAL UNDERSTANDING

In this section, we provide more details on the analytical understanding.

1. One-shot Hartree-Fock Hamiltonian

We develop the analytical understanding by using the one-shot Hartree-Fock Hamiltonian, which is derived as the follows. First, based on the initial state Eq. (65), we can derive the initial density matrices as

$$O^{ff}_{\mathrm{ini}} = \zeta^* \zeta^T \,,$$
$$\left[O^{cc}_{\mathrm{ini}}\right]_{\eta\beta s, \eta'\beta' s'} = \delta_{\eta\eta'} [Z_\eta]_{\beta\beta'} \delta_{ss'} \quad \text{with } [Z_\eta]_{\beta\beta} = 0 \,,$$

$$
\begin{aligned}
O^{dd}_{\text{ini}} &= 0\,, \\
O^{cf}_{\text{ini}} &= 0\,, \quad O^{fc}_{\text{ini}} = 0, \\
O^{df}_{\text{ini}} &= 0\,, \quad O^{fd}_{\text{ini}} = 0, \\
O^{cd}_{\text{ini}} &= 0\,, \quad O^{dc}_{\text{ini}} = 0\,,
\end{aligned} \tag{E1}
$$

where Z_η is a 4×4 matrix. Then, we can substitute Eq. (E1) into the Hartree-Fork Hamiltonian Eq. (D6), and the resultant Hartree-Fock Hamiltonian is the one-shot Hartree-Fock Hamiltonian, which reads

$$
H_{\text{HF,OS}} = \sum_\eta H^{\text{eff}}_{0,\eta} + H_{U,\text{OS}} + H_{V,c,\text{OS}} + H_{W,fc,\text{OS}} + H_{J,\text{OS}} + H_{V,d,\text{OS}} + H_{V,cd,\text{OS}} + H_{W,fd,\text{OS}} - E^{\text{OS}}_0 + \text{const.}\,, \tag{E2}
$$

where $H^{\text{eff}}_{0,\eta}$ is in Eq. (54),

$$
E^{\text{OS}}_0 = E_{U,\text{OS}} + E_{V,c,\text{OS}} + E_{W,fc,\text{OS}} + E_{J,\text{OS}} + E_{V,d,\text{OS}} + E_{V,cd,\text{OS}} + E_{W,fd,\text{OS}}, \tag{E3}
$$

$$
H_{U,\text{OS}} = \sum_{\boldsymbol{R}} f^\dagger_{\boldsymbol{R}} h_U f_{\boldsymbol{R}}\,, \tag{E4}
$$

$$
h_U = \frac{1}{2}U_1 + \nu U_1 + 6\nu U_2 - U_1 \zeta\zeta^\dagger\,, \tag{E5}
$$

$$
E_{U,\text{OS}} = \frac{N}{2}(4+\nu)^2(U_1+6U_2) - U_1\frac{N}{2}(4+\nu)\,, \tag{E6}
$$

$$
H_{V,c,\text{OS}} = 0\,, \tag{E7}
$$

$$
E_{V,c,\text{OS}} = 0\,, \tag{E8}
$$

$$
H_{W,fc,\text{OS}} = \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} c^\dagger_{\boldsymbol{p}} h_{W,fc} c_{\boldsymbol{p}}\,, \tag{E9}
$$

$$
h_{W,fc} = \begin{pmatrix} \nu W_1 \eta_0\tau_0 s_0 & \\ & \nu W_3 \eta_0\tau_0 s_0 \end{pmatrix}, \tag{E10}
$$

$$
E_{W,fc,\text{OS}} = 2N\nu n_{\Lambda_c}(2W_1+2W_3) + \text{const.}\,, \tag{E11}
$$

$$
H_{J,\text{OS}} = \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_c} c^\dagger_{\boldsymbol{p}} h_J c_{\boldsymbol{p}}\,, \tag{E12}
$$

$$
h_J = \begin{pmatrix} 0_{8\times 8} & \\ & \eta_z\tau_0 s_0 \zeta\zeta^\dagger \eta_z\tau_0 s_0 + \eta_0\tau_z s_0 \zeta\zeta^\dagger \eta_0\tau_z s_0 - \eta_0\tau_0 s_0 \end{pmatrix}, \tag{E13}
$$

$$
E_{J,\text{OS}} = -\frac{J}{2}\nu N n_{\Lambda_c} + \text{const.} \tag{E14}
$$

$$
H_{V,d,\text{OS}} = H_{V,cd,\text{OS}} = 0\,, \tag{E15}
$$

$$
E_{V,d,\text{OS}} = \text{const.} \quad E_{V,cd,\text{OS}} = \text{const.}, \tag{E16}
$$

$$
H_{W,fd,\text{OS}} = W_{fd}\nu \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|\leqslant\Lambda_d} d^\dagger_{\boldsymbol{p}} d_{\boldsymbol{p}}\,, \tag{E17}
$$

and

$$
E_{W,fd,\text{OS}} = 4\nu N W_{fd} n_{\Lambda_d} + \text{const.} \tag{E18}
$$

Here "const." consists of scalar terms that do not depend on the density matrices. It is clear that the dependence of E^{OS}_0 on the ground state is only through the filling ν, which is solely determined by the f modes at the one-shot level. Therefore the energy difference for different states with the same filling only comes from the operator part, which we will focus on in the following.

2. Simple rule for high-$\mathcal{E}$ states: high-$\mathcal{E}$ limit

Now we provide more details on the analytical understanding of the simple rule for high-$\mathcal{E}$ states, under the high-$\mathcal{E}$ limit. Here high-$\mathcal{E}$ limit mean that we choose $|\mathcal{E}|$ to be infinitely large compared with all other energy quantities. We also approximate

the chemical potential as

$$\mu = \nu(U_1 + 6U_2)\,, \tag{E19}$$

which is the correction of the chemical potential due to the density-density interaction of f modes [125]. The validity of these simplifications will be discussed in Appendix E 3. In Appendix E 3, we will demonstrate the validity of those approximations for $\nu = 0$.

Throughout this part, we choose $\nu \in \{0, -1, -2\}$. As an effective theory, we will focus on the Hartree-Fock Hamiltonian at $\pm K_M$ and Γ_M. We will first consider $\pm K_M$ and then consider Γ_M.

a. $\pm K_M$

The one-shot Hartree-Fock Hamiltonain around $\pm K_M$ reads

$$H^{\eta K_M}_{\mathrm{HF,OS}} = \sum_{\boldsymbol{p}}^{|\boldsymbol{p}|<\Lambda_d} (f^\dagger_{+,\eta K_M+\boldsymbol{p}}, f^\dagger_{-,\eta K_M+\boldsymbol{p}}, d^\dagger_{\eta,\boldsymbol{p}}) h^{\eta K_M}_{\mathrm{HF,OS}}(\boldsymbol{p}) \begin{pmatrix} f_{+,\eta K_M+\boldsymbol{p}} \\ f_{-,\eta K_M+\boldsymbol{p}} \\ d_{\eta,\boldsymbol{p}} \end{pmatrix}, \tag{E20}$$

$$h^{K_M}_{\mathrm{HF,OS}}(\boldsymbol{p}) = \left(\begin{array}{cc|c} \frac{1}{2}U_1 + \nu(U_1+6U_2) - U_1\zeta\zeta^\dagger & & M_1\mathcal{E}(\tau_0 + \mathrm{i}\tau_z)s_0 \\ & & 0_{4\times4} \\ \hline M_1\mathcal{E}(\tau_0 - \mathrm{i}\tau_z)s_0 & 0_{4\times4} & W_{fd}\nu + (p_x\sigma_x + p_y\sigma_y)s_0 \end{array}\right), \tag{E21}$$

and

$$h^{-K_M}_{\mathrm{HF,OS}}(\boldsymbol{p}) = \left(\begin{array}{cc|c} \frac{1}{2}U_1 + \nu(U_1+6U_2) - U_1\zeta\zeta^\dagger & & 0_{4\times4} \\ & & M_1\mathcal{E}(\tau_0 - \mathrm{i}\tau_z)s_0 \\ \hline 0_{4\times4} & M_1\mathcal{E}(\tau_0 + \mathrm{i}\tau_z)s_0 & W_{fd}\nu + (-p_x\sigma_x + p_y\sigma_y)s_0 \end{array}\right). \tag{E22}$$

By performing $d^\dagger_{\eta,\boldsymbol{p}} \to d^\dagger_{\eta,\boldsymbol{p}} e^{-\mathrm{i}\tau_z\frac{\pi}{4}} s_0$, we have

$$h^{K_M}_{\mathrm{HF,OS}}(\boldsymbol{p}) \to \widetilde{h}^{K_M}_{\mathrm{HF,OS}}(\boldsymbol{p}) = \left(\begin{array}{cc|c} \frac{1}{2}U_1 + \nu(U_1+6U_2) - U_1\zeta\zeta^\dagger & & \sqrt{2}M_1\mathcal{E}\tau_0 s_0 \\ & & 0_{4\times4} \\ \hline \sqrt{2}M_1\mathcal{E}\tau_0 s_0 & 0_{4\times4} & W_{fd}\nu + (-p_x\sigma_y + p_y\sigma_x)s_0 \end{array}\right) \tag{E23}$$

and

$$h^{-K_M}_{\mathrm{HF,OS}}(\boldsymbol{p}) \to \widetilde{h}^{-K_M}_{\mathrm{HF,OS}}(\boldsymbol{p}) = \left(\begin{array}{cc|c} \frac{1}{2}U_1 + \nu(U_1+6U_2) - U_1\zeta\zeta^\dagger & & 0_{4\times4} \\ & & \sqrt{2}M_1\mathcal{E}\tau_0 s_0 \\ \hline 0_{4\times4} & \sqrt{2}M_1\mathcal{E}\tau_0 s_0 & W_{fd}\nu + (-p_x\sigma_y - p_y\sigma_x)s_0 \end{array}\right), \tag{E24}$$

which are convenient to use.

Since we focus on $\pm K_M$ in this part, we only consider $\boldsymbol{p} = 0$ for $f^\dagger_{+,\eta K_M+\boldsymbol{p}}$, $f^\dagger_{-,\eta K_M+\boldsymbol{p}}$ and $d^\dagger_{\eta,\boldsymbol{p}}$. To proceed, let us define the following two unitary matrices:

$$\widetilde{U}_{+K_M} = \begin{pmatrix} \chi_{0,1} & \chi_{1,1} & \\ & & 1 \\ \chi_{0,2} & \chi_{1,2} & \end{pmatrix} \otimes \mathbb{1}_{4\times4} \quad \text{and} \quad \widetilde{U}_{-K_M} = \begin{pmatrix} & & 1 \\ \chi_{0,1} & \chi_{1,1} & \\ \chi_{0,2} & \chi_{1,2} & \end{pmatrix} \otimes \mathbb{1}_{4\times4}\,, \tag{E25}$$

where

$$\begin{pmatrix} \nu(U_1+6U_2) & \sqrt{2}M_1\mathcal{E} \\ \sqrt{2}M_1\mathcal{E} & W_{fd}\nu \end{pmatrix} \chi_\gamma = \epsilon_\gamma \chi_\gamma\,, \tag{E26}$$

$\gamma = 0, 1$, $\chi_\gamma = (\chi_{\gamma,1}, \chi_{\gamma,2})$ is real, and

$$\epsilon_\gamma = \frac{\nu(U_1+6U_2+W_{fd})}{2} + (-)^\gamma \sqrt{\left[\frac{\nu(U_1+6U_2-W_{fd})}{2}\right]^2 + 2M_1^2\mathcal{E}^2}\,. \tag{E27}$$

Then, we use $\widetilde{U}_{\eta K_M}$ to unitarily transformation $\widetilde{h}^{\eta K_M}_{\mathrm{HF,OS}}(0)$ to

$$\begin{aligned} &\widetilde{U}^\dagger_{\eta K_M} \widetilde{h}^{\eta K_M}_{\mathrm{HF,OS}}(0) \widetilde{U}_{\eta K_M} \\ &= \begin{pmatrix} \epsilon_0\mathbb{1}_{4\times4} & & \\ & \epsilon_1\mathbb{1}_{4\times4} & \\ & & \nu(U_1+6U_2)\mathbb{1}_{4\times4} \end{pmatrix} - U_1 \left(\begin{array}{cc|c} \begin{pmatrix} |\chi_{0,1}|^2 & \chi^*_{0,1}\chi_{1,1} \\ \chi^*_{1,1}\chi_{0,1} & |\chi_{1,1}|^2 \end{pmatrix} \otimes \left(\zeta_\eta\zeta^\dagger_\eta - \frac{1}{2}\right) & & \begin{matrix} \chi^*_{0,1}\zeta_\eta\zeta^\dagger_{-\eta} \\ \chi^*_{1,1}\zeta_\eta\zeta^\dagger_{-\eta} \end{matrix} \\ \hline \chi_{0,1}\zeta_{-\eta}\zeta^\dagger_\eta \quad \chi_{1,1}\zeta_{-\eta}\zeta^\dagger_\eta & & \zeta_{-\eta}\zeta^\dagger_{-\eta} - \frac{1}{2} \end{array}\right). \end{aligned} \tag{E28}$$

We perform the transformation in Eq. (E28) because (i) it gives a block-diagonal term (i.e., the first term) that has the three blocks with energies ϵ_0, ϵ_1 and $\nu(U_1 + 6U_2)$, and (ii) the gaps among ϵ_0, ϵ_1 and $\nu(U_1 + 6U_2)$ are of order $|\mathcal{E}|$ according to Eq. (E27),

which is much larger than U_1 in the high-$\mathcal{E}$ limit. Therefore, in Eq. (E28), the elements (of the second term) that couple different blocks in the first term can only change the eigenvalues at the order of $O(|U_1/\mathcal{E}|)$.

In the following, we will neglect all corrections to the energies that are of order $O(|U_1/\mathcal{E}^2|)$.

Then, we only need to consider the following Hamiltonian:

$$\widetilde{U}^{\dagger}_{\eta K_M}\widetilde{h}^{\eta K_M}_{\mathrm{HF,OS}}(0)\widetilde{U}_{\eta K_M} \approx \begin{pmatrix} \epsilon_0\mathbb{1}_{4\times4} - U_1|\chi_{0,1}|^2\left(\zeta_\eta\zeta_\eta^\dagger - \frac{1}{2}\right) & & \\ & \epsilon_1\mathbb{1}_{4\times4} - U_1|\chi_{1,1}|^2\left(\zeta_\eta\zeta_\eta^\dagger - \frac{1}{2}\right) & \\ & & \nu(U_1+6U_2)\mathbb{1}_{4\times4} - U_1\left(\zeta_{-\eta}\zeta_{-\eta}^\dagger - \frac{1}{2}\right) \end{pmatrix}. \tag{E29}$$

Recall that $\epsilon_1 < \nu(U_1+6U_2) < \epsilon_0$ and the gaps between them are of order $O(|\mathcal{E}|)$, which is much larger than U_1. Then, based on the expression of the chemical potential in Eq. (E19), the ϵ_0 block should be fully empty, while the ϵ_1 block should be fully occupied. Eventually, we know that the occupied states of Eq. (E29) are all eigenstates of

$$[\epsilon_1 - \nu(U_1+6U_2)]\mathbb{1}_{4\times4} - U_1|\chi_{1,1}|^2\left(\zeta_\eta\zeta_\eta^\dagger - \tfrac{1}{2}\right) \text{ for both } \eta = \pm, \tag{E30}$$

and all negative-energy states of

$$-U_1\left(\zeta_\eta\zeta_\eta^\dagger - \tfrac{1}{2}\right) \text{ for both } \eta = \pm. \tag{E31}$$

We have subtracted the chemical potential in Eqs. (E30) and (E31) compared to the corresponding block in Eq. (E29). We label the total energy of all those occupied states as $E_{\pm K_M}$.

We want to minimize $E_{\pm K_M}$. Recall that Eq. (E30) should be fully occupied for both $\eta = \pm$. To express the remaining contribution to $E_{\pm K_M}$, we use λ_i $(i = 1, 2, \ldots, 8)$ to label the eight eigenvalues of

$$\begin{pmatrix} \zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger \end{pmatrix}, \tag{E32}$$

since all negative-energy states of Eq. (E31) are all negative-energy states of

$$-U_1\left[\begin{pmatrix} \zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger \end{pmatrix} - \frac{1}{2}\right]. \tag{E33}$$

We choose $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_8$ without loss of generality, and choose n to be the largest integer that gives $\lambda_n \geqslant 1/2$. Then, according to Eqs. (E30) and (E31), we have

$$\begin{aligned} E_{\pm K_M} &= 8[\epsilon_1 - \nu(U_1+6U_2)] - U_1|\chi_{1,1}|^2\sum_\eta \mathrm{Tr}\left[\zeta_\eta\zeta_\eta^\dagger - \frac{1}{2}\right] - U_1\sum_{i=1}^{n}\left(\lambda_i - \frac{1}{2}\right) + O(|U_1|^2/|\mathcal{E}|) \\ &= 8[\epsilon_1 - \nu(U_1+6U_2)] - U_1|\chi_{1,1}|^2\nu - U_1\sum_{i=1}^{n}\left(\lambda_i - \frac{1}{2}\right) + O(|U_1|^2/|\mathcal{E}|), \end{aligned} \tag{E34}$$

where we have used

$$\sum_\eta \mathrm{Tr}[\zeta_\eta\zeta_\eta^\dagger] = \mathrm{Tr}[\zeta\zeta^\dagger] = 4+\nu. \tag{E35}$$

To proceed, let us derive the constraints on λ_i. First, as $\begin{pmatrix} \zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger \end{pmatrix}$ is positive semi-definite, $\lambda_i \geqslant 0$. Second, $\lambda_i \leqslant 1$. To see this, recall that ζ defined in Eq. (69) is a $8\times(4+\nu)$ matrix whose columns $(\zeta_1, \ldots, \zeta_{4+\nu})$ are orthonormal. Then, there exists $4-\nu$ 8-component vectors, $\bar{\zeta}_1$,...,$\bar{\zeta}_{4-\nu}$, such that $\zeta_1, \ldots, \zeta_{4+\nu}$ and $\bar{\zeta}_1$,...,$\bar{\zeta}_{4-\nu}$ form an orthonormal basis of $\mathbb{C}^8$. Let us define the $\bar{\zeta} = (\bar{\zeta}_1 \quad \ldots \quad \bar{\zeta}_{4-\nu})$ as a $8\times(4-\nu)$ matrix, whose columns are orthonormal and which satisfies $\bar{\zeta}^\dagger\zeta = 0$ and $\zeta\zeta^\dagger + \bar{\zeta}\bar{\zeta}^\dagger = \mathbb{1}_{8\times8}$. Then, we have

$$\begin{pmatrix} \zeta_+\zeta_+^\dagger & \zeta_+\zeta_-^\dagger \\ \zeta_-\zeta_+^\dagger & \zeta_-\zeta_-^\dagger \end{pmatrix} + \begin{pmatrix} \bar{\zeta}_+\bar{\zeta}_+^\dagger & \bar{\zeta}_+\bar{\zeta}_-^\dagger \\ \bar{\zeta}_-\bar{\zeta}_+^\dagger & \bar{\zeta}_-\bar{\zeta}_-^\dagger \end{pmatrix} = \mathbb{1}_{8\times8} \Rightarrow \zeta_\eta\zeta_\eta^\dagger + \bar{\zeta}_\eta\bar{\zeta}_\eta^\dagger = \mathbb{1}_{4\times4}. \tag{E36}$$

Combined with the fact that $\bar{\zeta}_\eta\bar{\zeta}_\eta^\dagger$ is also positive semidefinite, we can get $\lambda_i \leqslant 1$. Third,

$$\sum_{i=1}^{8}\lambda_i = \sum_\eta \mathrm{Tr}[\zeta_\eta\zeta_\eta^\dagger] = 4+\nu. \tag{E37}$$

In sum, we know $\lambda_i \in [0, 1]$ and $\sum_{i=1}^{8}\lambda_i = 4+\nu$.

With the constraints on λ_i, we have

$$\sum_{i=1}^{n}\lambda_i = 4+\nu \leqslant \frac{4+\nu}{2}\,. \tag{E38}$$

To see this, we first consider $n > 4+\nu$, which gives

$$\sum_{i=1}^{n}\left(\lambda_i-\frac{1}{2}\right)=\sum_{i=1}^{n}\lambda_i-\frac{n}{2}\leqslant\sum_{i=1}^{8}\lambda_i-\frac{n}{2}=4+\nu-\frac{n}{2}=\frac{4+\nu}{2}+\frac{4+\nu-n}{2}<\frac{4+\nu}{2}\,. \tag{E39}$$

For $n < 4+\nu$, we have

$$\sum_{i=1}^{n}\left(\lambda_i-\frac{1}{2}\right)\leqslant\sum_{i=1}^{n}\frac{1}{2}<\frac{4+\nu}{2}\,. \tag{E40}$$

For $n = 4+\nu$, we have

$$\sum_{i=1}^{n}\left(\lambda_i-\frac{1}{2}\right)=\sum_{i=1}^{n}\lambda_i-\frac{n}{2}\leqslant\sum_{i=8}^{n}\lambda_i-\frac{n}{2}=\frac{4+\nu}{2}\,. \tag{E41}$$

Therefore we proved Eq. (E38) and we know the equality in Eq. (E38) only happens when $n = 4+\nu$.

Equations (E38) and (E34) give

$$E_{\pm K_M}\geqslant 8[\epsilon_1-\nu(U_1+6U_2)]-U_1|\chi_{1,1}|^2\nu-U_1\frac{4+\nu}{2}+O(|U_1|^2/|\mathcal{E}|)\,. \tag{E42}$$

Then, we know that the lowest $E_{\pm K_M}$ is achieved if and only if $\sum_{i=1}^{n}\lambda_i = 4+\nu$, which only appears for $n = 4+\nu$. Owing to $\sum_{i=1}^{8}\lambda_i = 4+\nu$, we have

$$\begin{aligned}
&\sum_{i=1}^{n}\lambda_i = 4+\nu\\
\Leftrightarrow\ &\sum_{i=1}^{n}\lambda_i = 4+\nu \quad \text{and} \quad n = 4+\nu\\
\Leftrightarrow\ &\lambda_1=\lambda_2=\cdots=\lambda_{4+\nu}=1\\
\Leftrightarrow\ &\lambda_1=\lambda_2=\cdots=\lambda_{4+\nu}=1 \quad \text{and} \quad \lambda_{4+\nu+1}=\lambda_{4+\nu+2}=\cdots=\lambda_8=0\,.
\end{aligned} \tag{E43}$$

Therefore the lowest $E_{\pm K_M}$ is achieved if and only if

$$\begin{pmatrix}\zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger\end{pmatrix}\cong \mathrm{diag}(\underbrace{1,1,\ldots,1}_{4+\nu},\underbrace{0,0,\ldots,0}_{4-\nu})\,, \tag{E44}$$

where $\cong$ stands for matrix similarity defined by unitary transformations in U(8). Equation (E44) suggests that $\begin{pmatrix}\zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger\end{pmatrix}$ is a projection matrix. Then, we know

$$\mathrm{Tr}\left[\begin{pmatrix}\zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger\end{pmatrix}\begin{pmatrix}\zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger\end{pmatrix}\right]=\mathrm{Tr}\left[\begin{pmatrix}\zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger\end{pmatrix}\right]=\mathrm{Tr}[\zeta\zeta^\dagger]=\mathrm{Tr}[\zeta\zeta^\dagger\zeta\zeta^\dagger]\,, \tag{E45}$$

which results in

$$\mathrm{Tr}[\zeta_+\zeta_+^\dagger\zeta_+\zeta_+^\dagger]+\mathrm{Tr}[\zeta_-\zeta_-^\dagger\zeta_-\zeta_-^\dagger]=\mathrm{Tr}[\zeta_+\zeta_+^\dagger\zeta_+\zeta_+^\dagger]+\mathrm{Tr}[\zeta_-\zeta_-^\dagger\zeta_-\zeta_-^\dagger]+\mathrm{Tr}[\zeta_+\zeta_-^\dagger\zeta_-\zeta_+^\dagger]+\mathrm{Tr}[\zeta_-\zeta_+^\dagger\zeta_+\zeta_-^\dagger]\,, \tag{E46}$$

which results in

$$\mathrm{Tr}[\zeta_+\zeta_-^\dagger\zeta_-\zeta_+^\dagger]=0\Rightarrow\zeta_+\zeta_-^\dagger=0\,. \tag{E47}$$

Combined with

$$\zeta_+\zeta_-^\dagger=0\Rightarrow\begin{pmatrix}\zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger\end{pmatrix}=\zeta\zeta^\dagger\Rightarrow\begin{pmatrix}\zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger\end{pmatrix}\cong\mathrm{diag}(\underbrace{1,1,\ldots,1}_{4+\nu},\underbrace{0,0,\ldots,0}_{4-\nu})\,, \tag{E48}$$

we know $\zeta_+\zeta_-^\dagger = 0$ is equivalent to Eq. (E44). Therefore, in the high-$\mathcal{E}$ limit, the lowest $E_{\pm K_M}$ is achieved if and only if $\zeta_+\zeta_-^\dagger = 0$ (i.e., the intervalley coherence is zero), if we neglect all corrections to the energies that are of order $O(|U_1/\mathcal{E}^2|)$.

***b.* $\boldsymbol{\Gamma_M}$**

Now let us turn to the Γ_M point. The one-shot Hartree-Fock Hamiltonian at Γ_M reads

$$\begin{pmatrix} \frac{1}{2}U_1 + \nu(U_1+6U_2) - U_1\zeta\zeta^\dagger & \widetilde{\gamma}\mathbb{1}_8 & \\ \widetilde{\gamma}\mathbb{1}_8 & \nu W_1 & \\ & & \nu W_3 + h_{\Gamma_1\Gamma_2} \end{pmatrix}, \tag{E49}$$

where $\widetilde{\gamma} = \gamma + B_\gamma \mathcal{E}^2$,

$$h_{\Gamma_1\Gamma_2} = \widetilde{M}\eta_0\sigma_x s_0 - \frac{J}{2}\left(\eta_z\sigma_0 s_0 \zeta\zeta^\dagger \eta_z\sigma_0 s_0 + \eta_0\sigma_z s_0 \zeta\zeta^\dagger \eta_0\sigma_z s_0 - \mathbb{1}_8\right), \tag{E50}$$

and $\widetilde{M} = M + B_M\mathcal{E}^2$. Owing to

$$\frac{1}{2}U_1 + \nu(U_1+6U_2) - U_1\zeta\zeta^\dagger \cong \begin{pmatrix} \left(-\frac{1}{2}U_1 + \nu(U_1+6U_2)\right)\mathbb{1}_{(4+\nu)} & \\ & \left(\frac{1}{2}U_1 + \nu(U_1+6U_2)\right)\mathbb{1}_{(4-\nu)} \end{pmatrix}, \tag{E51}$$

the eigenvalues of

$$\begin{pmatrix} \frac{1}{2}U_1 + \nu(U_1+6U_2) - U_1\zeta\zeta^\dagger & \widetilde{\gamma}\mathbb{1}_8 \\ \widetilde{\gamma}\mathbb{1}_8 & \nu W_1 \end{pmatrix} \tag{E52}$$

does not depend on the ζ as long as ν is given. Therefore we will focus on $h_{\Gamma_1\Gamma_2}$.

Since we consider the high-$\mathcal{E}$ limit, we have $|\widetilde{M}| \gg J$. Then, the energy difference between different states given by $h_{\Gamma_1\Gamma_2}$ should be of order J, which is generally much smaller than the energy difference at $\pm K_M$ which is of the order U_1. Therefore we should only focus on the states with lowest $E_{\pm K_M}$, i.e., states with zero intervalley coherence. In other words, the discussion at $\pm K_M$ already suggests that only states without intervalley coherence are favored at large $\mathcal{E}$.

Now we show that $h_{\Gamma_1\Gamma_2}$ further picks out the favored high-$\mathcal{E}$ states among all states without intervalley coherence. Since we now only care about the states without intervalley coherence (i.e., $\zeta_+\zeta_-^\dagger = 0$), we have

$$\zeta\zeta^\dagger = \begin{pmatrix} \zeta_+\zeta_+^\dagger & \\ & \zeta_-\zeta_-^\dagger \end{pmatrix}. \tag{E53}$$

In general, $\zeta_\eta\zeta_\eta^\dagger$ has the following form:

$$\zeta_\eta\zeta_\eta^\dagger = \sum_{\mu\nu\in\{0,x,y,z\}} (y_\eta)_{\mu\nu}\sigma_\mu s_\nu, \tag{E54}$$

where $(y_\eta)_{\mu\nu}$ are the real coefficients. Owing to the spin-charge U(2) symmetries in each valley, namely, U(2) × U(2), we can always first rotate $\sum_{\nu\in\{0,x,y,z\}}(y_\eta)_{z\nu}\sigma_z s_\nu$ to $\sum_{\nu\in\{0,z\}}(y_\eta)_{z\nu}\sigma_z s_\nu$, and then rotate $\sum_{\nu\in\{0,x,y,z\}}(y_\eta)_{0\nu}\sigma_0 s_\nu$ to $\sum_{\nu\in\{0,x,z\}}(y_\eta)_{0\nu}\sigma_0 s_\nu$. Therefore we have

$$(y_\eta)_{zx} = (y_\eta)_{zy} = (y_\eta)_{0y} = 0 \tag{E55}$$

up to U(2) × U(2). With this observation, we have the following expression:

$$\zeta_\eta\zeta_\eta^\dagger = a_{0,\eta} + a_\eta\sigma_0 s_z + c_\eta\sigma_0 s_x + \frac{b_{2-\eta}+b_{3-\eta}}{2}\sigma_z s_0 + \frac{b_{2-\eta}-b_{3-\eta}}{2}\sigma_z s_z + \sum_{\mu\in\{0,z\},\nu\in\{0,x,y,z\}} (y_\eta)_{\mu\nu}\sigma_\mu s_\nu \tag{E56}$$

up to U(2) × U(2). Then,

$$h_{\Gamma_1\Gamma_2} = \begin{pmatrix} \widetilde{h}_+ & \\ & \widetilde{h}_- \end{pmatrix} \tag{E57}$$

up to U(2) × U(2), where

$$\widetilde{h}_\eta = \widetilde{M}\sigma_x s_0 - J\left[a_{0,\eta} + a_\eta\sigma_0 s_z + c_\eta\sigma_0 s_x + \frac{b_{2-\eta}+b_{3-\eta}}{2}\sigma_z s_0 + \frac{b_{2-\eta}-b_{3-\eta}}{2}\sigma_z s_z\right] + \frac{J}{2}. \tag{E58}$$

According to Eq. (E49), the eigenstates of $h_{\Gamma_1\Gamma_2}$ with energies lower than $\nu(U_1 + 6U_2 - W_3)$ are occupied. Before proceeding, we list some useful constraints derived from $\zeta\zeta^\dagger$ being a projection matrix of rank $4+\nu$. First, $\zeta_\eta\zeta_\eta^\dagger\zeta_\eta\zeta_\eta^\dagger = \zeta_\eta\zeta_\eta^\dagger$ gives

$$a_{0,\eta} = \frac{m_\eta}{4}, \quad a_{0,\eta}^2 + a_\eta^2 + c_\eta^2 + \frac{b_{2-\eta}^2 + b_{3-\eta}^2}{2} + \sum_{\mu\in\{0,z\},\nu\in\{0,x,y,z\}} (y_\eta)_{\mu\nu}^2 = \frac{m_\eta}{4}, \tag{E59}$$

where $m_\eta = \mathrm{Tr}[\zeta_\eta \zeta_\eta^\dagger] \in \mathbb{Z}$ and $m_+ + m_- = 4 + \nu$. Then, owing to the fact that the diagonal elements of $\zeta_\eta \zeta_\eta^\dagger$ are in [0,1], we have

$$\begin{aligned}
&a_{0,\eta} + a_\eta + b_{2-\eta}, a_{0,\eta} + a_\eta - b_{2-\eta}, a_{0,\eta} - a_\eta + b_{3-\eta}, a_{0,\eta} - a_\eta - b_{3-\eta} \in [0,1] \\
&\Rightarrow \begin{cases} a_\eta + b_{2-\eta} \in \left[-\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right] \,\&\, a_\eta - b_{2-\eta} \in \left[-\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right] \\ -a_\eta + b_{3-\eta} \in \left[-\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right] \,\&\, -a_\eta - b_{3-\eta} \in \left[-\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right] \end{cases} \\
&\Rightarrow \begin{cases} a_\eta \in \left[-\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right] \,\&\, \pm b_{2-\eta} \in \left[-\frac{m_\eta}{4} - a_\eta, 1 - \frac{m_\eta}{4} - a_\eta\right] \\ a_\eta \in \left[-1 + \frac{m_\eta}{4}, \frac{m_\eta}{4}\right] \,\&\, \pm b_{3-\eta} \in \left[-\frac{m_\eta}{4} + a_\eta, 1 - \frac{m_\eta}{4} + a_\eta\right] \end{cases} \\
&\Rightarrow |a_\eta| \in \left[0, \min\left(\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right)\right] \,\&\, |b_{2-\eta}| \in \left[0, \min\left(\frac{m_\eta}{4} + a_\eta, 1 - \frac{m_\eta}{4} - a_\eta\right)\right] \,\&\, |b_{3-\eta}| \\
&\in \left[0, \min\left(\frac{m_\eta}{4} - a_\eta, 1 - \frac{m_\eta}{4} + a_\eta\right)\right]. \qquad \text{(E60)}
\end{aligned}$$

Then, since we only care about $\nu \in \{-4, -3, -2, -1, 0\}$, we have

$$\begin{aligned}
&|a_\eta| = |b_{2-\eta}| = |b_{3-\eta}| = 0, \quad \text{for } m_\eta = 0; \\
&|a_\eta| \in \left[0, \frac{m_\eta}{4}\right], \quad |b_{2-\eta}| \in \left[0, \frac{m_\eta}{4} + a_\eta\right], \quad |b_{3-\eta}| \in \left[0, \frac{m_\eta}{4} - a_\eta\right], \quad \text{for } m_\eta = 1; \\
&|a_\eta| \in \left[0, \frac{1}{2}\right], \quad |b_{2-\eta}| \in \left[0, \frac{1}{2} - |a_\eta|\right], \quad |b_{3-\eta}| \in \left[0, \frac{1}{2} - |a_\eta|\right], \quad \text{for } m_\eta = 2; \\
&|a_\eta| \in \left[0, 1 - \frac{m_\eta}{4}\right], \quad |b_{2-\eta}| \in \left[0, 1 - \frac{m_\eta}{4} - a_\eta\right], \quad |b_{3-\eta}| \in \left[0, 1 - \frac{m_\eta}{4} + a_\eta\right], \quad \text{for } m_\eta = 3; \\
&|a_\eta| = |b_{2-\eta}| = |b_{3-\eta}| = 0\,, \quad \text{for } m_\eta = 4\,, \qquad \text{(E61)}
\end{aligned}$$

which leads to

$$b_{2-\eta}^2 + b_{3-\eta}^2 \leqslant \min\left(\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right). \qquad \text{(E62)}$$

In sum, the constraints that we will use are summarized as

$$\begin{aligned}
&a_{0,\eta} = \frac{m_\eta}{4}\,, \quad m_\eta \in \mathbb{Z}_{\geqslant 0}\,, \quad m_+ + m_- = 4 + \nu; \\
&a_{0,\eta}^2 + a_\eta^2 + c_\eta^2 + \frac{b_{2-\eta}^2 + b_{3-\eta}^2}{2} + \sum_{\mu \in \{0,z\}, \nu \in \{0,x,y,z\}} (y_\eta)_{\mu\nu}^2 = \frac{m_\eta}{4}; \\
&|a_\eta| \in \left[0, \min\left(\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right)\right], \quad |b_{2-\eta}| \in \left[0, \min\left(\frac{m_\eta}{4} + a_\eta, 1 - \frac{m_\eta}{4} - a_\eta\right)\right], \\
&|b_{3-\eta}| \in \left[0, \min\left(\frac{m_\eta}{4} - a_\eta, 1 - \frac{m_\eta}{4} + a_\eta\right)\right]; \quad b_{2-\eta}^2 + b_{3-\eta}^2 \leqslant \min\left(\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right). \qquad \text{(E63)}
\end{aligned}$$

We label the four eigenvalues of $\widetilde{h}_\eta$ as $E_{\eta,1}^{\Gamma_M} \leqslant E_{\eta,2}^{\Gamma_M} \leqslant E_{\eta,3}^{\Gamma_M} \leqslant E_{\eta,4}^{\Gamma_M}$. Owing to $|\widetilde{M}| \gg J$, we can solve the eigenvalues perturbatively to $O(M_J^{-2})$, where $M_J = |\widetilde{M}/J|$. To do so, we use the following unitary matrix:

$$\widetilde{U}_\eta = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes s_0\,, \qquad \text{(E64)}$$

$$\widetilde{U}_\eta^\dagger \frac{\widetilde{h}_\eta}{J} \widetilde{U}_\eta = \frac{2 - m_\eta}{4} + \begin{pmatrix} -M_J - a_\eta s_z - c_\eta s_x & \begin{pmatrix} -b_{2-\eta} & \\ & -b_{3-\eta} \end{pmatrix} \\ \begin{pmatrix} -b_{2-\eta} & \\ & -b_{3-\eta} \end{pmatrix} & M_J - a_\eta s_z - c_\eta s_x \end{pmatrix}, \qquad \text{(E65)}$$

where we used the fact that $\widetilde{M} < 0$ and $J > 0$. Then, we can project the off-diagonal $b_{2-\eta}$ and $b_{3-\eta}$ terms to the diagonal block, and get two effective Hamiltonians from $\widetilde{U}_\eta^\dagger \frac{\widetilde{h}_\eta}{J} \widetilde{U}_\eta$ as

$$\frac{2 - m_\eta}{4} \pm M_J - a_\eta s_z - c_\eta s_x \pm \frac{1}{2M_J} \begin{pmatrix} -b_{2-\eta} & \\ & -b_{3-\eta} \end{pmatrix}^2 + O\big(M_J^{-2}\big) \qquad \text{(E66)}$$

leading to

$$E^{\Gamma_M}_{\eta,i}/J = \frac{2-m_\eta}{4} + (-1)^{\lceil i/2\rceil}\left[M_J + \frac{1}{4}\left(b^2_{2-\eta}+b^2_{3-\eta}\right)M_J^{-1}\right] + (-1)^i\left[\sqrt{a^2_\eta + c^2_\eta} + O\left(M_J^{-1}\right)\right] + O\left(M_J^{-2}\right), \tag{E67}$$

where $i = 1, 2, 3, 4$ and $\lceil i/2\rceil$ is the smallest interger that is no smaller than $i/2$. Since we have $|\widetilde{M}| \gg |\nu(U_1 + 6U_2 - W_3)|$ in the high-$\mathcal{E}$ limit, $E^{\Gamma_M}_{\eta,i}$ with $i = 1, 2$ are the only occupied levels, leading to

$$E_{\Gamma_M}/J = -4M_J - \tfrac{1}{2}\left(b_1^2+b_2^2+b_3^2+b_4^2\right)M_J^{-1} + E_{\nu,\Gamma_M}/J + O\left(M_J^{-2}\right), \tag{E68}$$

where $E_{\nu=0,\Gamma_M}$ contains the other contribution that does not rely on ζ as long as ν is fixed. To lower E_{Γ_M}, we just need to maximize $b_1^2+b_2^2+b_3^2+b_4^2$. In the following, we will do it for $\nu = 0, -1, -2$, separately.

For $\nu = 0$, we have three cases distinguished by the values of $m_\pm$, i.e., there exists $\eta_0 \in \{+,-\}$ such that $(m_{\eta_0}, m_{-\eta_0}) = (4,0), (3,1), (2,2)$, which respectively leads to

$$\sum_\eta \min\left(\frac{m_\eta}{4}, 1-\frac{m_\eta}{4}\right) = 0, \frac{1}{2}, 1. \tag{E69}$$

Owing to Eq. (E63), we then have

$$E_{\Gamma_M}/J \geqslant -4M_J - \tfrac{1}{2}M_J^{-1} + E_{\nu=0,\Gamma_M}/J + O\left(M_J^{-2}\right). \tag{E70}$$

Then, by exploiting Eq. (E63),

$$\begin{aligned}
&E_{\Gamma_M}/J = -4M_J - \frac{1}{2}M_J^{-1} + E_{\nu=0,\Gamma_M}/J + O\left(M_J^{-2}\right), \text{ i.e., minimizing } E_{\Gamma_M} \text{ for } \nu = 0 \text{ states without intervalley coherence}\\
&\Leftrightarrow \begin{cases} m_+ = m_- = 2 \\ b_1^2+b_2^2+b_3^2+b_4^2 = 1 \end{cases}\\
&\Leftrightarrow \begin{cases} m_+ = m_- = 2 \\ b_1^2+b_2^2 = b_3^2+b_4^2 = \frac{1}{2} \end{cases}\\
&\Leftrightarrow \begin{cases} m_+ = m_- = 2 \\ |b_1| = |b_2| = |b_3| = |b_4| = \frac{1}{2} \\ a_\eta^2 + c_\eta^2 + \sum_{\mu\in\{0,z\},\nu\in\{0,x,y,z\}}(y_\eta)^2_{\mu\nu} = 0 \ \forall \eta \in \{+,-\} \end{cases}.
\end{aligned} \tag{E71}$$

Then, combined with Eqs. (E63) and (E54), it means that E_{Γ_M} is minimized if and only if $\zeta\zeta^\dagger$ is [up to U(2) × U(2)] spin-diagonal with each of the 4 valley-spin blocks being $\frac{1}{2}(1 \pm \sigma_z)$.

For $\nu = -1$, we have two cases distinguished by the values of $m_\pm$, i.e., there exists $\eta_0 \in \{+,-\}$ such that $(m_{\eta_0}, m_{-\eta_0}) = (3,0), (2,1)$, which respectively leads to

$$\sum_\eta \min\left(\frac{m_\eta}{4}, 1-\frac{m_\eta}{4}\right) = \frac{1}{4}, \frac{3}{4}. \tag{E72}$$

Owing to Eq. (E63), we then have

$$E_{\Gamma_M}/J \geqslant -4M_J - \tfrac{3}{8}M_J^{-1} + E_{\nu=-1,\Gamma_M}/J + O\left(M_J^{-2}\right). \tag{E73}$$

Then, by exploiting Eq. (E63),

$$\begin{aligned}
&E_{\Gamma_M}/J = -4M_J - \frac{3}{8}M_J^{-1} + E_{\nu=0,\Gamma_M}/J + O\left(M_J^{-2}\right), \text{ i.e., minimizing } E_{\Gamma_M} \text{ for } \nu = -1 \text{ states without intervalley coherence}\\
&\Leftrightarrow \begin{cases} m_{\eta_0} = 2,\ m_{-\eta_0} = 1 \\ b_1^2+b_2^2+b_3^2+b_4^2 = \frac{3}{4} \end{cases}\\
&\Leftrightarrow \begin{cases} m_{\eta_0} = 2,\ m_{-\eta_0} = 1 \\ b^2_{2-\eta_0} + b^2_{3-\eta_0} = \frac{1}{2},\ b^2_{2+\eta_0} + b^2_{3+\eta_0} = \frac{1}{4} \end{cases}\\
&\Leftrightarrow \begin{cases} m_{\eta_0} = 2,\ |b_{2-\eta_0}| = |b_{3-\eta_0}| = \frac{1}{2},\ a^2_{\eta_0} + c^2_{\eta_0} + \sum_{\mu\in\{0,z\},\nu\in\{0,x,y,z\}}(y_{\eta_0})^2_{\mu\nu} = 0 \\ m_{-\eta_0} = 1,\ |a_{-\eta_0}| \in [0, \frac{1}{4}],\ \frac{1}{4} \leqslant (\frac{1}{4}+a_{\eta_0})^2 + (\frac{1}{4}-a_{-\eta_0})^2,\ b^2_{2+\eta_0} + b^2_{3+\eta_0} = \frac{1}{4} \end{cases}\\
&\Leftrightarrow \begin{cases} m_{\eta_0} = 2,\ |b_{2-\eta_0}| = |b_{3-\eta_0}| = \frac{1}{2},\ a^2_{\eta_0} + c^2_{\eta_0} + \sum_{\mu\in\{0,z\},\nu\in\{0,x,y,z\}}(y_{\eta_0})^2_{\mu\nu} = 0 \\ m_{-\eta_0} = 1,\ a_{-\eta_0} = \pm\frac{1}{4}, |b_{2-\eta_0}| = a_{-\eta_0} + \frac{1}{4}, |b_{3-\eta_0}| = -a_{-\eta_0} + \frac{1}{4},\ c^2_{-\eta_0} + \sum_{\mu\in\{0,z\},\nu\in\{0,x,y,z\}}(y_{-\eta_0})^2_{\mu\nu} = 0 \end{cases}.
\end{aligned} \tag{E74}$$

Then, combined with Eqs. (E63) and (E54), it means that E_{Γ_M} is minimized if and only if $\zeta\zeta^\dagger$ is [up to U(2) × U(2)] spin-diagonal with 1 valley-spin block being zero and each of the remaining three valley-spin blocks being $\frac{1}{2}(1 \pm \sigma_z)$.

For $\nu = -2$, we have two cases distinguished by the values of $m_\pm$, i.e., there exists $\eta_0 \in \{+, -\}$ such that $(m_{\eta_0}, m_{-\eta_0}) = (2, 0), (1, 1)$, which both leads to

$$\sum_\eta \min\left(\frac{m_\eta}{4}, 1 - \frac{m_\eta}{4}\right) = \frac{1}{2} . \tag{E75}$$

Owing to Eq. (E63), we then have

$$E_{\Gamma_M}/J \geqslant -4M_J - \tfrac{1}{4}M_J^{-1} + E_{\nu=-2,\Gamma_M}/J + O\left(M_J^{-2}\right) . \tag{E76}$$

Then, by combining Eq. (E63) with the discussions for $\nu = 0$ and $\nu = -1$, we can get that E_{Γ_M} is minimized if and only if $\zeta\zeta^\dagger$ is [up to U(2) × U(2)] spin-diagonal with two valley-spin blocks being zero and each of the remaining two valley-spin blocks being $\frac{1}{2}(1 \pm \sigma_z)$.

Therefore we obtain the statement that for states without intervalley coherence, E_{Γ_M} is the lowest if and only of $\zeta\zeta^\dagger$ [up to U(2) × U(2)] has a spin-valley diagonal form with $4 + \nu$ blocks being $(\sigma_0 \pm \sigma_z)/2$ and $4 - \nu$ blocks being zero.

c. In sum

Let us summarize the whole procedure. In the high-$\mathcal{E}$ limit and given $\nu \in \{0, -1, -2\}$, we require the ground states to first minimize the total energy of the occupied levels of Eq. (E29), which make them have zero intervalley-coherence, and then minimize the total energy of the occupied levels of Eq. (E49). Then, we arrive at the proposition 1.

We numerate all initial states that satisfy proposition 1 for $\nu = 0, -1, -2$. All the states we found are [up to the symmetries of the total interacting Hamiltonian Eq. (32)] included in Appendix D 2.

Explicitly, at $\nu = 0$, the states that satisfy proposition 1 are $|\text{VH}_0^{\nu=0}\rangle$, $|\text{Chern}_0^{\nu=0}\rangle$, $|\text{half-Chern}_0^{\nu=0}\rangle$, $|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\rangle$ and their symmetry-related states.

At $\nu = -1$, the states that satisfy proposition 1 are $|\text{PVP}_0^{1,\nu=-1}\rangle$, $|\text{PVP}_0^{2,\nu=-1}\rangle$ and $|\text{PVP}_0^{3,\nu=-1}\rangle$ and their symmetry-related states.

At $\nu = -2$, the states that satisfy proposition 1 are $|\text{VP}_0^{1,\nu=-2}\rangle$, $|\text{VP}_0^{2,\nu=-2}\rangle$, $|\text{valley-unpolarized}_0^{1,\nu=-2}\rangle$, $|\text{valley-unpolarized}_0^{2,\nu=-2}\rangle$ and their symmetry-related states.

According to Sec. IV, these inital states, after performing the self-consistent calculations, give the high-$\mathcal{E}$ low-energy states with very similar energies (similar for a fixed ν).

3. Simple rule for high-$\mathcal{E}$ states: $\mathcal{E} = 300$ meV in EUS

(Recall that EUS is the unit system in which Å is the length unit and meV is the energy unit, as discussed at the beginning of Sec. II.)

In Appendix E 2, we analytically derive proposition 1 by looking at $\pm K_M$ and Γ_M in the high-$\mathcal{E}$ limit, i.e., assuming an infinitely large $\mathcal{E}$. Furthermore, we assume $\mu = \nu(U_1 + 6U_2)$. However, as shown in Fig. 1, we can only claim the validity of the f-c-d model in $\mathcal{E} \in [0, 300\text{ meV}]$ (EUS). Therefore, in this part, we will discuss the validity of Appendix E 2 for $\mathcal{E} = 300$ meV (EUS).

First, we note that $|\widetilde{M}| \gg |\nu(U_1 + 6U_2 - W_3)|$ need to hold in order to use Eq. (E68). For $\mathcal{E} = 300$ meV (EUS), we have $|\widetilde{M}| \approx 0.14$. However, for $\nu = -1$, we have $|\nu(U_1 + 6U_2 - W_3)| \approx 0.12$, which is close to $|\widetilde{M}|$; for $\nu = -2$, we have $|\nu(U_1 + 6U_2 - W_3)| \approx 0.24$, which is larger than $|\widetilde{M}|$. On the other hand, the same issue does not occur for $\nu = 0$ since $|\nu(U_1 + 6U_2 - W_3)| = 0$. Therefore the simplifications in Appendix E 2 are not all valid for $\nu = -1, -2$.

Now let us focus on $\nu = 0$, for which $\mu = \nu(U_1 + 6U_2) = 0$ is exactly correct. We discuss $\pm K_M$ first. We have $\epsilon_\gamma = (-)^\gamma|\sqrt{2}M_1\mathcal{E}|$ and $\chi_\gamma = \frac{1}{\sqrt{2}}(1, (-)^\gamma \text{sgn}(M_1\mathcal{E}))$ according to Eq. (E27). Then, Eq. (E28) becomes

$$\widetilde{U}^\dagger_{\eta K_M}\widetilde{h}^{\eta K_M}_{\text{HF,OS}}(0)\widetilde{U}_{\eta K_M} = \begin{pmatrix} \epsilon_0\mathbb{1}_{4\times4} & & \\ & \epsilon_1\mathbb{1}_{4\times4} & \\ & & 0\mathbb{1}_{4\times4} \end{pmatrix} - U_1 \left(\begin{array}{cc|c} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \otimes \left(\zeta_\eta\zeta_\eta^\dagger - \frac{1}{2}\right) & & \begin{matrix} \frac{1}{\sqrt{2}}\zeta_\eta\zeta_{-\eta}^\dagger \\ \frac{1}{\sqrt{2}}\zeta_\eta\zeta_{-\eta}^\dagger \end{matrix} \\ \hline \frac{1}{\sqrt{2}}\zeta_{-\eta}\zeta_\eta^\dagger & \frac{1}{\sqrt{2}}\zeta_{-\eta}\zeta_\eta^\dagger & \zeta_{-\eta}\zeta_{-\eta}^\dagger - \frac{1}{2} \end{array}\right) . \tag{E77}$$

Since $P_{ii} \in [0, 1]$ and $|P_{i\neq j}| \in [0, \frac{1}{2}]$ for any hermitian projectoin matrix P, the elements that couple different $\epsilon_0\mathbb{1}_{4\times4}$, $\epsilon_1\mathbb{1}_{4\times4}$ and $0\mathbb{1}_{4\times4}$ blocks have amplitudes no larger than $\frac{1}{2\sqrt{2}}U_1$, while the gaps among those blocks are no smaller than $|\sqrt{2}M_1\mathcal{E}|$. The energy contributions of the elements that couple different $\epsilon_0\mathbb{1}_{4\times4}$, $\epsilon_1\mathbb{1}_{4\times4}$ and $0\mathbb{1}_{4\times4}$ blocks are of the order $|\frac{U_1}{4M_1\mathcal{E}}|^2 \approx 0.3$, which can be neglected. Then, it is legitimate to only consider the Eq. (E29), which eventually leads to the fact that only states without intervalley coherent should be considered.

Now turn to Γ_M. In Eq. (E67), the terms that we neglect compared to the largest-order term are of order $M_J^{-3} \approx 0.29$, which is also reasonable. Then, the later derivation based on Eq. (E67) in Appendix E 2 should all be valid, leading to proposition 1. Therefore the derivation in Appendix E 2 should be valid for $\nu = 0$ even if $\mathcal{E} = 300$ meV (EUS).

TABLE III. This table shows the energies of the occupied levels of the one-shot Hamiltonian at $\pm K_M$ and Γ_M for the initial states in Appendix D 2 at $\nu = 0$ and $\mathcal{E} = 300$ meV (EUS). The first column specifies the initial states. The second, third, and fourth columns specifies the energies of the occupied levels of the one-shot Hamiltonian at Γ_M, K_M, and $-K_M$, respectively. The fifth column shows the total of the second, third, and fourth columns. The states with the lowest total are the lowest four, which are the high-$\mathcal{E}$ ground states found in the numerical calculations described in Appendix D.

Initial States	Γ_M	K_M	$-K_M$	Total
$\|\text{K-IVC}_0^{\nu=0}\rangle$	−1.30093	−1.15152	−1.15152	−3.60397
$\|\text{IVC}_0^{\nu=0}\rangle$	−1.27088	−1.15152	−1.15152	−3.57392
$\|\text{VP}_0^{\nu=0}\rangle$	−1.27088	−1.32885	−1.32885	−3.92858
$\|\text{PVP}_0^{1,\nu=0}\rangle$	−1.28591	−1.32885	−1.32885	−3.9436
$\|\text{PVP}_0^{2,\nu=0}\rangle$	−1.28591	−1.32885	−1.32885	−3.9436
$\|\text{VH}_0^{\nu=0}\rangle$	−1.30093	−1.32885	−1.32885	−3.95863
$\|\text{Chern}_0^{\nu=0}\rangle$	−1.30093	−1.32885	−1.32885	−3.95863
$\|\text{half-Chern}_0^{\nu=0}\rangle$	−1.30093	−1.32885	−1.32885	−3.95863
$\|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\rangle$	−1.30093	−1.32885	−1.32885	−3.95863

The numerical evidence for the validity for $\nu = 0$ and $\mathcal{E} = 300$ meV (EUS) is that if we only compare the energies of the occupied levels of the one-shot Hamiltonian at $\pm K_M$ and Γ_M for the initial states in Appendix D 2 at $\nu = 0$ and $\mathcal{E} = 300$ meV (EUS), we can get the right ground states, as shown in Table III.

4. One-shot Hartree-Fock energies for high-$\mathcal{E}$ states

The competing energies of the high-$\mathcal{E}$ Hartee-Fock ground states with fixed ν can also be understood analytically at one-shot level. At $\nu = 0$, the one-shot Hartree-Fock Hamiltonains for different high-$\mathcal{E}$ initial states are related with each other:

$$H^{\text{VH,OS}} = H^{\text{VH,OS}}_{+,\uparrow} + H^{\text{VH,OS}}_{+,\downarrow} + H^{\text{VH,OS}}_{-,\uparrow} + H^{\text{VH,OS}}_{-,\downarrow} - E_0^{\text{OS}} , \tag{E78}$$

$$H^{\text{Chern,OS}} = H^{\text{VH,OS}}_{+,\uparrow} + H^{\text{VH,OS}}_{+,\downarrow} + C_2\overline{\mathcal{T}}H^{\text{VH,OS}}_{-,\uparrow}(C_2\overline{\mathcal{T}})^{-1} + C_2\overline{\mathcal{T}}H^{\text{VH,OS}}_{-,\downarrow}(C_2\overline{\mathcal{T}})^{-1} - E_0^{\text{OS}} , \tag{E79}$$

$$H^{\text{half-Chern,OS}} = H^{\text{VH,OS}}_{+,\uparrow} + H^{\text{VH,OS}}_{+,\downarrow} + C_2\overline{\mathcal{T}}H^{\text{VH,OS}}_{-,\uparrow}(C_2\overline{\mathcal{T}})^{-1} + H^{\text{VH,OS}}_{-,\downarrow} - E_0^{\text{OS}} , \tag{E80}$$

$$H^{C_2\mathcal{T}\text{-invaraint,OS}} = C_2\overline{\mathcal{T}}H^{\text{VH,OS}}_{+,\uparrow}(C_2\overline{\mathcal{T}})^{-1} + H^{\text{VH,OS}}_{+,\downarrow} + C_2\overline{\mathcal{T}}H^{\text{VH,OS}}_{-,\uparrow}(C_2\overline{\mathcal{T}})^{-1} + H^{\text{VH,OS}}_{-,\downarrow} - E_0^{\text{OS}} , \tag{E81}$$

where

$$H^{\text{VH,OS}}_{\eta,s} = H_{0,\eta,s} + \left(-\frac{1}{2}U_1\right)\sum_{\boldsymbol{R}_M} f^\dagger_{\eta,\boldsymbol{R}_M,s}\tau_z f_{\eta,\boldsymbol{R}_M,s} + \left(-\frac{J}{2}\right)\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{\eta,\boldsymbol{k},\Gamma_1\Gamma_2,s}\tau_z c_{\eta,\boldsymbol{k},\Gamma_1\Gamma_2,s} , \tag{E82}$$

$H_{0,\eta,s}$ is the spin-s part of $H_{0,\eta}$ which is specified in Eq. (54), and $\overline{\mathcal{T}}$ is the spinless time-reversal symmetry.

At $\nu = -1$, the one-shot Hartree-Fock Hamiltonians for different high-$\mathcal{E}$ initial states are related:

$$H^{\text{PVP}_1,\text{OS}} = H^{\text{PVP}_1,\text{OS}}_{+,\uparrow} + H^{\text{PVP}_1,\text{OS}}_{+,\downarrow} + H^{\text{PVP}_1,\text{OS}}_{-,\uparrow} + H^{\text{PVP}_1,\text{OS}}_{-,\downarrow} - E_0^{\text{OS}} , \tag{E83}$$

$$H^{\text{PVP}_2,\text{OS}} = C_2\overline{\mathcal{T}}H^{\text{PVP}_1,\text{OS}}_{+,\uparrow}(C_2\overline{\mathcal{T}})^{-1} + C_2\overline{\mathcal{T}}H^{\text{PVP}_1,\text{OS}}_{+,\downarrow}(C_2\overline{\mathcal{T}})^{-1} + H^{\text{PVP}_1,\text{OS}}_{-,\uparrow} + H^{\text{PVP}_1,\text{OS}}_{-,\downarrow} - E_0^{\text{OS}} , \tag{E84}$$

$$H^{\text{PVP}_3,\text{OS}} = C_2\overline{\mathcal{T}}H^{\text{PVP}_1,\text{OS}}_{+,\uparrow}(C_2\overline{\mathcal{T}})^{-1} + H^{\text{PVP}_1,\text{OS}}_{+,\downarrow} + H^{\text{PVP}_1,\text{OS}}_{-,\uparrow} + H^{\text{PVP}_1,\text{OS}}_{-,\downarrow} - E_0^{\text{OS}} , \tag{E85}$$

where

$$\begin{aligned} H^{\text{PVP}_1,\text{OS}}_{+,\uparrow} = {} & H_{0,+,\uparrow} + \sum_{\boldsymbol{R}_M} f^\dagger_{+,\boldsymbol{R}_M,\uparrow}\left(-\frac{1}{2}U_1 - 6U_2 - \frac{1}{2}U_1 + \frac{1}{2}U_1\tau_z\right) f_{+,\boldsymbol{R}_M,\uparrow} + (-W_1)\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{+,\boldsymbol{k},\Gamma_3,\uparrow}\tau_z c_{+,\boldsymbol{k},\Gamma_3,\uparrow} \\ & + \sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{+,\boldsymbol{k},\Gamma_1\Gamma_2,\uparrow}\left(-W_3 + \frac{J}{2}\tau_z\right) c_{+,\boldsymbol{k},\Gamma_1\Gamma_2,\uparrow} + (-W_{fd})\sum_{\boldsymbol{p}}^{\Lambda_d} d^\dagger_{+,\boldsymbol{p},\uparrow} d_{+,\boldsymbol{p},\uparrow} , \end{aligned} \tag{E86}$$

$$H^{\text{PVP}_1,\text{OS}}_{+,\downarrow} = H_{0,+,\downarrow} + \sum_{\boldsymbol{R}_M} f^\dagger_{+,\boldsymbol{R}_M,\downarrow}\left(-\frac{1}{2}U_1 - 6U_2 - \frac{1}{2}U_1 + \frac{1}{2}U_1\tau_z\right) f_{+,\boldsymbol{R}_M,\downarrow} + (-W_1)\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{+,\boldsymbol{k},\Gamma_3,\downarrow}\tau_z c_{+,\boldsymbol{k},\Gamma_3,\downarrow}$$

$$+\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{+,\boldsymbol{k},\Gamma_1\Gamma_2,\downarrow}\left(-W_3+\frac{J}{2}\tau_z\right)c_{+,\boldsymbol{k},\Gamma_1\Gamma_2,\downarrow}+(-W_{fd})\sum_{\boldsymbol{p}}^{\Lambda_d} d^\dagger_{+,\boldsymbol{p},\downarrow}d_{+,\boldsymbol{p},\downarrow}\,, \tag{E87}$$

$$H^{\mathrm{PVP_1,OS}}_{-,\uparrow}=H_{0,-,\uparrow}+\sum_{\boldsymbol{R}_M} f^\dagger_{-,\boldsymbol{R}_M,\uparrow}\left(-\frac{1}{2}U_1-6U_2\right)f_{-,\boldsymbol{R}_M,\uparrow}+(-W_1)\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{-,\boldsymbol{k},\Gamma_3,\uparrow}\tau_z c_{-,\boldsymbol{k},\Gamma_3,\uparrow}$$
$$+\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{-,\boldsymbol{k},\Gamma_1\Gamma_2,\uparrow}\left(-W_3+\frac{J}{2}\right)c_{-,\boldsymbol{k},\Gamma_1\Gamma_2,\uparrow}+(-W_{fd})\sum_{\boldsymbol{p}}^{\Lambda_d} d^\dagger_{-,\boldsymbol{p},\uparrow}d_{-,\boldsymbol{p},\uparrow}\,, \tag{E88}$$

and

$$H^{\mathrm{PVP_1,OS}}_{-,\downarrow}=H_{0,-,\downarrow}+\sum_{\boldsymbol{R}_M} f^\dagger_{-,\boldsymbol{R}_M,\downarrow}\left(-\frac{1}{2}U_1-6U_2-\frac{1}{2}U_1-\frac{1}{2}U_1\tau_z\right)f_{-,\boldsymbol{R}_M,\downarrow}+(-W_1)\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{-,\boldsymbol{k},\Gamma_3,\downarrow}\tau_z c_{-,\boldsymbol{k},\Gamma_3,\downarrow}$$
$$+\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{-,\boldsymbol{k},\Gamma_1\Gamma_2,\downarrow}\left(-W_3-\frac{J}{2}\tau_z\right)c_{-,\boldsymbol{k},\Gamma_1\Gamma_2,\downarrow}+(-W_{fd})\sum_{\boldsymbol{p}}^{\Lambda_d} d^\dagger_{-,\boldsymbol{p},\downarrow}d_{-,\boldsymbol{p},\downarrow}\,. \tag{E89}$$

At $\nu=-2$, the one-shot Hartree-Fock Hamiltonains for different high-$\mathcal{E}$ initial states are related:

$$H^{\mathrm{VP_1,OS}}=H^{\mathrm{VP_1,OS}}_{+,\uparrow}+H^{\mathrm{VP_1,OS}}_{+,\downarrow}+H^{\mathrm{VP_1,OS}}_{-,\uparrow}+H^{\mathrm{VP_1,OS}}_{-,\downarrow}-E_0^{\mathrm{OS}}\,, \tag{E90}$$

$$H^{\mathrm{VP_2,OS}}=C_2\overline{\mathcal{T}}H^{\mathrm{VP_1,OS}}_{+,\uparrow}(C_2\overline{\mathcal{T}})^{-1}+H^{\mathrm{VP_1,OS}}_{+,\downarrow}+H^{\mathrm{VP_1,OS}}_{-,\uparrow}+H^{\mathrm{VP_1,OS}}_{-,\downarrow}-E_0^{\mathrm{OS}}\,, \tag{E91}$$

$$H^{\mathrm{VUP_1,OS}}=C_2\overline{\mathcal{T}}H^{\mathrm{VP_1,OS}}_{+,\uparrow}(C_2\overline{\mathcal{T}})^{-1}+\overline{\mathcal{T}}H^{\mathrm{VP_1,OS}}_{-,\downarrow}\overline{\mathcal{T}}^{-1}+\overline{\mathcal{T}}H^{\mathrm{VP_1,OS}}_{+,\uparrow}\overline{\mathcal{T}}^{-1}+H^{\mathrm{VP_1,OS}}_{-,\downarrow}-E_0^{\mathrm{OS}}\,, \tag{E92}$$

$$H^{\mathrm{VUP_2,OS}}=C_2\overline{\mathcal{T}}H^{\mathrm{VP_1,OS}}_{+,\uparrow}(C_2\overline{\mathcal{T}})^{-1}+\overline{\mathcal{T}}H^{\mathrm{VP_1,OS}}_{-,\downarrow}\overline{\mathcal{T}}^{-1}+H^{\mathrm{VP_1,OS}}_{-,\uparrow}+C_2H^{\mathrm{VP_1,OS}}_{+,\downarrow}C_2^{-1}-E_0^{\mathrm{OS}}\,, \tag{E93}$$

where

$$H^{\mathrm{VP_1,OS}}_{+,\uparrow}=H_{0,+,\uparrow}+\sum_{\boldsymbol{R}_M} f^\dagger_{+,\boldsymbol{R}_M,\uparrow}\left(-\frac{3}{2}U_1-12U_2-\frac{1}{2}U_1+\frac{1}{2}U_1\tau_z\right)f_{+,\boldsymbol{R}_M,\uparrow}+(-2W_1)\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{+,\boldsymbol{k},\Gamma_3,\uparrow}\tau_z c_{+,\boldsymbol{k},\Gamma_3,\uparrow}$$
$$+\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{+,\boldsymbol{k},\Gamma_1\Gamma_2,\uparrow}\left(-2W_3+\frac{J}{2}\tau_z\right)c_{+,\boldsymbol{k},\Gamma_1\Gamma_2,\uparrow}+(-2W_{fd})\sum_{\boldsymbol{p}}^{\Lambda_d} d^\dagger_{+,\boldsymbol{p},\uparrow}d_{+,\boldsymbol{p},\uparrow}\,, \tag{E94}$$

$$H^{\mathrm{VP_1,OS}}_{+,\downarrow}=H_{0,+,\downarrow}+\sum_{\boldsymbol{R}_M} f^\dagger_{+,\boldsymbol{R}_M,\downarrow}\left(-\frac{3}{2}U_1-12U_2-\frac{1}{2}U_1-\frac{1}{2}U_1\tau_z\right)f_{+,\boldsymbol{R}_M,\downarrow}+(-2W_1)\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{+,\boldsymbol{k},\Gamma_3,\downarrow}\tau_z c_{+,\boldsymbol{k},\Gamma_3,\downarrow}$$
$$+\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{+,\boldsymbol{k},\Gamma_1\Gamma_2,\downarrow}\left(-2W_3-\frac{J}{2}\tau_z\right)c_{+,\boldsymbol{k},\Gamma_1\Gamma_2,\downarrow}+(-2W_{fd})\sum_{\boldsymbol{p}}^{\Lambda_d} d^\dagger_{+,\boldsymbol{p},\downarrow}d_{+,\boldsymbol{p},\downarrow}\,, \tag{E95}$$

$$H^{\mathrm{VP_1,OS}}_{-,\uparrow}=H_{0,-,\uparrow}+\sum_{\boldsymbol{R}_M} f^\dagger_{-,\boldsymbol{R}_M,\uparrow}\left(-\frac{3}{2}U_1-12U_2\right)f_{-,\boldsymbol{R}_M,\uparrow}+(-2W_1)\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{-,\boldsymbol{k},\Gamma_3,\uparrow}\tau_z c_{-,\boldsymbol{k},\Gamma_3,\uparrow}$$
$$+\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{-,\boldsymbol{k},\Gamma_1\Gamma_2,\uparrow}\left(-2W_3+\frac{J}{2}\right)c_{-,\boldsymbol{k},\Gamma_1\Gamma_2,\uparrow}+(-2W_{fd})\sum_{\boldsymbol{p}}^{\Lambda_d} d^\dagger_{-,\boldsymbol{p},\uparrow}d_{-,\boldsymbol{p},\uparrow}\,, \tag{E96}$$

and

$$H^{\mathrm{VP_1,OS}}_{-,\downarrow}=H_{0,-,\downarrow}+\sum_{\boldsymbol{R}_M} f^\dagger_{-,\boldsymbol{R}_M,\downarrow}\left(-\frac{3}{2}U_1-12U_2\right)f_{-,\boldsymbol{R}_M,\downarrow}+(-2W_1)\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{-,\boldsymbol{k},\Gamma_3,\downarrow}\tau_z c_{-,\boldsymbol{k},\Gamma_3,\downarrow}$$
$$+\sum_{\boldsymbol{k}}^{\Lambda_c} c^\dagger_{-,\boldsymbol{k},\Gamma_1\Gamma_2,\downarrow}\left(-2W_3+\frac{J}{2}\right)c_{-,\boldsymbol{k},\Gamma_1\Gamma_2,\downarrow}+(-2W_{fd})\sum_{\boldsymbol{p}}^{\Lambda_d} d^\dagger_{-,\boldsymbol{p},\downarrow}d_{-,\boldsymbol{p},\downarrow}\,. \tag{E97}$$

Combined with the fact that the dependence of E_0^{OS} on the states is only through the filling ν, all the listed high-$\mathcal{E}$ initial states with same ν have exactly the same Hartree-Fock energies. The exact degeneracy will be broken in the self-consistent calculation. It is because the density matrix obtained from the self-consistent calculation will have nonzero O^{fc} [defined in Eq. (D5)], while

$O^{fc}=0$ for the initial states shown in Eq. (E1). To be concrete, let us consider the VH state and the Chern state for $\nu=0$, whose exact same energies at one-shot level require

$$H^{\mathrm{VH,OS}}_{+,\uparrow}+H^{\mathrm{VH,OS}}_{+,\downarrow}=H^{\mathrm{Chern,OS}}_{+,\uparrow}+H^{\mathrm{Chern,OS}}_{+,\downarrow} \tag{E98}$$

according to Eq. (E79), which further requires

$$\mathrm{Tr}\left[O^{fc}_{\mathrm{VH}}\begin{pmatrix}0_{8\times8}\\ \eta_z\tau_0s_0\end{pmatrix}\right]=\mathrm{Tr}\left[O^{fc}_{\mathrm{Chern}}\begin{pmatrix}0_{8\times8}\\ \eta_z\tau_0s_0\end{pmatrix}\right] \tag{E99}$$

according to Eq. (D13). As $\mathrm{Tr}[O^{fc}_{\mathrm{VH}}(\begin{smallmatrix}0_{8\times8}\\ \eta_z\tau_0s_0\end{smallmatrix})]$ is not necessarily equal to $\mathrm{Tr}[O^{fc}_{\mathrm{Chern}}(\begin{smallmatrix}0_{8\times8}\\ \eta_z\tau_0s_0\end{smallmatrix})]$ beyond one-shot level, the energies of the VH state and the Chern state for $\nu=0$ are not necessarily the same in the self-consistent calculation. However, the self-consistent calculation shows that the degeneracy breaking effect is very small.

5. Treating $\mathcal{E}$ as a perturbation at $\nu=0$

In this part, we will treat $\mathcal{E}$ perturbatively at $\nu=0$ in order to analytically answer two questions: (i) why there is a phase transition as we gradually increase $\mathcal{E}$? (ii) and what are the Chern numbers for the high-$\mathcal{E}$ ground states?

Let us discuss (i) first. To answer this question, we again try to develop effective models at $\pm K_M$ and at Γ_M. Since we consider $\mathcal{E}$ as a perturbation, the low-energy modes at $\pm K_M$ should be the d modes, since the f modes have energies $\pm\frac{1}{2}U_1$ at ηK_M. Then, based on Eqs. (E23) and (E24), we can project $\mathcal{E}$ to the d modes via the second-order perturbation and get the following effective Hamiltonian at ηK_M:

$$\frac{8M_1^2\mathcal{E}^2}{U_1}\left(\zeta_\eta\zeta_\eta^\dagger-\frac{1}{2}\mathbb{1}_4\right)-p_x\sigma_ys_0+\eta p_y\sigma_xs_0\ . \tag{E100}$$

The effective energy at $\pm K_M$ is given by occupying the lowest 4 bands in each valley. Then, by using Eqs. (D26) and (D30), the effective energy at $\pm K_M$ reads

$$\begin{aligned}E^{\text{K-IVC},\nu=0}_{\pm K_M}&=-4\sum_{\boldsymbol{p}}^{\Lambda_d}|\boldsymbol{p}| \quad \text{for the K-IVC state,}\\ E^{\mathrm{Ch},\nu=0}_{\pm K_M}&=-4\sum_{\boldsymbol{p}}^{\Lambda_d}\sqrt{|\boldsymbol{p}|^2+\frac{16M_1^4\mathcal{E}^4}{U_1^2}} \quad \text{for the Chern state.}\end{aligned} \tag{E101}$$

We mention that our Eq. (E100) is similar to Eq. (29) in Ref. [156], though Ref. [156] was only able to derive their Eq. (29) for Chern-diagonal states.

On the other hand, around Γ_M, the matrix form of the one-shot Hartree-Fock Hamiltonian in the basis $(f^\dagger_{\boldsymbol{p}},c^\dagger_{\boldsymbol{p},\Gamma_3},c^\dagger_{\boldsymbol{p},\Gamma_1\Gamma_2})$ reads

$$h^{\mathrm{OS}}_{\Gamma_M}(\boldsymbol{k})=\begin{pmatrix}\frac{1}{2}U_1-U_1\zeta\zeta^\dagger & \widetilde{\gamma}+v'_\star(p_x\eta_z\tau_x\tau_0+p_y\eta_0\tau_ys_0) & \\ \text{H.c.} & 0\mathbb{1}_{8\times8} & \\ & \text{H.c.} & \widetilde{M}\eta_0\tau_xs_0-\frac{J}{2}[\eta_z\tau_0s_0\zeta\zeta^\dagger\eta_z\tau_0s_0+\eta_0\tau_zs_0\zeta\zeta^\dagger\eta_0\tau_zs_0-\eta_0\tau_0s_0]\end{pmatrix}\begin{matrix}\\ \\ v(k_x\eta_z\tau_0s_0+k_y\mathrm{i}\eta_0\tau_zs_0)\end{matrix}, \tag{E102}$$

where $f^\dagger_{\boldsymbol{k}}=(\dots,f^\dagger_{\eta,\boldsymbol{k},\alpha,s},\dots)$, $c^\dagger_{\boldsymbol{p},\Gamma_3}=(\dots,c^\dagger_{\eta,\boldsymbol{p},\beta,s},\dots)$ with $\beta=1,2$, $c^\dagger_{\boldsymbol{p},\Gamma_1\Gamma_2}=(\dots,c^\dagger_{\eta,\boldsymbol{p},\beta,s},\dots)$ with $\beta=3,4$, $\widetilde{\gamma}=\gamma+B_\gamma\mathcal{E}^2$, $\widetilde{M}=M+B_M\mathcal{E}^2$, and we neglect $v''_\star$ and $B_{v''}\mathcal{E}^2$ since $v''_\star$ is small and we consider a perturbative $\mathcal{E}$. Note that

$$U(\theta)h^{\mathrm{OS}}_{\Gamma_M}(\boldsymbol{k})U^\dagger(\theta)=h^{\mathrm{OS}}_{\Gamma_M}(\boldsymbol{k})|_{\zeta\to U_{\bar{A}}(\theta)\zeta} \quad \text{when} \quad v'_\star=0\ , \tag{E103}$$

where $U(\theta)$ is a chiral U(4) operation [125] with the form

$$U(\theta)=\begin{pmatrix}U_{\bar{A}}(\theta) & & \\ & U_{\bar{A}}(\theta) & \\ & & U_{\bar{B}}(\theta)\end{pmatrix} \text{ with } U_{\bar{A}}(\theta)=\exp\left[\mathrm{i}\sum_{\mu\nu}\theta^{\mu\nu}\bar{A}_{\mu\nu}\right] \text{ and } U_{\bar{B}}(\theta)=\exp\left[\mathrm{i}\sum_{\mu\nu}\theta^{\mu\nu}\bar{B}_{\mu\nu}\right], \tag{E104}$$

and

$$\begin{aligned}\bar{A}_{\mu\nu}&=(\eta_0\tau_0s_\nu,\eta_x\tau_xs_\nu,\eta_y\tau_xs_\nu,\eta_z\tau_0s_\nu)_\mu,\\ \bar{B}_{\mu\nu}&=(\eta_0\tau_0s_\nu,-\eta_x\tau_xs_\nu,-\eta_y\tau_xs_\nu,\eta_z\tau_0s_\nu)_\mu\ .\end{aligned} \tag{E105}$$

Since $e^{\mathrm{i}\eta_y\tau_xs_0\frac{\pi}{4}}\zeta_{\text{K-IVC},\nu=0}=\zeta_{\mathrm{VH},\nu=0}$ derived from the initial states in Appendix D 2 and we know VH and Ch states have the same energies at the one-shot level, the energy difference between Ch and K-IVC states around Γ_M relies on $v'_\star$. Then, we focus on

$$\begin{pmatrix}\frac{1}{2}U_1-U_1\zeta\zeta^\dagger & \widetilde{\gamma}+v'_\star(p_x\eta_z\tau_x\tau_0+p_y\eta_0\tau_ys_0)\\ \text{H.c.} & 0\mathbb{1}_{8\times8}\end{pmatrix}. \tag{E106}$$

Filling the lowest eight bands would give the effective energy at Γ_M, resulting in

$$E^{\text{K-IVC},\nu=0}_{\Gamma_M} = -\sum_{\boldsymbol{k}}^{\Lambda_c}\left[\sqrt{U_1^2+16(|v'_\star \boldsymbol{p}|-|\widetilde{\gamma}|)^2}+\sqrt{U_1^2+16(|v'_\star \boldsymbol{p}|+|\widetilde{\gamma}|)^2}\right] \quad \text{for the K-IVC state,}$$

$$E^{\text{Ch},\nu=0}_{\Gamma_M} = -\sum_{\boldsymbol{k}}^{\Lambda_c}\sum_{z=\pm}\left|\sqrt{U_1^2+16|v'_\star \boldsymbol{p}|^2}+z\sqrt{U_1^2+16\widetilde{\gamma}^2}\right| \quad \text{for the Chern state.} \tag{E107}$$

Then, we know $E^{\text{K-IVC},\nu=0}_{\Gamma_M} \leqslant E^{\text{Ch},\nu=0}_{\Gamma_M}$, since

$$\begin{aligned}&\sqrt{a^2+(b-c)^2}+\sqrt{a^2+(b+c)^2} \geqslant 2\sqrt{a^2+b^2}\\ &\quad = |\sqrt{a^2+b^2}-\sqrt{a^2+c^2}|+|\sqrt{a^2+b^2}+\sqrt{a^2+c^2}| \quad \text{for } a>0 \quad \text{and} \quad b\geqslant c\geqslant 0\end{aligned} \tag{E108}$$

derived from

$$\frac{\partial}{\partial c}(\sqrt{a^2+(b-c)^2}+\sqrt{a^2+(b+c)^2}) = \sqrt{1-\frac{a^2}{a^2+(b+c)^2}}-\sqrt{1-\frac{a^2}{a^2+(b-c)^2}} > 0 \;\forall a>0, b\geqslant c>0\,. \tag{E109}$$

In sum, the total effective energy is

$$E^{\text{K-IVC},\nu=0}_{\text{eff}} = E^{\text{K-IVC},\nu=0}_{\Gamma_M}+E^{\text{K-IVC},\nu=0}_{\pm K_M}, \quad E^{\text{Ch},\nu=0}_{\text{eff}} = E^{\text{Ch},\nu=0}_{\Gamma_M}+E^{\text{Ch},\nu=0}_{\pm K_M}. \tag{E110}$$

Clearly, at $\mathcal{E}=0$, we have $E^{\text{K-IVC},\nu=0}_{\text{eff}} < E^{\text{Ch},\nu=0}_{\text{eff}}$ since $E^{\text{K-IVC},\nu=0}_{\Gamma_M} < E^{\text{Ch},\nu=0}_{\Gamma_M}$ and $E^{\text{K-IVC},\nu=0}_{\pm K_M} = E^{\text{Ch},\nu=0}_{\pm K_M}$. Moreover, at $\mathcal{E}=\mathcal{E}_c$ ($\approx$294.816 meV in EUS) that satisfies $\gamma+B_\gamma\mathcal{E}_c^2=0$, we have $E^{\text{K-IVC},\nu=0}_{\text{eff}} > E^{\text{Ch},\nu=0}_{\text{eff}}$ since $E^{\text{K-IVC},\nu=0}_{\Gamma_M} = E^{\text{Ch},\nu=0}_{\Gamma_M}$ and $E^{\text{K-IVC},\nu=0}_{\pm K_M} > E^{\text{Ch},\nu=0}_{\pm K_M}$, demonstrating the existence of the transition (as increasing $\mathcal{E}$ from $\mathcal{E}=0$ to $\mathcal{E}=\mathcal{E}_c$). Now we turn to the question (ii). Before answering the question, let us first specify our convention for the Berry connection. Given a isolated band with the cell-periodic part of its Bloch state being $|u_{\boldsymbol{k}}\rangle$, the Berry connection is defined as

$$\boldsymbol{A}(\boldsymbol{k}) = -\mathrm{i}\langle u_{\boldsymbol{k}}|\nabla_{\boldsymbol{k}}|u_{\boldsymbol{k}}\rangle\,. \tag{E111}$$

The berry curvature is just the curl of the Berry connection.

With the convention specified, let us answer the question (ii). As discussed in Appendix E 2, $|\text{VH}_0^{\nu=0}\rangle$, $|\text{Chern}_0^{\nu=0}\rangle$, $|\text{half-Chern}_0^{\nu=0}\rangle$, $|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\rangle$ and their symmetry-related states are the $\nu=0$ states that are energetically favored at high $\mathcal{E}$. If we plot the Hartree-Fock band structures of $|\text{VH}_0^{\nu=0}\rangle$, $|\text{Chern}_0^{\nu=0}\rangle$, $|\text{half-Chern}_0^{\nu=0}\rangle$ and $|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\rangle$, we find that their Hartree-Fock band structures remain gapped even for small $\mathcal{E}$. Therefore we are allowed to use determine the Chern numbers of $|\text{VH}_0^{\nu=0}\rangle$, $|\text{Chern}_0^{\nu=0}\rangle$, $|\text{half-Chern}_0^{\nu=0}\rangle$ and $|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\rangle$ at small $\mathcal{E}$ Since the one-shot Hartree-Fock Hamiltonians of $|\text{VH}_0^{\nu=0}\rangle$, $|\text{Chern}_0^{\nu=0}\rangle$, $|\text{half-Chern}_0^{\nu=0}\rangle$ and $|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\rangle$ are related with each other, let us consider the VH state first. Recall that the one-shot Hartree-Fock Hamiltonian of VH state is spin-valley diagonal as shown in Eq. (E78). Moreover, owing to the U(2) × U(2) symmetry and TR symmetry of the VH state, we have

$$H^{\text{VH,OS}}_{+,\downarrow} = H^{\text{VH,OS}}_{+,\uparrow}\big|_{\text{flipping spin}}, \quad H^{\text{VH,OS}}_{-,s} = \overline{\mathcal{T}}H^{\text{VH,OS}}_{+,s}\overline{\mathcal{T}}^{-1}\,, \tag{E112}$$

where $\overline{\mathcal{T}}$ is the spinless TR operation. Thus we only need to study the Chern number of $H^{\text{VH,OS}}_{+,\uparrow}$ at small $\mathcal{E}$.

Since we are considering the small $\mathcal{E}$, we can study the TBG part and the d modes separately. We first determine the Chern number of the TBG part of $H^{\text{VH,OS}}_{+,\uparrow}$, following Ref. [125]. The TBG part of $H^{\text{VH,OS}}_{+,\uparrow}$ has the following matrix form:

$$\begin{pmatrix} -\frac{1}{2}U_1\tau_z & \gamma+v'_\star(p_x\tau_x+p_y\tau_y) & \\ \text{H.c.} & 0\mathbb{1}_{2\times 2} & v(k_x\tau_0+k_y\mathrm{i}\tau_z)\\ & \text{H.c.} & \widetilde{M}\tau_x-\frac{J}{2}\tau_z\end{pmatrix} \tag{E113}$$

in the basis $(f^\dagger_{\boldsymbol{p}}, c^\dagger_{\boldsymbol{p},\Gamma_3}, c^\dagger_{\boldsymbol{p},\Gamma_1\Gamma_2})$, where $f^\dagger_{\boldsymbol{k}} = (\dots, f^\dagger_{\eta,\boldsymbol{k},\alpha,s},\dots)$, $c^\dagger_{\boldsymbol{p},\Gamma_3} = (\dots, c^\dagger_{\eta,\boldsymbol{p},\beta,s},\dots)$ with $\beta=1,2$, and $c^\dagger_{\boldsymbol{p},\Gamma_1\Gamma_2} = (\dots, c^\dagger_{\eta,\boldsymbol{p},\beta,s},\dots)$ with $\beta=3$ and 4. By projecting the f modes to c modes via second order perturbation (which is allowed since f modes have high energies ($\pm U_1/2$) and are topologically trivial), we get

$$\begin{pmatrix} \frac{2}{U_1}\gamma^2\tau_z & v(k_x\tau_0+k_y\mathrm{i}\tau_z)\\ \text{H.c.} & \widetilde{M}\tau_x-\frac{J}{2}\tau_z\end{pmatrix}, \tag{E114}$$

where we have neglected the k^2 term. Since the gap stays open as we tune $(\frac{2}{U_1}\gamma^2, -\frac{J}{2}, M)$ to $(m>0, -m, 0)$ based on Tables I and II, we can determine its Chern number by

$$\begin{pmatrix} m\tau_z & v(k_x\tau_0+k_y\mathrm{i}\tau_z)\\ \text{H.c.} & -m\tau_z\end{pmatrix}, \tag{E115}$$

which gives Ch = 1 owing to $v>0$.

Now we turn to the d modes. Around K_M, $H^{\mathrm{VH,OS}}_{+,\uparrow}$ has the following matrix form:

$$\begin{pmatrix} \frac{1}{2}U_1\sigma_z & M_1(\tau_0+\mathrm{i}\sigma_z)\mathcal{E} \\ M_1(\tau_0-\mathrm{i}\sigma_z)\mathcal{E} & p_x\sigma_x+p_y\sigma_y \end{pmatrix}. \tag{E116}$$

Again, by projecting the f modes to d modes via second order perturbation, we get

$$\frac{4}{U_1}M_1\mathcal{E}^2\sigma_z + p_x\sigma_x + p_y\sigma_y\,, \tag{E117}$$

giving $\mathrm{Ch}=-1/2$. Therefore we have $\mathrm{Ch}=1/2$ for $H^{\mathrm{VH,OS}}_{+,\uparrow}$. Owing to the $\mathrm{U}(2)\times\mathrm{U}(2)$ symmetry and TR symmetry of the VH state, we have

$$\mathrm{Ch}=\tfrac{1}{2}\eta \text{ for } H^{\mathrm{VH,OS}}_{\eta,s}\,. \tag{E118}$$

The Chern number is not an integer for $H^{\mathrm{VH,OS}}_{\eta,s}$ because $H^{\mathrm{VH,OS}}_{\eta,s}$ is built from three Dirac cones of at η valley. If such up both valleys (and include the trivial high-energy completion), we should have well-defined interger Chern numbers. By using the relations between different high-$\mathcal{E}$ states below Eq. (E78), we have

$$\begin{aligned}
&\mathrm{Ch}=\tfrac{1}{2}+\tfrac{1}{2}-\tfrac{1}{2}-\tfrac{1}{2}=0 \quad \text{for } \left|\mathrm{VH}_0^{\nu=0}\right\rangle,\\
&\mathrm{Ch}=\tfrac{1}{2}+\tfrac{1}{2}+\tfrac{1}{2}+\tfrac{1}{2}=2 \quad \text{for } \left|\mathrm{Chern}_0^{\nu=0}\right\rangle,\\
&\mathrm{Ch}=\tfrac{1}{2}+\tfrac{1}{2}+\tfrac{1}{2}-\tfrac{1}{2}=1 \quad \text{for } \left|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\right\rangle,\\
&\mathrm{Ch}=-\tfrac{1}{2}+\tfrac{1}{2}+\tfrac{1}{2}-\tfrac{1}{2}=0 \quad \text{for } \left|C_2\mathcal{T}\text{-invariant}_0^{\nu=0}\right\rangle.
\end{aligned} \tag{E119}$$

If we include the symmetry related states, we have

$$\begin{aligned}
&\mathrm{Ch}=0 \quad \text{for VH states},\\
&\mathrm{Ch}=\pm 2 \quad \text{for Chern states},\\
&\mathrm{Ch}=\pm 1 \quad \text{for half-Chern states},\\
&\mathrm{Ch}=0 \quad \text{for } C_2\mathcal{T}\text{-invariant states}\,.
\end{aligned} \tag{E120}$$

APPENDIX F: LOCAL TR-ODD C_2-even U(2) × U(2)-INVARIANT PERTURBATION ON HIGH-$\mathcal{E}$ STATES AT $\nu = 0$

In this section, we present general symmetry arguments on how local TR-odd and C_2-even perturbations affect high-$\mathcal{E}$ states at $\nu=0$ to the leading order. We also assume the local perturbation to preserve $\mathrm{U}(2)\times\mathrm{U}(2)$ symmetry.

For the VH states, if we keep the tensor-product nature of the states, we have the symmetry rep as

$$|\mathrm{VH}\rangle = (|\mathrm{VH},1\rangle, |\mathrm{VH},2\rangle) \tag{F1}$$

with

$$C_2|\mathrm{VH}\rangle = |\mathrm{VH}\rangle\sigma_x, \quad \mathcal{T}|\mathrm{VH}\rangle = |\mathrm{VH}\rangle, \tag{F2}$$

leading to

$$\langle\mathrm{VH}|H_\delta|\mathrm{VH}\rangle = 0. \tag{F3}$$

For the Chern states, if we keep the tensor-product nature of the states, we have the symmetry rep as

$$|\mathrm{Chern}\rangle = (|\mathrm{Chern},1\rangle, |\mathrm{Chern},2\rangle) \tag{F4}$$

with

$$\begin{aligned}
&C_2|\mathrm{Chern}\rangle = |\mathrm{Chern}\rangle,\\
&\mathcal{T}|\mathrm{Chern}\rangle = |\mathrm{Chern}\rangle\sigma_x\,,
\end{aligned} \tag{F5}$$

leading to

$$\langle\mathrm{Chern}|H_\delta|\mathrm{Chern}\rangle = b\sigma_z. \tag{F6}$$

For the half Chern states, if we keep the tensor-product nature of the states, we have the symmetry rep as

$$\begin{aligned}
|\text{half-Chern}\rangle = (&|\text{half-Chern},1\rangle, |\text{half-Chern},2\rangle,\\
&|\text{half-Chern},3\rangle, |\text{half-Chern},4\rangle)
\end{aligned} \tag{F7}$$

with

$$\begin{aligned}
&C_2|\text{half-Chern}\rangle = |\text{half-Chern}\rangle\tau_x\sigma_0,\\
&\mathcal{T}|\text{half-Chern}\rangle = |\text{half-Chern}\rangle\tau_0\sigma_x\,,
\end{aligned} \tag{F8}$$

leading to

$$\langle\text{half-Chern}|H_\delta|\text{half-Chern}\rangle = b_1\tau_0\sigma_z + b_2\tau_x\sigma_z\,. \tag{F9}$$

For the $C_2\mathcal{T}$-invariant states, if we keep the tensor-product nature of the states, we have the symmetry rep as

$$|C_2\mathcal{T}\text{-invariant}\rangle = (|C_2\mathcal{T}\text{-invariant},1\rangle, |C_2\mathcal{T}\text{-invariant},2\rangle \tag{F10}$$

with

$$\begin{aligned}
&C_2|C_2\mathcal{T}\text{-invariant}\rangle = |C_2\mathcal{T}\text{-invariant}\rangle\sigma_x,\\
&\mathcal{T}|C_2\mathcal{T}\text{-invariant}\rangle = |C_2\mathcal{T}\text{-invariant}\rangle\sigma_x\,,
\end{aligned} \tag{F11}$$

leading to

$$\langle C_2\mathcal{T}\text{-invariant}|H_\delta|C_2\mathcal{T}\text{-invariant}\rangle = 0\,. \tag{F12}$$

Here among the four types of states (i.e., VH, Chern, half-Chern and $C_2\mathcal{T}$), we neglect the mixing between different types of states induced by H_δ, since it is exponentially small due to the local nature of the perturbation H_δ. Moreover, the off-diagonal terms in Eq. (F9) are exponentially small for the same reason. As a result, we see that H_δ can shift the energy of certain Chern states by $-|b|$ energy, favoring it.

[1] R. Bistritzer and A. H. MacDonald, Moiré bands in twisted double-layer graphene, Proc. Natl. Acad. Sci. **108**, 12233 (2011).

[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 (London) **556**, 43 (2018).

[3] 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).

[4] 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 (London) **574**, 653 (2019).

[5] 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 (London) **583**, 375 (2020).

[6] 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).

[7] H. S. Arora, R. Polski, Y. Zhang, A. Thomson, Y. Choi, H. Kim, Z. Lin, I. Z. Wilson, X. Xu, J.-H. Chu *et al.*, Superconductivity in metallic twisted bilayer graphene stabilized by WSe_2, Nature (London) **583**, 379 (2020).

[8] Y. Cao, D. Rodan-Legrain, J. M. Park, N. F. Q. 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).

[9] F. K. de Vries, E. Portolés, G. Zheng, T. Taniguchi, K. Watanabe, T. Ihn, K. Ensslin, and P. Rickhaus, Gate-defined josephson junctions in magic-angle twisted bilayer graphene, Nat. Nanotechnol. **16**, 760 (2021).

[10] 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 (London) **600**240 (2021).

[11] G. D. Battista, P. Seifert, K. Watanabe, T. Taniguchi, K. C. Fong, A. Principi, and D. K. Efetov, Revealing the ultrasensitive calorimetric properties of supercon-ducting magic-angle twisted bilayer graphene, Nano Lett. **22**, 6465 (2022).

[12] H. Tian, S. Che, T. Xu, P. Cheung, K. Watanabe, T. Taniguchi, M. Randeria, F. Zhang, C. N. Lau, and M. W. Bockrath, Evidence for flat band dirac superconductor originating from quantum geometry, arXiv:2112.13401.

[13] Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras *et al.*, Correlated insulator behaviour at half-filling in magic-angle graphene superlattices, Nature (London) **556**, 80 (2018).

[14] 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).

[15] X. Liu, Z. Wang, K. Watanabe, T. Taniguchi, O. Vafek, and J. Li, Tuning electron correlation in magic-angle twisted bilayer graphene using coulomb screening, Science **371**, 1261 (2021).

[16] 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 moiré heterostructure, Science **367**, 900 (2019).

[17] Y. Xie, B. Lian, B. Jäck, 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 (London) **572**, 101 (2019).

[18] Y. Choi, J. Kemmer, Y. Peng, A. Thomson, H. Arora, R. Polski, Y. Zhang, H. Ren, J. Alicea, G. Refael *et al.*, Electronic correlations in twisted bilayer graphene near the magic angle, Nat. Phys. **15**, 1174 (2019).

[19] A. Kerelsky, L. J. McGilly, D. M. Kennes, L. Xian, M. Yankowitz, S. Chen, K. Watanabe, T. Taniguchi, J. Hone, C. Dean *et al.*, Maximized electron interactions at the magic angle in twisted bilayer graphene, Nature (London) **572**, 95 (2019).

[20] 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 (London) **573**, 91 (2019).

[21] H. Polshyn, M. Yankowitz, S. Chen, Y. Zhang, K. Watanabe, T. Taniguchi, C. R. Dean, and A. F. Young, Large linear-in-temperature resistivity in twisted bilayer graphene, Nat. Phys. **15**, 1011 (2019).

[22] Y. Cao, D. Chowdhury, D. Rodan-Legrain, O. Rubies-Bigorda, K. Watanabe, T. Taniguchi, T. Senthil, and P. Jarillo-Herrero, Strange Metal in Magic-Angle Graphene with near Planckian Dissipation, Phys. Rev. Lett. **124**, 076801 (2020).

[23] 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 (London) **582**, 198 (2020).

[24] U. Zondiner, A. Rozen, D. Rodan-Legrain, Y. Cao, R. Queiroz, T. Taniguchi, K. Watanabe, Y. Oreg, F. von Oppen, A. Stern *et al.*, Cascade of phase transitions and dirac revivals in magic-angle graphene, Nature (London) **582**, 203 (2020).

[25] K. P. Nuckolls, M. Oh, D. Wong, B. Lian, K. Watanabe, T. Taniguchi, B. A. Bernevig, and A. Yazdani, Strongly correlated Chern insulators in magic-angle twisted bilayer graphene, Nature (London) **588**, 610 (2020).

[26] Y. Choi, H. Kim, Y. Peng, A. Thomson, C. Lewandowski, R. Polski, Y. Zhang, H. S. Arora, K. Watanabe, T. Taniguchi *et al.*,

Correlation-driven topological phases in magic-angle twisted bilayer graphene, Nature (London) **589**, 536 (2021).

[27] Y. Saito, J. Ge, L. Rademaker, K. Watanabe, T. Taniguchi, D. A. Abanin, and A. F. Young, Hofstadter subband ferromagnetism and symmetry-broken Chern insulators in twisted bilayer graphene, Nat. Phys. **17**, 478 (2021).

[28] I. Das, X. Lu, J. Herzog-Arbeitman, Z.-D. Song, K. Watanabe, T. Taniguchi, B. A. Bernevig, and D. K. Efetov, Symmetry-broken Chern insulators and Rashba-like Landau-level crossings in magic-angle bilayer graphene, Nat. Phys. **17**, 710 (2021).

[29] S. Wu, Z. Zhang, K. Watanabe, T. Taniguchi, and E. Y. Andrei, Chern insulators, van Hove singularities and topological flat bands in magic-angle twisted bilayer graphene, Nat. Mater. **20**, 488 (2021).

[30] J. M. Park, Y. Cao, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, Flavour Hund's coupling, Chern gaps and charge diffusivity in moiré graphene, Nature (London) **592**, 43 (2021).

[31] Y. Saito, F. Yang, J. Ge, X. Liu, T. Taniguchi, K. Watanabe, J. Li, E. Berg, and A. F. Young, Isospin pomeranchuk effect in twisted bilayer graphene, Nature (London) **592**, 220 (2021).

[32] A. Rozen, J. M. Park, U. Zondiner, Y. Cao, D. Rodan-Legrain, T. Taniguchi, K. Watanabe, Y. Oreg, A. Stern, E. Berg *et al.*, Entropic evidence for a pomeranchuk effect in magic-angle graphene, Nature (London) **592**, 214 (2021).

[33] X. Lu, B. Lian, G. Chaudhary, B. A. Piot, G. Romagnoli, K. Watanabe, T. Taniguchi, M. Poggio, A. H. MacDonald, B. A. Bernevig, and D. K. Efetov, Multiple flat bands and topological hofstadter butterfly in twisted bilayer graphene close to the second magic angle, Proc. Natl. Acad. Sci. **118**, e2100006118 (2021).

[34] I. Das, C. Shen, A. Jaoui, J. Herzog-Arbeitman, A. Chew, C.-W. Cho, K. Watanabe, T. Taniguchi, B. A. Piot, B. A. Bernevig, and D. K. Efetov, Observation of Reentrant Correlated Insulators and Interaction-Driven Fermi-Surface Reconstructions at One Magnetic Flux Quantum per Moiré Unit Cell in Magic-Angle Twisted Bilayer Graphene, Phys. Rev. Lett. **128**, 217701 (2022).

[35] 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).

[36] J. Kang and O. Vafek, Symmetry, Maximally Localized Wannier States, and a Low-Energy Model for Twisted Bilayer Graphene Narrow Bands, Phys. Rev. X **8**, 031088 (2018).

[37] J. Kang and O. Vafek, Strong Coupling Phases of Partially Filled Twisted Bilayer Graphene Narrow Bands, Phys. Rev. Lett. **122**, 246401 (2019).

[38] 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).

[39] H. C. Po, L. Zou, T. Senthil, and A. Vishwanath, Faithful tight-binding models and fragile topology of magic-angle bilayer graphene, Phys. Rev. B **99**, 195455 (2019).

[40] O. Vafek and J. Kang, Lattice model for the coulomb interacting Chiral limit of the magic angle twisted bilayer graphene: symmetries, obstructions and excitations, Phys. Rev. B **104**, 075143 (2021).

[41] L. Zou, H. C. Po, A. Vishwanath, and T. Senthil, Band structure of twisted bilayer graphene: Emergent symmetries, commensurate approximants, and Wannier obstructions, Phys. Rev. B **98**, 085435 (2018).

[42] X. Y. Xu, K. T. Law, and P. A. Lee, Kekulé valence bond order in an extended hubbard model on the honeycomb lattice with possible applications to twisted bilayer graphene, Phys. Rev. B **98**, 121406(R) (2018).

[43] N. F. Q. Yuan and L. Fu, Model for the metal-insulator transition in graphene superlattices and beyond, Phys. Rev. B **98**, 045103 (2018).

[44] 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, Phys. Rev. B **103**, 205414 (2021).

[45] 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, Phys. Rev. B **103**, 205415 (2021).

[46] 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, Phys. Rev. B **103**, 205416 (2021).

[47] X. Zhang, G. Pan, Y. Zhang, J. Kang, and Z. Y. Meng, Momentum space quantum monte carlo on twisted bilayer graphene, Chin. Phys. Lett. **38**, 077305 (2021).

[48] 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).

[49] T. Cea and F. Guinea, Band structure and insulating states driven by coulomb interaction in twisted bilayer graphene, Phys. Rev. B **102**, 045107 (2020).

[50] Y. Zhang, K. Jiang, Z. Wang, and F. Zhang, Correlated insulating phases of twisted bilayer graphene at commensurate filling fractions: A Hartree-Fock study, Phys. Rev. B **102**, 035136 (2020).

[51] J. S. Hofmann, E. Khalaf, A. Vishwanath, E. Berg, and J. Y. Lee, Fermionic monte carlo study of a realistic model of twisted bilayer graphene, arXiv:2105.12112.

[52] D. K. Efimkin and A. H. MacDonald, Helical network model for twisted bilayer graphene, Phys. Rev. B **98**, 035404 (2018).

[53] X.-C. Wu, C.-M. Jian, and C. Xu, Coupled-wire description of the correlated physics in twisted bilayer graphene, Phys. Rev. B **99**, 161405 (2019).

[54] C. Xu and L. Balents, Topological Superconductivity in Twisted Multilayer Graphene, Phys. Rev. Lett. **121**, 087001 (2018).

[55] A. Thomson, S. Chatterjee, S. Sachdev, and M. S. Scheurer, Triangular antiferromagnetism on the honeycomb lattice of twisted bilayer graphene, Phys. Rev. B **98**, 075109 (2018).

[56] L. Classen, C. Honerkamp, and M. M. Scherer, Competing phases of interacting electrons on triangular lattices in moiré heterostructures, Phys. Rev. B **99**, 195120 (2019).

[57] P. M. Eugenio and C. B. Dağ, DMRG study of strongly interacting $\mathbb{Z}_2$ flatbands: a toy model inspired by twisted bilayer graphene, SciPost Phys. Core **3**, 015 (2020).

[58] C. Repellin, Z. Dong, Y.-H. Zhang, and T. Senthil, Ferromagnetism in Narrow Bands of Moiré Superlattices, Phys. Rev. Lett. **124**, 187601 (2020).

[59] R. M. Fernandes and J. W. F. Venderbos, Nematicity with a twist: Rotational symmetry breaking in a moiré superlattice, Sci. Adv. **6**, eaba8834 (2020).

[60] G. Tarnopolsky, A. J. Kruchkov, and A. Vishwanath, Origin of Magic Angles in Twisted Bilayer Graphene, Phys. Rev. Lett. **122**, 106405 (2019).

[61] J. Liu, J. Liu, and X. Dai, Pseudo landau level representation of twisted bilayer graphene: Band topology and implications on the correlated insulating phase, Phys. Rev. B **99**, 155415 (2019).

[62] 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).

[63] 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).

[64] B. Padhi, C. Setty, and P. W. Phillips, Doped twisted bilayer graphene near magic angles: proximity to Wigner crystallization, not Mott insulation, Nano Lett. **18**, 6175 (2018).

[65] B. Lian, F. Xie, and B. A. Bernevig, Landau level of fragile topology, Phys. Rev. B **102**, 041402(R) (2020).

[66] K. Hejazi, C. Liu, and L. Balents, Landau levels in twisted bilayer graphene and semiclassical orbits, Phys. Rev. B **100**, 035115 (2019).

[67] B. Padhi, A. Tiwari, T. Neupert, and S. Ryu, Transport across twist angle domains in moiré graphene, Phys. Rev. Res. **2**, 033458 (2020).

[68] M. Ochi, M. Koshino, and K. Kuroki, Possible correlated insulating states in magic-angle twisted bilayer graphene under strongly competing interactions, Phys. Rev. B **98**, 081102(R) (2018).

[69] F. Guinea and N. R. Walet, Electrostatic effects, band distortions, and superconductivity in twisted graphene bilayers, Proc. Natl. Acad. Sci. USA **115**, 13174 (2018).

[70] J. W. F. Venderbos and R. M. Fernandes, Correlations and electronic order in a two-orbital honeycomb lattice model for twisted bilayer graphene, Phys. Rev. B **98**, 245103 (2018).

[71] Y.-Z. You and A. Vishwanath, Superconductivity from valley fluctuations and approximate SO(4) symmetry in a weak coupling theory of twisted bilayer graphene, npj Quantum Mater. **4**, 16 (2019).

[72] F. Wu and S. Das Sarma, Collective Excitations of Quantum Anomalous Hall Ferromagnets in Twisted Bilayer Graphene, Phys. Rev. Lett. **124**, (2020).

[73] B. Lian, Z. Wang, and B. A. Bernevig, Twisted Bilayer Graphene: A Phonon-Driven Superconductor, Phys. Rev. Lett. **122**, 257002 (2019).

[74] F. Wu, A. H. MacDonald, and I. Martin, Theory of Phonon-Mediated Superconductivity in Twisted Bilayer Graphene, Phys. Rev. Lett. **121**, 257001 (2018).

[75] H. Isobe, N. F. Q. Yuan, and L. Fu, Unconventional Superconductivity and Density Waves in Twisted Bilayer Graphene, Phys. Rev. X **8**, 041041 (2018).

[76] C.-C. Liu, L.-D. Zhang, W.-Q. Chen, and F. Yang, Chiral Spin Density Wave and d+ i d Superconductivity in the Magic-Angle-Twisted Bilayer Graphene, Phys. Rev. Lett. **121**, 217001 (2018).

[77] N. Bultinck, S. Chatterjee, and M. P. Zaletel, Mechanism for Anomalous Hall Ferromagnetism in Twisted Bilayer Graphene, Phys. Rev. Lett. **124**, 166601 (2020).

[78] Y.-H. Zhang, D. Mao, Y. Cao, P. Jarillo-Herrero, and T. Senthil, Nearly flat Chern bands in moiré superlattices, Phys. Rev. B **99**, 075127 (2019).

[79] J. Liu, Z. Ma, J. Gao, and X. Dai, Quantum Valley Hall Effect, Orbital Magnetism, and Anomalous Hall Effect in Twisted Multilayer Graphene Systems, Phys. Rev. X **9**, 031021 (2019).

[80] J. F. Dodaro, S. A. Kivelson, Y. Schattner, X.-Q. Sun, and C. Wang, Phases of a phenomenological model of twisted bilayer graphene, Phys. Rev. B **98**, 075154 (2018).

[81] J. Gonzalez and T. Stauber, Kohn-Luttinger Superconductivity in Twisted Bilayer Graphene, Phys. Rev. Lett. **122**, 026801 (2019).

[82] K. Seo, V. N. Kotov, and B. Uchoa, Ferromagnetic Mott state in Twisted Graphene Bilayers at the Magic Angle, Phys. Rev. Lett. **122**, 246402 (2019).

[83] K. Hejazi, X. Chen, and L. Balents, Hybrid Wannier Chern bands in magic angle twisted bilayer graphene and the quantized anomalous Hall effect, Phys. Rev. Res. **3**, 013242 (2021).

[84] E. Khalaf, S. Chatterjee, N. Bultinck, M. P. Zaletel, and A. Vishwanath, Charged skyrmions and topological origin of superconductivity in magic-angle graphene, Sci. Adv. **7**, eabf5299 (2021).

[85] F. Xie, Z. Song, B. Lian, and B. A. Bernevig, Topology-Bounded Superfluid Weight in Twisted Bilayer Graphene, Phys. Rev. Lett. **124**, 167002 (2020).

[86] A. Julku, T. J. Peltonen, L. Liang, T. T. Heikkilä, and P. Törmä, Superfluid weight and Berezinskii-Kosterlitz-Thouless transition temperature of twisted bilayer graphene, Phys. Rev. B **101**, 060505 (2020).

[87] X. Hu, T. Hyart, D. I. Pikulin, and E. Rossi, Geometric and Conventional Contribution to the Superfluid Weight in Twisted Bilayer Graphene, Phys. Rev. Lett. **123**, 237002 (2019).

[88] 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).

[89] T. Soejima, D. E. Parker, N. Bultinck, J. Hauschild, and M. P. Zaletel, Efficient simulation of moiré materials using the density matrix renormalization group, Phys. Rev. B **102**, 205111 (2020).

[90] J. H. Pixley and E. Y. Andrei, Ferromagnetism in magic-angle graphene, Science **365**, 543 (2019).

[91] E. J. König, P. Coleman, and A. M. Tsvelik, Spin magnetometry as a probe of stripe superconductivity in twisted bilayer graphene, Phys. Rev. B **102**, 104514 (2020).

[92] M. Christos, S. Sachdev, and M. S. Scheurer, Superconductivity, correlated insulators, and Wess–Zumino–Witten terms in twisted bilayer graphene, Proc. Natl. Acad. Sci. USA **117**, 29543 (2020).

[93] C. Lewandowski, D. Chowdhury, and J. Ruhman, Pairing in magic-angle twisted bilayer graphene: Role of phonon and plasmon umklapp, Phys. Rev. B **103**, 235401 (2021).

[94] Y. H. Kwan, S. A. Parameswaran, and S. L. Sondhi, Twisted bilayer graphene in a parallel magnetic field, Phys. Rev. B **101**, 205116 (2020).

[95] Y. H. Kwan, Y. Hu, S. H. Simon, and S. A. Parameswaran, Exciton Band Topology in Spontaneous Quantum Anomalous

Hall Insulators: Applications to Twisted Bilayer Graphene, Phys. Rev. Lett. **126**, 137601 (2021).
[96] M. Xie and A. H. MacDonald, Nature of the Correlated Insulator States in Twisted Bilayer Graphene, Phys. Rev. Lett. **124**, 097601 (2020).
[97] 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).
[98] 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).
[99] Y. Da Liao, Z. Y. Meng, and X. Y. Xu, Valence Bond Orders at Charge Neutrality in a Possible Two-Orbital Extended Hubbard Model for Twisted Bilayer Graphene, Phys. Rev. Lett. **123**, 157601 (2019).
[100] Y. Da Liao, J. Kang, C. N. Breiø, X. Y. Xu, H.-Q. Wu, B. M. Andersen, R. M. Fernandes, and Z. Y. Meng, Correlation-Induced Insulating Topological Phases at Charge Neutrality in Twisted Bilayer Graphene, Phys. Rev. X **11**, 011014 (2021).
[101] D. M. Kennes, J. Lischner, and C. Karrasch, Strong correlations and $d + id$ superconductivity in twisted bilayer graphene, Phys. Rev. B **98**, 241407(R) (2018).
[102] Y. Huang, P. Hosur, and H. K. Pal, Quasi-flat-band physics in a two-leg ladder model and its relation to magic-angle twisted bilayer graphene, Phys. Rev. B **102**, 155429 (2020).
[103] T. Huang, L. Zhang, and T. Ma, Antiferromagnetically ordered mott insulator and d+id superconductivity in twisted bilayer graphene: A quantum monte carlo study, Sci. Bull. **64**, 310 (2019).
[104] H. Guo, X. Zhu, S. Feng, and R. T. Scalettar, Pairing symmetry of interacting fermions on a twisted bilayer graphene superlattice, Phys. Rev. B **97**, 235453 (2018).
[105] P. J. Ledwith, G. Tarnopolsky, E. Khalaf, and A. Vishwanath, Fractional Chern insulator states in twisted bilayer graphene: An analytical approach, Phys. Rev. Res. **2**, 023237 (2020).
[106] A. Abouelkomsan, Z. Liu, and E. J. Bergholtz, Particle-Hole Duality, Emergent Fermi Liquids, and Fractional Chern Insulators in Moiré Flatbands, Phys. Rev. Lett. **124**, 106803 (2020).
[107] C. Repellin and T. Senthil, Chern bands of twisted bilayer graphene: Fractional Chern insulators and spin phase transition, Phys. Rev. Res. **2**, 023238 (2020).
[108] O. Vafek and J. Kang, Renormalization Group Study of Hidden Symmetry in Twisted Bilayer Graphene with Coulomb Interactions, Phys. Rev. Lett. **125**, 257602 (2020).
[109] J. H. Wilson, Y. Fu, S. Das Sarma, and J. H. Pixley, Disorder in twisted bilayer graphene, Phys. Rev. Res. **2**, 023325 (2020).
[110] J. Wang, Y. Zheng, A. J. Millis, and J. Cano, Chiral approximation to twisted bilayer graphene: Exact intravalley inversion symmetry, nodal structure, and implications for higher magic angles, Phys. Rev. Res. **3**, 023155 (2021).
[111] B. A. Bernevig, Z.-D. Song, N. Regnault, and B. Lian, Twisted bilayer graphene. I. Matrix elements, approximations, perturbation theory, and a $k \cdot p$ two-band model, Phys. Rev. B **103**, 205411 (2021).
[112] Z.-D. Song, B. Lian, N. Regnault, and B. A. Bernevig, Twisted bilayer graphene. ii. stable symmetry anomaly, Phys. Rev. B **103**, 205412 (2021).
[113] B. A. Bernevig, Z.-D. Song, N. Regnault, and B. Lian, Twisted bilayer graphene. III. Interacting Hamiltonian and exact symmetries, Phys. Rev. B **103**, 205413 (2021).
[114] P. Cha, A. A. Patel, and E.-A. Kim, Strange Metals from Melting Correlated Insulators in Twisted Bilayer Graphene, Phys. Rev. Lett. **127**, 266601 (2021).
[115] D. V. Chichinadze, L. Classen, and A. V. Chubukov, Nematic superconductivity in twisted bilayer graphene, Phys. Rev. B **101**, 224513 (2020).
[116] Y. Sheffer and A. Stern, Chiral magic-angle twisted bilayer graphene in a magnetic field: Landau level correspondence, exact wavefunctions and fractional Chern insulators, Phys. Rev. B **104**, 121405 (2021).
[117] J. Kang, B. A. Bernevig, and O. Vafek, Cascades between light and heavy fermions in the normal state of magic angle twisted bilayer graphene, Phys. Rev. Lett. **127**, 266402 (2021).
[118] Y.-Z. Chou, F. Wu, J. D. Sau, and S. Das Sarma, Correlation-Induced Triplet Pairing Superconductivity in Graphene-Based Moiré Systems, Phys. Rev. Lett. **127**, 217001 (2021).
[119] M. J. Calderón and E. Bascones, Interactions in the 8-orbital model for twisted bilayer graphene, Phys. Rev. B **102**, 155149 (2020).
[120] A. Thomson and J. Alicea, Recovery of massless Dirac fermions at charge neutrality in strongly interacting twisted bilayer graphene with disorder, Phys. Rev. B **103**, 125138 (2021).
[121] J. Yu, M. Xie, F. Wu, and S. D. Sarma, Euler obstructed cooper pairing in twisted bilayer graphene: Nematic nodal superconductivity and bounded superfluid weight, arXiv:2202.02353 [cond-mat.supr-con].
[122] X. Wang, J. Finney, A. L. Sharpe, L. K. Rodenbach, C. L. Hsueh, K. Watanabe, T. Taniguchi, M. A. Kastner, O. Vafek, and D. Goldhaber-Gordon, Unusual magnetotransport in twisted bilayer graphene from strain-induced open Fermi surfaces, arXiv:2209.08204.
[123] F. Xie, J. Kang, B. A. Bernevig, O. Vafek, and N. Regnault, Phase diagram of twisted bilayer graphene at filling factor $\nu = -3$, Phys. Rev. B **107**, 075156 (2023).
[124] X.-F. Zhou, Y.-W. Liu, C.-Y. Hao, C. Yan, Q. Zheng, Y.-N. Ren, Y.-X. Zhao, K. Watanabe, T. Taniguchi, and L. He, Coexistence of reconstructed and unreconstructed structures in structural transition regime of twisted bilayer graphene, Phys. Rev. B **107**, 125410 (2023).
[125] Z.-D. Song and B. A. Bernevig, Magic-Angle Twisted Bilayer Graphene as a Topological Heavy Fermion Problem, Phys. Rev. Lett. **129**, 047601 (2022).
[126] Y.-Z. Chou and S. D. Sarma, Kondo Lattice Model in Magic-Angle Twisted Bilayer Graphene, Phys. Rev. Lett. **131**, 026501 (2023).
[127] H. Hu, B. A. Bernevig, and A. M. Tsvelik, Kondo lattice model of magic-angle twisted-bilayer graphene: Hund's rule, local-moment fluctuations, and low-energy effective theory, arXiv:2301.04669.
[128] H. Hu, G. Rai, L. Crippa, J. Herzog-Arbeitman, D. Călugăru, T. Wehling, G. Sangiovanni, R. Valenti, A. M. Tsvelik, and B. A. Bernevig, Symmetric Kondo lattice states in doped strained twisted bilayer graphene, arXiv:2301.04673.
[129] G.-D. Zhou and Z.-D. Song, Kondo phase in twisted bilayer graphene – A unified theory for distinct experiments, arXiv:2301.04661.

[130] H. Shi and X. Dai, Heavy fermion representation for twisted bilayer graphene systems, Phys. Rev. B **106**, 245129 (2022).
[131] H. Hu and Q. Si, Coupled topological flat and wide bands: Quasiparticle formation and destruction, arXiv:2209.10396.
[132] E. Khalaf, A. J. Kruchkov, G. Tarnopolsky, and A. Vishwanath, Magic angle hierarchy in twisted graphene multilayers, Phys. Rev. B **100**, 085109 (2019).
[133] X. Li, F. Wu, and A. H. MacDonald, Electronic structure of single-twist trilayer graphene, arXiv:1907.12338.
[134] Z. Zhu, P. Cazeaux, M. Luskin, and E. Kaxiras, Modeling mechanical relaxation in incommensurate trilayer van der waals heterostructures, Phys. Rev. B **101**, 224107 (2020).
[135] S. Carr, C. Li, Z. Zhu, E. Kaxiras, S. Sachdev, and A. Kruchkov, Ultraheavy and ultrarelativistic dirac quasiparticles in sandwiched graphenes, Nano Lett. **20**, 3030 (2020).
[136] G. A. Tritsaris, S. Carr, Z. Zhu, Y. Xie, S. B. Torrisi, J. Tang, M. Mattheakis, D. T. Larson, and E. Kaxiras, Electronic structure calculations of twisted multi-layer graphene superlattices, 2D Mater. **7**, 035028 (2020).
[137] G. A. Tritsaris, Y. Xie, A. M. Rush, S. Carr, M. Mattheakis, and E. Kaxiras, Lan: A materials notation for two-dimensional layered assemblies, J. Chem. Inf. Model. **60**, 3457 (2020).
[138] A. Lopez-Bezanilla and J. L. Lado, Electrical band flattening, valley flux, and superconductivity in twisted trilayer graphene, Phys. Rev. Res. **2**, 033357 (2020).
[139] Z. Zhu, S. Carr, D. Massatt, M. Luskin, and E. Kaxiras, Twisted Trilayer Graphene: A Precisely Tunable Platform for Correlated Electrons, Phys. Rev. Lett. **125**, 116404 (2020).
[140] C. Lei, L. Linhart, W. Qin, F. Libisch, and A. H. MacDonald, Mirror symmetry breaking and lateral stacking shifts in twisted trilayer graphene, Phys. Rev. B **104**, 035139 (2021).
[141] Z. Wu, Z. Zhan, and S. Yuan, Lattice relaxation, mirror symmetry and magnetic field effects on ultraflat bands in twisted trilayer graphene, Sci. China Phys. Mech. Astron. **64**, 267811 (2021).
[142] D. Călugăru, F. Xie, Z.-D. Song, B. Lian, N. Regnault, and B. A. Bernevig, Twisted symmetric trilayer graphene: Single-particle and many-body hamiltonians and hidden nonlocal symmetries of trilayer moiré systems with and without displacement field, Phys. Rev. B **103**, 195411 (2021).
[143] J. Gonzalez and T. Stauber, *p*-wave superconductivity induced from valley symmetry breaking in twisted trilayer graphene, arXiv:2110.11294.
[144] Y. W. Choi and H. J. Choi, Dichotomy of Electron-Phonon Coupling in Graphene Moiré Flat Bands, Phys. Rev. Lett. **127**, 167001 (2021).
[145] J. Shin, B. L. Chittari, and J. Jung, Stacking and gate-tunable topological flat bands, gaps, and anisotropic strip patterns in twisted trilayer graphene, Phys. Rev. B **104**, 045413 (2021).
[146] A. Fischer, Z. A. H. Goodwin, A. A. Mostofi, J. Lischner, D. M. Kennes, and L. Klebl, Unconventional superconductivity in magic-angle twisted trilayer graphene, npj Quantum Mater. **7**, 5 (2022).
[147] E. Lake and T. Senthil, Reentrant superconductivity through a quantum lifshitz transition in twisted trilayer graphene, Phys. Rev. B **104**, 174505 (2021).
[148] W. Qin and A. H. MacDonald, In-Plane Critical Magnetic Fields in Magic-Angle Twisted Trilayer Graphene, Phys. Rev. Lett. **127**, 097001 (2021).
[149] F. Xie, N. Regnault, D. Călugăru, B. A. Bernevig, and B. Lian, Twisted symmetric trilayer graphene. ii. projected Hartree-Fock study, Phys. Rev. B **104**, 115167 (2021).
[150] D. Guerci, P. Simon, and C. Mora, Higher-order van hove singularity in magic-angle twisted trilayer graphene, Phys. Rev. Res. **4**, L012013 (2022).
[151] V. o. T. Phong, P. A. Pantaleón, T. Cea, and F. Guinea, Band structure and superconductivity in twisted trilayer graphene, Phys. Rev. B **104**, L121116 (2021).
[152] M. Christos, S. Sachdev, and M. S. Scheurer, Correlated Insulators, Semimetals, and Superconductivity in Twisted Trilayer Graphene, Phys. Rev. X **12**, 021018 (2022).
[153] Z. Wu, X. Kuang, Z. Zhan, and S. Yuan, Magic angle and plasmon mode engineering in twisted trilayer graphene with pressure, Phys. Rev. B **104**, 205104 (2021).
[154] S. Li, G. Zheng, and J. Huang, Induced superconductivity in magic-angle twisted trilayer graphene through graphene-metal contacts, arXiv:2111.04451.
[155] B. Xie, R. Peng, S. Zhang, and J. Liu, Alternating twisted multilayer graphene: generic partition rules, double flat bands, and orbital magnetoelectric effect, npj Comput. Mater. **8**, 110 (2022).
[156] P. J. Ledwith, E. Khalaf, Z. Zhu, S. Carr, E. Kaxiras, and A. Vishwanath, Tb or not tb? Contrasting properties of twisted bilayer graphene and the alternating twist n-layer structures ($n = 3, 4, 5, \ldots$), arXiv:2111.11060.
[157] H. D. Scammell, J. Li, and M. S. Scheurer, Theory of zero-field superconducting diode effect in twisted trilayer graphene, 2D Mater. **9**, 025027 (2022).
[158] L. Classen, J. H. Pixley, and E. J. König, Interaction-induced velocity renormalization in magic-angle twisted multilayer graphene, 2D Mater. **9**, 031001 (2022).
[159] X. Lin, C. Li, K. Su, and J. Ni, Energetic stability and spatial inhomogeneity in the local electronic structure of relaxed twisted trilayer graphene, Phys. Rev. B **106**, 075423 (2020).
[160] R. Samajdar, Y. Teng, and M. S. Scheurer, Moiré phonons and impact of electronic symmetry breaking in twisted trilayer graphene, Phys. Rev. B **106**, L201403 (2022).
[161] N. Leconte, Y. Park, J. An, A. Samudrala, and J. Jung, Electronic structure of lattice relaxed alternating twist tng-multilayer graphene: from few layers to bulk at-graphite, arXiv:2206.09412.
[162] Y. Li, S. Zhang, F. Chen, L. Wei, Z. Zhang, H. Xiao, H. Gao, M. Chen, S. Liang, D. Pei, L. Xu, K. Watanabe, T. Taniguchi, L. Yang, F. Miao, J. Liu, B. Cheng, M. Wang, Y. Chen, and Z. Liu, Observation of coexisting dirac bands and moiré flat bands in magic-angle twisted trilayer graphene, Adv. Mater. **34**, 2205996 (2022).
[163] K. Shin, Y. Jang, J. Shin, J. Jung, and H. Min, Electronic structure of biased alternating twist multilayer graphene, Phys. Rev. B **107**, 245139 (2023).
[164] N. J. Zhang, Y. Wang, K. Watanabe, T. Taniguchi, O. Vafek, and J. I. A. Li, Electronic anisotropy in magic-angle twisted trilayer graphene, arXiv:2211.01352.
[165] P. J. Ledwith, E. Khalaf, and A. Vishwanath, Strong coupling theory of magic-angle graphene: A pedagogical introduction, Ann. Phys. **435**, 168646 (2021).

[166] X.-F. Li, R.-X. Sun, S.-Y. Wang, X. Li, Z.-B. Liu, and J.-G. Tian, Recent advances in moiré superlattice structures of twisted bilayer and multilayer graphene, Chin. Phys. Lett. **39**, 037301 (2022).

[167] J. M. Park, Y. Cao, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, Tunable strongly coupled superconductivity in magic-angle twisted trilayer graphene, Nature (London) **590**, 249 (2021).

[168] Z. Hao, A. M. Zimmerman, P. Ledwith, E. Khalaf, D. H. Najafabadi, K. Watanabe, T. Taniguchi, A. Vishwanath, and P. Kim, Electric field–tunable superconductivity in alternating-twist magic-angle trilayer graphene, Science **371**, 1133 (2021).

[169] Y. Cao, J. M. Park, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, Pauli-limit violation and re-entrant superconductivity in moiré graphene, Nature (London) **595**, 526 (2021).

[170] S. Turkel, J. Swann, Z. Zhu, M. Christos, K. Watanabe, T. Taniguchi, S. Sachdev, M. S. Scheurer, E. Kaxiras, C. R. Dean, and A. N. Pasupathy, Orderly disorder in magic-angle twisted trilayer graphene, Science **376**, 193 (2022).

[171] H. Kim, Y. Choi, C. Lewandowski, A. Thomson, Y. Zhang, R. Polski, K. Watanabe, T. Taniguchi, J. Alicea, and S. Nadj-Perge, Evidence for unconventional superconductivity in twisted trilayer graphene, Nature (London) **606**, 494 (2022).

[172] X. Liu, N. J. Zhang, K. Watanabe, T. Taniguchi, and J. I. A. Li, Isospin order in superconducting magic-angle twisted trilayer graphene, Nat. Phys. **18**, 522 (2022).

[173] S.-J. Yang, J.-H. Jung, E. Lee, E. Han, M.-Y. Choi, D. Jung, S. Choi, J.-H. Park, D. Oh, S. Noh, K.-J. Kim, P. Y. Huang, C.-C. Hwang, and C.-J. Kim, Wafer-scale programmed assembly of one-atom-thick crystals, Nano Lett. **22**, 1518 (2022).

[174] Y. Zhang, R. Polski, C. Lewandowski, A. Thomson, Y. Peng, Y. Choi, H. Kim, K. Watanabe, T. Taniguchi, J. Alicea, F. von Oppen, G. Refael, and S. Nadj-Perge, Ascendance of superconductivity in magic-angle graphene multilayers, arXiv:2112.09270.

[175] C. Shen, P. J. Ledwith, K. Watanabe, T. Taniguchi, E. Khalaf, A. Vishwanath, and D. K. Efetov, Dirac cone spectroscopy of strongly correlated phases in twisted trilayer graphene, Nat. Mater. **22**, 316 (2023).

[176] A. Ramires and J. L. Lado, Emulating Heavy Fermions in Twisted Trilayer Graphene, Phys. Rev. Lett. **127**, 026401 (2021).

[177] Bilbao crystallographic server, www.cryst.ehu.es.

[178] R. Winkler, S. Papadakis, E. De Poortere, and M. Shayegan, *Spin-Orbit Coupling in Two-Dimensional Electron and Hole Systems* (Springer, Berlin, Heidelberg, 2003), Vol. 41, pp. 211–223.