Editors' Suggestion

Moiré fractional Chern insulators. I. First-principles calculations and continuum models of twisted bilayer $MoTe_2$

Yujin Jia,[1,2,*] Jiabin Yu,[3,*] Jiaxuan Liu,[1,2,*] Jonah Herzog-Arbeitman,[3] Ziyue Qi,[1,2] Hanqi Pi,[1,2] Nicolas Regnault,[3,4] Hongming Weng,[1,2,5] B. Andrei Bernevig,[3,6,7,†] and Quansheng Wu[1,2,‡]

[1]*Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*

[2]*University of Chinese Academy of Sciences, Beijing 100049, China*

[3]*Department of Physics, Princeton University, Princeton, New Jersey 08544, USA*

[4]*Laboratoire de Physique de l'Ecole normale supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris-Diderot, Sorbonne Paris Cité, 75005 Paris, France*

[5]*Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China*

[6]*Donostia International Physics Center, P. Manuel de Lardizabal 4, 20018 Donostia-San Sebastian, Spain*

[7]*IKERBASQUE, Basque Foundation for Science, Bilbao, Spain*

(Received 19 November 2023; revised 20 February 2024; accepted 22 February 2024; published 7 May 2024)

Recent experiments observed fractional Chern insulators (FCI) in twisted bilayer $MoTe_2$ at zero magnetic field, yet even the single-particle model of this material is controversial, leading to unreliable predictions of the experimental phase diagram as discussed in [Yu *et al.*, Phys. Rev. B **109**, 045147 (2024)]. In this light, we revisit the single-particle model of twisted bilayer $MoTe_2$. Utilizing large-scale density functional theory, we calculate the band structure of twisted AA-stacked bilayer $MoTe_2$ at various twist angles relevant to experiment. We find that a band inversion occurs near 4.41° between the second and third bands in one valley. Our *ab initio* band structure is in qualitative agreement with [Wang *et al.*, Phys. Rev. Lett. **132**, 036501 (2024)], but shows important differences in the remote bands and in the Γ valley. We incorporate two higher harmonic terms into the continuum model to capture the highest three valence bands per valley. We confirm that the two highest valence bands per valley have opposite Chern numbers with $|C| = 1$ for experimentally relevant angles, and also use our model to predict a variety of Chern states in the remote bands accessible by displacement field. We also perform DFT calculations and build models for the AB-stacking configuration. Our paper serves as a foundation for accurate determination of the correlated phases in twisted bilayer $MoTe_2$.

DOI: 10.1103/PhysRevB.109.205121

I. INTRODUCTION

Fractional Chern insulators (FCI) [1–3] were shown to naturally appear in zero magnetic field when nearly flat Chern bands [4,5] are fractionally filled. Over the last several years, there have been extensive theoretical [6–33] and experimental [34–38] studies on zero-field FCIs in moiré materials [39,40], as well as studies of fractional quantum Hall (FQH)-like states under nonzero external magnetic field [41] (which were called FCIs because the Chern bands where they appeared did not have flat Berry curvature and quantum geometry), and small B-field-induced FCIs [42]. Remarkably, FCIs without any external magnetic fields have recently been observed in twisted bilayer $MoTe_2$ ($tMoTe_2$) [34–37] and in the pentalayer-graphene/hBN moiré superlattice [38]. These *moiré fractional Chern insulator* (mFCI) states provide the best experimental platform to date for the physics proposed in Refs. [1–3].

In this paper, we will focus on $tMoTe_2$, where mFCIs are observed at fractional fillings $\nu = -2/3, -3/5$, as well as a Chern insulator (CI) at $\nu = -1$. Throughout, ν is the electron filling measured from the charge neutrality point. FCIs are expected to appear [1–3] in the CI bands found in the $tMoTe_2$ model first proposed in Ref. [43]. However, an accurate model of the material is required to correctly reproduce the FCI states and competing nontopological states. Otherwise serious disagreement with experiment can occur, as has been somewhat overlooked in the flurry of recent literature. For example, the spin polarization at $\nu = -1/3, -4/3$ cannot be reproduced theoretically unless band mixing is taken into account [32], showing that bands beyond the lowest valence manifold play an integral part in the physics. Recent theoretical studies [20,21,23,24,27,29,30,32,33,44], some of which predicts the existence of fractional states not seen in the experiment, might then be subject to change when band mixing is properly included. Even interaction-driven band mixing beyond the two-band Hilbert space considered in Ref. [32] could be important if the single-particle bands are energetically close. Hence it is of the utmost importance to obtain a continuum model capturing the *ab initio* band structure over the range of energies accessible by the Coulomb potential.

Most importantly, there are two different sets of *ab initio* parameters Ref. [20,21] used [20,21,23,24,27,29,30,32,33,44] for interacting calculations in the continuum model of Ref. [43]. These sets of parameters give rise to different

*These authors contributed equally to this work.
†bernevig@princeton.edu
‡quansheng.wu@iphy.ac.cn

 ©2024 American Physical Society

single-particle phase diagrams, many-body spin polarizations, and stabilities of 2/3 and 1/3 mFCI states Ref. [32]. Settling the single-particle model of $t\text{MoTe}_2$ is the subject of this first paper in the mFCI series.

We perform large-scale density functional theory (DFT) calculations to study twisted bilayer MoTe_2 at various twist angles for AA stacking, which is the stacking configuration relevant to the experiments [34–37]. To accurately capture the impact of crystal structure relaxation, we rigorously test 19 different van der Waals (vdW) exchange-correlation functionals, and determine that DFT-D2 yields the most reliable lattice parameters when compared with experimental results. With DFT-D2, we further use a highly efficient two-step method to obtain the structure. Our first step is to use the DFT-D2 functional to train a machine learning algorithm that generates relaxed structures for a set of twist angles. The second step is to use these relaxed structures as the initial configurations, and further perform the full DFT relaxation to obtain final, accurate relaxed structures. The moiré bands are eventually calculated with the final relaxed structures. Our DFT results show that the ±K-valley valence bands dominate the low-energy physics while the Γ-valley valence bands are about 80 meV away from the valence band maximum (VBM) for the experimentally relevant angle 3.89°. Furthermore, the second and the third valence bands in K valley (or −K valley related by time-reversal symmetry) undergo a gap closing around 4.41°, changing the Chern number of the second valence band from −1 at smaller angles to 1 at larger angles. Our *ab initio* results for the ±K-valley bands are closer to those of Ref. [21] than to those of Ref. [20], although our second and third band per valley are closer to each other than those of Ref. [21] at 3.89°, which shifts their gap closing to smaller angles. Our Γ-valley band is also lower in energy than that of Ref. [21], and hence will not contribute to many-body physics near $\nu = -1$. The band structure in Ref. [20] looks different from ours due to the different selection of the vdW functionals.

We then use the moiré model to capture the low-energy DFT bands. If we only keep the first harmonics (FH) as in Ref. [43], we can only manage to capture the top two valence bands in each valley. To capture the top three valence bands in each valley, we add two extra second harmonic (SH) terms, and obtain a good match with the DFT band structure and symmetry representations (reps). We propose that this more accurate model be used in many-body calculations. In our model with SH terms, the gap closing between the second and third top valence bands in one valley happens around 4.2° at zero displacement field, which is consistent with our DFT calculation. We find that adding a displacement field can change the Chern number of the top valence band from 1 to 0 in K valley, and can also achieve a variety of Chern numbers (from –2 to 2) for the second and third top valence bands in one valley. Accessing these bands provides another route to integer Chern physics seen in twisted transition metal dichalcogenides [45–49].

In addition to the experimentally relevant AA-stacking configuration, we also study the AB-stacking configuration. Our DFT results show that the Γ-valley bands in the AB-stacking case is closer to the VBM (only about 30 meV away) and are extremely flat. We build a FH model capturing these Γ-valley bands, and show that they are extremely localized atomic bands whose flatness comes from zero hopping among atomic orbitals on the triangular lattice. We also used the ±K-valley model in Ref. [43] to match the top two valence bands in each valley.

In the rest of this paper, we discuss the DFT calculations at a range of twist angles and stacking configurations in Sec. II, and the continuum models we employ to faithfully reproduce these calculations in Sec. III. We conclude the paper in Sec. IV, and provide more details in a series of appendices.

II. DFT CALCULATIONS

In this section, we discuss the large-scale DFT calculations on the $t\text{MoTe}_2$ at various twist angles and different stacking configurations. We will mainly discuss the results for 3.89°, which is the closest one to the structure in recent experiments [34–37], as well as the topological phase transition from large angle to 3.89°. A complete discussion can be found in Appendix A.

A. 3.89° $t\text{MoTe}_2$

Utilizing the coincidence lattice method [50,51], we construct twisted bilayer crystal structures of MoTe_2 at various commensurate angles: 13.2°, 9.43°, 7.34°, 5.09°, 4.41°, 3.89°, and 3.48°, considering both AA and AB stacking. Here, AA (AB) stacking implies that, without any twist, the Mo/Te atoms of the top layer respectively align with the Mo/Te (Te/Mo) atoms of the bottom layer. Subsequently, large-scale DFT calculations using the Vienna *Ab Initio* Simulation Package (VASP) [52–55] are performed on these structures.

To capture the van der Waals (vdW) interactions between top and bottom layers, we test 19 exchange-correlation functionals using the experimental bulk crystal structure [56] as a benchmark (see Table III in Appendix A 1). Ultimately, we find that DFT-D2 gives the lattice parameter closest to the experimental value [56], and we use this functional throughout the paper. Secondly, different pseudopotential (PP) combinations have also been tested. It has been found that the energy differences between different PPs are negligible (see Fig. 7 of Appendix A 1). Considering the computational cost, we choose the PAW pseudopotential and PBE exchange-correlation functional.

Based on these functionals, we develop a highly efficient two-step relaxation scheme, which combines machine learning and DFT. First, we construct the machine-learned force field (MLFF) using the relaxation data generated by DFT-D2. The MLFF method has a much lower computational cost than the direct relaxation with DFT since the moiré unit cells are very large with 1302 atoms. MLFF can produce a relaxed structure quickly. Using this structure as an initial guess, we further perform the full DFT relaxation. It turns out that the MLFF is reasonably good, making the DFT relaxation quite fast. To achieve total-force convergence with an accuracy of 5m eV/Å and energy convergence of 1×10^{-5} eV, the DFT+MLFF method requires 17 ionic steps, which take 7.5 hours using 16 NVIDIA A100 GPUs. In contrast, direct relaxation using DFT requires 178 ionic steps and take 55 hours when using eight NVIDIA A100 GPUs, demanding

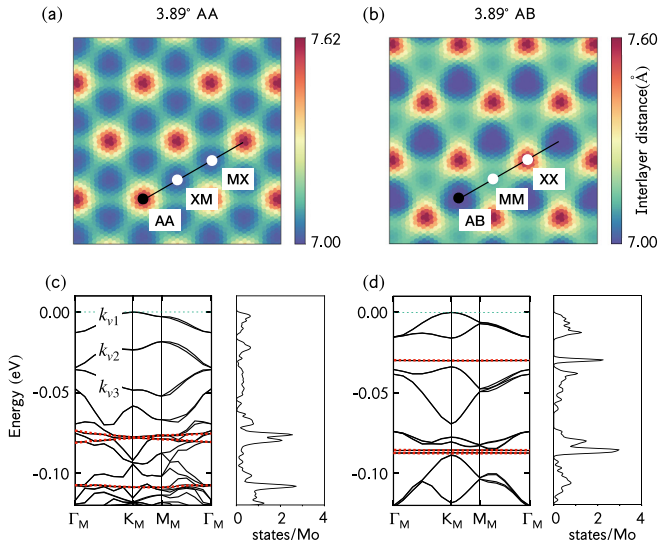

FIG. 1. Relaxation and band structure of 3.89° tMoTe$_2$. (a) Relaxation of the 3.89° AA stacking. (b) Relaxation of the 3.89° AB stacking. (c) and (d) show the band structure and density of states (DOS) of the structures in (a) and (b) respectively, with red lines indicating bands from the Γ valley and the green line marking the Fermi level.The notations k_{v1}, k_{v2}, and k_{v3} represent the three highest pairs of valence bands of the AA-stacked configuration in the ±K valleys. The DOS is normalized per molybdenum atom.

approximately four times more resources than the DFT+MLFF method.

The faster convergence of the DFT+MLFF method (compared to the DFT direct relaxation) does not sacrifice precision in the band structure. To show it, we present the band structures obtained from three different approaches: MLFF relaxation, DFT+MLFF relaxation, and DFT direct relaxation, as shown in Fig. 17 of Appendix A 3 d. The band structures obtained from the DFT+MLFF and DFT relaxed methods have qualitatively the same shapes, which is consistent with the fact that the DFT relaxation upon the MLFF structures is quite fast. In particular, the bands from the DFT+MLFF and the direct DFT relaxation methods are extremely similar, verifying the validity of the DFT+MLFF method. Moreover, the MLFF-initialized relaxation leads to a more stable configuration, likely due to the maintenance of C_{3z} symmetry throughout the relaxation process. MLFF+DFT relaxation results in an approximately 0.9 eV lower energy compared to direct DFT relaxation, indicating the advantage of MLFF initial guess of relaxed structure in finding the lowest energy configurations. Therefore, we use the DFT+MLFF method in this paper instead of the DFT direct relaxation. The stacking-dependent corrugated moiré structures of AA and AB configuration generated by the DFT+MLFF method are shown in Figs. 1(a) and 1(b).

AA-stacked tMoTe$_2$ has a twofold rotational symmetry axis along y axis, C_{2y}, and the C_3 symmetry. The relaxed structure exhibits a maximum interlayer distance $d = 7.62$ Å in the AA region, where the metal atoms in top layer is aligned with metal atoms in bottom layer, while a minimum interlayer distance $d = 7.0$ Å in the MX (XM) region, where the top layer metal (chalcogen) atoms are aligned with chalcogen (metal) atoms of bottom layer, as shown in Fig. 6 of Appendix A 1.

For AA tMoTe$_2$ at 3.89 degree, the VBM is located at the K_M point in the moiré Brillouin zone (BZ), which is folded from the K point in the untwisted bilayer structure. The top three pairs of valence bands, labeled as k_{v1}, k_{v2}, and k_{v3}, originate from the ±K valleys, exhibiting bandwidths of 12.8 meV, 16.2 meV, and 16.5 meV, respectively. The combined effects of lattice relaxation and SOC lead to the Γ-valley bands shifting downward by about 80 meV from the VBM. The Γ-valley bands, illustrated in Fig. 1(c) and marked by red-dashed lines, contribute to two distinct peaks in the density of states (DOS).

The AB-stacking configuration has a twofold rotational symmetry C_{2x} with the axis along x as well as the threefold rotation symmetry C_3 with axis along z. For the twist angle 3.89°, the top two pairs of valence bands also come from ±K valley with band width of 16 meV, while the first pair of ultra-flat bands from Γ valley is only 30 meV below the VBM. The charge density of these ultra-flat bands is highly localized in the AB region of the moiré lattice (see Fig. 20 of Appendix A 3 d).

B. Topological phase of valence bands from 5.09° to 3.89°

In this part, we will discuss the topological phase transition around 4.41° of the valence bands in AA-stacked tMoTe$_2$. Since the full DFT calculations are performed in huge unit cells, it is impractical to calculate the Chern number directly, and we turn to symmetry eigenvalues to efficiently deduce the topology [57].

Figure 2 shows the band structure and the C_3 eigenvalues at the high-symmetry points Γ_M and M_M, as calculated using the IRVSP software Ref. [58]. Our convention is to use the spinful C_3 eigenvalues, which are labeled by $\omega = e^{i\pi/3}$, $\omega^* = e^{-i\pi/3}$, and $\bar{1} = e^{-i\pi}$. Since ±K-valley bands are valley-spin locked, we distinguish two valleys by their spins (spin up/down for K/−K valley).

For 5.09°, we observe that there is a gap of 2.8 meV between k_{v2} and k_{v3} bands at the moiré Γ_M point. This gap is closed around 4.41°, and reopens at angle 3.89°. From the C_3 eigenvalues labeled in Fig. 2, we see that this band crossing exchanges the symmetry representations at Γ_M for k_{v2} and k_{v3} bands. Explicitly, for 5.09°, the k_{v2} band at Γ_M has two spin-polarized states: $|\uparrow, \omega^*\rangle$ and $|\downarrow, \omega\rangle$, while the k_{v3} band has $|\uparrow, \omega\rangle$ and $|\downarrow, \omega^*\rangle$. When energy gap reopening occurs, the symmetry representations of k_{v2} and k_{v3} at 3.89° switch in comparison to those at 5.09°.

The exchange of symmetry eigenvalues is proof of a band inversion and causes a change in the spin/valley Chern number. Specifically, recall that the Chern number can be determined from symmetry via $e^{i\frac{2\pi}{3}C} = -\xi_{\Gamma_M}\xi_{K_M}\xi_{K'_M}$ where $\xi_{\mathbf{k}}$ is the spinful C_3 eigenvalue at high-symmetry point $\mathbf{k}$ [57]. Since $C_{2y}\mathcal{T}$ relates the moiré K_M and K'_M points and is antiunitary, $\xi_{K_M} = \xi_{K'_M}$. Thus for θ slightly larger than 4.41°, we find $e^{i\frac{2\pi}{3}C} = -(-1)(-1)\omega^*$ so that $C = 1$ mod 3 for the second top spin-↑ band, but for θ slightly smaller than 4.41°, we find $e^{i\frac{2\pi}{3}C} = -(-1)(-1)\omega$ so that $C = -1$ mod 3. This topological phase transition is matched by the continuum model as we will show in Sec. III.

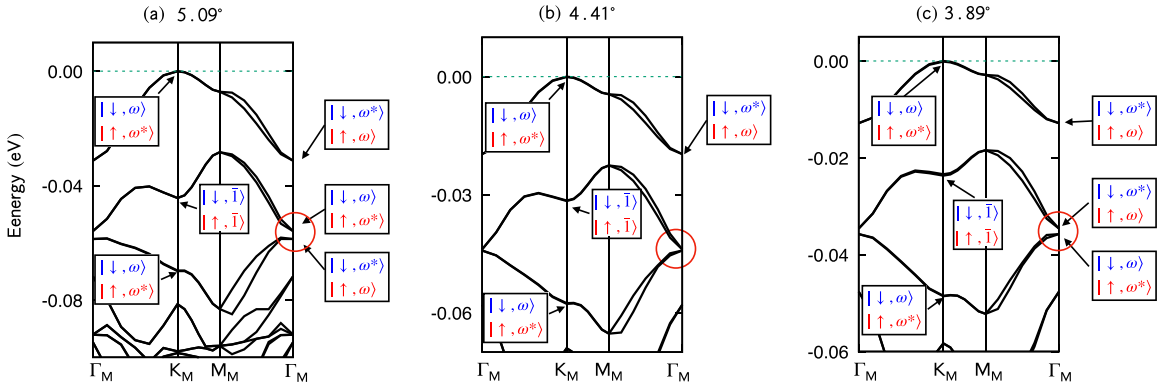

FIG. 2. Irreducible representations at high-symmetry points Γ and K of valence bands of monolayer and twist $MoTe_2$ with twist angle 5.09°, 4.41°, and 3.89° for relaxed AA-stacking configuration with SOC. The eigenvalues of the C_3 are denoted as $\omega = e^{i\pi/3}$, $\omega^* = e^{-i\pi/3}$, and $\bar{1} = e^{-i\pi}$. The critical point of band inversion between the second and third valence bands at γ point of moire BZ happens at around twist angle 4.41°.

C. Comparison with Ref. [20,21]

Two recent papers [20,21] have also studied the relaxation and band structure of tMoTe$_2$. The band structures in Ref. [20,21] are different from our result, mainly due to the different relaxed structures. Reference [20] employed the SCAN density functional with dDsC dispersion correction to perform crystal structure relaxation using VASP, while Ref. [21] used SIESTA with DFT-D2 functional to perform the relaxation. The comparison of the relaxation results of AA-tMoTe$_2$ between this paper and Refs. [20,21] is shown in Fig. 3. As shown in Fig. 3, the interlayer distance in our relaxed structure has the qualitatively the same shape as those in Refs. [20,21]—largest interlayer distance at MM and the smallest interlayer distance at MX/XM. However, the interlayer distance in our relaxed structure has smaller spatial fluctuations than that in Ref. [20] as shown in Figs. 3(a) and 3(b), while our relaxed structure has larger interlayer distance than that of Ref. [21] [Fig. 3(c)]. The maximum interlayer distances of this paper, Ref. [20], and Ref. [21] are about 7.6Å, 7.8 Å, and 7.4 Å, respectively, and the minimum interlayer distances are respectively about 7.0Å, 7.0Å, and 6.9Å.

Our relaxation result is consistent with the AA and AB stacking untwisted bilayer structure. In the AA region, the stacking configuration is close to that of AA untwisted bilayer structure, and thus the maximum interlayer distance should be close to but slightly smaller than (due to the corrugation effect due to the connection to other stacking configurations in the moiré structure) the interlayer distance of AA-stacking untwisted bilayer (7.7 Å), which is consistent with our results but not with Ref. [20]. Furthermore, the MX region has the stacking configuration akin to that of AB untwisted bilayer structure. As a result, the smallest interlayer distance in tMoTe$_2$ structures should be close to but slightly larger than (due to the corrugation) 7 Å, which is consistent with our results but not with Ref. [21]. More details about the relaxation and its influence on band structure are discussed in Appendix A 3 e.

III. CONTINUUM MODELS

In this section, we use the continuum model to fit the DFT results. Up to now, FCI states were only found for hole doping experimentally [34–37]; thus, we will focus on the model for

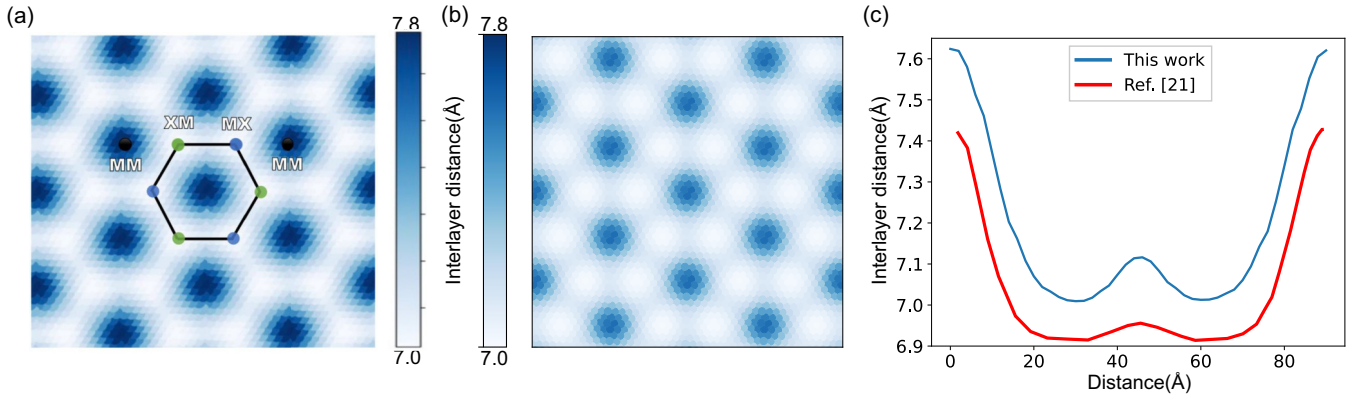

FIG. 3. Relaxation results of AA twisted structure. (a) Relaxation result of 4.41° AA $MoTe_2$ from Fig. 2(a) of Ref. [20]. (b) Our relaxation result of 4.41° AA $MoTe_2$. (c) The interlayer distance of 3.89° AA $MoTe_2$ along the black line in Fig. 1(a). The red curve is the result from Fig. 1(d) from Ref. [21], and the blue curve is our result.

the valence bands. We will first discuss the AA stacking and then discuss the AB stacking.

A. AA-Stacking

According to the DFT results, the low-energy valence bands mainly originate from the ±K valleys in the monolayer. The symmetry group of AA-stacking $t\mathrm{MoTe}_2$ (AA-$t\mathrm{MoTe}_2$) is generated by C_3, C_{2y}, and $\mathcal{T}$, in addition to the moiré lattice translations [43]. We pick a convention where the top layer is rotated by $-\theta/2$ and the bottom layer by $\theta/2$. The moiré lattice constant is

$$a_M = \frac{a_0}{2\sin\left(\frac{\theta}{2}\right)}, \tag{1}$$

where $a_0 = 3.52$ Å is the lattice constant of monolayer MoTe_2.

The continuum model for the monolayer ±K valleys in general reads

$$H^{AA}_{\eta,0} = \int d^2r (c^\dagger_{\eta,b,\boldsymbol{r}}, c^\dagger_{\eta,t,\boldsymbol{r}}) h^{AA}_{\eta,0}(\boldsymbol{r}) \begin{pmatrix} c_{\eta,b,\boldsymbol{r}} \\ c_{\eta,t,\boldsymbol{r}} \end{pmatrix}, \tag{2}$$

where $c^\dagger_{\eta,l,\mathbf{r}}$ labels the basis of the continuum model, $\eta = \pm$ labels the ±K valleys (or equivalently spins), $l = t, b$ labels the layer, and $\boldsymbol{r}$ labels the position. Reference [43] proposed a model with only the first harmonics (FH); however, here we add the certain second harmonics (SH) terms in order to accurately match the higher bands. [See the definition of the first and second Harmonics in Eqs. (B17) and (B20) of Appendix B 1.] Expanding the potential terms and using symmetry, we find the form

$$\begin{aligned} V_{\eta,l}(\boldsymbol{r}) &= Ve^{-(-)^l i\psi} \sum_{i=1,2,3} e^{i\boldsymbol{g}_i\cdot\boldsymbol{r}} + Ve^{(-)^l i\psi} \sum_{i=1,2,3} e^{-i\boldsymbol{g}_i\cdot\boldsymbol{r}} \\ &\quad + 2V_2 \sum_{i=1}^{3} \cos(\boldsymbol{g}_{2i}\cdot\boldsymbol{r}), \\ t_\eta(\boldsymbol{r}) &= w \sum_{i=1,2,3} e^{-\eta i\boldsymbol{q}_i\cdot\boldsymbol{r}} + w_2 \sum_{i=1,2,3} e^{-\eta i\boldsymbol{q}_{2i}\cdot\boldsymbol{r}}, \end{aligned} \tag{3}$$

where $l = t, b$ corresponds to $l = 0, 1$ for $(-)^l$, respectively, $\boldsymbol{g}_i = C_3^{i-1}\boldsymbol{b}_{M,1}$, $\boldsymbol{b}_{M,1} = \frac{4\pi}{\sqrt{3}a_M}(1,0)^T$ and $\boldsymbol{b}_{M,2} = \frac{4\pi}{\sqrt{3}a_M}(\frac{1}{2}, \frac{\sqrt{3}}{2})^T$ are the basis moiré reciprocal lattice vectors, $\boldsymbol{q}_1 = \frac{4\pi}{3a_0} 2\sin(\frac{\theta}{2})(0,1)^T$, $\boldsymbol{q}_2 = C_3\boldsymbol{q}_1$, $\boldsymbol{q}_3 = C_3^2\boldsymbol{q}_1$, $\boldsymbol{g}_{21} = \boldsymbol{b}_{M,1} + \boldsymbol{b}_{M,2}$, $\boldsymbol{g}_{2i} = C_3^{i-1}\boldsymbol{g}_{21}$, $\boldsymbol{q}_{21} = \boldsymbol{b}_{M,1} + \boldsymbol{q}_1$, and $\boldsymbol{q}_{2i} = C_3^{i-1}\boldsymbol{q}_{21}$. V, ψ, and w characterize the FH terms, while V_2 and w_2 belong to the SH. The AA-stacking FH ±K-valley model [Eq. (2)] has effective inversion symmetry that makes the two bands from the two valleys identical (in accord with the DFT results, which show small splitting about 1.2 meV). The effective inversion symmetry is natural with only FH terms, and we only include the SH terms that preserve the effective inversion symmetry. In total, the model [Eq. (2)] has six real parameters m^*, V, ψ, w, V_2, and w_2.

We fit to the DFT band structure at $\theta = 3.89°$ in two ways. (See Appendix B 1 for details.) First, we set $V_2 = w_2 = 0$, which corresponds to the FH model. In this case, we manage to fit the top four valence bands (two in each valley) with the corresponding FH parameters in Table I, as shown in Fig. 4(a).

TABLE I. Values of the parameters in the ±K-valley continuum model [Eq. (3)] for the AA-stacking $t\mathrm{MoTe}_2$. V, w, V_2, w_2 are in meV. "FH" means we only include the first harmonics, whereas "FH+SH" means that we include both the first harmonics and the effective-inversion-invariant second harmonics.

Model	m^* (m_e)	V	ψ (deg)	w	V_2	w_2
FH	0.60	16.5	–105.9	–18.8	0	0
FH+SH	0.62	7.94	–88.43	–10.77	20.00	10.21

Then, we allow nonzero V_2 and w_2, i.e., adding the SH terms. We are now able to fit the top six valence bands (three in each valley) with the corresponding FH+SH parameters in Table I, as shown in Fig. 4(b). The match is not only good along the high-symmetry line but also good in the full BZ as shown in Figs. 21(a) and 21(b) in Appendix B 3.

As shown in Figs. 22(a) and 22(b) in Appendix B 3, we can see that the C_3 eigenvalues for the top six valence bands match the DFT calculation in both FH and FH+SH cases. Furthermore, the Chern numbers of the top three bands (in decreasing order of energy) in K valley are (1,–1,0) in both cases, which are consistent with the C_3 eigenvalues [57]. The C_3 eigenvalues and the Chern numbers of the top two bands per valley are the same as those in Ref. [21]. It is clear that adding the SH terms improves the reliability of the model across a wider range of energies. We expect more remote bands to be accessible in future experiments, and hence our FH+SH model is an essential improvement.

At last, we discuss the evolution of the bands of the AA-stacking moiré model in Eq. (2) with the FH+SH parameters values in Table I as a function of the twist angle θ and the displacement field ε. We will focus on the K valley. As shown in Fig. 5(a), the gap between second top and third top bands closes around 4.2° for zero-displacement field, which is close the the DFT's 4.41° in Fig. 2; the gap closing will change the Chern numbers of the second and third top bands in K valley from (–1,0) to (1,–2). Further increasing the angle at zero displacement field will cause a band inversion between the third and fourth top valence bands, which changes the Chern number of the third top band from –2 to –1 as shown in Fig. 5(d). Figure 5(b) shows that increasing the displacement field can trivialize the top valence band, while a variety of Chern numbers (ranging from –2 to 2) can arise for nonzero displacement field for the second and third top bands as shown in Figs. 5(c) and 5(d).

B. AB-Stacking

The generators of the symmetry group of AB-stacking $t\mathrm{MoTe}_2$ (AB-$t\mathrm{MoTe}_2$), which can be thought of as twisting the top layer of AA-$t\mathrm{MoTe}_2$ by another 180°, are C_3, C_{2x}, and $\mathcal{T}$. Note that C_{2x} is local to the monolayer K point (unlike C_{2y} in AA-$t\mathrm{MoTe}_2$), and thus it preserves the valley quantum number in the moiré model. This difference in the valley symmetry group is important and, as we now show, leads to different behavior with the potential for interesting many-body spin and Hubbard physics [59–65].

The DFT results show that the low-energy valence bands of AB-$t\mathrm{MoTe}_2$ come from both the ±K valleys and the Γ

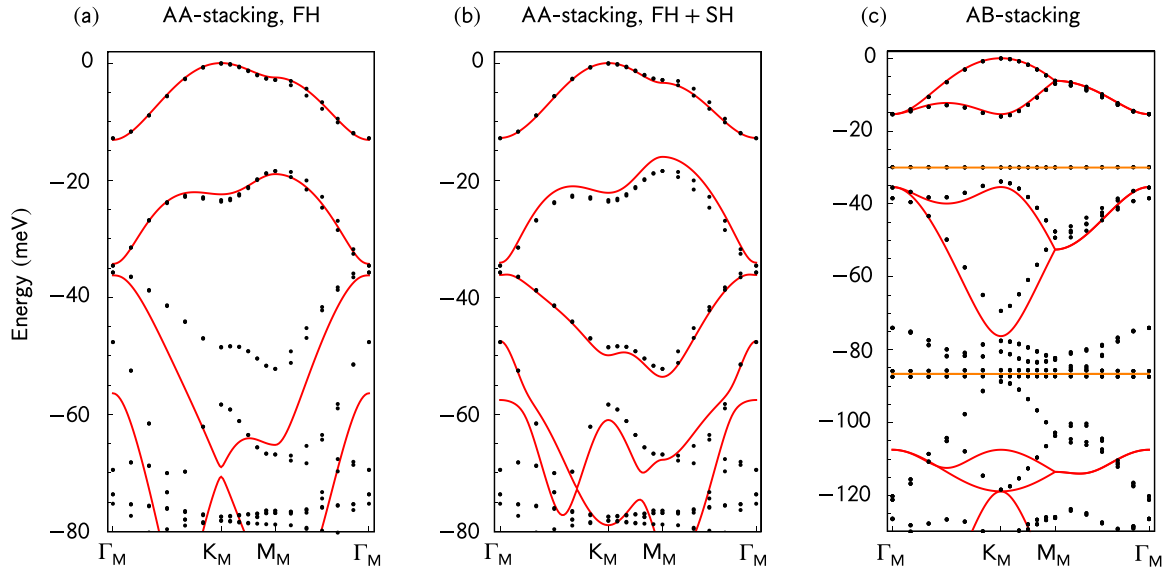

FIG. 4. The comparison between the valence bands from the DFT calculation (black dots) and from the continuum model (red/orange line) for twist angle 3.89° for (a) (b) AA stacking and (c) AB stacking along the high-symmetry line. The red line comes from the ±K-valley model, while the orange line comes from the Γ-valley model. Each red line is doubly degenerate [except along $\Gamma_M - M_M$ in (c) where it is fourfold degenerate]. The orange line in (c) has double degeneracy around −30 meV and fourfold degeneracy around −86 meV. The fitting in (a) is done with only FH terms, while SH terms are added in (b).

valleys in the monolayer. The ±K-valley model was proposed in Ref. [43], which reads

$$H^{AB}_{\eta,0} = \sum_l \int d^2r c^\dagger_{\eta,l,\boldsymbol{r}} \left[\frac{\hbar^2\nabla^2}{2m^*} + V_{\eta,l}(\boldsymbol{r}) + (-)^l \frac{\varepsilon}{2} \right] c_{\eta,l,\boldsymbol{r}} + \left[\int d^2r c^\dagger_{\eta,b,\boldsymbol{r}} t_\eta(\boldsymbol{r}) c_{\eta,t,\boldsymbol{r}} + \text{H.c.} \right], \tag{4}$$

where $c^\dagger_{\eta,l,\boldsymbol{r}}$ labels the basis of the continuum model with $\eta = \pm$ labeling the ±K valleys (or equivalently spins), $l = t, b$ labels the layer, and $\boldsymbol{r}$ labels the position. Since AB-tMoTe$_2$ is given by rotating the top layer of AA-tMoTe$_2$ by an extra 180°, the two layers in K or −K valley now have opposite spin. With first harmonics, the forms of $V_{\eta,l}(\boldsymbol{r})$ and $t_\eta(\boldsymbol{r})$ derived from the symmetries read

$$V_{\eta,l}(\boldsymbol{r}) = Ve^{-i\psi} \sum_{i=1,2,3} e^{i\boldsymbol{g}_i\cdot\boldsymbol{r}} + Ve^{i\psi} \sum_{i=1,2,3} e^{-i\boldsymbol{g}_i\cdot\boldsymbol{r}},$$
$$t_\eta(\boldsymbol{r}) = w \sum_{i=1,2,3} e^{\eta i(i-1)\frac{2\pi}{3}} e^{-\eta i\boldsymbol{q}_i\cdot\boldsymbol{r}}, \tag{5}$$

where V is real, and w can made real by choosing the relative phase between the two layers. For $\varepsilon = 0$, the minimal continuum model has the effective TR symmetry within each valley that flips layer (or equivalently spin), making the bands from the two valleys identical.

On the other hand, the Γ-valley model has the following form:

$$H_\Gamma = \int d^2r (\psi^\dagger_{\boldsymbol{r},b} \;\; \psi^\dagger_{\boldsymbol{r},t}) h^{AB}_\Gamma(\boldsymbol{r}) \begin{pmatrix} \psi_{\boldsymbol{r},b} \\ \psi_{\boldsymbol{r},t} \end{pmatrix}, \tag{6}$$

where we choose the kinetic term in the continuum model as the intralayer spin-independent ∇^2 term,

$$h^{AB}_\Gamma(\boldsymbol{r}) = \frac{\hbar^2\nabla^2}{2m^*_\Gamma} + E_\Gamma + \begin{pmatrix} V_{\Gamma,b}(\boldsymbol{r}) - \epsilon/2 & t_\Gamma(\boldsymbol{r}) \\ t^\dagger_\Gamma(\boldsymbol{r}) & V_{\Gamma,t}(\boldsymbol{r}) + \epsilon/2 \end{pmatrix}, \tag{7}$$

$$\psi^\dagger_{\boldsymbol{r},l} = (\psi^\dagger_{\boldsymbol{r},l,\uparrow}, \psi^\dagger_{\boldsymbol{r},l,\downarrow}), \tag{8}$$

t and b correspond to the top and bottom layers, respectively, E_Γ accounts for the energy difference between the Γ-valley and ±K-valley bands, and $V_{\Gamma,l}(\boldsymbol{r})$ and $t_\Gamma(\boldsymbol{r})$ are 2×2 matrix functions. For zero-displacement field, the Γ-valley model has effective TR symmetry within each spin subspace, which makes the bands from the two spins identical (resulting in at least double degeneracy of each band).

We fit the low-energy bands at $\theta = 3.89°$, and the resulting parameter values are summarized in Table II. As the illustration, we show the good match between the DFT bands and those from the models along the high-symmetry line in Fig. 4(b). The bands match well also in the full BZ, and the C_3 eigenvalues also match the DFT calculation as discussed in Appendix C 3.

For the ±K-valley model, we notice that the interlayer coupling can be set to zero $w = 0$ in the ±K-valley model while keeping the match of the bands good, which can be understood as the follows. The two layers in one valley now have opposite spins; owing to the spin U(1) symmetry for the low-energy states near ±K valleys in monolayer MoTe$_2$, we expect the spin $U(1)$ symmetry is approximately preserved in tMoTe$_2$, which means the interlayer coupling is very small for the AB stacking. The zero interlayer coupling makes the eigenstates have well-defined valley and layer/spin. As a result, the two states with the same spin in the ±K-valley model are degenerate at K_M, since the combination of the effective TR symmetry and the TR symmetry leaves the spin

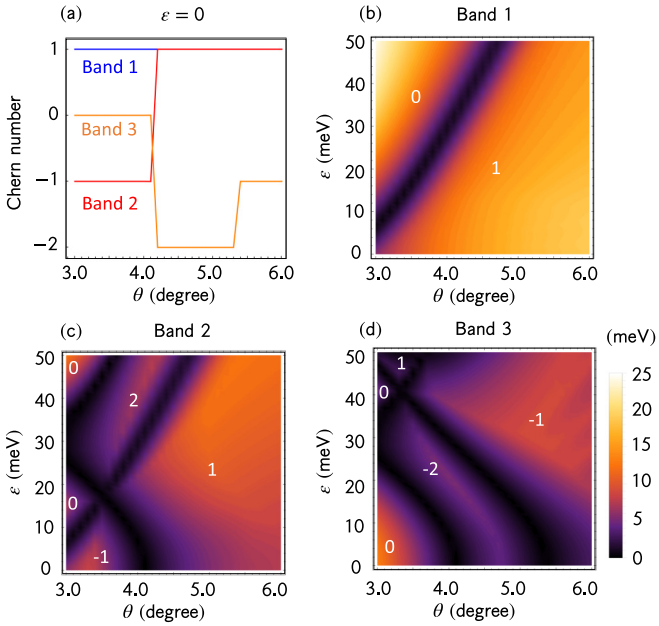

FIG. 5. Phase diagram of AA-stacking moiré model in Eq. (2) with the FH+SH parameters values in Table I. In this plot, band 1,2,3 refer to the top, second top and third top valence bands in the K valley. In (a), we show the Chern numbers of the three bands (blue for 1, red for 2 and orange for 3) for zero-displacement field, which shows a band inversion between the second and third top bands around 4.2°. In (b), (c), and (d), we show the direct gaps and the Chern numbers of the top (a), second top (b). and third top (c) bands in the ±K valley. Explicitly, the direct gaps are illustrated by the color (the color bar is the same across plots) and the Chern numbers are labeled in white.

invariant; this is consistent with DFT calculations as shown in Fig. 19 of Appendix C 3. The top four valence bands are in the $A_1@1a$ atomic limit, representing an atomic s orbital at the MM stacking positions; the band structure can be thought of as arising from hopping on this moiré triangular lattice.

The Γ-valley bands are extremely flat as shown in Fig. 4(c), owing to the large $m_\Gamma^* = 10m_e$ ($-\frac{\hbar^2|\boldsymbol{q}_1|^2}{2m_\Gamma^*} = -2.49$ meV) compared to the potential $V_\Gamma = 72$ meV. As a result, those flat bands are extremely localized atomic states localized around the minima of the intralayer potential (i.e., 1a positions due to $\psi = 0$), and their energies can be approximately calculated from an array of decoupled harmonic oscillators related by moiré translations. (See Appendix C 3.)

At last, we discuss the effect of the displacement field. As shown in Fig. 25 in Appendix C 3, the effect of the displacement field on the ±K-valley bands is just to shift the bands from different layers relative to each other, since the layer is a good quantum number due to the zero interlayer coupling (see Table II). On the other hand, the effect of the displacement field on the low-energy Γ-valley bands is negligible, which is consistent with the fact that the very large interlayer coupling makes the eigenstates equally distributed between the two layers (see Table II).

TABLE II. Values of the parameters in the ±K-valley continuum model in Eq. (5) (first and second rows) and the Γ-valley continuum model in Eq. (7) (third and fourth rows) for the AB-stacking tMoTe$_2$.

m^* (m_e)	V (meV)	ψ (deg)	w (meV)
0.62	53	−56	0
m_Γ^* (m_e)	V_Γ (meV)	ψ_Γ (deg)	w_Γ (meV)
10	72	0	300

IV. CONCLUSIONS

Our extensive DFT study, accelerated by machine learning, has confirmed (along with our continuum model analysis) that that the lowest two bands in the K valley have Chern numbers +1 and −1 respectively, at a twist angle of 3.89° for the experimentally relevant AA stacking. A phase transition to K valley have Chern numbers +1 and +1 occurs at slightly larger angle 4.41°. These Chern numbers mod 3 are accessible directly from the DFT data, and we compute the exact value from the continuum model fit to the DFT bands. The Chern numbers are consistent with Ref. [21] but do not agree with those computed by Ref. [20] at larger twist angles and extrapolated by us to 3.89°, although the phase boundaries are relatively close.

We succeeded in matching the top three valence bands (as well as much of the fourth band) per valley by adding only two higher harmonic terms to the moiré Hamiltonian [see Eq. (2)]. This Hamiltonian still preserves the effective intravalley inversion symmetry, and will serve as a faithful model of the dispersion and topology for a wide range of electron fillings. The matching of the three bands (rather than the previous matching of two bands) might be necessary as the band mixing turns out to be an important characteristic of these systems. Excitingly, our model predicts a rich topological phase diagram accessible through displacement fields in the remote bands. Our forthcoming paper will study the many-body physics of this model, contributing to the broader study of correlations and topology [49,66–81] now accessible in experiment.

Note: It's worth mentioning our other work [82] that performs first-principles calculations and builds continuum models of rhombohedral graphene/h-BN moiré superlattices. Focusing on the pentalayer case, we analytically explain the robust $|C| = 0, 5$ Chern numbers seen in the low-energy single-particle bands and their flattening with a displacement field. We then predict nonzero valley Chern numbers at the $\nu = 4, 0$ insulators observed in experiments.

ACKNOWLEDGMENTS

The authors thank Allan H. MacDonald, Kin-Fai Mak, Jie Shan, Oskar Vafek, Zhijun Wang, and Xiaodong Xu for helpful discussions, with special thanks to D. Xiao, Y. Zhang, and Jianpeng Liu for their helpful cross-checking of the details in the DFT calculations. J.Y. and J. H.-A. are grateful for conversations with Pok Man Tam. This work was supported by the National Key Research and Development Program of China (Grants No. 2023YFA1607400, No. 2022YFA1403800), the Science Center of the National Natural Science Foundation of China (Grant No. 12188101) and the National Natural Science Foundation of China (Grant No. 12274436).

H.W. acknowledge support from the Informatization Plan of the Chinese Academy of Sciences (CASWX2021SF-0102) and the New Cornerstone Science Foundation through the XPLORER PRIZE. B.A.B.'s work was primarily supported by the DOE Grant No. DE-SC0016239 and the Simons Investigator Grant No. 404513. N.R. also acknowledges support from the QuantERA II Programme that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 101017733 and from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 101020833). J.H.-A. is supported by a Hertz Fellowship, with additional support from DOE Grant No. DE-SC0016239 by the Gordon and Betty Moore Foundation through Grant No. GBMF8685 towards the Princeton Theory Program, the Gordon and Betty Moore Foundation's EPiQS Initiative (Grant No. GBMF11070), Office of Naval Research (ONR Grant No. N00014-20-1-2303), BSF Israel US Foundation No. 2018226 and NSF-MERSEC DMR-2011750. J.Y. is supported by the Gordon and Betty Moore Foundation through Grant No. GBMF8685 towards the Princeton Theory Program.

APPENDIX A: DFT RESULTS

1. Atomic structures of tMoTe$_2$

Bulk 2H-MoTe$_2$, as sketched in Fig. 6(b), has a hexagonal structure with space group 194 generated by the inversion $\mathcal{P}$, a threefold symmetry C_3 with axis along z axis, a twofold symmetry C_{2x} with axis along x axis, and lattice translations. Experimentally, the lattice constant is found to be a = 3.519 Å and c = 13.976 Å [56].

We used VASP with Perdew-Burke-Ernzerhof (PBE) functionals to perform the DFT calculation. Before performing DFT calculations for band structure, we need to choose the suitable exchange-correlation functionals to describe the electron-electron interaction and the suitable pseudopotentials to deal with the interactions between electrons and nucleus. To pick the best exchange-correlation functionals, we try 19 different ones to calculate the relaxed bulk crystal structure, and show the results in Table III. Here, we choose the energy cutoff of 300 eV and energy convergence condition of less than 1×10^{-6} eV in the self consistent calculation. The MoTe$_2$ system is an insulator, so the smearing parameter for VASP should be chosen as ISMEAR = 0 to use Gaussian smearing. We find that the DFT-D2 functional (IVDW = 10) provides a lattice constant of $a = b = 3.518$ Å and $c = 13.976$ Å for the bulk MoTe$_2$ primitive crystal structure, which is closest to the experimental values in Ref. [56]. Therefore, we conclude that the DFT-D2 functional (IVDW = 10) is the most appropriate choice for the MoTe$_2$ structural relaxation among the 19 functionals.

After selecting the exchange-correlation functionals, we test the different projector augmented wave pseudopotentials—Mo-Te, Mo_pv-Te, Mo_sv-Te—where the pseudopotentials without suffices are general pseudopotential, and the suffix "_sv" ("_pv") means that the inner s (p) electrons are considered as valence electrons. These combinations show little difference in the relaxed lattice parameters. This is consistent with experimental data. In addition, we also test different pseudopotentials by calculating the band structure of monolayer, bulk, AB-stacking bilayer and AA-stacking bilayer MoTe$_2$, as shown in Fig. 7. The band structures calculated by different pseudopotentials coincides with each other. Therefore, it is legitimate to choose any of them. Besides, we also test these pseudopotentials by doing the relaxation of bulk MoTe$_2$, see in Table IV. The "Mo-Te" combination gives the most closest lattice parameters to the experimental result, meaning that the "Mo-Te" combination, which containing six electronics for Mo atoms and Te atoms, can capture the relaxation effect of bulk MoTe$_2$ well. Considering the balance of accuracy and computational cost,

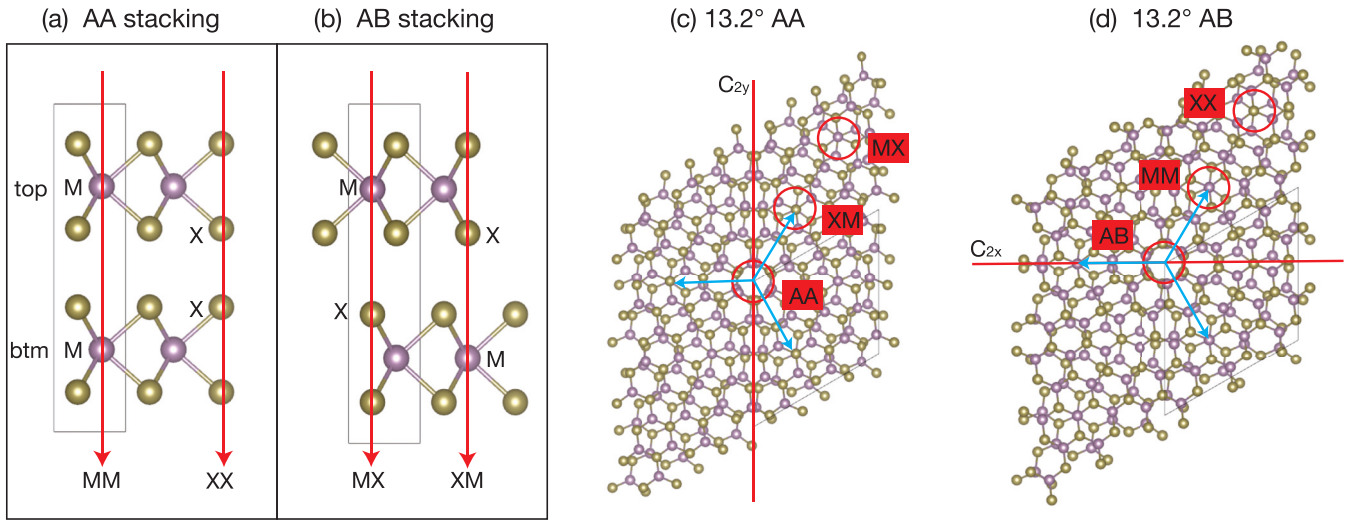

FIG. 6. Rigid atomic structures of 13.2° tMoTe$_2$. The purple one is Mo atom and yellow-brown one is Te atom. (a) Bilayer AA-stacking configuration. (b) Bilayer AB-stacking configuration. (c) 13.2° tMoTe$_2$ AA configuration with in-plane twofold rotational symmetry axis along the y axis. (b) 13.2° tMoTe$_2$ AB configuration with in-plane twofold rotational symmetry axis along the x axis. Both AA and AB have the C_3 threefold rotation symmetry axis perpendicular to the plane. Mo atom is labeled by M, while chalcogen atom is labeled by X. MX represents the top layer Mo atom is aligned with bottom layer chalcogen atom. Similarly with XM and MM.

TABLE III. Relaxed lattice constant of bulk $MoTe_2$ using different vdW functionals. The IVDW number corresponds to different vdW functionals provided in VASP. Here IVDW=10 is the DFT-D2 method of Grimme (marked by gray). IVDW=11 is the DFT-D3 method of Grimme with zero-damping function. IVDW=13 is the DFT-D4 method. IVDW=20 is the Tkatchenko-Scheffler method. IVDW=21 is the Tkatchenko-Scheffler method with iterative Hirshfeld partitioning. IVDW=263 is the Many-body dispersion energy with fractionally ionic model for polarizability method. IVDW=4 is the dDsC dispersion correction method. IVDW=3 is the DFT-ulg method. Experimental data is from [56].

	Experiment	optB86	optB88	vdW-DF	vdW-DF-cx
a,b(Å)	3.519	3.527	3.567	3.631	3.502
c(Å)	13.964	14.032	14.213	15.007	13.867
	PBE	optPBE-vdW	rVV10	SCAN+rVV10	r^2SCAN+rVV10
a,b(Å)	3.551	3.580	3.546	3.503	3.542
c(Å)	15.095	14.475	13.949	14.223	14.187
	rev-vdW-DF2	vdW-DF2	IVDW=10	IVDW=11	IVDW=12
a,b(Å)	3.529	3.711	3.519	3.512	3.490
c(Å)	14.028	14.891	13.976	13.984	13.649
	IVDW=20	IVDW=21	IVDW=263	IVDW=4	IVDW=3
a,b(Å)	3.514	3.516	3.490	3.515	3.531
c(Å)	13.923	13.826	13.709	14.047	14.222

here we choose the "Mo-Te" combination, which has the smaller computational cost due to fewer electrons while has the better accuracy.

Before studying the electronic structures of the $t\text{MoTe}_2$ system, we first generated their crystal structures. We start from both the AA and AB stacking [83]. As illustrated in Fig. 6, AA stacking means that, when there is no twist, the Mo/Te atoms of the top layer aligns with the Mo/Te atoms of the bottom layer. AB stacking means that, when there is no twist, the Mo/Te atoms of the top layer is directly above the Te/Mo atoms of the bottom layer [see Fig. 6(a)]. The bilayer untwisted crystal structure relaxed by DFT-D2 functional gives lattice parameter $a = 3.5228$ Å, and thus two primitive lattice vectors are $a_1 = a(1, 0, 0)$ and $a_2 = \frac{a}{2}(-\frac{1}{2}, \frac{\sqrt{3}}{2}, 0)$. For twisted homobilayer with both top and bottom layer being $MoTe_2$, a commensurate structure occurs when the moiré lattice vector of top layer and bottom layer satisfy the commensurate lattice condition $\boldsymbol{a}_{Mb} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 = m_1\boldsymbol{a}_1^r + m_2\boldsymbol{a}_2^r = \boldsymbol{a}_{Mt}$ for certain integers n_1, n_2, m_1, m_2, where $\boldsymbol{a}_1^r$ is the primitive lattice vectors rotated by an angle θ. In this way, we can obtain the rigid $t\text{MoTe}_2$ structures with different angles. All the rigid structures were generated using a homemade software 2DTwist. In the DFT calculation for $t\text{MoTe}_2$, when the twist angle is smaller than 9.43°, we use Γ point sampling. While for 13.2° $t\text{MoTe}_2$, we use $4 \times 4 \times 1$ k-point sampling.

The moiré structures twisted from AA stacking has a hexagonal structure in space group 150, with C_3, C_{2y} [see Fig. 6(a)] as well as the time-reversal (TR) symmetry. Thus, the little group in the K valley is generated by C_3, $C_{2y}\mathcal{T}$, and $\pm$K valleys are exchanged by C_{2y} and $\mathcal{T}$. In contrast, Fig. 6(b) shows the moiré structures twisted from AB stacking, which has a hexagonal structure of space group 149, with C_3, C_{2x}, and $\mathcal{T}$. For the AB configuration, the little group in the K valley is generated by C_3 and C_{2x}, and the two valleys are exchanged by $\mathcal{T}$.

As the twist angle becomes smaller but nonzero, a U(1) valley symmetry emerges due to the exponential suppression of intervalley scattering off the moiré potential. This will enable us to build continuum models around the monolayer $\pm$K points using the little group symmetries to constrain the low-order terms.

TABLE IV. Relaxed lattice constant of bulk $MoTe_2$ using different pseudopotentials. Three pseudopotentials combinations show little difference. Experimental data is from [56].

	Experiment	Mo-Te	Mo_pv-Te	Mo_sv-Te
a,b(Å)	3.519	3.519	3.521	3.523
c(Å)	13.964	13.976	13.985	13.997

2. Relaxation of $t\text{MoTe}_2$

Since the relaxation will greatly affect the band structure, it is necessary to perform relaxation on $t\text{MoTe}_2$. However, there are 1302 atoms for 3.89° and 1626 atoms for 3.48° in the moire cell of twisted structures, making the relaxation process difficult to converge. We decided to construct a Machine Learning Force Field (MLFF) to obtain relaxed structures in an efficient way (see below). MLFF is a machine learning algorithm that will "learn" energies and forces of atoms from *ab initio* calculation and can be applied to predict forces and energies for similar systems. We note that MLFF is not a relaxed structure—it is a function that maps structures to forces/energies; it can be used to efficiently generate the relaxed structures. We firstly constructed a MLFF and applied it to obtain an MLFF-relaxed structure. Then, we performed DFT relaxation on the MLFF-relaxed structures. It only took around 20 DFT steps in relaxing the largest moire structure to converge in this strategy. In comparison, the relaxation from rigid 3.89° AA-$t\text{MoTe}_2$ takes 178 steps to converge.

a. Construction of Machine Learning Force field

During the construction of MLFF, two software packages are used. One is the VASP together with its integrated MLFF module [84], and the second is NequIP [85], which is an MLFF built on an E(3) equivariant neural network.

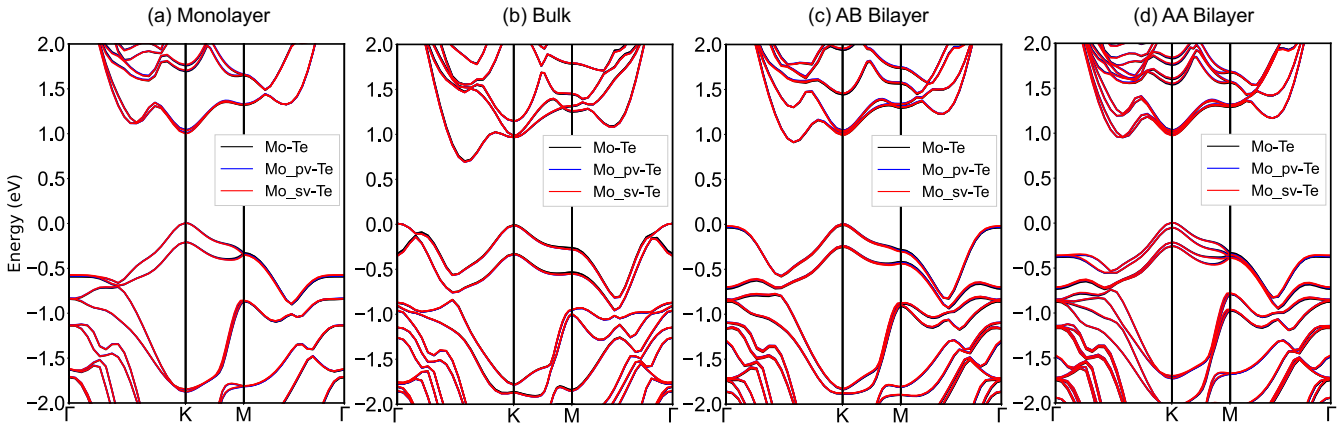

FIG. 7. Band structure of bulk, monolayer, AA bilayer, and AB bilayer $MoTe_2$ calculated with SOC and different pseudopotentials. The crystal structures are relaxed with corresponding pseudopotentials and IVDW = 10(DFT-D2) functional.

We use the VASP MLFF module to generate the *ab initio* data needed for training a high-precision MLFF. The VASP MLFF module itself is a way to generate MLFF and accelerate the molecular dynamics (MD) simulation. It firstly runs several *ab initio* MD steps, collects all the energies, forces and structure data into a dataset, then trains a MLFF, and finally estimates the error by doing MLFF-based MD calculations. Here the MLFF-based MD calculation is based on Bayesian linear regression, and the error is directly estimated by the spread of the Gaussian distribution. If the estimated error of MLFF is large, an additional *ab initio* MD step will be performed to enlarge dataset, the MLFF is retrained with the new dataset, and estimate the error again. The procedure will be repeated until the total number of MD steps exceeds the preset.

However, the MLFF algorithm in VASP is lightweight. It helps accelerate the MD simulation, but the MLFF generated by the algorithm is not accurate enough. Therefore, we will not directly use the generated MLFF; instead, we just run the VASP MLFF module for tens of thousands of steps to generate a set of *ab initio* MD data.

With the set of *ab initio* MD data generated from the VASP MLFF module, we use the NequIP software to train an accurate MLFF. NequIP is based on an E(3) equivariant neural network, meaning that the input and output of each neural network layer are equivariant (in other words, covariant) under the rotation, reflection and translation in 3D space. The NequIP software is reported to outperform several other MLFF algorithms in both data efficiency and accuracy [85].

The training process is summarized as the follows. We started from small supercells of untwisted AA and AB bilayer $MoTe_2$ with different in-plane shift between top and bottom layer. We ran MD simulation using VASP MLFF module on those structures and collected the data from all *ab initio* MD steps. We merged all the collected data and trained a NequIP MLFF, which is then used in Atomic Simulation Environment [86] to relax the tMoTe$_2$ at various angles.

b. Relaxation results

The relaxation results of AA and AB $MoTe_2$ in different twist angles are listed in Figs. 8–10. As we mentioned before, the DFT+MLFF-relaxed results are obtained by further DFT relaxation based on the MLFF-relaxed structures. In Figs. 8 and 9, the MLFF-relaxed structures are close to the DFT+MLFF-relaxed structure. The extra DFT relaxation in the DFT+MLFF method only modified the quantitative details, keeping the qualitative shape of the MLFF-relaxed structures unchanged. It indicates that the MLFF can reproduce the main part of DFT relaxation, making the DFT relaxation easier to converge.

In structures of smaller twist angles, the local conformation is more similar to untwisted structures. Because our MLFF are constructed from untwisted structures, more accurate results in smaller angles are expected. Comparing Fig. 9 with Fig. 8, the MLFF-relaxed results are indeed better at 3.89° than that at 7.34°. Figure 10 shows the relaxation results of AA and AB tMoTe$_2$ in different twist angles. In both AA and AB stacking, interlayer distance comes to the lowest point in MX/XM/AB region and becomes higher in MM/XX/AA region. In AA stacking, the interlayer distance of MX region is the same as XM region since MX and XM configurations are related by C_2 symmetry. In the case of AB stacking, however, the MM region is different from XX region, as they are not symmetry related. The interlayer distance of XX region is always higher than MM region, which results in the asymmetry in Figs. 10(b) and 10(d).

3. Electronic structures of tMoTe$_2$

In this part, we discuss the electron band structures of tMoTe$_2$ obtained from the DFT+MLFF relaxed structure.

a. Energy bands of monolayer, AA bilayer and AB bilayer $MoTe_2$

Before diving into electronic structure of the tMoTe$_2$, we first study the monolayer and bilayer structures. For the monolayer structure, the inclusion of SOC brings a clear splitting for top valence band at K point. Additionally, at the Γ point, SOC causes a downward shift of the top valence band by 70 meV in comparison to the scenario without SOC, as depicted in Fig. 11(a).

Figure 11(b) shows that lattice relaxation has negligible influence on the band structure of AB stacking. The reason is that the functional we used is optimized for the bulk $MoTe_2$,

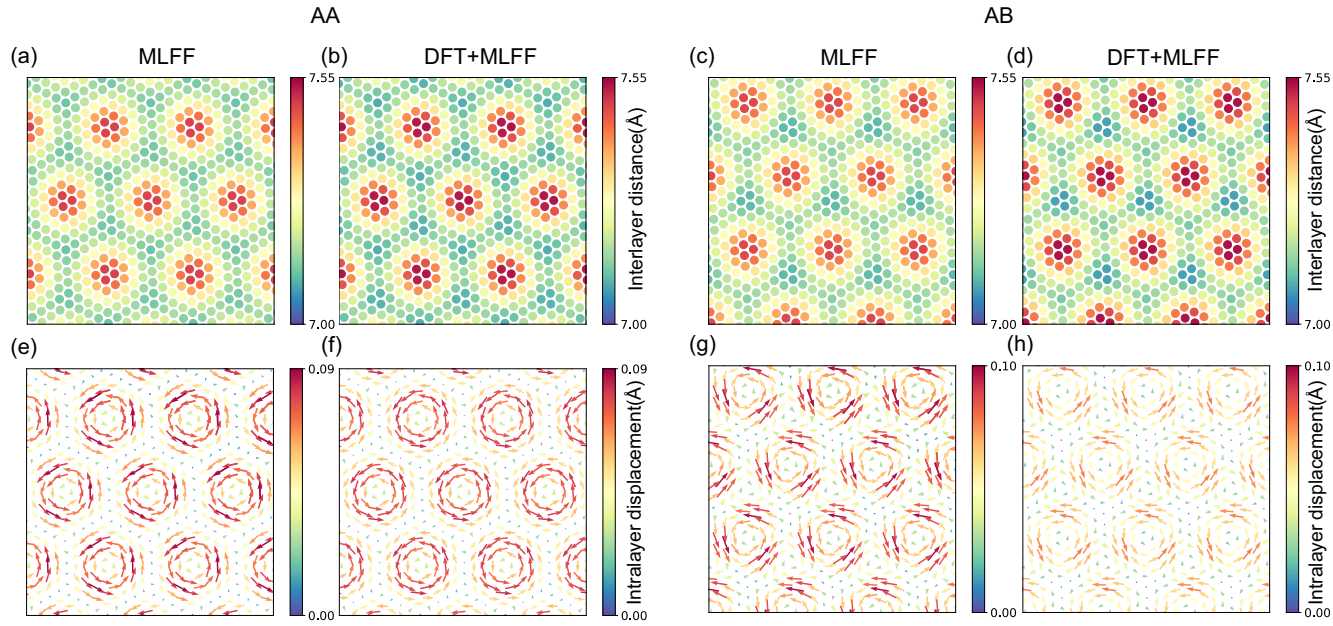

FIG. 8. Relaxation results of 7.34° AA and AB tMoTe$_2$. (a), (b), (c), and (d) are interlayer distances of MLFF-relaxed AA structure, DFT+MLFF-relaxed AA structure, MLFF-relaxed AB structure, and DFT+MLFF-relaxed AB structure respectively. (e), (f), (g), and (h) are intralayer displacements of MLFF-relaxed AA structure, DFT+MLFF-relaxed AA structure, MLFF-relaxed AB structure, and DFT+MLFF-relaxed AB structure. Interlayer distance is the distance between the top and bottom layer, while the intralayer displacement indicates the in-plane displacement from rigid positions to relaxed positions of a Mo atom in top layer. In the interlayer distance plots, rigid structures are selected to have the same lattice constant as relaxed structures.

which is also AB stacked. Conversely, Fig. 11(c) demonstrates that lattice relaxation significantly alters the band structure of AA stacking. This is due to the fact that the relaxed interlayer distance measures approximately 7.7 Å, which is around 10% greater than the interlayer distance of the rigid structure at 7 Å. Relaxation causes a downward shift of the first valence band at the Γ point, resulting in an energy about 380 meV lower at Γ compared to the K point. For conduction bands, upon

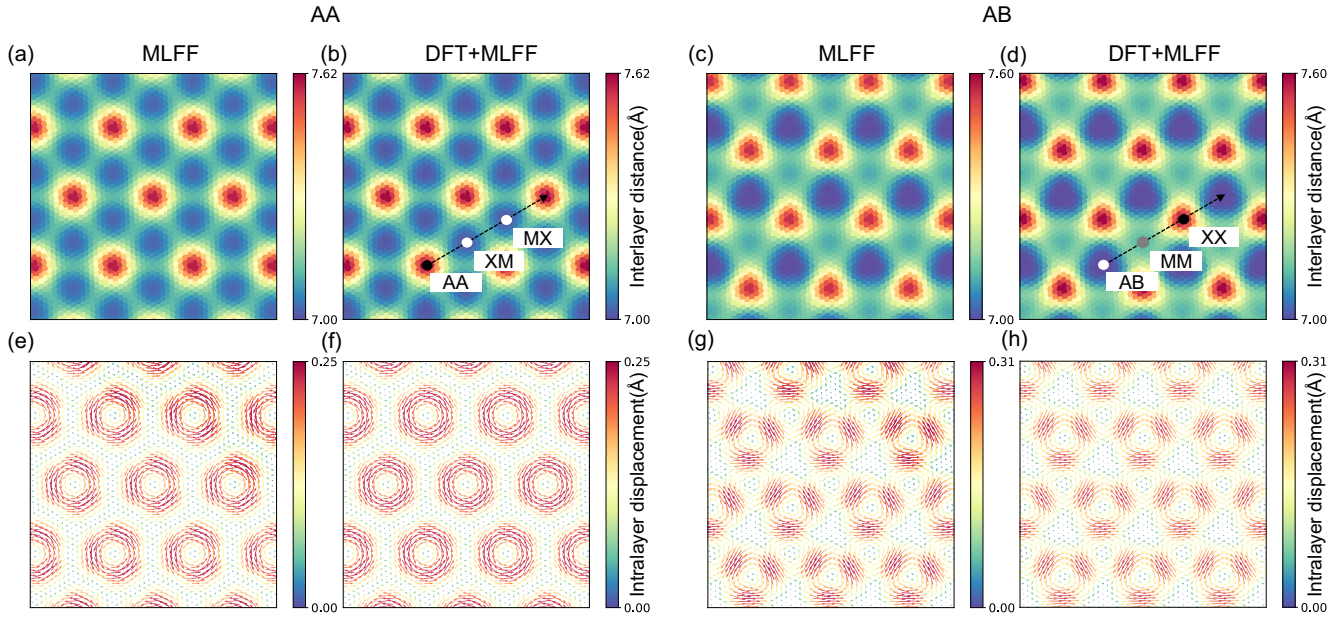

FIG. 9. Relaxation results of 3.89° AA and AB tMoTe$_2$. (a), (b), (c), and (d) are interlayer distances of MLFF-relaxed AA structure, DFT+MLFF-relaxed AA structure, MLFF-relaxed AB structure, and DFT+MLFF-relaxed AB structure respectively. (e), (f), (g), and (h) are intralayer displacements of MLFF-relaxed AA structure, DFT+MLFF-relaxed AA structure, MLFF-relaxed AB structure and DFT+MLFF-relaxed AB structure respectively.

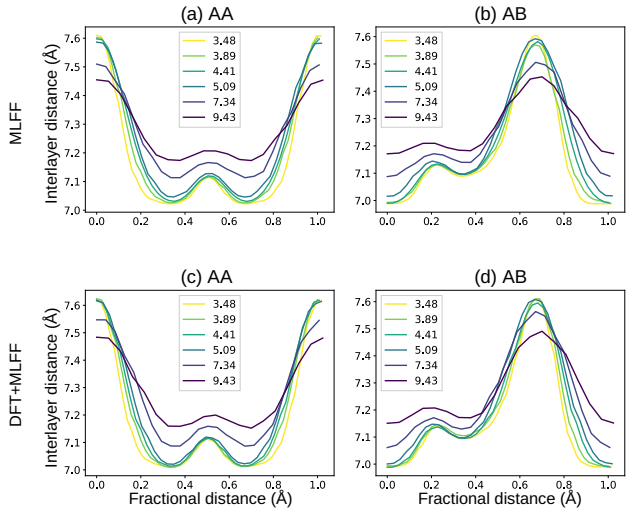

FIG. 10. Relaxation results of different angles along black arrows inFigs. 9(b) and 9(d). (a) and (b) interlayer distance of AA and AB structures given by MLFF. (c) and (d) interlayer distance of AA and AB structures given by DFT+MLFF.

relaxation, the bottom bands along Γ-K path are elevated to a similar energy level as those at the K point.

b. 13.2° tMoTe$_2$

We start from the 13.2° tMoTe$_2$ and gradually decrease the twist angle to follow the evolution of the band structure. In Fig. 12(a), when spin-orbit coupling (SOC) is not considered, there is an isolated narrow valence band (NVB) located at the top of the valence bands for AA stacking 13.2° tMoTe$_2$ without relaxation. A very similar band is also present for the AB stacking with negligible difference [Fig. 12(d)]. After considering the SOC, the NVB start to go down and entangle with other bands. [See Figs. 12(b) and 12(e).] Further including the relaxations, we obtain stacking-dependent corrugated moiré structures. Relaxation pushes the NVB further below as shown in Figs. 12(c) and 12(f).

By projecting the moiré bands into atoms' orbitals as shown in Figs. 12(l)–12(n), it is clear that the NVB is consist of Mo atoms d_{z^2} orbitals, which is the same band character of the VBM at Γ point of monolayer MoTe$_2$ [see Fig. 14(a) below]. So we can identify the NVB as the Γ-valley band. Using the orbital nature, we can see that upon considering both SOC and lattice relaxation effects, the Γ-valley bands are pushed down by 250 meV below the valence band maximum (VBM) for AA configuration, while about 180 meV for AB configuration. Then, the state around the VBM are composed of molybdenum (Mo) $d_{x^2-y^2}$ and d_{xy} orbitals, which have the same orbital characteristics as the VBM at the ±K points in a monolayer MoTe$_2$ [see Fig. 14(a)]. Therefore, the VBM of the moiré energy bands primarily originates from the ±K valley. We note that the downward shift of the Γ bands, caused by the SOC and lattice relaxation effects, is consistent with the lower Γ-valley bands after including the relaxation in the AA-stacking untwisted case in Fig. 11(c).

c. AA stacking: Evolution of band structure from 9.43° *to* 3.48°

We also calculate the band structures of several relaxed AA-stacking tMoTe$_2$ structure at 9.43°, 7.34°, 5.09°, 4.41°, 3.89°, and 3.48° as shown in Fig. 13.

We discuss the valence bands first. The valence band maximum is at the K$_M$ points of the moire BZ. The top pair of valence bands remains quasi degenerate in $\Gamma_M - \mathrm{K}_M - \mathrm{M}_M$ high-symmetry line, suggesting the presence of an extra symmetry than TR symmetry, which would map $-\boldsymbol{k}$ to $\boldsymbol{k}$. As shown in Fig. 14 on the orbital natures of the bands at 9.43°,

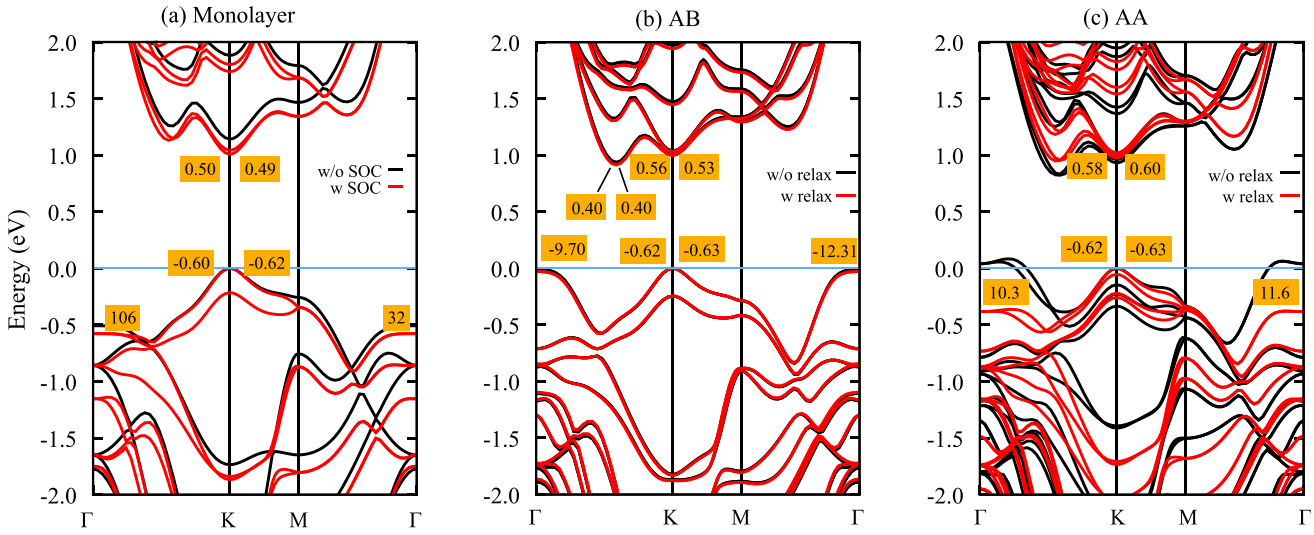

FIG. 11. Band structure of monolayer, AB bilayer, and AA bilayer MoTe$_2$ with and without relaxation. (a) Band structure of monolayer, the effective mass of the monolayer MoTe$_2$ of the VBM at the Γ and K points, and the CBM at the K point, along two different directions, are indicated in the orange boxes with units m_e. (b) Band structure of AB bilayer, where the rigid and relaxed) bands of AA and AB bilayer MoTe$_2$ are indicated by black and red lines, respectively. (c) Band structure of AA bilayer. All the effective masses are calculated for relaxed structures with SOC.

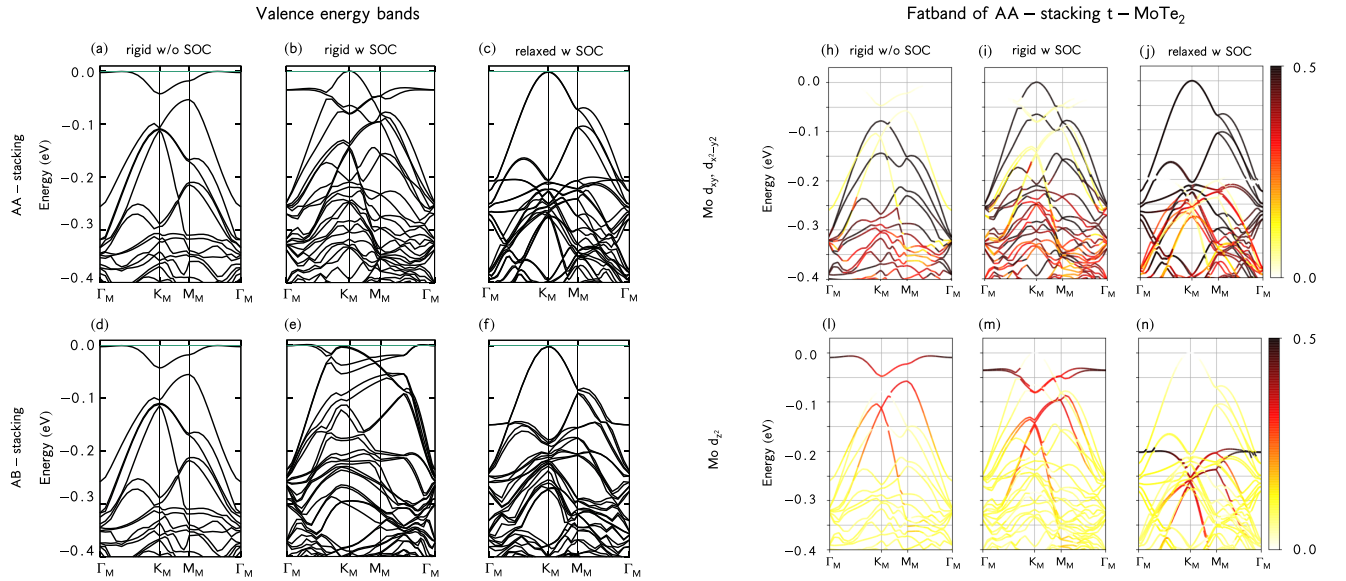

FIG. 12. Band structure and band character analysis (fatband) of 13.2° tMoTe$_2$. (a) The band structure for rigid AA without SOC, showing negligible difference with rigid AB without SOC in (d). Considering SOC, the top isolated valence band tangles with other valence bands both for rigid AA and AB. SOC brings larger splitting in $\Gamma - M$ k path for rigid AA in (b) while larger splitting around K point for rigid AB in (e). After the relaxation, the bands from ±K valley are lifted up to the top of valence bands, both for relaxed AA in (c) and relaxed AB in (f).

the valence bands near the VBM mainly come from ±K valleys, owing to their $d_{x^2-y^2}$ and d_{xy} nature, similar to the discussion of Appendix A 3 b; the expectation is the bands labeled by red dashed lines in Fig. 13, which comes form the Γ valley owing to its d_{z^2} orbital nature.

Analyzing the evolution of the valence bands in Fig. 13 from panels (g) to (l), it becomes evident that the bandwidth of the top two pairs of valence bands narrows as the twist angle decreases—a characteristic commonly observed in twisted systems. The band width of first top pairs of valence bands ($Kv1$) and second top pairs of valence bands ($Kv2$) are shown in Table V. Moreover, there is a discernible indication of band inversion between k_{v2} and k_{v3} at Γ_M, occurring as the twist angle decreases from 5.09° and 3.48°. Specifically, at the Γ_M point for 5.09°, there is a gap of approximately 2.8 meV separating the second and third pairs of top valence

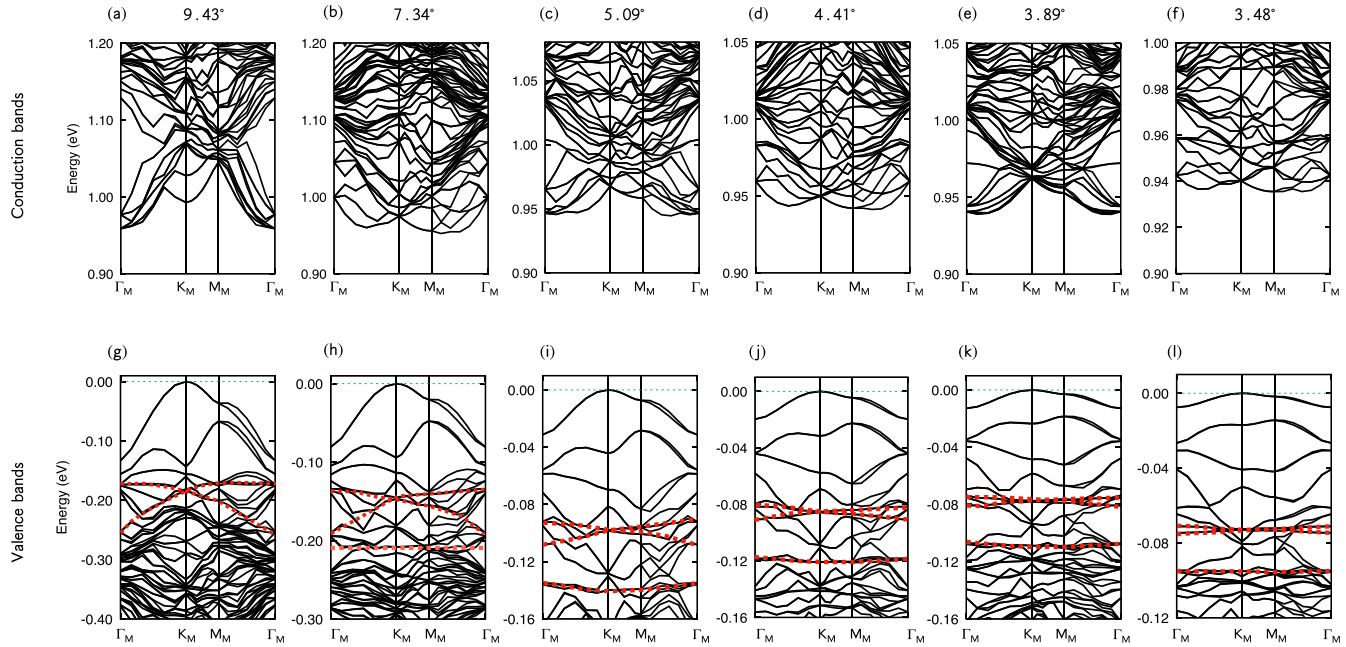

FIG. 13. Evolution of band structures of AA-stacking configuration tMoTe$_2$ with twist angles ranging from 9.43° to 3.48°. The six band structures are calculated from DFT+MLFF relaxed moiré structure with the consideration of SOC effects. The dashed-red lines show the flat bands coming from Γ valley.

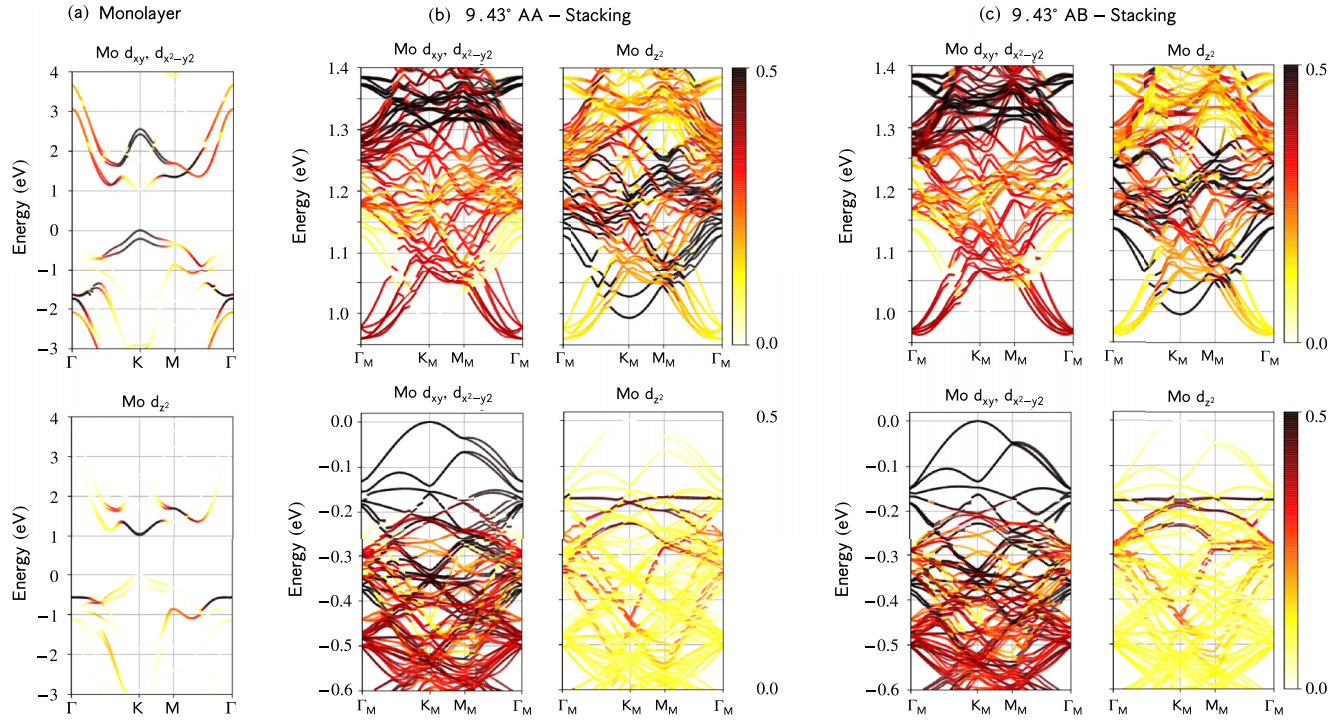

FIG. 14. Orbital-projected band structure of monolayer layer $MoTe_2$ and AA (AB) tMoTe$_2$ with twisted angle 9.43°. Bands are projected on d_{xy}, $d_{x^2-y^2}$ and d_{z^2} of Mo atoms. Figure obtained with the open-source code PYPROCAR [87].

bands. This gap closes at a twist angle of 4.41° and reopens to about 1.2 meV at 3.89°. This cycle of gap closing and reopening suggests the possibility of a band inversion. To investigate this, we calculated the irreducible representations of the six highest valence bands at the K_M and Γ_M points for twist angles ranging from 7.34° to 3.89° and have presented the findings in Fig. 2. As shown in Fig. 13 from panels (a) to (f), the conduction bands are quite messy. Based on the decomposition of orbital content depicted in Fig. 14, we determine that the conduction band minimum (CBM) at the Γ_M point for a 9.43° twist exhibits the same orbital characteristics—specifically Mo d_xy and $d_{x^2-y^2}$—as those found in the electron pocket along the Γ-K path in the monolayer band structure, also illustrated in Fig. 14. The conduction band minimum of tMoTe$_2$is around the Γ_M point rather than K_M point can be understood from the band structures of untwisted AB and AA bilayers in Fig. 11. In the AB bilayer band structure, the dip of conduction bands along the $\Gamma - K$ path is lower than that at the K point. Conversely, in the AA bilayer, the $\Gamma - K$ path's minimum aligns with the level at the K point. Given that the moiré unit cell is consist of AA, AB stacking and the intermediate stacking states between AA and AB types, the moiré conduction band minimum consequently shifts away from the K_M point.

We discuss the erratic evolution of the electron bands in Appendix C 4, which we can attribute to the lowest energy states coming from an electron-like pocket. This pocket occurs at generic momentum $\mathbf{p}_i$ in the untwisted BZ, so that the bands are folded around a generic point $\mathbf{p}_i \bmod \mathbf{G}_M$ in the moiré BZ that depends extremely sensitively on the angle through $\mathbf{G}_M$.

d. 3.89° *AA stacking* tMoTe$_2$ *electronic structures*

Let us focus on the 3.89° twisted AA stacking. In Fig. 15, we present the band structure calculations for both the rigid and relaxed configurations, with and without SOC. Before relaxation, the Γ-valley ultra-flat bands are located at the top of the valence bands when SOC is omitted. The introduction of SOC raises some dispersive bands; however, even with SOC, the rigid structure maintains flat Γ-valley bands at the top. Relaxation dependent on stacking has a significant impact on the band structure. When SOC is absent, the relaxation raises

TABLE V. Band width of the top two pairs of valence band of AA-stacking tMoTe$_2$.

Twist angle	9.43°	7.34°	5.09°	4.41°	3.89°	3.48°
k_{v1} (meV)	133.3	80.7	31.3	19.7	12.8	7.6
k_{v2} (meV)	88.7	56.9	27.6	21.3	16.2	12.2
k_{v3} (meV)	None	None	24.5	21.1	16.5	11.3

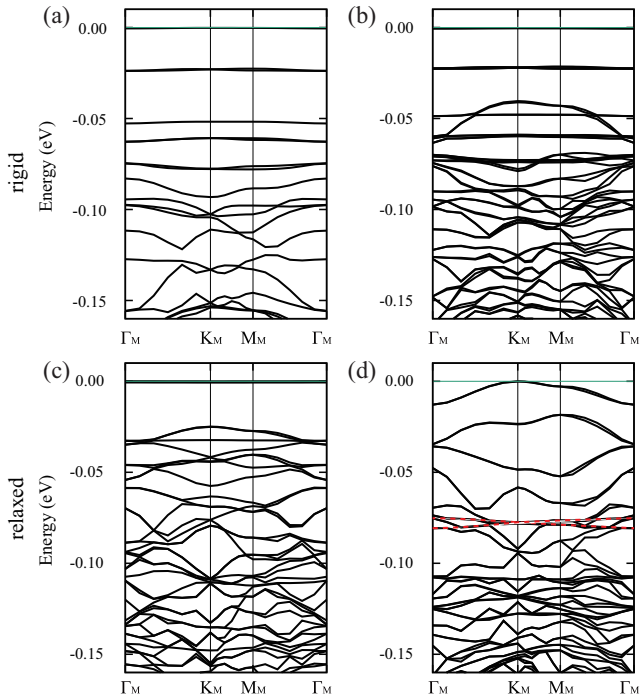

FIG. 15. AA-stacking configuration band structure of 3.89° tMoTe$_2$. (a) and (b) show the band structures without SOC and with SOC bands for rigid structure, respectively. (c) and (d) show the band structures without SOC and with SOC bands for relaxed structure, respectively.

the ±K bands. Further incorporating SOC, ±K bandsnow becomes the VBM, and the red-dashed bands originating from the Γ valley, as denoted in Fig. 15(d), are markedly diminished, sinking approximately 80 meV below the VBM. Our results are different from Ref. [21], where the Γ bands are close to the top two valence bands from ±K valley, with only 30 meV away from the VBM. The top valence band in Ref. [21] has the bandwidth of about 9 meV, while the width of the same band in our results is about 12.8 meV.

The density of states (DOS) calculations depicted in Fig. 16(c) reveal two distinct peaks within the valence bands around –75 meV and –105 meV, which correspond to the Γ-valley band embedded in the backdrop of dispersive bands. Additionally, several minor peaks, around 0 meV and 20 meV, are observed, which originate from the valence bands of the ±K valleys.

In Fig. 17, we show the band structure given by three different relaxation methods: (a) only using MLFF, (b) two-step MLFF+DFT relaxation, and (c) relaxation directly from rigid structure using DFT. The accuracy of MLFF can closely approach that of DFT calculations, especially the top three pairs of ±K-valley valence bands and the highest Γ-valley bands as shown in Fig. 17(a). However, we find the forces of MLFF relaxed structure is not small enough (with a mean absolute force of 7.8×10^{-2} eV/Å), meaning that the MLFF result can be relaxed further using the more accurate DFT relaxation method. After the further DFT relaxation, the mean absolute force reached 2.1×10^{-3} eV/Å, and the band structure changes quantitatively [see in Figs. 17(a) and 17(b)]. If the relaxation is performed from the rigid structure using DFT only, we will get almost the same band structure as the two-step relaxation band. The two-step relaxation leads to a more stable configuration, likely due to the maintenance of C_{3z} symmetry throughout the relaxation process. MLFF+DFT relaxation results in an approximately 0.9 eV lower energy compared to direct DFT relaxation, indicating the advantage of MLFF initial guess of relaxed structure in finding the lowest energy configurations.

The squared wavefunctions $|\psi_{\Gamma_M}|^2$ of the top four valence bands at the Γ_M point in the K valley is illustrated in Fig. 20(a) below. The wavefunctions of the first and second topmost valence bands in the K valley are predominantly localized in the XM and MX regions, collectively manifesting a hexagonal lattice pattern in real space. In contrast, the wavefunction of the third topmost valence band primarily occupies the AA region, delineating a triangular lattice. Meanwhile, the fourth band's wavefunction is concentrated at the XM, MX, and AA regions, constituting another hexagonal lattice configuration.

e. Comparision with Ref. [20,21]

References [20,21] have studied the relaxation and band structures of tMoTe$_2$. Reference [20] employed the SCAN density functional with dDsC dispersion correction to perform crystal structure relaxation. In our assessment of 19 different functionals, as shown in Table III, we find that the SCAN functional and dDsC dispersion correction (IVDW = 4 in Table III) yielded a larger c-axis lattice parameter and smaller a(b)-axis lattice parameters, whereas the DFT-D2 functional that we use (IVDW = 10 in Table III) provide the lattice parameters closest to the experimental results. After crystal structure relaxation, Ref. [20] showed a larger interlayer distance compared to our relaxation.

In contrast to both our approach and Refs. [20,21] used SIESTA with DFT-D2 functional to perform the DFT calculations. After the relaxation, Ref. [21] obtains a smaller interlayer distance (ILD) of which the minimum is about 6.9 Å. This Mo-Mo ILD is smaller than our relaxation result shown in Fig. 3, even smaller than ILD of bulk MoTe$_2$ crystal structure with 7.0 Å.

Our relaxation results shows that, in the AA region, where the stacking configuration is close to that of AA untwisted bilayer structure, and thus the maximum interlayer distance should be close to but slightly smaller than (due to the drag of surrounding environment, which is not aligned in AA form) the interlayer distance of AA-stacking untwisted bilayer (7.7 Å), which is not satisfied by Ref. [20]. Furthermore, the MX region has the stacking configuration akin to that of AB untwisted bilayer structure. As a result, the smallest interlayer distance in tMoTe$_2$ structures should be close to but slightly larger than (due to the surrounding environment's influence) 7 Å, which is violated by Ref. [21].

From Fig. 15, it is clear that relaxation will push the Γ-valley bands down as well as lift the bands from ±K valleys up. The relaxation in Ref. [21] gets the smaller ILD, resulting that their Γ-valley bands are not pushed down significantly and close to the ±K-valley bands. The relaxed Γ-valley bands in Ref. [21] are only 30 meV below the valence band maximum, while our Γ-valley bands are 80 meV below the VBM. This implies, likely, lack of band-mixing from the Γ-valley

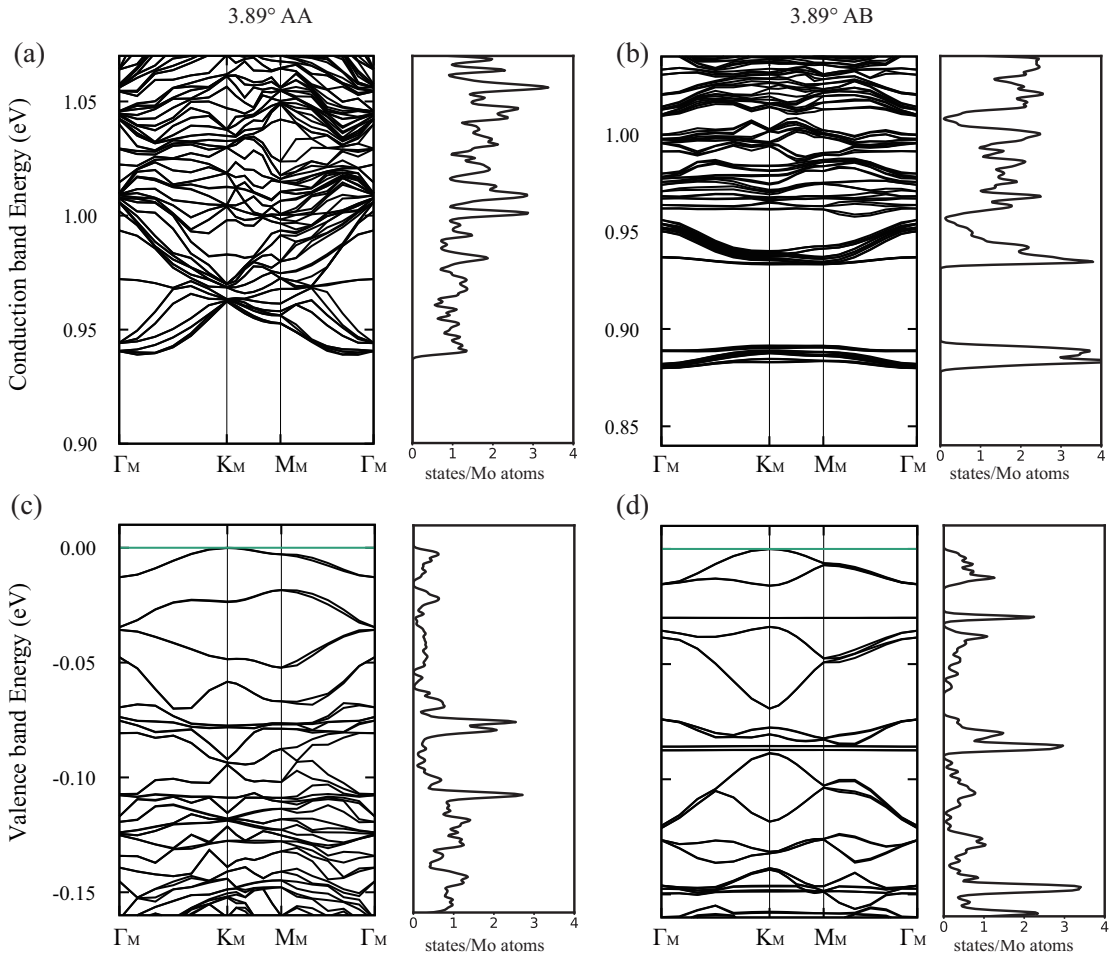

FIG. 16. Conduction bands and valence bands with their density of states (DOS) in unit of states per Mo atom of 3.89° for relaxed AA-stacking and AB-stacking configuration with SOC. a and b are the conduction bands of AA- and AB-stacking configuration with DOS. c and d are the valence bands of AA- and AB-stacking configuration. The green line represents the Fermi level.

band in the many-body calculations. Meanwhile, the relaxation in Ref. [20] gives the maximal ILD of 7.8 Å, which is larger that us, due to the overestimated c-axis lattice parameter by SCAN functional. Reference [20] calculated the band structure of 4.4° moiré structure (1014 atoms per unit cell). Due to Ref. [20] only calculated 4 k point along the

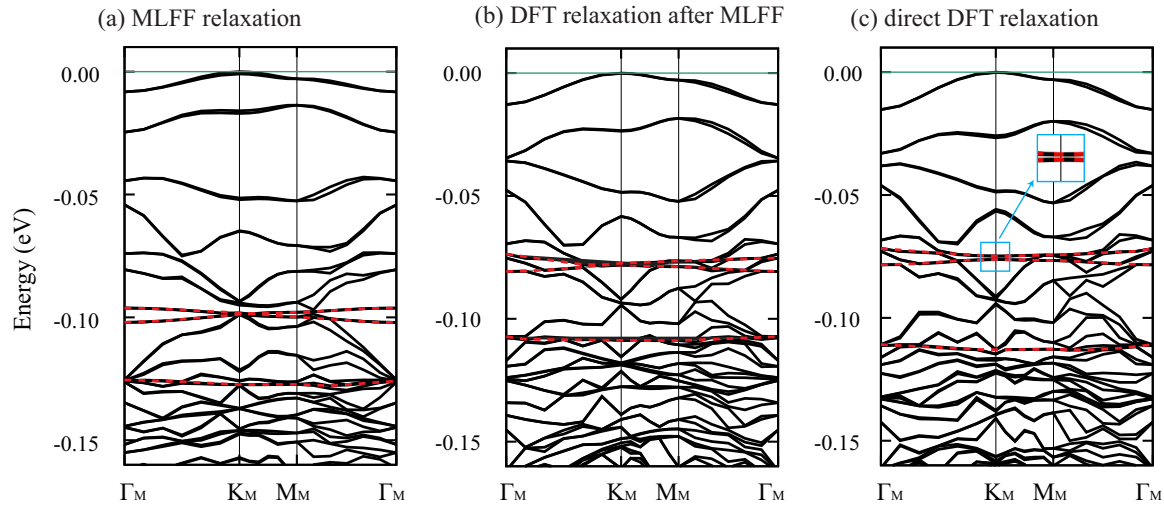

FIG. 17. Comparison of different relaxation band structures. (a) Band structure of MLFF relaxation. (b) Band structure of MLFF+DFT relaxation. (c) Band structure of direct DFT relaxation from rigid structure.

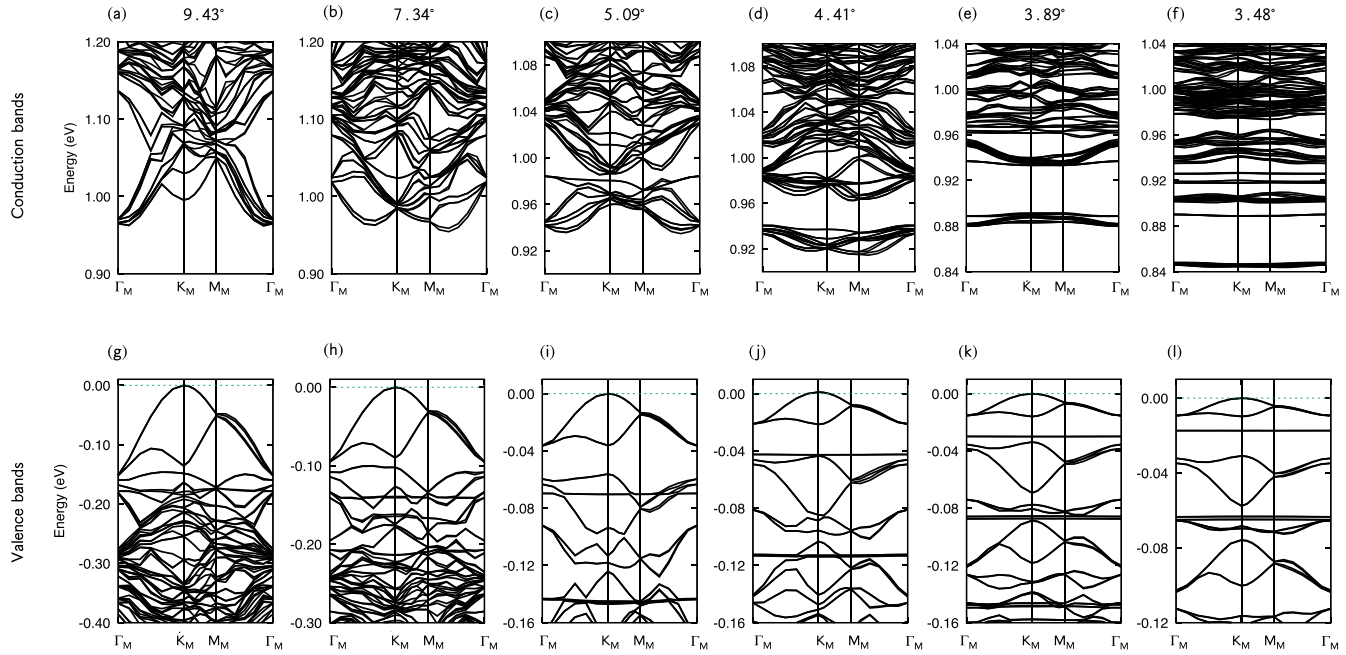

FIG. 18. Evolution of band structures for various angles range from 9.43° to 3.48° AB-stacking configuration $t\text{MoTe}_2$. Top (bottom) row is for conduction (valence) bands. The lattice was fully relaxed with DFT-D2 vdW functional. SOC are considered in the calculation.

high-symmetry line, the band width is estimated from their fitted continuum model. From the model, the top valence band from ±K valleys has band width of about 30 meV at 4.41°, while our 4.41° structure (1014 atoms per unit cell) ±K-valley band has a smaller bandwidth. For 3.89° structure in Ref. [21], the top pair of the valence bands have band width of about 9 meV, which is smaller than our result of 12.8 meV.

f. AB stacking: Evolution of band structure from 9.43° *to* 3.48°

The symmetry group of the AB stacking can be given by replacing C_{2y} in the group for *AA* stacking by C_{2x}. At large twist angle, such as 13.2°, the electronic structures of the two stacking configurations appear similar, as illustrated in Fig. 12. However, they exhibit notable differences at smaller twist angles (most notable for angles smaller than 5°, as seen in Figs. 18 and 13).

As listed in Table VI, the bandwidth of the top two pairs of bands narrows with decreasing twist angle, specifically to 16 meV for 3.89° and 9.8 meV for 3.48°. As the twist angle decreases, a pair of ultra-flat valence bands gradually moves up, and become isolated for $\theta < 4.41°$ as shown in Fig. 18. When the twist angle decreases to 3.89°, the two set of ultra-flat valence bands are separated with about 56.0 meV, bringing two distinct peak in the DOS as shown in Fig. 16.

Similar with that of AA, the top two pairs of valence bands of AB configuration consist of $d_{x^2-y^2}$ and d_{xy} orbitals of Mo atoms as shown in Fig. 14, indicating these bands come from the ±K valleys. We also find that the flat valence bands of AB come from Γ valley because they consist of Mo d_{z^2} orbitals.

Distinct from from AA stacking, there are isolated moiré bands on the conduction band side for the AB stacking. As the twist angle decreases to 4.41° [see Fig. 18(d)], there are 12 bands at the bottom of the conduction bands that are isolated from the higher-energy bands. The degeneracy of these bands suggests they likely originate from the pockets along the Γ-K line of the conduction bands, as depicted in Figs. 11(b) and 11(c). Compared with AA stacking, the valence bands from the Γ valley in AB stacking are extremely flat [see Figs. 18(k) and 18(l)], resulting in pronounced peaks in the density of states (DOS) (see Fig. 16). To facilitate further comprehensive analysis in Appendixes B and C, we have also calculated the irreducible representations of the top four valence bands for AB stacking, which can be found in Fig. 19. In Fig. 20(b), we plot the charge densities of first four pairs of valence bands' wavefunctions at Γ_M. The top ±K-valley bands wavefunctions are localized at MX region, similar with the top Γ-valley bands. The third pair valence bands' wavefunctions are localized at MM region, forming triangular lattice, while the fourth pair valence bands' wavefunctions form triangular rings with minimum at MX region.

TABLE VI. Band width of the top two pairs of valence band of AB-stacking $t\text{MoTe}_2$.

Twist angle	9.43°	7.34°	5.09°	4.41°	3.89°	3.48°
Bandwidth (meV)	135.4	89.8	36.1	22.5	16.0	9.8

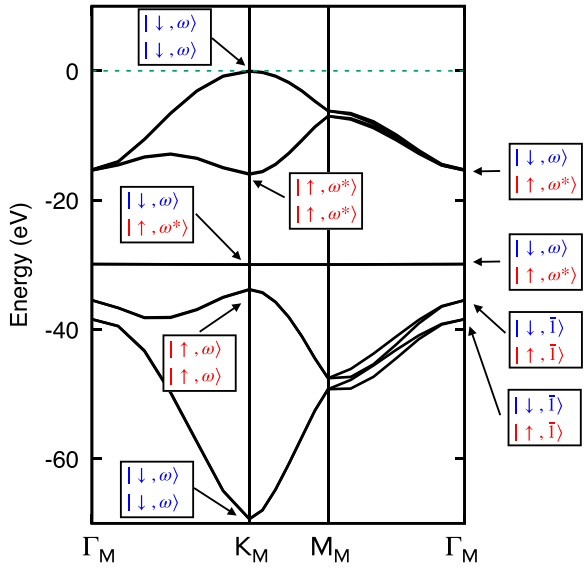

FIG. 19. Irreducible representations at high-symmetry points Γ and K of valence bands of 3.89° for relaxed AB-stacking configuration with SOC. The eigenvalues of the C_3 are denoted as $\omega = e^{i\pi/3}$, $\omega^* = e^{-i\pi/3}$ and $\bar{1} = e^{-i\pi}$.

APPENDIX B: SINGLE-PARTICLE CONTINUUM MODEL AND FITTING: AA-STACKING

The symmetry group of AA-stacked tMoTe$_2$ (AA-tMoTe$_2$) is generated by C_3, C_{2y}, and $\mathcal{T}$, in addition to the moiré lattice translations $T_{\boldsymbol{R}_M}$ [43], where $\boldsymbol{R}_M$ labels the moiré lattice vectors. We label the moiré lattice basis vectors as $\boldsymbol{a}_{M,1} = a_M(\frac{\sqrt{3}}{2}, -\frac{1}{2})^T$ and $\boldsymbol{a}_{M,2} = C_3\boldsymbol{a}_{M,1}$, and the moiré reciprocal lattice basis vectors as $\boldsymbol{b}_{M,1} = \frac{4\pi}{\sqrt{3}a_M}(1,0)^T$ and $\boldsymbol{b}_{M,2} = \frac{4\pi}{\sqrt{3}a_M}(\frac{1}{2}, \frac{\sqrt{3}}{2})^T$, where

$$a_M = \frac{a_0}{2\sin\left(\frac{\theta}{2}\right)}, \tag{B1}$$

and $a_0 = 3.52$ Å is the lattice constant of the monolayer MoTe$_2$.

Up to now, FCI states were only found for hole doping experimentally [34–37]. From Fig. 13, it is immediately apparent that isolated, nearly flat bands (understood to be an important precursor to the FCI phase [1–3]) appear only in the valence bands, which are accessible through hole doping. In contrast the conduction bands do not have well-separated bands. As such, we focus on building a model for the valence bands in this section.

We will discuss the continuum models for both ±K valleys and the Γ valley.

1. AA-Stacking: ±K Valleys

a. Microscopic Basis, Symmetries, and Inter-layer coupling

We now derive the moiré states that make up the continuum model basis for the AA-stacked twisted heterostructure. This derivation follows [88,89] by expanding the tight-binding states around the monolayer K point. (K can be straightforwardly obtained via time-reversal symmetry.) We consider a two-layer system where the $l = t, b$ layer is twisted via the linear transformation $M_\ell = R(-(-)^l\frac{\theta}{2})$ to leading order $\theta \ll 1$, where $l = t, b$ corresponds to $l = 0, 1$ for $(-)^l$. We will derive the symmetry representations as well as the form of the interlayer coupling within the two-center approximation.

From our first-principles calculations in Fig. 14, we see the valence band maximum around K is spanned by $d_{x^2-y^2} + id_{xy}$ orbitals on the Mo atoms, which we will refer to as d for brevity in this part. The d_{z^2} Mo orbitals and Te orbitals do not contribute significantly to the density of states near the active bands. We write $M_\ell\mathbf{R}$ as the positions of the Mo atoms on the ℓth layer, where $\mathbf{R}$ is an untwisted lattice vector. The states carried by these orbitals are

$$|\mathbf{p}, \ell\rangle = \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{iM_\ell\mathbf{R}\cdot\mathbf{p}}|M_\ell\mathbf{R}', \ell\rangle \tag{B2}$$

for $\ell = 0, 1$ corresponding to top/bottom, and $\mathbf{p} = \mathbf{K} + \delta\mathbf{p}$ is a momentum near the $\mathbf{K}$ point. The intralayer C_3 symmetry acts as (define $\mathbf{K}_\ell = M_\ell\mathbf{K}$)

$$\begin{aligned} C_3|\mathbf{K}_\ell + \delta\mathbf{p}, \ell\rangle &= \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{iM_\ell\mathbf{R}\cdot C_3(\mathbf{K}_\ell+\delta\mathbf{p})}|M_\ell\mathbf{R}, \ell\rangle e^{i\lambda_d} \\ &= \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{iM_\ell\mathbf{R}\cdot(\mathbf{K}_\ell+\mathbf{G}_\ell+C_3\delta\mathbf{p})}|M_\ell\mathbf{R}, \ell\rangle e^{i\lambda_d} \\ &= |\mathbf{K}_\ell + C_3\delta\mathbf{p}, \ell\rangle e^{i\lambda_d}, \end{aligned} \tag{B3}$$

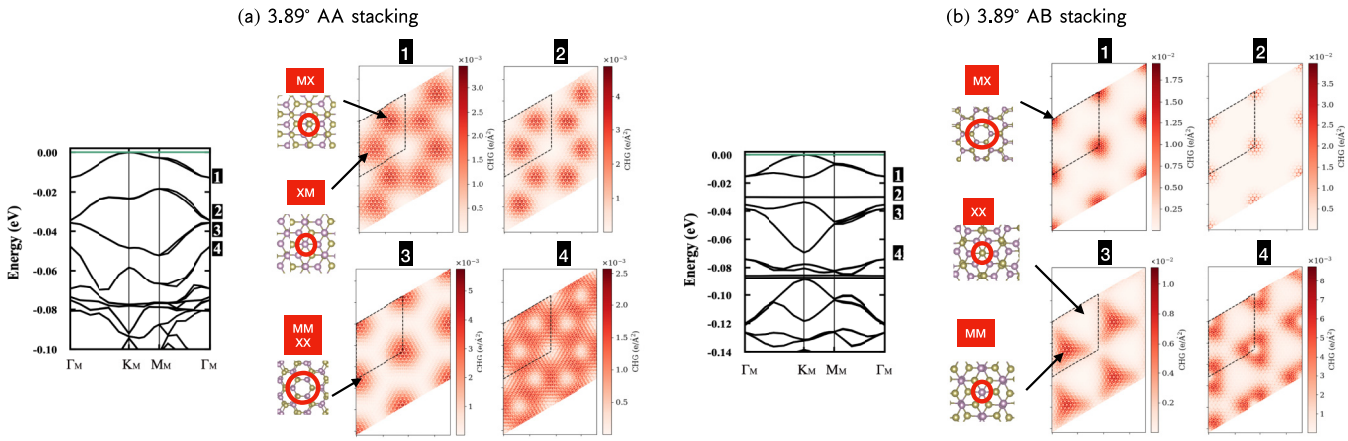

FIG. 20. Charge densities of the valence-band wavefunctions at Γ_M point in the 2 × 2 moiré cell of relaxed 3.89° (a) AA-twist and (b) AB-twist MoTe$_2$. The color bar represents the value of charge density.

where $e^{i\lambda_d} = e^{i\frac{2\pi}{3}}$ is the (spinless) C_3 eigenvalue of d orbital, and $\mathbf{G}_\ell = C_3\mathbf{K}_\ell - \mathbf{K}_\ell$ is a reciprocal lattice vector obeying $\mathbf{G}_\ell \cdot M_\ell \mathbf{R} = 0 \bmod 2\pi$. We also used that the rotation matrix C_3 commutes with M_ℓ, since both are rotations. Note that $e^{i\lambda_d}$ is simply an overall phase of the rotation representation. Secondly, the $C_{2y}\mathcal{T}$ representation is

$$\begin{aligned} C_{2y}\mathcal{T}|\mathbf{K}_\ell + \delta\mathbf{p}, \ell\rangle &= \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{-iM_\ell\mathbf{R}\cdot(\mathbf{K}_\ell+\delta\mathbf{p})}|C_{2y}M_\ell\mathbf{R}, -\ell\rangle \\ &= \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{-iM_\ell\mathbf{R}\cdot(\mathbf{K}_\ell+\delta\mathbf{p})}|M_{-\ell}C_{2y}\mathbf{R}, -\ell\rangle \\ &= \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{-iM_\ell C_{2y}\mathbf{R}\cdot(\mathbf{K}_\ell+\delta\mathbf{p})}|M_{-\ell}\mathbf{R}, -\ell\rangle \\ &= \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{iM_{-\ell}\mathbf{R}\cdot(-C_{2y})(\mathbf{K}_\ell+\delta\mathbf{p})}|M_{-\ell}\mathbf{R}, -\ell\rangle \\ &= \sum_{\ell'}|\mathbf{K}_{\ell'} + C_{2y}(-\delta\mathbf{p}), \ell'\rangle[\sigma_x]_{\ell\ell'} \quad \text{(B4)} \end{aligned}$$

using the fact that $\mathcal{T}$ is antiunitary and C_{2y} flips layer, the reflection property $C_{2y}M_\ell C_{2y}^{-1} = M_{-\ell}$, the symmetry $-C_{2y}\mathbf{K}_\ell = \mathbf{K}_{-\ell}$ since the untwisted K point is at the x axis, and the trivial transformation of d orbital under $C_{2y}\mathcal{T}$ [32]. In this discussion, we have neglected the spin degree of freedom, which can be easily reintroduced because of the spin-valley locking in $MoTe_2$. Including the angular momentum of the spin $\uparrow$ states in the K valley, we have $\lambda_d \to \lambda_{d,K}e^{\frac{\pi i}{3}}$. In the K′ valley, the band is composed of the Mo $d_{x^2-y^2} - id_{xy}$ with spin down, giving $\lambda_{d,\mathrm{K}'} = e^{-\frac{\pi i}{3}}$. These form a spinful representation of C_3. Lastly, we note that the emergent intravalley inversion (which will be discussed in Appendix B 1 b) has no representation on the microscopic basis since it is not a true symmetry of the model, much like the emergent particle-hole symmetry in twisted bilayer graphene [89,90].

We now derive the form of the interlayer Hamiltonian. Formally, we compute the overlap

$$\begin{aligned} H^{\text{inter}}_{l,-l}(\mathbf{p}, \mathbf{p}') = \frac{1}{N}\sum_{\mathbf{R},\mathbf{R}'} e^{-iM_l\mathbf{R}\cdot\mathbf{p}+iM_{-l}\mathbf{R}'\cdot\mathbf{p}'} \\ \times \langle M_l\mathbf{R}, l|H|M_{-l}\mathbf{R}', -l\rangle \,. \quad \text{(B5)} \end{aligned}$$

It is convenient to shift the sum so that the bottom layer is rotated by θ and the top layer is unrotated,

$$H^{\text{inter}}_{b,t}(\mathbf{p}, \mathbf{p}') = \frac{1}{N}\sum_{\mathbf{R},\mathbf{R}'} e^{-iM\mathbf{R}\cdot\mathbf{p}+i\mathbf{R}'\cdot\mathbf{p}'}\langle M\mathbf{R}, b|H|\mathbf{R}', t\rangle, \quad \text{(B6)}$$

where $M = R(\theta)$. To proceed, we assume that the matrix element of the Hamiltonian is only dependent on the distance between orbitals (the "two-center" approximation) leading to

$$\langle M\mathbf{R}, b|H|\mathbf{R}', t\rangle = \frac{1}{N\Omega}\sum_{\mathbf{q}\in BZ}\sum_{\mathbf{G}} t_{\mathbf{q}+\mathbf{G}}e^{i(\mathbf{q}+\mathbf{G})\cdot(M\mathbf{R}-\mathbf{R}')} \,. \quad \text{(B7)}$$

Plugging this expression into the interlayer Hamiltonian and using $M\mathbf{r}\cdot\mathbf{k} = (M\mathbf{r})^T\mathbf{k} = \mathbf{r}\cdot M^T\mathbf{k}$ gives

$$H^{\text{inter}}_{b,t}(\mathbf{p}, \mathbf{p}') = \sum_{\mathbf{G}_1,\mathbf{G}_2}\frac{t_{\mathbf{p}+\mathbf{G}_1}}{\Omega}\delta_{\mathbf{p}+\mathbf{G}_1,\mathbf{p}'+M\mathbf{G}_2} \,. \quad \text{(B8)}$$

Using $\mathbf{p} = \mathbf{K} + \delta\mathbf{p}$ (and similarly for $\mathbf{p}'$), we see that $H^{\text{inter}}_{b,t}(\mathbf{p}, \mathbf{p}')$ connects momenta apart by $M\mathbf{K} - \mathbf{K} = \mathbf{q}_1$. We now keep only the lowest order $\mathbf{G}_1$ since $t_\mathbf{p}$ is rapidly decaying. Thus in real space, this Hamiltonian can be written

$$H^{\text{inter}}_{b,t}(\mathbf{K} + \delta\mathbf{p}, M\mathbf{K} + \delta\mathbf{p}') = \sum_{j=1}^{3} w\delta_{\delta\mathbf{p},\delta\mathbf{p}'+\mathbf{q}_i} \quad \text{(B9)}$$

where $w = t_K$. We will find agreement with this term and the symmetry-based approach in the following section.

The intralayer Hamiltonian, in the two-center approximation, is given by expanding the monolayer dispersion $h_{\text{mono}}(\mathbf{K} + \delta\mathbf{p}) = -\frac{\hbar^2\delta\mathbf{p}^2}{2m^*}$ to leading order. Note that the two-center approximation cannot capture the intralayer moiré potential, which arises due to relaxation within the moiré unit cell.

b. Continuum Model

In this part, we discuss the continuum model for the low-energy states around the ±K valleys based on symmetries. Reference [43] proposed a model with first harmonics, and we will introduce more terms into it. In monolayer $MoTe_2$, the strong spin-orbit coupling locks the spin degree of freedom to the valley degree of freedom. Explicitly, the highest valence band around the K valley is made up of $d_{x^2-y^2} + id_{xy}$ Mo orbitals with spin $\uparrow$ [91]. The highest and second highest valence bands are separated by the large energy 200 meV [43], and thus we only consider the highest electron valence band around K and −K (in both layers) to construct the moiré model. The basis of the continuum model is labeled by $c^\dagger_{\eta,l,\boldsymbol{r}}$ with $\eta = \pm$ labels the ±K valleys (or equivalently spins), $l = t, b$ labels the layer, and $\boldsymbol{r}$ labels the position. The wavefunction of $c^\dagger_{\eta,l,\boldsymbol{r}}$ is a continuous approximation to the microscopic basis in Eq. (B2). Using Eqs. (B4) and (B3) and after accounting for the electron spin, the *spinful* symmetry representations furnished by $c^\dagger_{\eta,l,\boldsymbol{r}}$ read

$$\begin{aligned} C_3 c^\dagger_{\eta,l,\boldsymbol{r}}C_3^{-1} &= c^\dagger_{\eta,l,C_3\boldsymbol{r}}e^{i\eta\frac{\pi}{3}}, \\ (C_{2y}\mathcal{T})c^\dagger_{\eta,l,\boldsymbol{r}}(C_{2y}\mathcal{T})^{-1} &= c^\dagger_{\eta,\bar{l},C_{2y}\boldsymbol{r}}, \\ \mathcal{T}c^\dagger_{\eta,l,\boldsymbol{r}}\mathcal{T}^{-1} &= c^\dagger_{-\eta,l,\boldsymbol{r}}(-\eta), \\ T_{\boldsymbol{R}_M}c^\dagger_{\eta,l,\boldsymbol{r}}T^{-1}_{\boldsymbol{R}_M} &= c^\dagger_{\eta,l,\boldsymbol{r}+\boldsymbol{R}_M}e^{-i\eta\boldsymbol{R}_M\cdot\mathrm{K}_l} \,, \end{aligned} \quad \text{(B10)}$$

where $\bar{l} = b, t$ for $l = t, b$ respectively, and K_l is the rotated K point of the lth layer. For AA-$t$$MoTe_2$, we rotate the top layer by $-\theta/2$ and the bottom layer by $\theta/2$, and thus

$$\begin{aligned} \mathrm{K}_b &= \frac{4\pi}{3a_0}\left(\cos\left(\frac{\theta}{2}\right), \sin\left(\frac{\theta}{2}\right)\right)^T, \\ \mathrm{K}_t &= \frac{4\pi}{3a_0}\left(\cos\left(\frac{\theta}{2}\right), -\sin\left(\frac{\theta}{2}\right)\right)^T. \end{aligned} \quad \text{(B11)}$$

For the convenience of the discussions in the rest of this part, we define

$$\begin{aligned} \boldsymbol{q}_1 &= \mathrm{K}_b - \mathrm{K}_t = \frac{4\pi}{3a_0}2\sin\left(\frac{\theta}{2}\right)\begin{pmatrix}0\\1\end{pmatrix}, \\ \boldsymbol{q}_2 &= C_3\boldsymbol{q}_1 \,,\ \boldsymbol{q}_3 = C_3^2\boldsymbol{q}_1 \,. \end{aligned} \quad \text{(B12)}$$

The general form of the single-particle Hamiltonian in the η valley reads

$$H^{AA}_{\eta,0} = \int d^2r (c^\dagger_{\eta,b,\boldsymbol{r}}, c^\dagger_{\eta,t,\boldsymbol{r}})(h_{\eta,b}(\boldsymbol{r})\ t_\eta(\boldsymbol{r})\ \ t^*_\eta(\boldsymbol{r}) h_{\eta,t}(\boldsymbol{r})) \begin{pmatrix} c_{\eta,b,\boldsymbol{r}} \\ c_{\eta,t,\boldsymbol{r}} \end{pmatrix}, \tag{B13}$$

where $h_{\eta,l}(\boldsymbol{r})$ is Hermitian. The symmetry requirements on the terms in the Hamiltonian read

$$\begin{aligned} C_3 :\ & h_{\eta,l}(C_3\boldsymbol{r}) = h_{\eta,l}(\boldsymbol{r})\ ,\ t_\eta(C_3\boldsymbol{r}) = t_\eta(\boldsymbol{r}), \\ C_{2y}\mathcal{T} :\ & h_{\eta,\bar{l}}(C_{2y}\boldsymbol{r}) = h^*_{\eta,l}(\boldsymbol{r})\ ,\ t_\eta(C_{2y}\boldsymbol{r}) = t_\eta(\boldsymbol{r}), \\ \mathcal{T} :\ & h_{-\eta,l}(\boldsymbol{r}) = h^*_{\eta,l}(\boldsymbol{r})\ ,\ t_{-\eta}(\boldsymbol{r}) = t^*_\eta(\boldsymbol{r}), \\ T_{\boldsymbol{R}_M} :\ & h_{\eta,l}(\boldsymbol{r}+\boldsymbol{R}_M) = h_{\eta,l}(\boldsymbol{r})\ , \\ & t_\eta(\boldsymbol{r}+\boldsymbol{R}_M) = t_\eta(\boldsymbol{r}) e^{-\mathrm{i}\eta\boldsymbol{q}_1\cdot\boldsymbol{R}_M}. \end{aligned} \tag{B14}$$

Owing to C_3 symmetry, the dispersion around ±K in the monolayer $MoTe_2$ is quadratic. Thus, the spatial derivatives are kept to the second order in the intralayer term,

$$h_{\eta,l}(\boldsymbol{r}) = \frac{\hbar^2\nabla^2}{2m^*} + V_{\eta,l}(\boldsymbol{r})\ , \tag{B15}$$

where m^* is real and positive as we are considering the valence electron bands, the $C_{2y}\mathcal{T}$ and $\mathcal{T}$ symmetries require m^* to be the same for all values of η, l, and Hermiticity requires $V_{\eta,l}(\boldsymbol{r})$ to be real. The kinetic term follows from the two-center approximation [see the discussion after Eq. (B9)]. Owing to the constraints imposed by moiré lattice translations on the potential terms [see Eq. (B14)], $V_{\eta,l}(\boldsymbol{r})$ and $e^{\mathrm{i}\eta\boldsymbol{q}_1\cdot\boldsymbol{r}}t_\eta(\boldsymbol{r})$ can be expanded in series of the moiré lattice vector $\boldsymbol{G}_M$,

$$\begin{aligned} V_{\eta,l}(\boldsymbol{r}) &= \sum_{\boldsymbol{G}_M} e^{-\eta\mathrm{i}\boldsymbol{G}_M\cdot\boldsymbol{r}} V_{\eta,l,\boldsymbol{G}_M}, \\ t_\eta(\boldsymbol{r}) &= \sum_{\boldsymbol{G}_M} e^{-\eta\mathrm{i}(\boldsymbol{q}_1+\boldsymbol{G}_M)\cdot\boldsymbol{r}} t_{\eta,\boldsymbol{q}_1+\boldsymbol{G}_M}\ , \end{aligned} \tag{B16}$$

where the TR symmetry requires $V_{\eta,l,\boldsymbol{G}_M} = V^*_{-\eta,l,\boldsymbol{G}_M}$ and $t_{\eta,\boldsymbol{q}_1+\boldsymbol{G}_M} = t^*_{-\eta,\boldsymbol{q}_1+\boldsymbol{G}_M}$.

In Ref. [43], only the first harmonics of $V_{\eta,l}(\boldsymbol{r})$ and $t_\eta(\boldsymbol{r})$ is kept, i.e., only including $\boldsymbol{G}_M \in \{\pm\boldsymbol{g}_i | i = 1, 2, 3\}$ with $\boldsymbol{g}_i = C_3^{i-1}\boldsymbol{b}_{M,1}$ for the intralayer terms, and only include $\boldsymbol{q}_1 + \boldsymbol{G}_M \in \{-\boldsymbol{q}_i | i = 1, 2, 3\}$ for the interlayer terms. Combined with C_3, $C_{2y}\mathcal{T}$, and TR symmetries, $V_{+,l}(\boldsymbol{r})$ and $t_+(\boldsymbol{r})$ take the forms of

$$\begin{aligned} V^{FH}_{\eta,l}(\boldsymbol{r}) &= Ve^{-(-)^l\mathrm{i}\psi} \sum_{i=1,2,3} e^{\mathrm{i}\boldsymbol{g}_i\cdot\boldsymbol{r}} + Ve^{(-)^l\mathrm{i}\psi} \sum_{i=1,2,3} e^{-\mathrm{i}\boldsymbol{g}_i\cdot\boldsymbol{r}}, \\ t^{FH}_\eta(\boldsymbol{r}) &= w \sum_{i=1,2,3} e^{-\eta\mathrm{i}\boldsymbol{q}_i\cdot\boldsymbol{r}}\ , \end{aligned} \tag{B17}$$

where "FH" labels the first harmonics, $l = t, b$ respectively correspond to $l = 0, 1$ for $(-)^l$, V and ψ are real, and w is chosen to be real nonpositive by tuning the relative phase between the two layers. The lowest-harmonics model has effective inversion symmetry

$$\mathcal{I} c^\dagger_{\eta,l,\boldsymbol{r}} \mathcal{I}^{-1} = c^\dagger_{\eta,\bar{l},-\boldsymbol{r}}\ . \tag{B18}$$

Adding displacement field would induce an energy difference between the two layers, i.e.,

$$H_{\eta,\varepsilon} = \int d^2r \sum_l c^\dagger_{\eta,l,\boldsymbol{r}} c_{\eta,l,\boldsymbol{r}} (-)^l \frac{\varepsilon}{2}\ , \tag{B19}$$

which breaks the effective inversion symmetry, as well as the C_{2y} symmetry.

As discussed in the previous work (e.g., Ref. [21]) and in Appendix B 3, the lowest-harmonics model can only well describe the highest two valence bands per valley and the gap between second and third valence bands at Γ. In our paper, we want to capture the third band in each valley by including one more harmonics. In general, including higher harmonics may break the effective inversion symmetry. However, according to the DFT band structure in Fig. 13, the third pairs of valence bands are still approximately degenerate for angles below 5.09 degrees; therefore, we will keep the effective inversion symmetry when adding extra harmonics. The second harmonics (SH) have $|\boldsymbol{G}_M| = |\boldsymbol{b}_{M,1} + \boldsymbol{b}_{M,2}|$ for the intralayer potential and have $|\boldsymbol{q}_1 + \boldsymbol{G}_M| = \boldsymbol{q}_1 + \boldsymbol{b}_{M,1}$ for the interlayer potential; combined with C_3, $C_{2y}\mathcal{T}$, TR and the effective inversion symmetries, the form of the SH for the intralayer potential read

$$\begin{aligned} V^{SH}_{\eta,l}(\boldsymbol{r}) &= 2V_2 \sum_{i=1}^{3} \cos(\boldsymbol{g}_{2i}\cdot\boldsymbol{r}), \\ t^{SH}_\eta(\boldsymbol{r}) &= w_2 \sum_{i=1,2,3} e^{-\eta\mathrm{i}\boldsymbol{q}_{2i}\cdot\boldsymbol{r}}\ , \end{aligned} \tag{B20}$$

where $\boldsymbol{g}_{21} = \boldsymbol{b}_{M,1} + \boldsymbol{b}_{M,2}$, $\boldsymbol{g}_{2i} = C_3^{i-1}\boldsymbol{g}_{21}$, $\boldsymbol{q}_{21} = \boldsymbol{b}_{M,1} + \boldsymbol{q}_1$, $\boldsymbol{q}_{2i} = C_3^{i-1}\boldsymbol{q}_{21}$, and V_2 and w_2 are real. We note that here we already include the effective inversion, which constraints the form of Eq. (B20); without the effective inversion, we would have additional terms. With the extra terms in Eq. (B20), the final form of $V_\eta(\boldsymbol{r})$ and $t_\eta(\boldsymbol{r})$ reads

$$\begin{aligned} V_{\eta,l}(\boldsymbol{r}) &= Ve^{-(-)^l\mathrm{i}\psi} \sum_{i=1,2,3} e^{\mathrm{i}\boldsymbol{g}_i\cdot\boldsymbol{r}} + Ve^{(-)^l\mathrm{i}\psi} \sum_{i=1,2,3} e^{-\mathrm{i}\boldsymbol{g}_i\cdot\boldsymbol{r}} \\ &\quad + 2V_2 \sum_{i=1}^{3} \cos(\boldsymbol{g}_{2i}\cdot\boldsymbol{r}), \\ t_\eta(\boldsymbol{r}) &= w \sum_{i=1,2,3} e^{-\eta\mathrm{i}\boldsymbol{q}_i\cdot\boldsymbol{r}} + w_2 \sum_{i=1,2,3} e^{-\eta\mathrm{i}\boldsymbol{q}_{2i}\cdot\boldsymbol{r}}\ . \end{aligned} \tag{B21}$$

The moiré translations allow us to express the Hamiltonian in the momentum space. The Fourier transformation of the basis reads

$$c^\dagger_{\eta,l,\boldsymbol{r}} = \frac{1}{\sqrt{\mathcal{V}}} \sum_{\boldsymbol{k}} \sum_{\boldsymbol{Q}\in\mathcal{Q}^\eta_l} e^{-\mathrm{i}(\boldsymbol{k}-\boldsymbol{Q})\cdot\boldsymbol{r}} c^\dagger_{\eta,l,\boldsymbol{k}-\boldsymbol{Q}}\ , \tag{B22}$$

where

$$\mathcal{Q}^\eta_l = \{\boldsymbol{G}_M + \eta(-)^l\boldsymbol{q}_1\}\ . \tag{B23}$$

As a result, $H_{\eta,0}$ reads

$$H^{AA}_{\eta,0} = \sum_{\boldsymbol{k}} \sum_{\boldsymbol{Q},\boldsymbol{Q}'\in\mathcal{Q}} c^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q}} h^{AA}_{\eta,\boldsymbol{Q}\boldsymbol{Q}'}(\boldsymbol{k}) c_{\eta,\boldsymbol{k},\boldsymbol{Q}'}\ , \tag{B24}$$

where we have included the displacement field, $\mathcal{Q} = \{\boldsymbol{G}_M + \boldsymbol{q}_1\} \cup \{\boldsymbol{G}_M - \boldsymbol{q}_1\}$,

$$c^\dagger_{\eta,\boldsymbol{k},\boldsymbol{Q}} = c^\dagger_{\eta,l^\eta_{\boldsymbol{Q}},\boldsymbol{k}-\boldsymbol{Q}}\,, \tag{B25}$$

and $l^\eta_{\boldsymbol{Q}} = l$ for $\boldsymbol{Q} \in \mathcal{Q}^\eta_l$. The general form of $h^{AA}_{\eta,\boldsymbol{Q}\boldsymbol{Q}'}(\boldsymbol{k})$ reads

$$\begin{aligned} h^{AA}_{\eta,\boldsymbol{Q}\boldsymbol{Q}'}(\boldsymbol{k}) = \delta_{\boldsymbol{Q}\boldsymbol{Q}'}\left(\frac{-\hbar^2(\boldsymbol{k}-\boldsymbol{Q})^2}{2m^*} + \frac{\varepsilon}{2}\eta(-)^{\boldsymbol{Q}}\right) \\ + \sum_{\boldsymbol{G}_M} \delta_{\boldsymbol{Q}_1,\boldsymbol{Q}_2+\eta\boldsymbol{G}_M} V_{\eta,l^\eta_{\boldsymbol{Q}_1},\boldsymbol{G}_M} \\ + \sum_{\boldsymbol{G}_M} \delta_{\boldsymbol{Q}_1,\boldsymbol{Q}_2+\eta(\boldsymbol{q}_1+\boldsymbol{G}_M)} t_{\eta,\boldsymbol{q}_1+\boldsymbol{G}_M} \\ + \sum_{\boldsymbol{G}_M} \delta_{\boldsymbol{Q}_1,\boldsymbol{Q}_2-\eta(\boldsymbol{q}_1+\boldsymbol{G}_M)} t^*_{\eta,\boldsymbol{q}_1+\boldsymbol{G}_M}\,. \end{aligned} \tag{B26}$$

With the FH and SH terms in Eq. (B21), the matrix Hamiltonian $h^{AA}_{\eta,\boldsymbol{Q}\boldsymbol{Q}'}(\boldsymbol{k})$ has the form

$$\begin{aligned} h^{AA}_{\eta,\boldsymbol{Q}\boldsymbol{Q}'}(\boldsymbol{k}) = \delta_{\boldsymbol{Q}\boldsymbol{Q}'}\left(\frac{-\hbar^2(\boldsymbol{k}-\boldsymbol{Q})^2}{2m^*} + \frac{\varepsilon}{2}\eta(-)^{\boldsymbol{Q}}\right) \\ + V\sum_{i=1}^{3}\left[e^{-\eta(-)^{\boldsymbol{Q}}\mathrm{i}\psi}\delta_{\boldsymbol{Q}+\boldsymbol{g}_i,\boldsymbol{Q}'} + e^{\eta(-)^{\boldsymbol{Q}}\mathrm{i}\psi}\delta_{\boldsymbol{Q}-\boldsymbol{g}_i,\boldsymbol{Q}'}\right] \\ + V_2\sum_{i=1}^{3}\left[\delta_{\boldsymbol{Q}+\boldsymbol{g}_{2i},\boldsymbol{Q}'} + \delta_{\boldsymbol{Q},\boldsymbol{Q}'+\boldsymbol{g}_{2i}}\right] \\ + w\sum_{i=1}^{3}\left[\delta_{\boldsymbol{Q}+\boldsymbol{q}_i,\boldsymbol{Q}'} + \delta_{\boldsymbol{Q},\boldsymbol{Q}'+\boldsymbol{q}_i}\right] \\ + w_2\sum_{i=1}^{3}\left[\delta_{\boldsymbol{Q}+\boldsymbol{q}_{2i},\boldsymbol{Q}'} + \delta_{\boldsymbol{Q},\boldsymbol{Q}'+\boldsymbol{q}_{2i}}\right], \end{aligned} \tag{B27}$$

where $(-)^{\boldsymbol{Q}} = \pm 1$ for $\boldsymbol{Q} \in \{\boldsymbol{G}_M \pm \boldsymbol{q}_1\}$, and we used $\eta(-)^{\boldsymbol{Q}} = (-)^{l^\eta_{\boldsymbol{Q}}}$.

2. AA-Stacking: Γ Valley

In this part, we discuss the continuum model for the Γ valley. In the monolayer $MoTe_2$, the bands around Γ mainly come from the d_{z^2} orbital, and time-reversal symmetry at Γ leads to a Kramers degeneracy between spin-up and spin-down states. Thus, the basis of the moiré model at Γ valley is labeled as $\psi^\dagger_{\boldsymbol{r},l,s}$, where l is the layer index, and $s = \uparrow / \downarrow$ labels the spin. Then, the continuum model in general has the following form:

$$H^{AA}_\Gamma = \int d^2r (\psi^\dagger_{\boldsymbol{r},b}\ \psi^\dagger_{\boldsymbol{r},t}) \begin{pmatrix} h_{\Gamma,b}(\boldsymbol{r}) & t_\Gamma(\boldsymbol{r}) \\ t^\dagger_\Gamma(\boldsymbol{r}) & h_{\Gamma,t}(\boldsymbol{r}) \end{pmatrix} \begin{pmatrix} \psi^\dagger_{\boldsymbol{r},b} \\ \psi^\dagger_{\boldsymbol{r},t} \end{pmatrix}, \tag{B28}$$

where

$$\psi^\dagger_{\boldsymbol{r},l} = (\psi^\dagger_{\boldsymbol{r},l,\uparrow}, \psi^\dagger_{\boldsymbol{r},l,\downarrow})\,, \tag{B29}$$

t and b correspond to the top and bottom layers, respectively, and $h_{\Gamma,l}(\boldsymbol{r})$ and $t_\Gamma(\boldsymbol{r})$ are 2×2 moiré-periodic matrix functions.

The spinful representations of the d_{z^2} orbital symmetries of the moiré model are

$$\begin{aligned} C_3\psi^\dagger_{\boldsymbol{r},l}C_3^{-1} &= \psi^\dagger_{C_3\boldsymbol{r},l}e^{-\mathrm{i}s_z\frac{\pi}{3}}, \\ C_{2y}\mathcal{T}\psi^\dagger_{\boldsymbol{r},l}(C_{2y}\mathcal{T})^{-1} &= \psi^\dagger_{C_{2y}\boldsymbol{r},\bar{l}}, \\ \mathcal{T}\psi^\dagger_{\boldsymbol{r},l}\mathcal{T}^{-1} &= \psi^\dagger_{\boldsymbol{r},l}\mathrm{i}s_y, \\ T_{\boldsymbol{R}_M}\psi^\dagger_{\boldsymbol{r},l}T^{-1}_{\boldsymbol{R}_M} &= \psi^\dagger_{\boldsymbol{r}+\boldsymbol{R}_M,l}\,, \end{aligned} \tag{B30}$$

where s_x, s_y, s_z are the Pauli matrices for the spin index, and $\bar{l} = t, b$ if $l = b, t$. Based on symmetries, we know

$$\begin{aligned} e^{-\mathrm{i}s_z\frac{\pi}{3}}h_{\Gamma,l}(\boldsymbol{r})e^{\mathrm{i}s_z\frac{\pi}{3}} &= h_{\Gamma,l}(C_3\boldsymbol{r}), \\ e^{-\mathrm{i}s_z\frac{\pi}{3}}t_\Gamma(\boldsymbol{r})e^{\mathrm{i}s_z\frac{\pi}{3}} &= t_\Gamma(C_3\boldsymbol{r}), \\ h^*_{\Gamma,l}(\boldsymbol{r}) = h_{\Gamma,\bar{l}}(C_{2y}\boldsymbol{r})\,,\ t^*_\Gamma(\boldsymbol{r}) &= t^\dagger_\Gamma(C_{2y}\boldsymbol{r}), \\ \sigma_y h^*_{\Gamma,l}(\boldsymbol{r})\sigma_y = h_{\Gamma,l}(\boldsymbol{r})\,,\ \sigma_y t^*_\Gamma(\boldsymbol{r})\sigma_y &= t_\Gamma(\boldsymbol{r}), \\ h_{\Gamma,l}(\boldsymbol{r}) = h_{\Gamma,l}(\boldsymbol{r}+\boldsymbol{R}_M)\,,\ t_\Gamma(\boldsymbol{r}) &= t_\Gamma(\boldsymbol{r}+\boldsymbol{R}_M)\,. \end{aligned} \tag{B31}$$

The intralayer kinetic energy term must take the form

$$h_{\Gamma,l}(\boldsymbol{r}) = \frac{\hbar^2\nabla^2}{2m^*_\Gamma}s_0 + V_{\Gamma,l}(\boldsymbol{r}), \tag{B32}$$

since a linear Rashba-like kinetic term is forbidden by C_3, $\mathcal{T}$, and M_z symmetry in the monolayer. The effective mass m^*_Γ is the same in both layers because of the C_{2y} symmetry. We note that the value of m^*_Γ in Eq. (B32) might not be equal to that for monolayer $MoTe_2$ due to the interlayer coupling renormalizing the bands, due to the fact that the bands near Γ in the monolayer is quite flat as shown in Fig. 11(a). The potentials are periodic and can be expanded as

$$\begin{aligned} V_{\Gamma,l}(\boldsymbol{r}) &= \sum_{\boldsymbol{G}} V_{\Gamma,l,\boldsymbol{G}}e^{\mathrm{i}\boldsymbol{G}\cdot\boldsymbol{r}}, \\ t_\Gamma(\boldsymbol{r}) &= \sum_{\boldsymbol{G}} t_{\Gamma,\boldsymbol{G}}e^{\mathrm{i}\boldsymbol{G}\cdot\boldsymbol{r}}, \end{aligned} \tag{B33}$$

whose components are restricted by symmetry to obey

$$\begin{aligned} e^{-\mathrm{i}s_z\frac{\pi}{3}}V_{\Gamma,l,\boldsymbol{G}}e^{\mathrm{i}s_z\frac{\pi}{3}} = V_{\Gamma,l,C_3\boldsymbol{G}}\,,\ e^{-\mathrm{i}s_z\frac{\pi}{3}}t_{\Gamma,\boldsymbol{G}}e^{\mathrm{i}s_z\frac{\pi}{3}} &= t_{\Gamma,C_3\boldsymbol{G}}, \\ V^*_{\Gamma,l,\boldsymbol{G}} = V_{\Gamma,\bar{l},-C_{2y}\boldsymbol{G}}\,,\ t^*_{\Gamma,\boldsymbol{G}} &= t^\dagger_{\Gamma,-C_{2y}\boldsymbol{G}}, \\ s_yV^*_{\Gamma,l,\boldsymbol{G}}s_y = V_{\Gamma,l,-\boldsymbol{G}}\,,\ s_yt^*_{\Gamma,\boldsymbol{G}}s_y &= t_{\Gamma,-\boldsymbol{G}}, \\ V^\dagger_{\Gamma,l,\boldsymbol{G}} &= V_{\Gamma,l,-\boldsymbol{G}}\,. \end{aligned} \tag{B34}$$

To capture the dominant contribution, we only include the first harmonics in $V_{\Gamma,l}(\boldsymbol{r})$, and the zeroth harmonic in the $t_\Gamma(\boldsymbol{r})$, i.e.,

$$\begin{aligned} V_{\Gamma,l}(\boldsymbol{r}) &= \sum_{\boldsymbol{G}\in\{\pm\boldsymbol{g}_i|i=1,2,3\}} V_{\Gamma,l,\boldsymbol{G}}e^{\mathrm{i}\boldsymbol{G}\cdot\boldsymbol{r}}, \\ t_\Gamma(\boldsymbol{r}) &= t_{\Gamma,0}\,. \end{aligned} \tag{B35}$$

We neglect the zeroth harmonic in $V_{\Gamma,l}(\boldsymbol{r})$ since it is just a total shift of energy (although it must be included to compare the energies of the in-active Γ valley and active ±K valleys). According to Eq. (B34), the first harmonic terms in $V_{\Gamma,l}(\boldsymbol{r})$ can

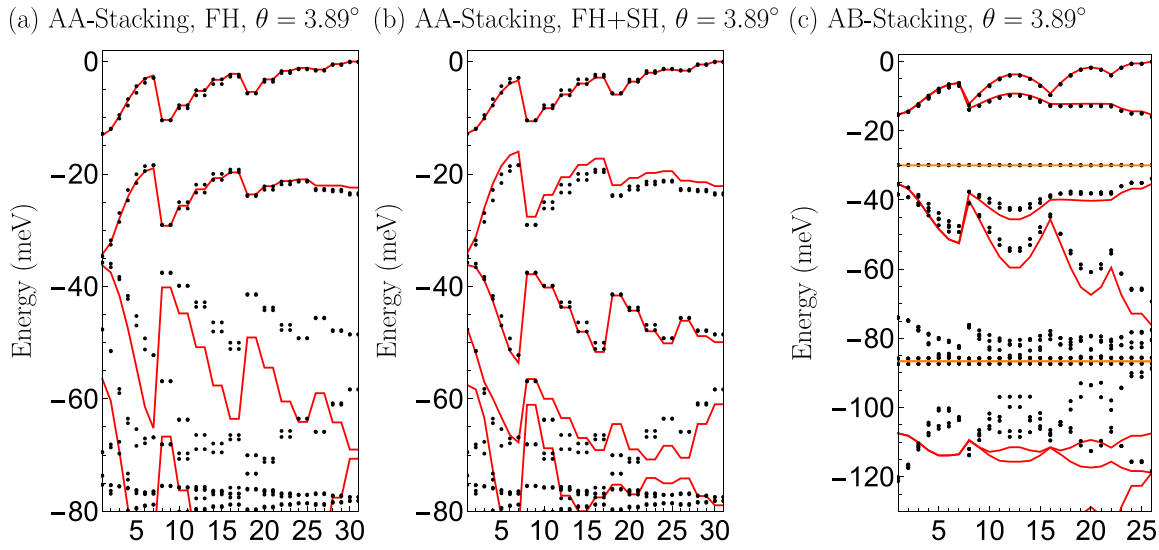

FIG. 21. The comparison between the valence bands from the DFT calculation (black dots) and from the continuum model (red/orange line) for twist angle 3.89° in the fundamental zone for (a) (b) AA stacking and (c) AB stacking. The red line comes from the ±K-valley model, while the orange line comes from the Γ-valley model. The horizontal axes labels the 31 momenta in the k mesh of the fundamental zone that we choose. Each red line is at least doubly degenerate. The orange line in (c) has double degeneracy around −30 meV and fourfold degeneracy around −86 meV. The fitting in (a) is done with FH terms, while SH terms are added in (b).

be written

$$V_{\Gamma,b,\pm\boldsymbol{g}_i} = V_\Gamma e^{\pm i\psi_\Gamma} s_0,$$
$$V_{\Gamma,t,\pm\boldsymbol{g}_i} = V_\Gamma e^{\mp i\psi_\Gamma} s_0\ , \tag{B36}$$

and similarly, $t_{\Gamma,0}$ has the form

$$t_{\Gamma,0} = w_\Gamma e^{i s_z \phi_\Gamma}\ . \tag{B37}$$

Here V_Γ, ψ_Γ, w_Γ, and ϕ_Γ are real. We set $\phi_\Gamma = 0$ as a gauge choice of the relative phase between the layers. As a result, Eq. (B35) is further simplified to

$$V_{\Gamma,l}(\boldsymbol{r}) = \sum_{i=1}^{3} V_\Gamma e^{i(-)^l \psi_\Gamma} e^{i\boldsymbol{g}_i\cdot\boldsymbol{r}} + \sum_{i=1}^{3} V_\Gamma e^{-i(-)^l \psi_\Gamma} e^{-i\boldsymbol{g}_i\cdot\boldsymbol{r}},$$
$$t_\Gamma(\boldsymbol{r}) = w_\Gamma s_0\ . \tag{B38}$$

With the simplification in Eqs. (B32) and (B38), the Γ-valley continuum model in Eq. (B28) has effective inversion symmetry, i.e.,

$$\mathcal{I}\psi^\dagger_{\boldsymbol{r},l}\mathcal{I}^{-1} = \psi^\dagger_{-\boldsymbol{r},\bar{l}}\ . \tag{B39}$$

Similar to the ±K-valley case, adding displacement breaks the effective inversion symmetry, as well as the C_{2y} symmetry, where the displacement field term reads

$$H_{\Gamma,\varepsilon} = \int d^2r \sum_l \psi^\dagger_{\boldsymbol{r},l}\psi_{\boldsymbol{r},l}(-)^l \frac{\varepsilon}{2}\ . \tag{B40}$$

Including the first harmonics in $t_\Gamma(\boldsymbol{r})$ and the higher harmonics in $V_{\Gamma,l}(\boldsymbol{r})$ may also break the effective inversion symmetry.

3. AA-Stacking: Fitting to the DFT Data

As shown in Fig. 13, the Γ-valley valence bands are below the 6th highest valence bands for $\theta \leqslant 5.09°$ and zero displacement field $\varepsilon = 0$. Therefore, we only use the ±K-valley model [Eq. (B27)] in the fitting.

We fit the DFT bands at 3.89° in two ways. First, we set $V_2 = w_2 = 0$, which corresponds to the FH model. In this case, we manage to fit the top four valence bands (two in each valley) with the corresponding FH parameters in Table I, as shown in Fig. 4(a) along the high-symmetry lines and in Fig. 21(a) for the $\boldsymbol{k}$ points in the fundamental zone. Then, we allow nonzero V_2 and w_2, which means we add the effective-inversion-symmetric SH terms. We are now able to fit the top six valence bands (three in each valley) with the corresponding FH+SH parameters in Table I, as shown in Fig. 4(b) along the high-symmetry lines and in Fig. 21(b) for the $\boldsymbol{k}$ points in the fundamental zone.

As shown in Table I, the value of the effective mass m^* is similar to the monolayer and untwisted AA-stacking bilayer structures masses, as shown in Fig. 11. We note that all bands from the −K-valley continuum model are the same as those from the K-valley continuum model, due to the combination of the effective inversion and the TR symmetry. By comparing Figs. 22(a) and 22(b) to Fig. 2(c), we can see that the C_3 eigenvalues for the top 6 valence bands from the model are the same as those from the DFT calculation in both FH and FH+SH cases at 3.89°. At 3.89°, the Chern numbers of the top three bands in K valley are (1,–1,0) for the highest, the second highest and the third highest valence bands in both FH and FH+SH cases, respectively, which are consistent with the C_3 eigenvalues. The Chern numbers and the C_3 eigenvalues of the top two bands per valley are the same as those in Ref. [21] at 3.89°. The quantum geometry on the moiré Brillouin zone is shown in Figs. 23 and 24, with Chern number C and integrated Fubini-Study metric $G/2\pi$ (see Ref. [32]). We observe that the Berry curvature and Fubini-Study metric are more uniform in the first band than in the second (remote) band, with the second remote band showing a peak around Γ where the gap

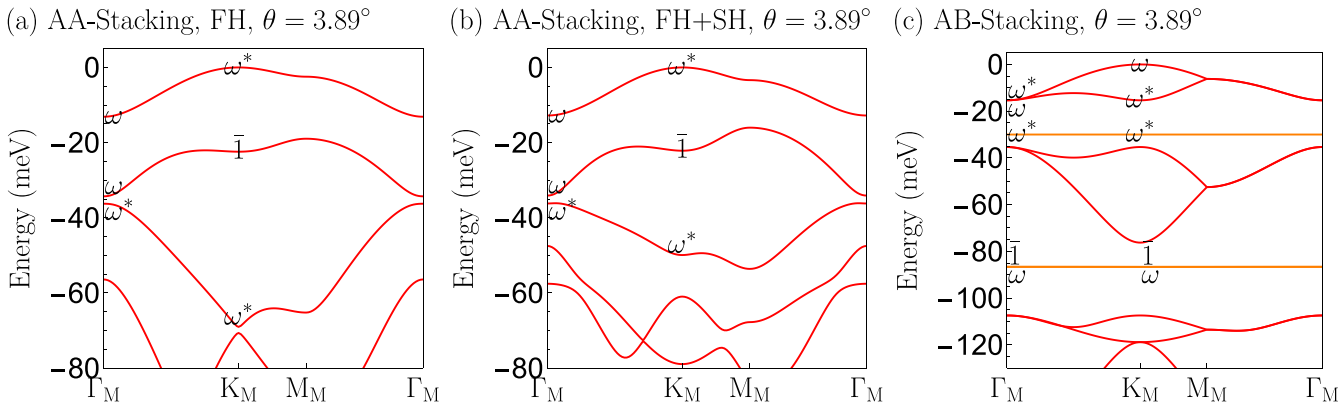

FIG. 22. The C_3 eigenvalues of the bands from the continuum model for twist angle 3.89° for (a) (b) AA stacking and (c) AB stacking. The plot in (a) is done with FH terms, while SH terms are added in (b). In (a) and (b), we only include C_3 eigenvalues for the K-valley bands (red) at Γ_M and K_M, since K'_M point is related to the K_M point by $C_{2y}\mathcal{T}$ and the $-$K valley can be obtained from the TR symmetry. In (c), we only include C_3 eigenvalues for the K-valley bands (red) and the Γ-valley spin-up bands at Γ_M and K_M, since K'_M point is related to the K_M point by the effective TR symmetry, and those for the $-$K-valley and the Γ-valley spin down can be obtained from the TR symmetry. Like in Fig. 2, $\omega = e^{i\pi/3}$ and $\bar{1} = -1$.

to the third band is smallest. Note also that the effect of second harmonic terms is weak on the first ($C = -1$) valence band, whereas the second valence band becomes significantly more strongly peaked at the Γ point and flatter elsewhere on the BZ. This can be understood from the decrease in the gap at Γ between the second and third valence bands due to the improved accuracy of the higher-harmonic model. Since the FH+SH model captures more bands than the FH model, we use the FH+SH model to discuss the phase diagram under the twist angle θ and the displacement field ε. Figure 5 in the main text shows the single-particle phase diagram for the top three valence bands in K valley. The highest valence band—most relevant to the filling factors where FCIs have been observed. The top valence K-valley band displays a Chern number $C = 1$ throughout the phase diagram at $\varepsilon = 0$, and shows a phase transition into a trivial insulator as ε is turned on. The phase diagram of the remote bands is richer, showing all Chern numbers $C = -2, \ldots, 2$ throughout the θ, ε plane. In particular, at $\varepsilon = 0$, the second and third top bands have a gap closing around $\theta = 4.2°$, which is close to the gap closing around 4.41° shown in Fig. 2.

APPENDIX C: SINGLE-PARTICLE CONTINUUM MODEL AND FITTING: AB-STACKING (2H-STACKING)

Replacing C_{2y} by C_{2x} in the generators of the symmetry group of AA-tMoTe$_2$ gives the generators of the symmetry group of AB-stacking tMoTe$_2$ (AB-tMoTe$_2$), which can be thought of as twisting the top layer of AA-tMoTe$_2$ by another 180°. The change of symmetry has a dramatic effect on the model. Since C_{2x} is local to the monolayer K point (unlike C_{2y}), it preserves the valley quantum in the moiré model.

Different from the AA-tMoTe$_2$, the DFT results show that the low-energy valence bands of AB-tMoTe$_2$ come from both the $\pm$K valleys and the Γ valleys. Therefore, in the following, we will review the continuum model for the $\pm$K valleys proposed in Ref. [43], and will also discuss the continuum model for the Γ valley.

1. AB-Stacking: ±K Valleys

In this part, we review the continuum model of AB-tMoTe$_2$ for the $\pm$K valleys proposed in Ref. [43]. The basis of the continuum model is still labeled by $c^\dagger_{\eta,l,\boldsymbol{r}}$ with $\eta = \pm$ labeling

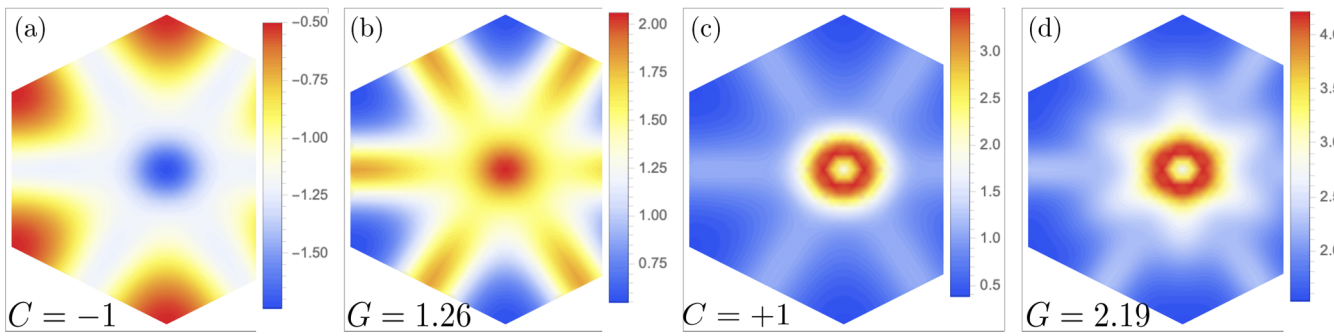

FIG. 23. Quantum geometry of the AA-stacked continuum FH model ($\theta = 3.89°$) in (B27) (without higher harmonics). (a) and (b) show the Berry curvature and Fubini-Study metric of the highest valence band, respectively, and (c) and (d) show the corresponding plots for the second highest valence band. The Chern numbers are $C = -1, +$ for (a) and (c) respectively, and integrated Fubini-Study metrics are marked.

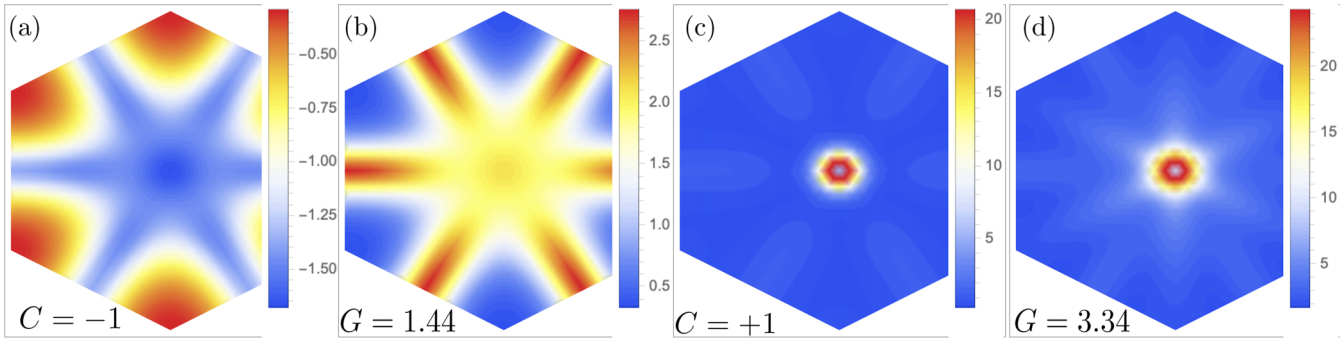

FIG. 24. Quantum geometry of the AA-stacked continuum model ($\theta = 3.89°$) including second-shell harmonics in (B27). (a) and (b) show the Berry curvature and Fubini-Study metric of the highest valence band, respectively, and (c) and (d) show the corresponding plots for the second highest valence band. The Chern numbers are $C = -1, +$ for (a) and (c) respectively.

the $\pm$K valleys (or equivalently spins), $l = t, b$ labels the layer, and $\boldsymbol{r}$ labels the position. Since AB-tMoTe$_2$ is given by rotating the top layer of AA-tMoTe$_2$ by an extra 180°, the two layers in one valley now have opposite spin. Then, the symmetry reps furnished by $c^\dagger_{\eta,l,\boldsymbol{r}}$ now read

$$\begin{aligned} C_3 c^\dagger_{\eta,l,\boldsymbol{r}} C_3^{-1} &= c^\dagger_{\eta,l,C_3\boldsymbol{r}} e^{-\mathrm{i}\eta(-)^l \frac{\pi}{3}}, \\ C_{2x} c^\dagger_{\eta,l,\boldsymbol{r}} C_{2x}^{-1} &= c^\dagger_{\eta,\bar{l},C_{2x}\boldsymbol{r}}(-\mathrm{i}), \\ \mathcal{T} c^\dagger_{\eta,l,\boldsymbol{r}} \mathcal{T}^{-1} &= c^\dagger_{-\eta,l,\boldsymbol{r}}(-\eta), \\ T_{\boldsymbol{R}_M} c^\dagger_{\eta,l,\boldsymbol{r}} T^{-1}_{\boldsymbol{R}_M} &= c^\dagger_{\eta,l,\boldsymbol{r}+\boldsymbol{R}_M} e^{-\mathrm{i}\eta \boldsymbol{R}_M \cdot \mathrm{K}_l}, \end{aligned} \tag{C1}$$

where $\bar{l} = b, t$ for $l = t, b$ respectively, K_l is the rotated K point of the lth layer, and $l = t, b$ respectively correspond to $l = 0, 1$ for $(-)^l$.

The general form of the single-particle Hamiltonian in the η valley reads

$$\begin{aligned} H^{AB}_{\eta,0} = \sum_l \int d^2 r c^\dagger_{\eta,l,\boldsymbol{r}} \left[\frac{\hbar^2 \nabla^2}{2m^*} + V_{\eta,l}(\boldsymbol{r}) \right] c_{\eta,l,\boldsymbol{r}} \\ + \left[\int d^2 r c^\dagger_{\eta,b,\boldsymbol{r}} t_\eta(\boldsymbol{r}) c_{\eta,t,\boldsymbol{r}} + \mathrm{H.c.} \right], \end{aligned} \tag{C2}$$

where $V_{\eta,l}(\boldsymbol{r})$ is real, and we have used the fact that the dispersion around $\pm$K in the monolayer MoTe$_2$ is quadratic and the fact that m^* is the same for two valleys and two layers owing to C_{2x} and $\mathcal{T}$ symmetries. The symmetry requirements on $V_{\eta,l}(\boldsymbol{r})$ and $t_\eta(\boldsymbol{r})$ read

$$\begin{aligned} C_3 &: V_{\eta,l}(C_3\boldsymbol{r}) = V_{\eta,l}(\boldsymbol{r}), \; t_\eta(C_3\boldsymbol{r}) = t_\eta(\boldsymbol{r}) e^{-\mathrm{i}\eta \frac{2\pi}{3}}, \\ C_{2x} &: V_{\eta,l}(C_{2x}\boldsymbol{r}) = V_{\eta,\bar{l}}(\boldsymbol{r}), \; t^\dagger_\eta(C_{2x}\boldsymbol{r}) = t_\eta(\boldsymbol{r}), \\ \mathcal{T} &: V_{-\eta,l}(\boldsymbol{r}) = V^*_{\eta,l}(\boldsymbol{r}), \; t_{-\eta}(\boldsymbol{r}) = t^*_\eta(\boldsymbol{r}), \\ T_{\boldsymbol{R}_M} &: V_{\eta,l}(\boldsymbol{r}+\boldsymbol{R}_M) = V_{\eta,l}(\boldsymbol{r}), \\ & \quad t_\eta(\boldsymbol{r}+\boldsymbol{R}_M) = t_\eta(\boldsymbol{r}) e^{\mathrm{i}\eta \boldsymbol{q}_1 \cdot \boldsymbol{R}_M}. \end{aligned} \tag{C3}$$

By keeping the first harmonics, the symmetry requirements lead to

$$\begin{aligned} V_{\eta,l}(\boldsymbol{r}) &= V e^{-\mathrm{i}\psi} \sum_{i=1,2,3} e^{\mathrm{i}\boldsymbol{g}_i \cdot \boldsymbol{r}} + V e^{\mathrm{i}\psi} \sum_{i=1,2,3} e^{-\mathrm{i}\boldsymbol{g}_i \cdot \boldsymbol{r}}, \\ t_\eta(\boldsymbol{r}) &= w \sum_{i=1,2,3} e^{\eta \mathrm{i}(i-1)\frac{2\pi}{3}} e^{-\eta \mathrm{i}\boldsymbol{q}_i \cdot \boldsymbol{r}}. \end{aligned} \tag{C4}$$

As can be checked from the expressions above, this lowest-harmonics model has effective antiunitary symmetry

$$\mathcal{K} c^\dagger_{\eta,l,\boldsymbol{r}} \mathcal{K}^{-1} = c^\dagger_{\eta,\bar{l},\boldsymbol{r}}, \tag{C5}$$

which flips the layers and acts like time reversal. This symmetry can be anticipated since the opposite layers have opposite spins in the twisted AB stacking. Adding displacement field would induce an energy difference between the two layers, i.e.,

$$H_{\eta,\varepsilon} = \int d^2 r \sum_l c^\dagger_{\eta,l,\boldsymbol{r}} c_{\eta,l,\boldsymbol{r}} (-)^l \frac{\varepsilon}{2}, \tag{C6}$$

which breaks the effective TR symmetry, as well as the C_{2x} symmetry. Including higher harmonics for the intralayer and interlayer terms may break the effective TR symmetry.

Similar to the discussion in Appendix B 1, we can express the Hamiltonian in the momentum space based on the Fourier transformation in Eq. (B22), leading to

$$H^{AB}_{\eta,0} = \sum_{\boldsymbol{k}} \sum_{\mathcal{Q},\mathcal{Q}' \in \mathcal{Q}} c^\dagger_{\eta,\boldsymbol{k},\mathcal{Q}} h^{AB}_{\eta,\mathcal{Q}\mathcal{Q}'}(\boldsymbol{k}) c_{\eta,\boldsymbol{k},\mathcal{Q}'}, \tag{C7}$$

where $c^\dagger_{\eta,\boldsymbol{k},\mathcal{Q}}$ is defined in Eq. (B25), and $\mathcal{Q}$ is defined right below Eq. (B24). Specifically, $h^{AB}_{\eta,\mathcal{Q}\mathcal{Q}'}(\boldsymbol{k})$ reads

$$\begin{aligned} h^{AB}_{\eta,\mathcal{Q}\mathcal{Q}'}(\boldsymbol{k}) = \delta_{\mathcal{Q}\mathcal{Q}'} \left(\frac{-\hbar^2 (\boldsymbol{k} - \mathcal{Q})^2}{2m^*} + \frac{\varepsilon}{2} \eta (-)^{\mathcal{Q}} \right) \\ + V \sum_{i=1}^{3} \sum_{\alpha=\pm} \left[e^{-\alpha \mathrm{i}\psi} \delta_{\mathcal{Q}+\alpha\boldsymbol{g}_i,\mathcal{Q}'} \right] \\ + w \sum_{i=1}^{3} \sum_{\alpha=\pm} \left[e^{\mathrm{i}\alpha \frac{2\pi(i-1)}{3}} \delta_{\mathcal{Q},\mathcal{Q}'+\alpha\boldsymbol{q}_i} \right], \end{aligned} \tag{C8}$$

where we have included the displacement field, and w is chosen to be real by picking the relative phase between the two layers, and $(-)^{\mathcal{Q}} = \pm 1$ for $\boldsymbol{Q} \in \{\boldsymbol{G}_M \pm \boldsymbol{q}_1\}$. We can see that without the displacement field, $h^{AB}_{\eta,\boldsymbol{Q}\boldsymbol{Q}'}(\boldsymbol{k})$ is the same for two valleys due to the effective time-reversal symmetry.

2. AB-Stacking: Γ Valley

In this part, we discuss the continuum model for the Γ valley. Thus, the basis of the moire model at Γ valley for AB-tMoTe$_2$ is the same as that for AA-tMoTe$_2$, which is labeled as $\psi^\dagger_{\boldsymbol{r},l,s}$ with l being the layer index and $s = \uparrow/\downarrow$ labeling the spin. The important difference is that, in the Γ valley, the momentum space lattice is triangular instead of a honeycomb at ±K points. The reps of the symmetries of the moiré model read

$$\begin{aligned} C_3\psi^\dagger_{\boldsymbol{r},l}C_3^{-1} &= \psi^\dagger_{C_3\boldsymbol{r},l}e^{-\mathrm{i}s_z\frac{\pi}{3}}, \\ C_{2x}\psi^\dagger_{\boldsymbol{r},l}(C_{2x})^{-1} &= \psi^\dagger_{C_{2x}\boldsymbol{r},\bar{l}}(-\mathrm{i}s_x), \\ \mathcal{T}\psi^\dagger_{\boldsymbol{r},l}\mathcal{T}^{-1} &= \psi^\dagger_{\boldsymbol{r},l}\mathrm{i}s_y, \\ T_{\boldsymbol{R}_M}\psi^\dagger_{\boldsymbol{r},l}T^{-1}_{\boldsymbol{R}_M} &= \psi^\dagger_{\boldsymbol{r}+\boldsymbol{R}_M,l}\,, \end{aligned} \tag{C9}$$

where s_x, s_y, s_z are the Pauli matrices for the spin index, and $\bar{l} = t, b$ if $l = b, t$.

After choosing the kinetic term in the continuum model as the intralayer spin-independent ∇^2 term similar to the AA-stacking case, the form of the continuum model reads

$$H_\Gamma = \int d^2r(\psi^\dagger_{\boldsymbol{r},b}\psi^\dagger_{\boldsymbol{r},t})h^{AB}_\Gamma(\boldsymbol{r})\begin{pmatrix}\psi^\dagger_{\boldsymbol{r},b}\\ \psi^\dagger_{\boldsymbol{r},t}\end{pmatrix}, \tag{C10}$$

where

$$h^{AB}_\Gamma(\boldsymbol{r}) = \begin{pmatrix}\frac{\hbar^2\nabla^2}{2m^*_\Gamma}s_0 + V_{\Gamma,b}(\boldsymbol{r}) & t_\Gamma(\boldsymbol{r}) \\ t^\dagger_\Gamma(\boldsymbol{r}) & \frac{\hbar^2\nabla^2}{2m^*_\Gamma}s_0 + V_{\Gamma,t}(\boldsymbol{r})\end{pmatrix} + E_\Gamma\,, \tag{C11}$$

$$\psi^\dagger_{\boldsymbol{r},l} = (\psi^\dagger_{\boldsymbol{r},l,\uparrow}, \psi^\dagger_{\boldsymbol{r},l,\downarrow})\,, \tag{C12}$$

t and b correspond to the top and bottom layers, respectively, E_Γ accounts for the energy difference between the Γ-valley and ±K-valley bands, and $V_{\Gamma,l}(\boldsymbol{r})$ and $t_\Gamma(\boldsymbol{r})$ are 2×2 matrix functions. Similar to the discussion for AA-tMoTe$_2$, we only include the terms up to the first harmonics in $V_{\Gamma,l}(\boldsymbol{r})$, and the zeroth harmonics in the $t_\Gamma(\boldsymbol{r})$. Then, combined with the symmetry reps in Eq. (C9), we arrive at

$$\begin{aligned} V_{\Gamma,l}(\boldsymbol{r}) &= \left[\sum_{i=1}^{3} V_\Gamma e^{\mathrm{i}\psi_\Gamma}e^{\mathrm{i}\boldsymbol{g}_i\cdot\boldsymbol{r}} + \sum_{i=1}^{3} V_\Gamma e^{-\mathrm{i}\psi_\Gamma}e^{-\mathrm{i}\boldsymbol{g}_i\cdot\boldsymbol{r}}\right]s_0, \\ t_\Gamma(\boldsymbol{r}) &= w_\Gamma s_0\,. \end{aligned} \tag{C13}$$

With the simplification in Eqs. (C11) and (C13), the Γ-valley continuum model in Eq. (C10) has an effective TR symmetry

$$\mathcal{K}\psi^\dagger_{\boldsymbol{r},l}\mathcal{K}^{-1} = \psi^\dagger_{\boldsymbol{r},\bar{l}} \tag{C14}$$

arising because of the spin-layer locking in the AB stacked structure. Similar to the ±K-valley case, adding displacement breaks the effective inversion symmetry, as well as the C_{2x} symmetry, where the displacement field term reads

$$H_{\Gamma,\varepsilon} = \int d^2r \sum_l \psi^\dagger_{\boldsymbol{r},l}\psi_{\boldsymbol{r},l}(-)^l\frac{\varepsilon}{2}\,. \tag{C15}$$

Including the first harmonics in $t_\Gamma(\boldsymbol{r})$ and the higher harmonics in $V_{\Gamma,l}(\boldsymbol{r})$ can also break the effective TR symmetry.

To express the Hamiltonian in the momentum space, we use the Fourier transformation of the basis, which reads

$$\psi^\dagger_{\boldsymbol{r},l,s} = \frac{1}{\sqrt{\mathcal{V}}}\sum_{\boldsymbol{k}}\sum_{\boldsymbol{G}_M} e^{-\mathrm{i}(\boldsymbol{k}-\boldsymbol{G}_M)\cdot\boldsymbol{r}}\psi^\dagger_{\boldsymbol{k}-\boldsymbol{G}_M,l,s}\,. \tag{C16}$$

As shown in Eq. (C13), the Hamiltonian has the same form in the two spin subspaces. As a result, $H_{\eta,0}$ reads

$$H^{AB}_{\Gamma,0} = \sum_{\boldsymbol{k},\boldsymbol{G}_M,\boldsymbol{G}'_M}\sum_{l,l',s}\psi^\dagger_{\boldsymbol{k}-\boldsymbol{G}_M,l,s}h^{AB}_{ll',\boldsymbol{G}_M\boldsymbol{G}'_M}(\boldsymbol{k})\psi_{\boldsymbol{k}-\boldsymbol{G}'_M,l',s}\,, \tag{C17}$$

where

$$\begin{aligned} h^{AB}_{ll,\boldsymbol{G}_M\boldsymbol{G}'_M}(\boldsymbol{k}) &= \left[-\frac{\hbar^2(\boldsymbol{k}-\boldsymbol{G})^2}{2m^*_\Gamma} + (-)^l\frac{\varepsilon}{2} + E_\Gamma\right]\delta_{\boldsymbol{G}_M,\boldsymbol{G}'_M} \\ &\quad + V_\Gamma\sum_{i=1}^{3}\left[e^{\mathrm{i}\psi_\Gamma}\delta_{\boldsymbol{G}_M,\boldsymbol{G}'_M-\boldsymbol{g}_i} + e^{-\mathrm{i}\psi_\Gamma}\delta_{\boldsymbol{G}_M,\boldsymbol{G}'_M+\boldsymbol{g}_i}\right] \\ h^{AB}_{l\bar{l},\boldsymbol{G}_M\boldsymbol{G}'_M}(\boldsymbol{k}) &= w_\Gamma\delta_{\boldsymbol{G}_M,\boldsymbol{G}'_M}\,, \end{aligned} \tag{C18}$$

and we have included the displacement field.

3. AB-Stacking: Fitting to the DFT Data

As shown in Fig. 18, both the ±K-valley and the Γ-valley valence bands are close to the charge neutrality. Therefore, we only use both the ±K-valley model [Eq. (C8)] and the Γ-valley model [Eq. (C18)] in the fitting. We mainly fit to the top four valence bands (two in each valley) for the ±K-valley model [Eq. (C8)] and top six valence bands in the Γ valley; similar to the case of AA stacking, we fit to the DFT bands at $\theta = 3.89°$. The good match between those DFT bands and those from the models is shown in Figs. 4(b) and 21(b), where the model parameter values obtained from the fitting are shown in Table II. In the following, we provide more details for the fitting for the ±K-valley model [Eq. (C8)] and the Γ-valley model, separately.

a. ±K-valley

We choose the value of m^* to be the effective mass ($\sim 0.62m_e$) of the top valence band for the untwisted AB-stacking bilayer structure around K, as shown in Fig. 11. The value of effective mass around K is similar for the monolayer MoTe$_2$ and the untwisted AB-stacking bilayer MoTe$_2$.

Owing to the effective TR symmetry Eq. (C5) and the true microscopic TR symmetry, the −K valley has the same bands as the +K valley, leading to at least double degeneracy of each band from the ±K-valley continuum model. Along the $\Gamma_M - M_M$ line, the bands from the ±K-valley continuum model have fourfold degeneracy because we set the interlayer coupling to be zero ($w = 0$), which is justified by the nearly fourfold degeneracy for the top four DFT valence bands along

$\Gamma_M - M_M$, as shown in Fig. 4. The small $w = 0$ can be understood as the follows. The two layers in one valley now have opposite spin; owing to the spin U(1) symmetry for the low-energy states near ±K valleys in monolayer $MoTe_2$, we expect the spin $U(1)$ symmetry is approximately preserved in $tMoTe_2$, which means the interlayer coupling is very small for the AB stacking. The zero interlayer coupling makes the eigenstates have well-defined valley and layer/spin. As a result, the two states with the same spin are degenerate at K_M, since the combination of the effective TR symmetry and the TR symmetry leaves the spin invariant. This is consistent with the DFT result in Fig. 19. In Fig. 22(b), we show the C_3 eigenvalues of the bands from the ±K-valley model, which is consistent with Fig. 19 and which shows that each of the top four valence bands is in the $A_1@1a$ atomic limit, representing an atomic s orbital at the origin of the center of moiré Wigner-Seitz unit cell.

b. Γ*-valley*

In our fitting, we choose the values of m^*_Γ to be the effective mass of the top valence band for the untwisted AB-stacking bilayer structure around Γ in Fig. 11. (There is some anisotropy, so in practice we round the effective mass to $10m_e$. The resulting bands show little dependence on the precise value of the mass.) Although the value of effective mass around K is similar for the monolayer $MoTe_2$ and the untwisted AB-stacking bilayer $MoTe_2$, it differs a lot around Γ as shown in Fig. 11. The change of the Γ effective mass from the monolayer $MoTe_2$ to the untwisted AB-stacking bilayer $MoTe_2$ comes from the interlayer coupling between the monolayer conduction and valence bands. This effect can only be taken into account in the Γ-valley continuum model by changing the value of m^*, since the the monolayer conduction bands are not explicitly included in the Γ-valley continuum model. Therefore, we do not take the values of m^*_Γ from the monolayer; instead, we round it to the Γ effective mass in the untwisted AB-stacking bilayer $MoTe_2$. Furthermore, we determine the interlayer coupling w_Γ for the Γ-valley model based on the interlayer coupling at Γ for the untwisted AB-stacking bilayer structure, as shown in Fig. 11.

The Γ-valley continuum model in Eq. (C10) preserves spin U(1). Owing to the effective TR symmetry Eq. (C5) and the TR symmetry, the spin-down bands are the same as the spin-up bands, leading to at least double degeneracy of each band from the Γ-valley continuum model. As shown in Fig. 22(c), the C_3 eigenvalues of the bands from the ±K-valley model are consistent with the DFT calculation shown in Fig. 19.

The Γ-valley bands are extremely flat as shown in Fig. 4. To understand it, first we note that the interlayer coupling w_Γ is very large (shown in Table II), allowing us to project the model into $\psi^\dagger_{\boldsymbol{r},s} = \frac{1}{\sqrt{2}}(\psi^\dagger_{t,\boldsymbol{r},s} + \psi^\dagger_{b,\boldsymbol{r},s})$ basis. (Note that we have rotated the relative phase between $\psi^\dagger_{t,\boldsymbol{r},s}$ and $\psi^\dagger_{b,\boldsymbol{r},s}$ to make w_Γ positive.) Then, the Γ-valley continuum model in Eq. (C11) with zero displacement field becomes

$$\tilde{h}^{AB}(\boldsymbol{r}) = \frac{\hbar^2\nabla^2}{2m^*_\Gamma} + w_\Gamma + E_\Gamma + V_\Gamma(\boldsymbol{r}) \,, \tag{C19}$$

where the spin degeneracy is implicit, and

$$\begin{aligned} V_\Gamma(\boldsymbol{r}) &= \sum_{i=1}^{3} V_\Gamma e^{i\psi_\Gamma} e^{i\boldsymbol{g}_i\cdot\boldsymbol{r}} + \sum_{i=1}^{3} V_\Gamma e^{-i\psi_\Gamma} e^{-i\boldsymbol{g}_i\cdot\boldsymbol{r}} \\ &= 2V_\Gamma \sum_{i=1}^{3} \cos(\boldsymbol{g}_i\cdot\boldsymbol{r} + \psi_\Gamma) \,. \end{aligned} \tag{C20}$$

For convenience of the discussion, let us consider the hole Hamiltonian

$$-\tilde{h}^{AB}(\boldsymbol{r}) = -\frac{\hbar^2\nabla^2}{2m^*_\Gamma} - w_\Gamma - E_\Gamma - V_\Gamma(\boldsymbol{r}) \,; \tag{C21}$$

the highest bands of $\tilde{h}^{AB}(\boldsymbol{r})$ are lowest bands of $-\tilde{h}^{AB}(\boldsymbol{r})$. With $\psi = 0$ and $V_\Gamma > 0$ in Table II, the minimum of the potential $-V_\Gamma(\boldsymbol{r})$ is at the moiré lattice points $\boldsymbol{r} = \boldsymbol{R}_M$, which has value $-V_\Gamma(\boldsymbol{R}_M) = -6V_\Gamma$. To go from one moiré lattice point $\boldsymbol{R}_M$ to its nearest neighbors $\boldsymbol{R}_M$ to $\boldsymbol{R}_M + \boldsymbol{a}_{M,1}$, the minimum energy barrier is $8V_\Gamma = 576$ meV. The energy barrier is much larger than the kinetic energy term estimated for $m^*_\Gamma = 10m_e$: $\frac{\hbar^2|\boldsymbol{q}_1|^2}{2m^*_\Gamma} = 2.49$ meV. As a result, the lowest-energy states of $-\tilde{h}^{AB}(\boldsymbol{r})$ are bounded around $\boldsymbol{R}_M$; then we can expand $-V_\Gamma(\boldsymbol{r})$ around $\boldsymbol{R}_M$ to the second order of $\boldsymbol{r} - \boldsymbol{R}_M$ (the first order of $\boldsymbol{r} - \boldsymbol{R}_M$ is forbidden by the C_3 symmetry),

$$\begin{aligned} &-V_\Gamma(\boldsymbol{r}) \\ &= -V(\boldsymbol{R}_M) + V_\Gamma \sum_i (\boldsymbol{g}_i\cdot(\boldsymbol{r} - \boldsymbol{R}_M))^2 + O(|\boldsymbol{r} - \boldsymbol{R}_M|^4) \\ &= -6V_\Gamma + \frac{3}{2}V_\Gamma|b_{M,1}|^2|\boldsymbol{r} - \boldsymbol{R}_M|^2 + O(|\boldsymbol{r} - \boldsymbol{R}_M|^4) \end{aligned} \tag{C22}$$

for small $|\boldsymbol{r} - \boldsymbol{R}_M|$. Then, the effective Hamiltonian for those lowest-energy states can be approximated by an array of decoupled harmonic oscillators centered around the $\boldsymbol{R}_M$. Harmonic oscillators centered around different $\boldsymbol{R}_M$'s are related by the moiré lattice translations, and thus have the same spectrum. As a result, we have a set of flat bands with energies given by solving (choosing $\boldsymbol{R}_M = 0$ as an example)

$$h^{AB}_{\Gamma,\text{harmonic}}(\boldsymbol{r}) = -\frac{\hbar^2\nabla^2}{2m^*_\Gamma} + w_\Gamma - E_\Gamma + \frac{3}{2}V_\Gamma|b_{M,1}|^2|\boldsymbol{r}|^2, \tag{C23}$$

which yields

$$E^{AB}_{\Gamma,\text{harmonic}}(n_x, n_y) = w_\Gamma E_\Gamma - \hbar\omega - (n_x + n_y)\frac{\hbar\omega}{2} \,, \tag{C24}$$

where $n_x, n_y = 0, 1, 2, 3, \ldots$ and

$$\hbar\omega = \hbar\sqrt{\frac{3V_\Gamma|b_{M,1}|^2}{m^*_\Gamma}} = 56.76 \text{ meV} \,. \tag{C25}$$

This spectrum tells us (i) after including spin, the highest valence band $(n_x, n_y) = (0, 0)$ is doubly degenerate and the second highest valence band $(n_x, n_y) = (1, 0), (0, 1)$ is four-fold degenerate, (ii) the highest and second valence bands have spinless C_3 eigenvalues 1 and $e^{\pm i2\pi/3}$, respectively, and (iii) the energy difference between the highest valence band and the second highest valence is $\hbar\omega = 56.76$ meV. All of the features are consistent with the DFT calculation. In

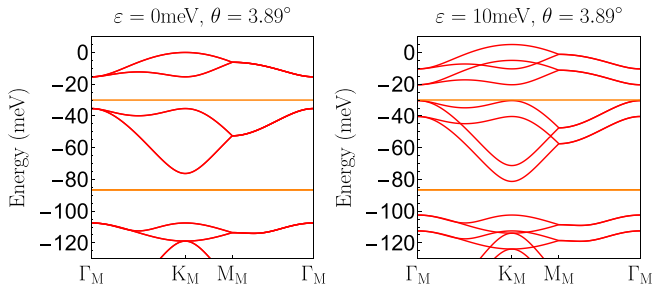

FIG. 25. The evolution of the bands of the AB-stacking $t\mathrm{MoTe}_2$ based on the continuum model with parameter values in Table II. The red and orange lines come from the ±K valley and the Γ valley, respectively.

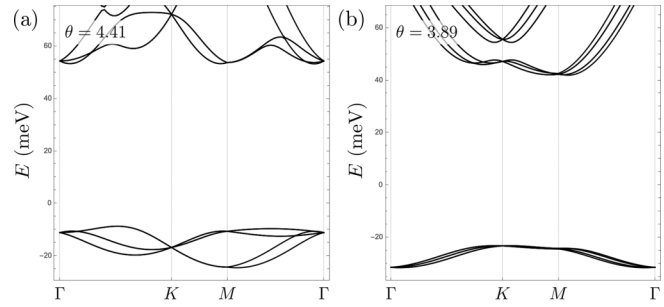

FIG. 26. Conduction bands of the AB-stacked $t\mathrm{MoTe}_2$using $s = 0.68$ for $\theta = 4.41°$ (a) and $\theta = 3.89°$ (b). Note that there are six low energy bands (per valley) in each case, although in (b) they are nearly degenerate. The features of the band structures are well matched by the *lowest* 12 bands in *ab initio* in Fig. 18 between both angles. One should note that the *ab initio* band structures also show bands from the K valley at slightly higher energies.

Fig. 21(b), the flat bands from the Γ valley around −30 meV and −86 meV have near twofold degeneracy (spinless C_3 eigenvalue 1) and near fourfold degeneracy (spinless C_3 eigenvalue $e^{\pm i2\pi/3}$), and the energy difference between them is 56.74 meV.

At last, we discuss the effect of the displacement field. As shown in Fig. 25, the effect of the displacement field on the ±K-valley bands are just to shift the bands from different layers relative to each other, since the valley is a good quantum number due to the zero interlayer coupling (see Table II). On the other hand, the effect of the displacement field on the low-energy Γ-valley bands is negligible, which is consistent with the fact that the very large interlayer (see Table II) coupling make the eigenstates equally distributed between the two layers.

4. AB-Stacking Electron Bands

For completeness, we also briefly discuss the conduction bands although they are not the focus of this paper. The Fermi surface of the electron bands in the untwisted structure is more complex than the valence bands (see Fig. 11), exhibiting an electron-like pocket along the ΓK line, which is below the ±K-valley states in both bilayer configurations. Let us denote the center of the pocket by $s\mathbf{K}$, $s \sim 0.7$, along with its C_3- and $\mathcal{T}$-related partners. Thus in total there are six pockets in the untwisted BZ. Figure 11 shows that there are only two low-energy states (spin ↑ and spin ↓) at $s\mathbf{K}$, unlike the four total states (spin ↑ and spin ↓ in both layers) at $\mathbf{K}$. This can be understood from the simple model $\frac{\mathbf{k}^2}{2m_p}\tau_0 - w_p\tau_1$ where τ is a layer Pauli matrix, $\mathbf{k}$ is the momentum measured from the Fermi pockets with mass m_p, and w_p is the effective interlayer coupling, which splits the two layers into bonding/antibonding states with energy $\mathbf{k}^2/2m_p \mp w_p$.

Comparing the monolayer and bilayer band structures in Fig. 11, we estimate $w_p \sim 200$ meV. Since this is a large hybridization that removes the antibonding state from the low-energy bands, an effective moiré model on the bonding state is

$$H_{\mathbf{p}_i} = \frac{(-i\nabla + \mathbf{p}_i)^2}{2m_p} + V_p(\mathbf{r}), \tag{C26}$$

where $V_p(\mathbf{r})$ is a moiré-periodic potential and $\mathbf{p}_j = R(\frac{2\pi j}{6})s\mathbf{K}$ is the center of the pocket. Note that the momentum-space origin of $H_{\mathbf{p}_i}$, $\mathbf{p}_i$ mod $\mathbf{G}_M$, depends sensitively on the value of s and θ and is generically off the moiré high-symmetry points. The resulting shifting and folding of the $\mathbf{p}_i$ pocket onto the moiré BZ explains the nonuniform evolution of the conduction bands in Figs. 13 and 18 as a function of twist angle. In comparison, the valence bands, which originate at the K points converge smoothly.

We show band structures at $\theta = 4.41°, 3.89°$ in Fig. 26 using $s = .68$ as an example to illustrate the strong effect of $\mathbf{p}_i$. Here we chose similar parameters to the K-valley model in Table II, namely $m_p = 0.62m_e$, $V = 26.5$ meV (half the K-valley value), and $\psi = -52°$. We see that Fig. 26 qualitatively matches the lowest bands in Fig. 18. Since the only symmetry preserved at $\mathbf{p}_i$ is C_{2x}, symmetry does not strongly constrain the form of Eq. (C26). Nevertheless, the minimal model here successfully reproduces many features of the bands.

[1] T. Neupert, L. Santos, C. Chamon, and C. Mudry, Phys. Rev. Lett. **106**, 236804 (2011).
[2] D. N. Sheng, Z.-C. Gu, K. Sun, and L. Sheng, Nat. Commun. **2**, 389 (2011).
[3] N. Regnault and B. A. Bernevig, Phys. Rev. X **1**, 021014 (2011).
[4] E. Tang, J.-W. Mei, and X.-G. Wen, Phys. Rev. Lett. **106**, 236802 (2011).
[5] K. Sun, Z. Gu, H. Katsura, and S. Das Sarma, Phys. Rev. Lett. **106**, 236803 (2011).
[6] E. J. Bergholtz and Z. Liu, Int. J. Mod. Phys. B **27**, 1330017 (2013).
[7] S. A. Parameswaran, R. Roy, and S. L. Sondhi, C. R. Phys. **14**, 816 (2013).
[8] A. Abouelkomsan, Z. Liu, and E. J. Bergholtz, Phys. Rev. Lett. **124**, 106803 (2020).

[9] P. J. Ledwith, G. Tarnopolsky, E. Khalaf, and A. Vishwanath, Phys. Rev. Res. **2**, 023237 (2020).

[10] C. Repellin and T. Senthil, Phys. Rev. Res. **2**, 023238 (2020).

[11] D. Parker, P. Ledwith, E. Khalaf, T. Soejima, J. Hauschild, Y. Xie, A. Pierce, M. P. Zaletel, A. Yacoby, and A. Vishwanath, arXiv:2112.13837.

[12] P. Wilhelm, T. C. Lang, and A. M. Läuchli, Phys. Rev. B **103**, 125406 (2021).

[13] Y. Sheffer and A. Stern, Phys. Rev. B **104**, L121405 (2021).

[14] H. Li, U. Kumar, K. Sun, and S.-Z. Lin, Phys. Rev. Res. **3**, L032070 (2021).

[15] V. Crépel and L. Fu, Phys. Rev. B **107**, L201109 (2023).

[16] J. Dong, J. Wang, and L. Fu, arXiv:2208.10516.

[17] A. Abouelkomsan, K. Yang, and E. J. Bergholtz, Phys. Rev. Res. **5**, L012015 (2023).

[18] N. Morales-Durán, J. Wang, G. R. Schleder, M. Angeli, Z. Zhu, E. Kaxiras, C. Repellin, and J. Cano, Phys. Rev. Res. **5**, L032022 (2023).

[19] Y.-M. Wu, D. Shaffer, Z. Wu, and L. H. Santos, Phys. Rev. B **109**, 115111 (2024).

[20] A. P. Reddy, F. F. Alsallom, Y. Zhang, T. Devakul, and L. Fu, Phys. Rev. B **108**, 085117 (2023).

[21] C. Wang, X.-W. Zhang, X. Liu, Y. He, X. Xu, Y. Ran, T. Cao, and D. Xiao, Phys. Rev. Lett. **132**, 036501 (2024).

[22] W.-X. Qiu, B. Li, X.-J. Luo, and F. Wu, Phys. Rev. X **13**, 041026 (2023).

[23] J. Dong, J. Wang, P. J. Ledwith, A. Vishwanath, and D. E. Parker, Phys. Rev. Lett. **131**, 136502 (2023).

[24] H. Goldman, A. P. Reddy, N. Paul, and L. Fu, Phys. Rev. Lett. **131**, 136501 (2023).

[25] H. Yu, M. Chen, and W. Yao, Natl. Sci. Rev. **7**, 12 (2020).

[26] N. Morales-Durán, N. Wei, and A. H. MacDonald, Phys. Rev. Lett. **132**, 096602 (2024).

[27] A. P. Reddy and L. Fu, Phys. Rev. B **108**, 245159 (2023).

[28] X.-Y. Song, Y.-H. Zhang, and T. Senthil, Phys. Rev. B **109**, 085143 (2024).

[29] C. Xu, J. Li, Y. Xu, Z. Bi, and Y. Zhang, PNAS **121**, e2316749121 (2024).

[30] T. Wang, T. Devakul, M. P. Zaletel, and L. Fu, arXiv:2306.02501.

[31] X. Liu, C. Wang, X.-W. Zhang, T. Cao, and D. Xiao, arXiv:2308.07488.

[32] J. Yu, J. Herzog-Arbeitman, M. Wang, O. Vafek, B. A. Bernevig, and N. Regnault, Phys. Rev. B **109**, 045147 (2024).

[33] A. Abouelkomsan, A. P. Reddy, L. Fu, and E. J. Bergholtz, Phys. Rev. B **109**, L121107 (2024).

[34] J. Cai, E. Anderson, C. Wang, X. Zhang, X. Liu, W. Holtzmann, Y. Zhang, F. Fan, T. Taniguchi, K. Watanabe *et al.*, Nature (London) **622**, 63 (2023).

[35] Y. Zeng, Z. Xia, K. Kang, J. Zhu, P. Knüppel, C. Vaswani, K. Watanabe, T. Taniguchi, K. F. Mak, and J. Shan, Nature (London) **622**, 69 (2023).

[36] H. Park, J. Cai, E. Anderson, Y. Zhang, J. Zhu, X. Liu, C. Wang, W. Holtzmann, C. Hu, Z. Liu *et al.*, Nature (London) **622**, 74 (2023).

[37] F. Xu, Z. Sun, T. Jia, C. Liu, C. Xu, C. Li, Y. Gu, K. Watanabe, T. Taniguchi, B. Tong *et al.*, Phys. Rev. X **13**, 031037 (2023).

[38] Z. Lu, T. Han, Y. Yao, A. P. Reddy, J. Yang, J. Seo, K. Watanabe, T. Taniguchi, L. Fu, and L. Ju, arXiv:2309.17436.

[39] Y. Cao, V. Fatemi, A. Demir, S. Fang, S. L. Tomarken, J. Y. Luo, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, E. Kaxiras *et al.*, Nature (London) **556**, 80 (2018).

[40] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Nature (London) **556**, 43 (2018).

[41] E. M. Spanton, A. A. Zibrov, H. Zhou, T. Taniguchi, K. Watanabe, M. P. Zaletel, and A. F. Young, Science **360**, 62 (2018).

[42] Y. Xie, A. T. Pierce, J. M. Park, D. E. Parker, E. Khalaf, P. Ledwith, Y. Cao, S. H. Lee, S. Chen, P. R. Forrester *et al.*, Nature (London) **600**, 439 (2021).

[43] F. Wu, T. Lovorn, E. Tutuc, I. Martin, and A. H. MacDonald, Phys. Rev. Lett. **122**, 086402 (2019).

[44] B. Li, W.-X. Qiu, and F. Wu, Phys. Rev. B **109**, L041106 (2024).

[45] T. Li, S. Jiang, B. Shen, Y. Zhang, L. Li, Z. Tao, T. Devakul, K. Watanabe, T. Taniguchi, L. Fu, J. Shan, and K. F. Mak, Nature (London) **600**, 641 (2021).

[46] M. Yankowitz and K. F. Mak, APL Mater. **10**, 080401 (2022).

[47] W. Zhao, K. Kang, L. Li, C. Tschirhart, E. Redekop, K. Watanabe, T. Taniguchi, A. Young, J. Shan, and K. F. Mak, Nat. Phys. **20**, 275 (2024).

[48] K. F. Mak and J. Shan, Nat. Nanotechnol. **17**, 686 (2022).

[49] P. Mai, E. W. Huang, J. Yu, B. E. Feldman, and P. W. Phillips, npj Quantum Mater. **8**, 14 (2023).

[50] K. Uchida, S. Furuya, J.-I. Iwata, and A. Oshiyama, Phys. Rev. B **90**, 155451 (2014).

[51] D. S. Koda, F. Bechstedt, M. Marques, and L. K. Teles, J. Phys. Chem. C **120**, 10895 (2016).

[52] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).

[53] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[54] G. Kresse and J. Hafner, J. Phys.: Condens. Matter **6**, 8245 (1994).

[55] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[56] D. Puotinen and R. E. Newnham, Acta Crystallogr. **14**, 691 (1961).

[57] C. Fang, M. J. Gilbert, and B. A. Bernevig, Phys. Rev. B **86**, 115112 (2012).

[58] J. Gao, Q. Wu, C. Persson, and Z. Wang, Comput. Phys. Commun. **261**, 107760 (2021).

[59] Y. Xu, K. Kang, K. Watanabe, T. Taniguchi, K. F. Mak, and J. Shan, Nat. Nanotechnol. **17**, 934 (2022).

[60] Y. Xu, S. Liu, D. A. Rhodes, K. Watanabe, T. Taniguchi, J. Hone, V. Elser, K. F. Mak, and J. Shan, Nature (London) **587**, 214 (2020).

[61] C. R. Kometter, J. Yu, T. Devakul, A. P. Reddy, Y. Zhang, B. A. Foutty, K. Watanabe, T. Taniguchi, L. Fu, and B. E. Feldman, Nat. Phys. **19**, 1861 (2023).

[62] N. Morales-Durán, P. Potasz, and A. H. MacDonald, Phys. Rev. B **107**, 235131 (2023).

[63] Q. Tan, A. Rasmita, Z. Zhang, H. Cai, X. Cai, X. Dai, K. Watanabe, T. Taniguchi, A. H. MacDonald, and W. Gao, Nat. Mater. **22**, 605 (2023).

[64] N. Morales-Durán, N. C. Hu, P. Potasz, and A. H. MacDonald, Phys. Rev. Lett. **128**, 217202 (2022).

[65] B. A. Foutty, C. R. Kometter, T. Devakul, A. P. Reddy, K. Watanabe, T. Taniguchi, L. Fu, and B. E. Feldman, arXiv:2304.09808.
[66] C. Setty, F. Xie, S. Sur, L. Chen, M. G. Vergniory, and Q. Si, arXiv:2309.14340.
[67] B. A. Bernevig, B. Lian, A. Cowsik, F. Xie, N. Regnault, and Z.-D. Song, Phys. Rev. B **103**, 205415 (2021).
[68] Z.-D. Song and B. A. Bernevig, Phys. Rev. Lett. **129**, 047601 (2022).
[69] N. Bultinck, E. Khalaf, S. Liu, S. Chatterjee, A. Vishwanath, and M. P. Zaletel, Phys. Rev. X **10**, 031034 (2020).
[70] M. O. Soldini, N. Astrakhantsev, M. Iraola, A. Tiwari, M. H. Fischer, R. Valentí, M. G. Vergniory, G. Wagner, and T. Neupert, Phys. Rev. B **107**, 245145 (2023).
[71] J. Herzog-Arbeitman, B. A. Bernevig, and Z.-D. Song, arXiv:2212.00030.
[72] P. Mai, B. E. Feldman, and P. W. Phillips, Phys. Rev. Res. **5**, 013162 (2023).
[73] N. Wagner, L. Crippa, A. Amaricci, P. Hansmann, M. Klett, E. König, T. Schäfer, D. Di Sante, J. Cano, A. Millis *et al.*, Nat Commun **14**, 7531 (2023).
[74] P. Mai, J. Zhao, B. E. Feldman, and P. W. Phillips, Nat. Commun. **14**, 5999 (2023).
[75] J. K. Ding, L. Yang, W. O. Wang, Z. Zhu, C. Peng, P. Mai, E. W. Huang, B. Moritz, P. W. Phillips, B. E. Feldman *et al.*, arXiv:2309.07876.
[76] Y. Fu, J. H. Wilson, and J. H. Pixley, Phys. Rev. B **104**, L041106 (2021).
[77] V. Crépel, D. Guerci, J. Cano, J. Pixley, and A. Millis, Phys. Rev. Lett. **131**, 056001 (2023).
[78] X. Hu, T. Hyart, D. I. Pikulin, and E. Rossi, Phys. Rev. Lett. **123**, 237002 (2019).
[79] F. Xie, Z. Song, B. Lian, and B. A. Bernevig, Phys. Rev. Lett. **124**, 167002 (2020).
[80] A. Julku, T. J. Peltonen, L. Liang, T. T. Heikkilä, and P. Törmä, Phys. Rev. B **101**, 060505(R) (2020).
[81] E. Rossi, Curr. Opin. Solid State Mater. Sci. **25**, 100952 (2021).
[82] J. Herzog-Arbeitman, Y. Wang, J. Liu, P. M. Tam, Z. Qi, Y. Jia, D. K. Efetov, O. Vafek, N. Regnault, H. Weng, Q. Wu, B. A. Bernevig, and J. Yu, following paper, Phys. Rev. B **109**, 205122 (2024).
[83] S. Olin, E. Jmukhadze, A. H. MacDonald, and W.-C. Lee, arXiv:2310.17824.
[84] R. Jinnouchi, F. Karsai, and G. Kresse, Phys. Rev. B **100**, 014105 (2019).
[85] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, Nat. Commun. **13**, 2453 (2022).
[86] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dulak, J. Friis, M. N. Groves, B. Hammer, C. Hargus *et al.*, J. Phys.: Condens. Matter **29**, 273002 (2017).
[87] U. Herath, P. Tavadze, X. He, E. Bousquet, S. Singh, F. Muñoz, and A. H. Romero, Comput. Phys. Commun. **251**, 107080 (2020).
[88] R. Bistritzer and A. H. MacDonald, Proc. Natl. Acad. Sci. USA **108**, 12233 (2011).
[89] Z.-D. Song, Z. Wang, W. Shi, G. Li, C. Fang, and B. A. Bernevig, Phys. Rev. Lett. **123**, 036401 (2019).
[90] J. Herzog-Arbeitman, A. Chew, and B. A. Bernevig, Phys. Rev. B **106**, 085140 (2022).
[91] D. Xiao, G.-B. Liu, W. Feng, X. Xu, and W. Yao, Phys. Rev. Lett. **108**, 196802 (2012).