# Long-Range Machine Learning of Electron Density for Twisted Bilayer Moiré Materials

Zekun Lou [1], Alan M. Lewis [2] and Mariana Rossi [1,3,*]

[1] *MPI for the Structure and Dynamics of Matter, Luruper Chaussee 149, 22761 Hamburg, Germany*

[2] *Department of Chemistry, University of York, York YO10 5DD, United Kingdom*

[3] *Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

(Received 12 February 2026; accepted 22 July 2026; published 27 August 2026)

Moiré superlattices in two-dimensional materials exhibit rich quantum phenomena, but *ab initio* modeling of these systems remains computationally prohibitive. Existing machine learning methods for accelerating density-functional theory can target the prediction of different quantities and often rely on the locality assumption. Here, we train a Gaussian process regression model for symmetry-adapted learning of three-dimensional electron densities exclusively on the electron densities of small displaced bilayer structures and then extrapolate electron density prediction to the large supercells required to describe small twist angles between these bilayers. We show the necessity of long-range descriptors to yield reliable band structures and electrostatic properties of large twisted bilayer structures when these are derived from predicted densities. We demonstrate that the choice of descriptor determines the distribution of residual density errors, which in turn affects the downstream electronic properties. We apply our models to twisted bilayer graphene, hexagonal boron nitride, and transition metal dichalcogenides, focusing on the model's capacity to predict complex phenomena, including flat-band formation, bandwidth narrowing, domain wall electric fields, and spin-orbit coupling effects. Beyond moiré materials, this approach provides a general methodology for electronic structure prediction in large-scale systems with substantial long-range phenomena related to nonlocal geometric information.

DOI: 10.1103/4575-9cmx

## I. INTRODUCTION

The electronic structure of two-dimensional (2D) materials exhibits significant quantum confinement effects, making them different from their bulk counterparts. When two or more monolayers are stacked with a small twist angle, large moiré superlattices emerge with extended periodic interlayer modulation, inducing dramatic changes in electronic properties. Such engineered structures have unveiled a rich landscape of quantum phases, including unconventional superconductivity, correlated insulating states, ferroelectricity, ferromagnetism, and nontrivial topology, making moiré materials a highly tunable platform for exploring these phenomena and potential applications in quantum devices [1–8].

While density-functional theory (DFT) remains the cornerstone approach for electronic structure calculations, its computational cost limits direct simulations of large moiré-scale systems. Continuum models and tight-binding approaches that fit parameters to DFT data have been developed to capture the essential low-energy physics at significantly reduced cost [9–12]. However, their parametrization and fixed functional form limit transferability. For example, for twisted bilayer (TB) transition metal dichalcogenides (TMDCs), continuum model parameters vary substantially across works targeting different twist angles and physical phenomena [13,14], restricting their predictive power for unexplored configurations or physics beyond the fitted data. Tight-binding models also often require refitting when applied to different stacking configurations or materials [12,15]. These limitations highlight the advantages that an *ab initio* based efficient method for accurate large-scale moiré simulation could have.

Machine learning (ML) techniques that predict a system's electronic structure offer a promising path forward by learning directly from complete *ab initio* data and predicting unseen structures, avoiding the fixed analytical forms and the fitting limitations of traditional models, while maintaining access to the full set of ground-state electronic properties. For DFT-based methods (ML-DFT), two main ML approaches have been explored: Hamiltonian-based methods that directly predict real-space Hamiltonian matrix elements [16–24], and density-based methods that predict the electron density distribution and derive electronic properties [25–35].

State-of-the-art Hamiltonian-based methods [18,19] achieved impressive sub-meV accuracy for Hamiltonian matrix elements in training datasets of 2D materials and medium-sized twisted bilayers [36]. They exploit the nearsightedness principle of electronic matter [37] to allow efficient subquadratic scaling with system sizes. This assumption can limit their applicability to systems with important long-range interactions, such as systems with weak electronic screening, significant Coulomb interactions, and charge rearrangements [38]. Direct prediction of the electron density offers an alternative ML approach, since the density uniquely determines all ground-state properties in the DFT formalism [39]. In

*Contact author: mariana.rossi@mpsd.mpg.de

particular, using the density fitting (DF) technique [40,41] to describe the electron density, as opposed to representing it on a grid, has been shown to yield a storage-efficient and transferable model with excellent accuracy in representing both core and bonding regions of real space [25,27]. Still, the challenge of the locality assumption for representing atomic environments persists, and the density fitting approach can introduce noise, which further complicates the intricate balance between density accuracy in training and the accuracy of downstream electronic properties.

Our work addresses these challenges by extending the SALTED (symmetry-adapted learning of three-dimensional electron densities) model [25,27], which leverages a density fitting technique [40], an equivariant kernel method [42,43], and optimized Gaussian process regression (GPR) with prediction cost linear in training set size [27]. We introduce (1) long-range descriptors to capture nonlocal structural features, (2) a numerical stabilization technique to reduce DF noise, and (3) explicit validation of downstream electronic properties, which together ensure reliable extrapolation of the electronic structure of 2D TB materials, even down to small twist angles.

By predicting the electron densities of five TB-2D materials [graphene, hexagonal boron nitride (hBN), $TiS_2$, $ZrS_2$, and $MoS_2$], we demonstrate that local descriptors fail to capture the long-range physics of moiré systems, while long-range descriptors extending beyond the locality assumption enable accurate extrapolation of band structures and electrostatic properties to large twisted superlattices. We achieve robust predictions, with low-energy band error below 5 meV on twisted bilayer structures containing more than 1000 atoms. The resulting SALTED models successfully predict flat-band formation, spin-orbit coupling (SOC) effects, and real-space observables, while achieving speedup between 1 and 2 orders of magnitude over fully converged DFT calculations. This framework thus provides a generalizable approach for ML-accelerated electronic structure prediction in systems dominated by long-range interactions. Our approach is promising for accelerating the design of quantum materials and expanding the scope of first-principles simulations for complex quantum phenomena.

## II. RESULTS

### A. Extrapolation to moiré superlattices

We first evaluated the ability of SALTED models to predict the electronic structure of 2D materials by applying them to validation sets of aligned (displaced) bilayer structures for each 2D bilayer material (graphene, hBN, $TiS_2$, $ZrS_2$, and $MoS_2$). To enable controlled comparisons, we fixed all descriptor hyperparameters across experiments, varying only the Gaussian process regression regularization $\eta$ (see Sec. IV E). The descriptors considered are described in more detail in Sec. IV B, and comprise the short-range smooth overlap of atomic positions (SOAP), and the long-range long-distance equivariant and descriptors (LODE and LOVV). We refer the reader to Sec. IV B for an explanation of the chosen acronyms for these descriptors. Table I summarizes the accuracy of SALTED models for the electron density (direct output) and band structure (derived output). Definitions of the error metrics used can be found in Sec. IV D.

TABLE I. The error in the density and band structure predictions averaged across the validation set for each of the five 2D materials, using each of the three descriptors. The error metrics are defined in Sec. IV D. Hyperparameters are optimized for density prediction accuracy (see Sec. S2 of the Supplemental Material [47] for details).

| | Density (%) | | | Band structure (meV) | | |
|---|---|---|---|---|---|---|
| Material | LOVV | LODE | SOAP | LOVV | LODE | SOAP |
| Graphene | 3.63 | 0.77 | 0.48 | 7.1 | 2.5 | 1.86 |
| hBN | 2.12 | 0.95 | 0.65 | 4.0 | 2.1 | 1.51 |
| $TiS_2$ | 2.12 | 2.20 | 1.71 | 9.6 | 16.4 | 49 |
| $ZrS_2$ | 0.45 | 0.46 | 0.40 | 6.6 | 13.4 | 27 |
| $MoS_2$ | 0.04 | 0.04 | 0.09 | 10.2 | 52 | 330 |

Consistent with previous work by some of us [25,44], we observed that the density error does not reliably correlate with the error in the prediction of derived observables, in this case the band structure. For graphene and hBN, the errors in both the density and band structure follow the more common relation: Descriptors that yield larger errors for density also result in larger errors on the band structure. However, for $TiS_2$, $ZrS_2$, and $MoS_2$, SOAP descriptors consistently achieve low density error, yet fail catastrophically in band structure predictions, while LOVV performs very well. The contrast is especially clear for $ZrS_2$, where SOAP achieves the lowest density error yet produces the largest band error, while LOVV with comparable density accuracy yields smallest band error. For $MoS_2$, LOVV and LODE are almost identical in density accuracy yet differ for 5 times in band error. For $TiS_2$, a larger error in the density is also seen when using LODE, in comparison to SOAP, but smaller errors in the band structure are observed.

Such behavior reveals a nuanced relationship between the electron density and electronic structure fidelity. Perfect density prediction guarantees accurate band structures. However, when density predictions contain errors, the resulting electronic structure accuracy also depends on the spatial distribution of errors and their coupling to the physics governing the observables at hand. Therefore, while this initial validation of the models provides strong evidence that SALTED can be effective in predicting both densities and band structures, further tests are necessary to identify the descriptors that will produce the most accurate band structures when extrapolating to large supercells of twisted 2D materials.

We thus directly assess the framework's ability to systematically extrapolate to small twist angles (i.e., large moiré length scales), a regime where conventional DFT calculations become computationally prohibitive because the unit cells contain thousands of atoms. We consider an independent test set containing twisted bilayer systems. For graphene and hBN, structures with ten different twist angles are included in the test dataset, corresponding to twist angles from 21.79° to 3.15°. To describe these small twist angles, supercells with lattice vectors ranging in size from 6 to 45 Å were required, which contain up to 1324 atoms at the smallest twist angles. For the TMDCs, structures with eight different twist

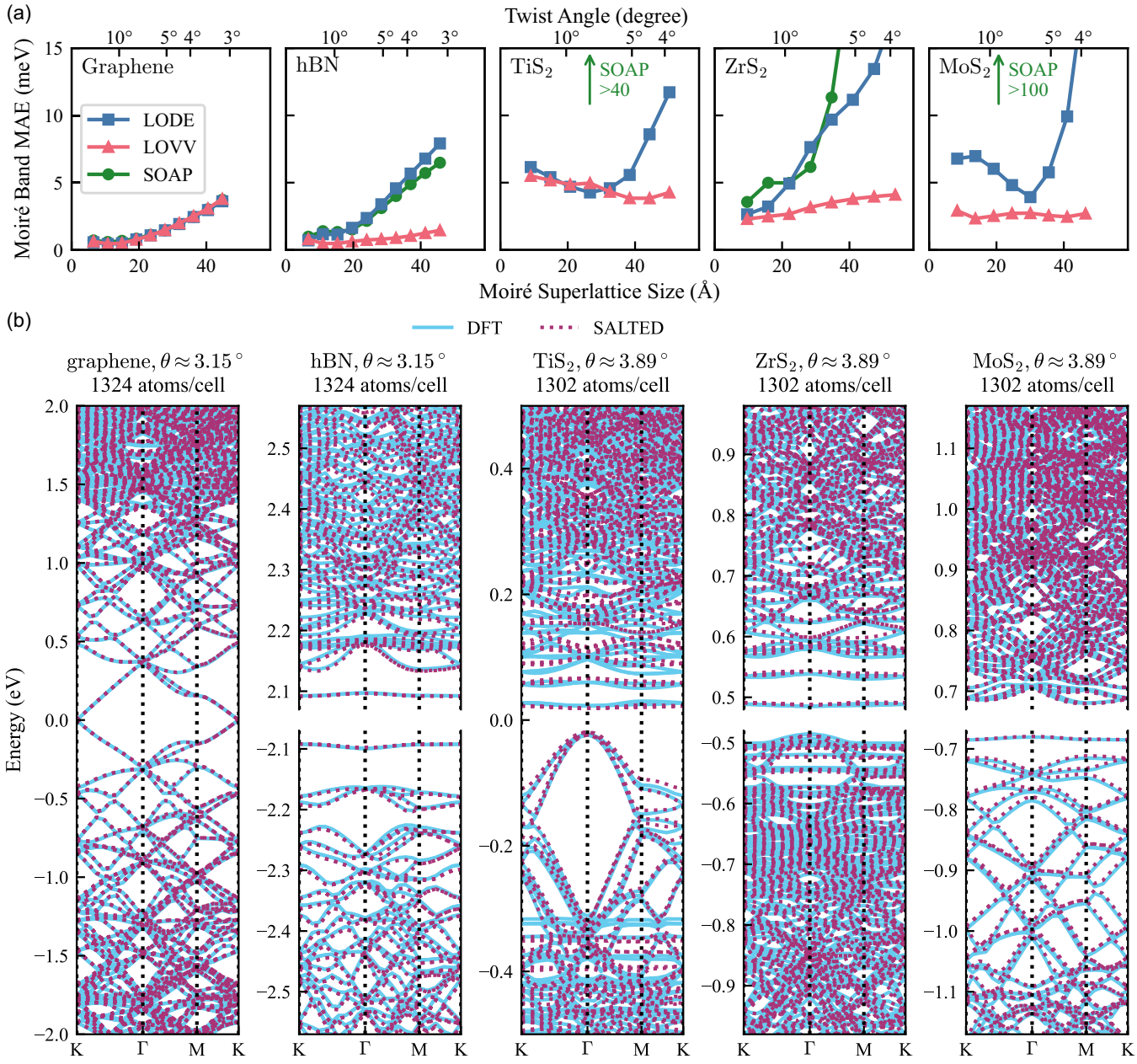


FIG. 1. Moiré band predictions across diverse twisted bilayer materials. (a) Low-energy moiré band error vs system size for twisted bilayers. SALTED models used here are those with the best band structure prediction performance for each material and descriptor, obtained by following the model optimization workflow (see Sec. IV F and Sec. S2 of the Supplemental Material [47]). SOAP fails catastrophically for $TiS_2$ (errors >40 meV, off scale) and $MoS_2$ (errors >100 meV, off scale), thus not shown here. (b) Comparisons between DFT-converged and SALTED-predicted moiré band structure for graphene, hBN, $TiS_2$, $ZrS_2$, and $MoS_2$, respectively. These five representative twisted bilayer systems have small twist angles <4° and contain >1300 atoms per superlattice. Only the graphene prediction is based on SOAP; the others are based on LOVV. Energy alignment: band gap centered at zero for all materials, with gap central region omitted for hBN, $ZrS_2$, and $MoS_2$. The predictions accurately match DFT band structures across all materials, demonstrating the framework's capability to handle various 2D systems.

angles were generated, ranging from 21.78° to 3.89°, which correspond to supercells with lattice vectors of 9–50 Å, with 1302 atoms in the largest structure. For every material other than graphene, structures with both parallel and antiparallel stacking symmetries were generated at each twist angle, with reported band structure errors for a given twist angle (or equivalently, supercell size) being the average of the errors for these two structures.

Figure 1(a) illustrates the dependence of the accuracy of the extrapolated band structures on both the choice of descriptor and the system size. In Fig. 1(a), we plot the low-energy moiré band error against the superlattice size (equivalently, decreasing twist angle) for each material. For graphene, all descriptors maintain reasonable accuracy (<5 meV) even at large system sizes. However, the other materials display descriptor-dependent performance: SOAP and LODE exhibit

degraded accuracy with increasing moiré length scales, with SOAP failing completely for $TiS_2$ and $MoS_2$; only LOVV maintains robust extrapolation with errors $\lesssim$5 meV across all materials and system sizes exceeding 1000 atoms and superlattice size of $\sim$50 Å. For a better visual comparison of the prediction accuracy, we also show the DFT band structures of the smallest twist-angle systems in the test set for each material in Fig. 1(b), along with the optimal SALTED prediction of the band structure.

This material- and descriptor-dependent performance reveals systematic differences in moiré extrapolation capability. SOAP uses only the local atomic density within its cutoff radius (6 Å), which proves sufficient for graphene but fails for other materials at large superlattice sizes. LODE combines the local atomic density and an electrostatic representation, extending the effective range beyond that of SOAP and showing moderate extrapolation for some materials ($TiS_2$, $MoS_2$), but failing for others ($ZrS_2$, hBN) as moiré length scales increase. LOVV has a different structure, using purely long-range representations, and is shown to achieve consistent accuracy across all materials and system sizes tested. The physical and methodological origins of these performance differences are discussed in Sec. III.

Having established and validated the SALTED framework's accuracy and identified LOVV's superior extrapolation capability for moiré superlattices, we proceed to demonstrate its practical utility for investigating phenomena where conventional DFT is prohibitively expensive. These case studies span three classes of problems: (1) systematic trend mapping toward small twist angles, requiring many large structures; (2) incorporating other physical effects at minimal additional cost; and (3) predicting real-space observables not directly accessible to Hamiltonian-based methods.

### B. Bandwidth extrapolation in hBN and TMDCs

Small twist angles in 2D materials produce large-scale moiré superlattices that flatten electronic bands and enhance electron localization and correlations. Simulating twisted bilayers at very small angles ($<3^\circ$) is crucial for understanding these effects. We leveraged the efficiency of SALTED and the extrapolative power of the LOVV descriptor to map out frontier bandwidths of twisted bilayer hBN, $TiS_2$, $ZrS_2$, and $MoS_2$.

Figure 2 shows the evolution of the bandwidth while decreasing the twist angle (and therefore increasing the number of atoms in the superlattice) for these materials. For this dataset, the structures with the smallest twist angles contain 4564 atoms for hBN and 3768 atoms for TMDCs. We only perform validation DFT calculations for structures with lattice constants of up to $\approx$50 Å and $\approx$1300 atoms per superlattice.

Both the frontier bandwidths and the band gap show a decreasing behavior with decreasing angle for each of these systems. Panel (a) of Fig. 2 shows representative band structures at $\theta \approx 4.41^\circ$ with analyzed bands highlighted. For hBN and $MoS_2$, we track the highest valence band (HVB) due to their topological character, while for $TiS_2$ and $ZrS_2$ we focus on the group of the three lowest conduction bands dominated by transition-metal $d$ orbitals mixed with chalcogen $p$ orbitals. Panel (b) demonstrates the systematic bandwidth narrowing with decreasing twist angle: hBN's flat HVB narrows to $\sim$5 meV, and the TMDC bands approach $\sim$1 meV, limited by the numerical precision of DFT band structure calculations. Panel (c) reveals the band gap evolution: hBN maintains large gap ($\sim$4 eV) albeit narrowing about 0.5 eV from $21^\circ$ to $1.8^\circ$, and our prediction suggests potential gap closure in $TiS_2$ below $\theta \sim 3^\circ$.

For all materials, the LOVV-based SALTED model successfully extrapolates the trend of full DFT calculations, with meV accuracy. In contrast, models based on LODE or SOAP, while accurate at large angles, fail as the moiré length increases and exceeds their training effective range, leading to unphysical predictions. We do not show the predictions with SOAP descriptors in Fig. 2 because they fail completely for band structure predictions, especially for small twist angles, as shown in Fig. 1(a). While models based on LOVV exhibit a diverging behavior for $TiS_2$ at twist angles below $2^\circ$, where gap closure and band merging occur, it is always stable for larger system sizes than models based on LODE. This trend highlights a limitation: SALTED may struggle to accurately predict band structures near metal-insulator transitions where the small energy gaps are sensitive to subtle density errors and minute doping of the system. Nevertheless, this extrapolation application highlights the advantages of the LOVV descriptor in capturing long-range interactions and moiré physics.

### C. Spin-orbit coupling in TMDCs

SOC is a critical factor in TMDCs, leading to significant band splitting that governs their topological and spintronic properties [6,45]. SOC is often included via a non-self-consistent perturbation applied after the scalar-relativistic DFT calculation has converged [46]. This approach integrates seamlessly with SALTED: The Kohn-Sham Hamiltonian is constructed from the predicted density, the SOC perturbation is added, and the combined Hamiltonian is diagonalized. This is a key feature of density-based prediction: The same predicted density serves as input for SOC perturbation, electrostatic field extraction, and charge analysis, without any modification to the ML model or retraining.

As an illustration, we included SOC in our SALTED prediction of the band structures of twisted bilayers of $TiS_2$ and $ZrS_2$, with the results shown in Fig. 3. The results for $MoS_2$ can be found in the Supplemental Material [47]. The predictions are in excellent agreement with full DFT calculations, accurately capturing the SOC-induced splitting and modifications to the band structure. This confirms that the density predicted by SALTED is of sufficient quality to serve as a foundation for subsequent calculations of properties beyond the initial training. The accurate reproduction of SOC splitting is critical for predicting spin-valley coupling, Berry curvature, and topological properties in moiré TMDCs, which govern their potential for valleytronic and spintronic applications.

### D. Structural relaxation of twisted bilayer graphene

We now focus on the relationship between atomic structure and electronic structure. The emergence of flat bands and correlated states near the "magic angle" ($\theta \approx 1.05^\circ$) in twisted bilayer graphene (TBG) depends critically on structural relaxation. Interlayer van der Waals (vdW) forces induce

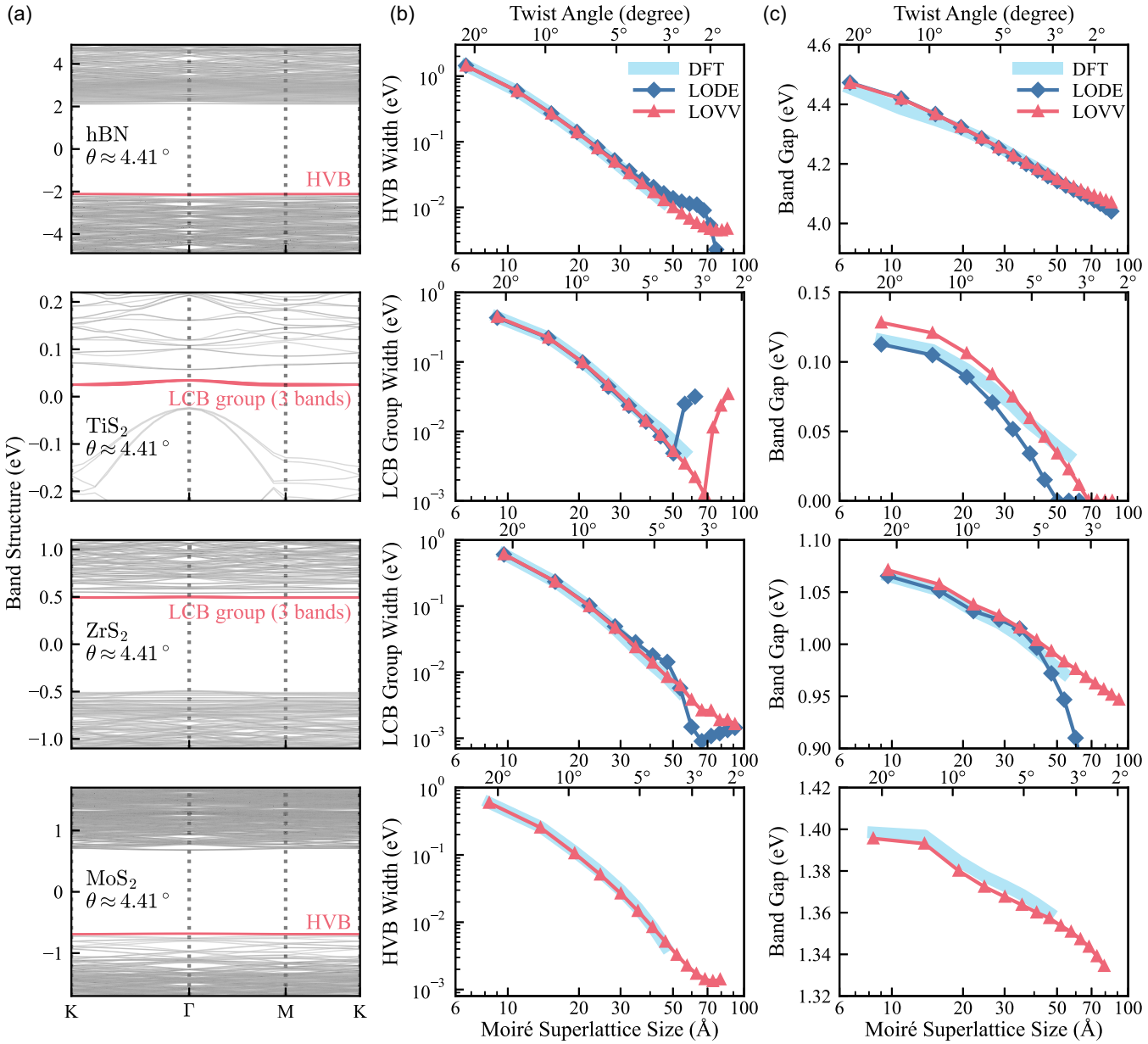


FIG. 2. Behavior of band gaps and bandwidths with decreasing twist angle. (a) Highlighted flat band(s) in the band structure, with materials' names and twist angles. (b) Bandwidth vs moiré superlattice size (equivalently, twist angle). (c) Band gap vs moiré superlattice size. LODE descriptor predictions for $MoS_2$ are not shown because the prediction error was too large. All structures use commensurate supercells with parallel unit cells in each layer. HVB: highest valence band. LCB: lowest conduction band.

out-of-plane and in-plane atomic displacements that significantly modify the electronic band structure compared to an idealized rigid model, especially for small twist angles around and below the magic angle [48–50]. Obtaining these relaxed geometries is extremely challenging, since a large supercell is needed to describe such a small twist angle and high energy and force accuracy is needed to capture the buckling of the 2D sheets.

We trained a bespoke MACE machine learning interatomic potential (MLIP) [51] on structures spanning twist angles from 21.79° to 3.48° sampled from molecular dynamics simulations and calculated with the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional including many-body vdW corrections [52] to perform these relaxations (see Sec. S3C of the Supplemental Material [47] for training details). The relaxation patterns, illustrated in Fig. 4(b) for a TBG with $\theta = 2.13^\circ$, share the same features at small twist angles: Energetically unfavorable AA-stacking regions are reduced, deforming in a curl pattern, while the areas of energetically favorable AB-stacking regions increase. The interlayer distance shows periodic behavior, with maximum ∼3.63 Å at AA-stacking and minimum ∼3.43 Å at AB-stacking, modifying local electronic interlayer hopping probabilities and ultimately changing the electronic structure [49,50].

When we apply SALTED to these relaxed geometries, it demonstrates robust predictive capability, benefiting from training on randomly displaced bilayers that naturally sample interlayer distances (detailed analysis in Sec. S3C of the

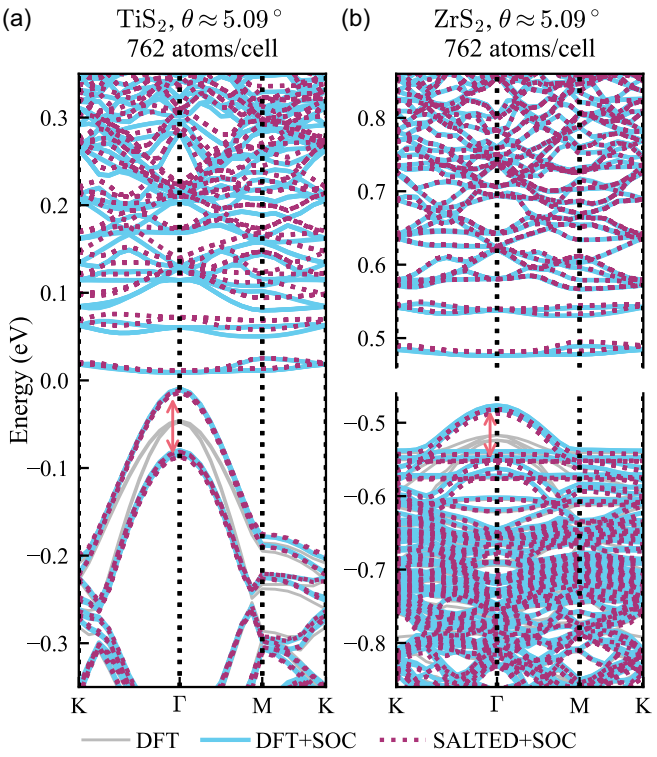


FIG. 3. Prediction of band structure of twisted bilayer TMDCs with spin-orbit coupling. Band structures for (a) $TiS_2$ and (b) $ZrS_2$ at a twist angle $\theta \approx 5.09°$. Gray lines and cyan lines show scalar-relativistic DFT without and with perturbative SOC, respectively. Purple dotted lines show SALTED predictions with perturbative SOC constructed from the predicted density. Bands are aligned to the conduction band region to emphasize the valence band splitting indicated by the red double arrows. The excellent agreement between SALTED + SOC and DFT + SOC demonstrates that SALTED-predicted electron density accurately captures the orbital character necessary for perturbative SOC calculations.

Supplemental Material [47]). Figure 4(a) shows that the model accurately predicts the band structure for low twisting angles. At the magic angle $\theta = 1.05°$ with 11 908 atoms per unit cell, SALTED reproduces the flat bands with a mean absolute error (MAE) of approximately 15 meV relative to a full DFT calculation. The predicted magic-angle flat bands do exhibit residual dispersion resembling hole doping [53]. For magic-angle TBG, the electronic density error amounts to about $10^{-4}\,e$ per electron. We find that normalizing the deficient predicted density does not alter the band structure prediction in this case, suggesting that the observed disagreement with DFT comes rather from the distribution of errors in space. The density errors cluster in AA-stacking domains (see Sec. S3C of the Supplemental Material [47]), where the density is underestimated by approximately $2.5 \times 10^{-4}\,e\,\text{Å}^{-3}$. Because flat-band states sit spatially in AA regions [54], the localized errors produce a hole-doping-like dispersion. The spatial distribution of density errors relative to the electronic states determines band structure accuracy.

Nevertheless, Fig. 4(c) demonstrates that SALTED reliably captures the relaxation-affected band gap across the full angle range down to the magic angle, serving as a quantitative computational reference for experiments. The MLIP + SALTED pipeline also establishes a first-principle workflow of including both structural and electronic properties in moiré systems, extending the reach of *ab initio* methods to previously inaccessible scales.

### E. TB-hBN electric field at domain boundaries

Finally, we turn our attention to emerging physical properties of these systems. TB-hBN exhibits ferroelectricity [55], where alternating AB- and BA-stacking domains have opposite out-of-plane polarizations, generating in-plane electric field at the domain boundaries. These domain wall electric fields have attracted significant interest due to their ability to modulate excitons in adjacent materials [56,57]. However, the spatial profile of these fields and their dependence on twist angle remain challenging to characterize experimentally. Previous theoretical studies have predicted moiré electrostatic potential in TB-hBN [58], but assumed rigid bilayer structures without atomic relaxation. This approximation becomes increasingly unrealistic at small twist angles, where interlayer forces drive significant atomic deregistration and fundamentally alter the domain wall geometry and thus the electric field.

We address this challenge by combining structure relaxation using MLIPs and electronic structure prediction with SALTED. The relaxed TB-hBN structures were obtained using a MACE model [51] trained on 512 displaced $5 \times 5 \times 1$ bilayer hBN supercells with random sampling and PBE including many-body vdW corrections [52] (see Sec. S3D of the Supplemental Material [47]). We then predicted the electron density for these relaxed structures using SALTED, computed the Hartree potential, and obtained the electric fields by computing its gradient numerically on a real-space grid.

Panel (a) of Fig. 5 reveals domain-dependent structural relaxation in TB-hBN. The energetically unfavorable AA-stacking region exhibits out-of-plane buckling and counterclockwise in-plane displacement pattern, reducing its spatial extent while increasing interlayer separation. Conversely, the energetically favored AB and BA domains exhibit reduced interlayer distances and correspondingly expanded area vacated by AA regions. These domains determine the electrostatic landscape shown in panel (b), where the in-plane electric field component is evaluated 1.5 nm above the bilayer center. BA and AB domains possess opposite out-of-plane polarizations, functioning as field sources and drains, respectively. The electric field intensity reaches its maximum within this plane at the center of the AB-BA domain wall, marked as the "Measurement Point" in Fig. 5(b). The AA domain is expected to exhibit spatially uniform weak field due to zero polarization, while the prediction exhibits residual field divergence. This numerical artifact does not propagate to domain wall field measurements due to hexagonal symmetry (see Sec. S3D of the Supplemental Material [47] for detailed analysis).

Our results, presented in Fig. 5(c), quantitatively reveal the electric field at the AB-BA domain wall. The predicted field intensity agrees well with full DFT results, confirming SALTED accurately captures polarization-induced electric fields. The electric field intensity saturates as the twist angle decreases: The field strength increases sharply from large twist angles down to approximately $\theta \approx 1°$, then plateaus at

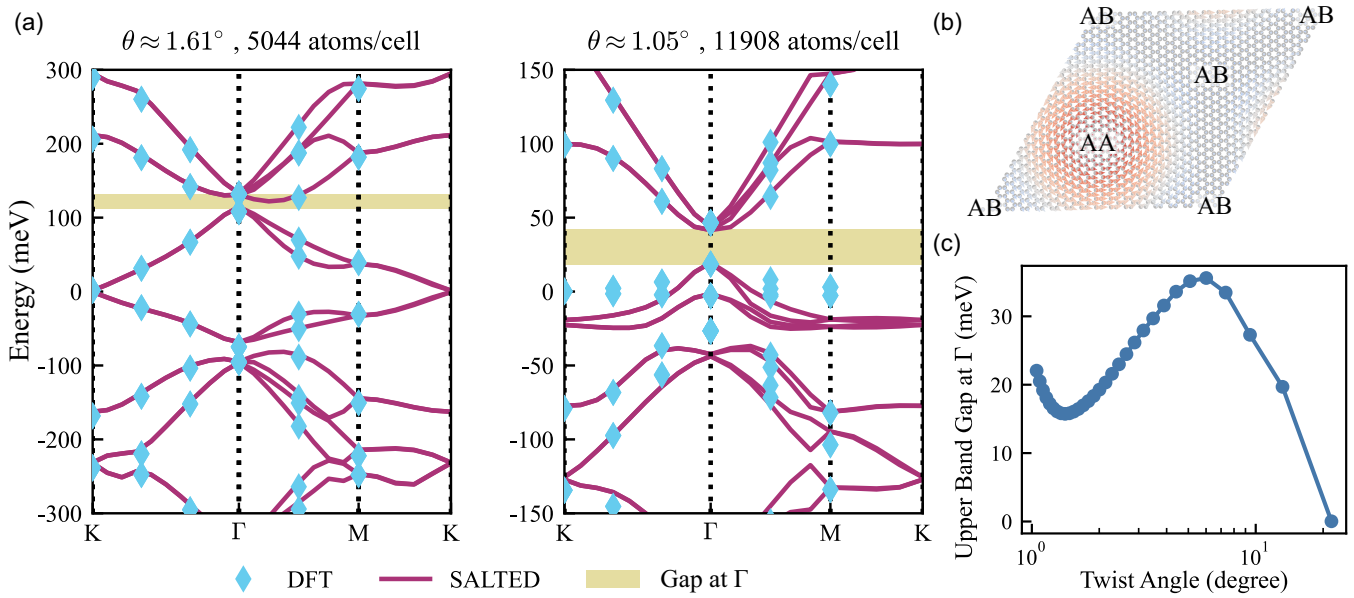


FIG. 4. SALTED predictions for relaxed TBG. (a) SALTED-predicted TBG band structure (maroon lines) at 1.61° and 1.05°, compared to full DFT results (blue diamonds). The band structures are predicted by models based on SOAP with GPR regularization $\eta = 10^{-3}$, guided by the band structure metric. (b) The relaxation effect of TBG ($\theta \approx 2.13°$, 2884 atoms/cell). The upper layer deregistration is shown with arrows (in-plane displacement) and color map (out-of-plane displacement: red for upward, blue for downward, and gray for minimal). The energetically unfavorable AA-stacking exhibits counterclockwise curl with upward displacement (red), while favorable AB-stacking regions show downward displacement (blue). These relaxations minimize energy by reducing AA-stacking area with increased interlayer distance and expanding AB-stacking area with decreased interlayer distance. (c) The upper band gap [between flat bands and the conduction bands above, shown in yellow in panel (a)] at Γ point against twist angles.

smaller angles. For example, at a probe distance of 2 nm above the bilayer center, the field intensity saturates near 9 mV/nm for $\theta < 1°$.

This saturation can be understood through the formation of structural solitons [59–61]. The AB-BA domain boundary is a structural soliton, which is a stable, localized deformation arising from the competition between interlayer registry energy (favoring AB or BA stacking) and elastic shearing energy (opposing sharp deformations). As the moiré superlattice size $a$ increases with decreasing twist angle $\theta$, structural relaxation reaches an equilibrium where the domain wall adopts an equilibrium width $w_{\text{eq}}$ independent of further twist angle reduction. We extracted domain wall widths from our relaxed structures (see Sec. S3D of the Supplemental Material [47]), finding saturation at $w_{\text{eq}} \approx 4.23 \pm 0.01$ nm for characteristic superlattice length $a \geqslant 15.9 \pm 0.1$ nm. This characteristic length scale corresponds to twist angles $\theta \approx 1°$, precisely where we observe electric field saturation in Fig. 5(c). The electric field saturation arises because our measurement probe distance ($z \sim 1.5$–4 nm) is small compared to the superlattice size at these twist angles ($a > 15$ nm). At this local scale, the measured electric field is dominated by the polarization gradient at the domain wall itself. Once the domain wall structure saturates, this local polarization gradient stabilizes. At larger probe distances, minor contributions from the extended polarization pattern across the larger superlattice still contribute. The domain wall $E$ field is sufficiently local, being dominated by the short-range interlayer B-N charge transfer (see Sec. S3D of the Supplemental Material [47]), to justify using a SOAP model in this case. These results demonstrate that structural relaxation substantially affects domain wall geometry and electric field behavior at small twist angles, confirming that relaxation effects cannot be neglected for accurate predictions.

## III. DISCUSSION

The four case studies described above demonstrate SALTED's general applicability across the three problem classes outlined at the outset: Systematic parameter mapping (bandwidth extrapolation), incorporating additional physics without retraining (SOC), and computing real-space observables (electric fields at domain walls). The TBG and TB-hBN relaxation studies further demonstrate robustness for large systems with complex structural distortions. The framework's computational efficiency enables systematic investigations of moiré physics previously inaccessible to *ab initio* methods, opening avenues for first-principles exploration and design of correlated moiré systems.

The necessity of long-range descriptors like LOVV for accurate electronic structure predictions in moiré materials is demonstrated empirically in Sec. II A. The importance of long-range representation emerges most clearly when comparing graphene against hBN, materials with identical lattice structures but contrasting electronic properties. As shown in Fig. 1, all three descriptors achieve comparable accuracy for graphene, while for hBN, LOVV significantly outperforms both SOAP and LODE at small twist angles.

This raises the question: Why does long-range geometric information matter for electronic structure prediction

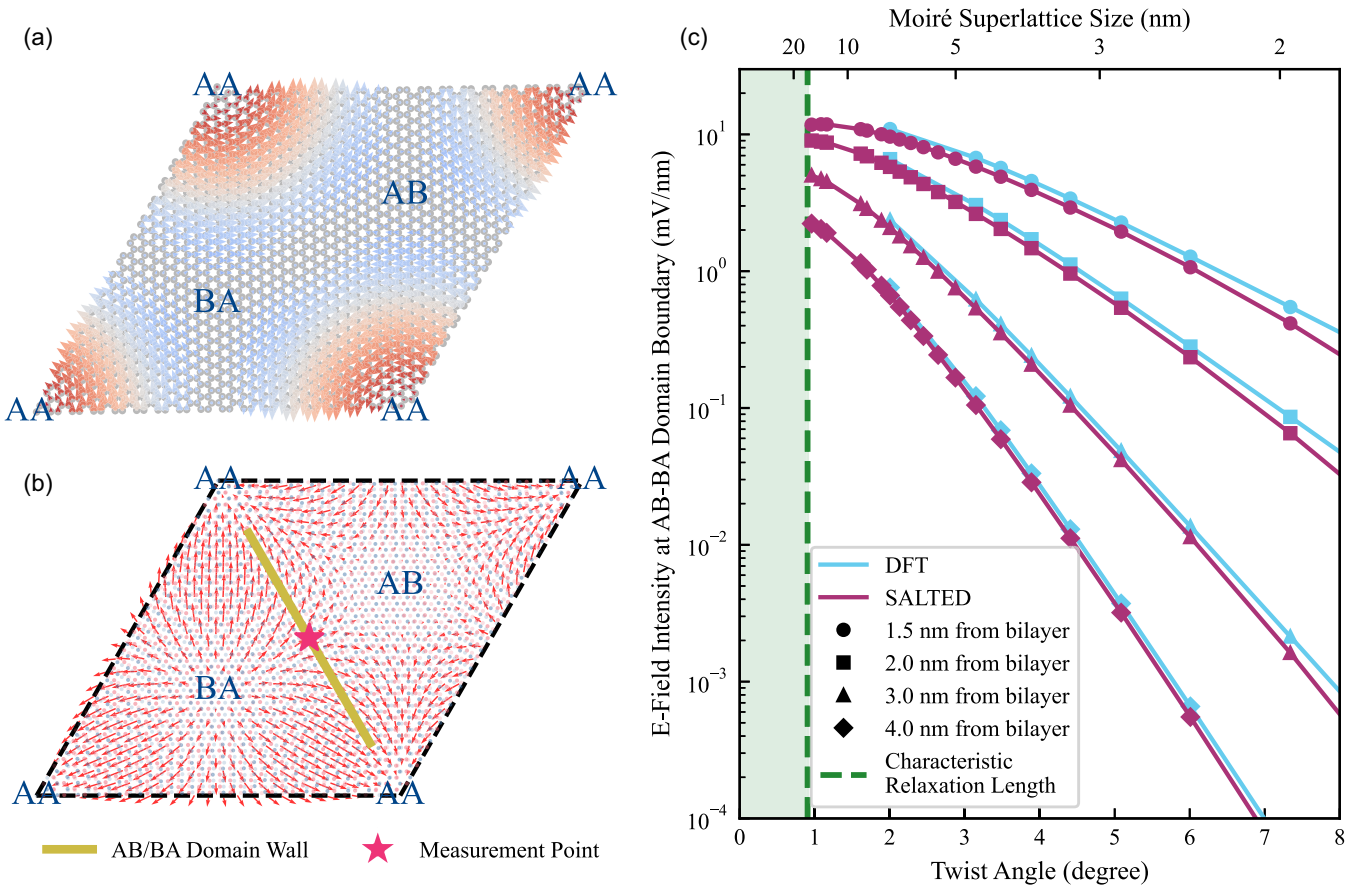


FIG. 5. Twisted bilayer hBN relaxation patterns and electric fields. (a) Relaxed TB-hBN structure ($\theta \approx 1.89°$, 3676 atoms/cell) showing atomic position deregistration. The upper layer deregistration is shown with arrows (in-plane displacement) and color map (out-of-plane displacement: red for upward, blue for downward, and gray for minimal). (b) Schematic view of the relaxed TB-hBN structure showing stacking regions (AA, AB, and BA) and predicted in-plane electric field 1.5 nm above the bilayer center. Blue and light red dots represent B and N nuclei, respectively, with red arrows indicating the electric field direction and magnitude. The AB-BA domain wall is marked with a yellow line, and the star indicates the in-plane location of the electric field values shown in panel (c). (c) In-plane electric field intensity perpendicular to the AB-BA domain boundary as a function of twist angle, evaluated at various heights above the bilayer center. SALTED predictions (SOAP-based model, $\eta = 10^{-6}$, optimized for density accuracy) show excellent agreement with full DFT results. The characteristic relaxation length represents reaching equilibrium for the domain wall.

in these systems? The most important factor is that polar bilayer systems exhibit interlayer charge rearrangement that creates out-of-plane polarization and long-range electrostatic potentials, which affect the electronic structure. The paradigmatic example is hBN. Despite having the same $sp^2$ hybridization as homoatomic graphene bilayers, hBN bilayers exhibit significant interlayer charge rearrangement between the B and N atoms at different layers, with the magnitude depending on interatomic distance. This charge rearrangement creates out-of-plane dipole moments and long-range electrostatic fields that modulate the band edge [55,62]. Similar polarization effects are observed or predicted in TMDCs, where interlayer charge redistribution creates spatially varying electrostatic environments [8,63]. In moiré systems, local stacking configurations vary continuously across the moiré supercell, creating ferroelectric domains with alternating polarization in TB-hBN [55,62] and in-plane electric field vortices in TB-$MoS_2$ [8,63]. The electrostatic potential experienced by electrons at any site depends on the cumulative charge rearrangement and consequent polarization from extended stacking domains spanning nanometer length scales. The empirical comparison of band errors between graphene and hBN supports this argument.

Interestingly, the LOVV descriptor's superior performance cannot be attributed solely to its spatial extent. Quantitative sensitivity analysis (see Sec. S5A of the Supplemental Material [47]) reveals that LODE and LOVV exhibit similar distance-dependent decay characteristics, with effective ranges of approximately 13 and 17 Å, respectively, at a sensitivity threshold of $10^{-2}$. This small range difference contrasts with LOVV's substantial moiré extrapolation accuracy over LODE. Previous work on molecular systems [64] demonstrated that LODE's mixed $|\rho \otimes V\rangle$ structure provides an optimal balance between short- and long-range description for diverse molecular interactions at length scales around 10 Å, while our observation shows that LOVV's $|V \otimes V\rangle$ structure performs better for extended moiré systems. We propose that LOVV's $|V \otimes V\rangle$ structure, which is capable of correlating two independent points that lie outside of the cutoff radius of the central atom, better captures the geometric features

relevant to long-range physics in bilayers and moiré superlattices, enabling better representation of the connection between displaced bilayer training structures and moiré bilayer test structures. This interpretation is supported by descriptor space coverage analysis via uniform manifold approximation and projection (UMAP) dimensionality reduction (see Sec. IV E and Sec. S5B of the Supplemental Material [47]), which shows LOVV providing superior manifold overlap between displaced and twisted bilayers compared to both LODE and SOAP for certain materials, particularly TMDCs. However, UMAP analysis shows better coverage for all descriptors in simpler systems. The specific mathematical structure of the descriptor representation, i.e., how it encodes correlations between atomic environments, likely plays an important role that warrants further investigation.

Regarding training costs, dominated by Hessian matrix construction [27], they are comparable across descriptors. However, prediction exhibits distinct scaling behavior: In the current implementation, LOVV and LODE show quadratic scaling with system size, due to all-to-all electrostaticlike descriptor evaluations, whereas SOAP maintains near-linear scaling from its finite cutoff (see Sec. S4 of the Supplemental Material [47] for details). Despite this scaling, LOVV-based SALTED models remain 10–100 times faster than fully converged DFT for moiré structures. We further show in Sec. S4 of the Supplemental Material [47] that tuning the Gaussian kernel width $\sigma$ can reduce LOVV prediction cost with only a small impact on band gap extrapolation accuracy. We expect that implementation improvements, including multipole approximation for long-range descriptor contributions from distant nuclei, graphics processing unit (GPU) acceleration and message passing interface (MPI) parallelization, will speed up the models substantially. This computational efficiency could enable future applications such as unfolding *ab initio* moiré band structure [65] for direct comparison with angle-resolved photoemission spectroscopy experiments [66,67].

We further comment on our model's performance in comparison to other reports in the literature. Recent neural network methods for Hamiltonian prediction achieve sub-meV accuracy for Hamiltonian matrix elements [18,20,21]. These methods achieve efficiency by relying on the locality assumption and the nearsightedness principle. Band structures can also be derived from these predictions. For relaxed TBG at $\theta \approx 2.0^\circ$ (3268 atoms), the pioneering work reported with DeepH [21] achieves 1.3 meV low-energy band error while SALTED achieves 1.5 meV (see Sec. S3C of the Supplemental Material [47]), with both methods demonstrating $<0.3\%$ error over the 600 meV low-energy band range. For TBG with larger twisting angles, the band structures derived from SALTED predictions in this work can yield up to an order of magnitude lower error than DeepH results published a couple of years ago [21], as summarized in Sec. S2E of the Supplemental Material [47]. For magic-angle TBG, DeepH produces narrower flat bandwidth matching the expected low-energy profile more closely than SALTED's prediction. However, graphene represents a case where the locality assumption is valid. We expect this assumption to become limiting for other twisted bilayer systems, as discussed above. Indeed, comparing the performance of the band structure presented here for $MoS_2$ and the ones derived from previously published DeepH predictions [21], we observe an order of magnitude better accuracy with our model, as detailed in Sec. S2E of the Supplemental Material [47]. Given the pioneering character of the publications using DeepH in this area, we expect that predictions with that method can also be improved by incorporating long-range interactions. To our knowledge, published Hamiltonian prediction methods have not yet demonstrated robust band structure performance for hBN or other TMDC systems at small twist angles below $5^\circ$ [36].

Many recent methods that predict the electron density directly report impressive global metrics (global density error $\leqslant 10^{-3}\, e\,\text{Å}^{-3}$ or $\leqslant 1\%$ if normalized by total density), but do not verify that predicted densities yield accurate downstream electronic properties [29,30,32,34]. Consistent with our previous observations [25,44], we have found again that although accurate density prediction is necessary for accurate electronic structure prediction, the quality of the prediction of certain properties is strongly dependent on the spatial distribution of density errors. Our results emphasize that global density error, being a single scalar quantity, is an insufficient validation metric for density prediction methods intended for electronic structure applications, and careful verification of downstream observables is essential.

We are not aware of other electronic density-prediction methods that have been employed to compute the band structure of twisted bilayer systems. Because such models are not straightforward to use, it has proved difficult to directly perform such a comparison in this work. The authors of this paper nevertheless believe that a collaboration targeting widespread comparison of predicted properties among electronic-structure methods is of immense importance to the community.

We further note the generality of SALTED. Training separate models for each material is a pragmatic choice because of GPR's data scaling but not a constraint of SALTED: Previous works [25,27] have shown multimaterial SALTED models, but they need larger training datasets. The SALTED models are not property specific because they first predict electron densities and then derive all downstream electronic properties. The descriptor choice between long range and short range is physics driven and is analogous to neural network architectural choices such as range-separate message passing [21] or $k$-space aggregation [68]. The optimal regularization $\eta$ may differ slightly between density- and band-targeted models, but the predictions are accurate and robust across a wide range of $\eta$ values (see Sec. S2B of the Supplemental Material [47]). Using a multiproperty loss function could in principle automate this balance, but the current framework works reliably without adjusting too much for each property.

We conclude by discussing perhaps the main limitation of SALTED: As it is based on Gaussian process regression, there are intrinsic limitations on the size of the training sets that can be used and, ultimately, on the model's accuracy and breadth. While active-learning strategies for the electronic density can be explored, in order to increase the model accuracy without increasing the training set size, ultimately an implementation of this model within a neural network architecture would be desirable. Because of the numerical instabilities of

SALTED discussed in this paper, related to the density representation, we expect such an implementation to face some challenges.

In summary, we established a systematic and transferable framework for predicting electronic structures in large moiré superlattices, which overcomes the locality constraint limiting existing machine learning approaches. We have established that the locality assumption, applied to most existing machine learning methods for electronic structure, is not universally valid for polar twisted bilayer systems. While training only on computationally tractable displaced bilayers ($3 \times 3$ supercells), our method achieves accurate extrapolation to twisted moiré structures with supercell sizes up to $a \sim 100$ Å ($\sim$4000 atoms), at computational costs 10–100$\times$ faster than direct DFT calculations, within the current implementation of SALTED. This enables *ab initio* exploration of quantum phenomena at twist angles and system sizes previously inaccessible due to computational constraints.

This advance builds upon three key innovations addressing distinct physical and methodological challenges. First, we used LOVV descriptors, which are built upon atomic electrostaticlike representations and encode distant geometric information through $1/r$ Coulomb form, to train our SALTED models. As a result, these models were able to capture the essential physics that govern moiré materials: Interlayer charge transfer and polarization and moiré-scale modulation of electronic potentials, which require structural information far beyond limited cutoffs of local descriptors. Second, we systematically resolved numerical instabilities from ill-conditioned density fitting auxiliary basis sets in heavy-element systems through singular value truncation, suppressing noise-dominated subspaces while preserving density fitting accuracy. This finding is transferable to other density-fitting-based machine learning methods, which faced difficulties with heavy elements or larger density fitting basis sets. Third, we demonstrated that spatial density prediction accuracy alone is insufficient for reliable electronic structure modeling: The spatial distribution of density errors determines their coupling to target properties. This necessitates explicit validation on downstream observables (band structures, real-space fields) rather than relying solely on global density metrics.

We validated the framework systematically across five diverse 2D bilayer materials (graphene, hBN, $TiS_2$, $ZrS_2$, and $MoS_2$) with different electronic characters. The framework's practical utility was demonstrated through diverse applications, previously only barely accessible to *ab initio* methods. We revealed low-energy isolated bandwidth narrowing trends with smaller twist angles, predicted geometrically relaxed magic-angle twisted bilayer graphene band structures, and computed real-space electric field profiles at ferroelectric domain walls in relaxed twisted bilayer hBN, unraveling the reason behind the plateau value of in-plane electric fields at low twisting angles. The characterization of the impact of long-range descriptors, the singular value truncation for training stabilization, and the demonstration that spatial distribution of density errors determines downstream electronic property accuracy enable a robust first-principles machine learning framework for polar moiré materials, supporting future research in this promising area.

These capabilities enable systematic materials screening, accelerating the discovery of correlated quantum phases, and supporting inverse design of moiré structures with targeted electronic properties. This work provides both conceptual foundation and practical tools for accurate, efficient, and transferable electronic structure prediction in moiré materials and beyond.

## IV. METHODS

### A. SALTED foundations

SALTED predicts electron densities using symmetry-adapted GPR [25,27]. The electron density is represented using the density-fitting technique, where the density is expanded in atom-centered auxiliary basis functions:

$$\rho(\mathbf{r}) \approx \rho^{\mathrm{DF}}(\mathbf{r}) = \sum_{i,n\lambda\mu,\mathbf{T}} c_i^{n\lambda\mu} \phi_i^{n\lambda\mu}(\mathbf{r} - \mathbf{R}_i + \mathbf{T}), \qquad (1)$$

where $\mathbf{r}$ is the spatial coordinate, $\mathbf{R}_i$ is the position of $i$th atom in the unit cell, $\mathbf{T}$ is the lattice translation vector, $\phi_i^{n\lambda\mu}$ are density fitting auxiliary basis functions with radial index $n$, angular momentum $\lambda$, and magnetic quantum number $\mu$, and $c_i^{n\lambda\mu}$ are the corresponding expansion coefficients (collectively denoted as $\mathbf{c}^{\mathrm{DF}}$).

SALTED employs symmetry-adapted kernels [42] based on representations of atomic environments $A_i$. These are most commonly the SOAP [69] descriptors, which result in

$$\mathbf{k}^{\lambda}(A_i, A_j) = \int_{\hat{R}\in \mathrm{SO}(3)} \mathrm{d}\hat{R} \mathbf{D}^{\lambda}(\hat{R}) \left| \int \mathrm{d}\mathbf{r} \rho_i(\mathbf{r}) \rho_j(\hat{R}\mathbf{r}) \right|^2, \qquad (2)$$

where $i$, $j$ index atoms in the structure, $\mathbf{D}^{\lambda}$ is the Wigner-$D$ matrix for spherical harmonics $\lambda$. $\rho_i$ is the representation of the atomic environment around atom $i$. This kernel can be equivalently written using Dirac notation for descriptors [70],

$$\mathbf{k}^{\lambda}(A_i, A_j) = \left\langle \chi_j^{(2)}, \lambda \middle| \chi_i^{(2)}, \lambda \right\rangle, \qquad (3)$$

where the general feature vector $|\chi_i^{(2)}\rangle$ is given in this case by the tensor product of atomic density representation of the atomic environment with itself, $|\chi_i^{(2)}\rangle = |\rho_i \otimes \rho_i\rangle$. Different representations of the atomic environment yield distinct descriptors with various radial range characteristics [64], which are summarized in Sec. IV B.

The GPR optimization minimizes the loss function

$$\mathcal{L}(\tilde{\mathbf{w}}) = (\mathbf{c}^{\mathrm{ML}} - \mathbf{c}^{\mathrm{DF}})^{\top} \mathbf{S} (\mathbf{c}^{\mathrm{ML}} - \mathbf{c}^{\mathrm{DF}}) + \eta \tilde{\mathbf{w}}^{\top} \tilde{\mathbf{w}}, \qquad (4)$$

where the first term measures the squared spatial density error between predicted DF coefficients $\mathbf{c}^{\mathrm{ML}}$ and reference DF coefficients $\mathbf{c}^{\mathrm{DF}}$ weighted by the overlap matrix $\mathbf{S}$ of the auxiliary basis, $\tilde{\mathbf{w}}$ represents the regression weights, and $\eta$ is the regularization parameter controlling the trade-off between fitting accuracy and model generalization ability.

Applying SALTED to moiré superlattice prediction introduces specific technical challenges. First, previous SALTED implementations [25,27] employed the SOAP descriptor when calculating symmetry-adapted kernels, as in Eq. (2). This descriptor relies on the locality assumption: Atomic environments are represented through exponentially decaying Gaussian densities with a cutoff radius of a few angstroms.

However, this locality constraint can fail for moiré superlattices with periodicities reaching tens to hundreds of Å. This point is addressed in more depth in Sec. IV B.

Second, the density fitting technique used in SALTED employs intentionally overcomplete auxiliary basis sets to ensure accurate density representation across diverse chemical environments. However, this overcompleteness results in ill-conditioned overlap matrices that introduce numerical instability and degrade model performance, particularly for heavy elements requiring large and diffuse basis sets. This challenge is addressed in Sec. IV C.

Third, downstream electronic structure properties depend not only on density prediction accuracy measured by density error metric, but also on the spatial distribution of density errors [25,44]. This can create a validation gap between density-based loss function optimization and reliable electronic property prediction. Therefore, we evaluate the model explicitly on the accuracy of these predicted electronic properties, as well as the density itself. Details of these evaluation metrics are given in Sec. IV D.

### B. Beyond locality by long-range representation

Twisted bilayer systems exhibit electronic structure governed by moiré-scale interlayer interactions spanning length scales far exceeding the range of conventional local descriptor cutoffs and graph neural network message-passing distances. Local SOAP descriptors [69] rely on exponentially decaying Gaussian atomic density representations with cutoff radius limited to a few Å to avoid oversmoothness. A different family of descriptors has been proposed by Grisafi and coworkers, which we employ here. This is the LODE [71] descriptor family that uses atomic electrostatic representations $V_i$ to naturally encode nonlocal geometric information through a Coulomb potentiallike descriptor. $V_i$ is defined as

$$V_i(\mathbf{r}) = \int \mathrm{d}\mathbf{r}' \frac{\rho_i(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|}, \tag{5}$$

where $\rho_i$ is the atomic density representation used in SOAP. The $1/r$ decay of $V_i$ enables descriptors to capture structural information beyond local neighborhoods without requiring prohibitively large cutoffs.

Symmetry-adapted kernels sensitive to structural features at different length scales can be constructed by choosing different representations for the atomic environments, $|\chi_i^{(2)}\rangle$ in Eq. (3). SOAP uses $|\chi_i^{(2)}\rangle = |\rho_i \otimes \rho_i\rangle$, relying solely on the local atomic density representation. Several different LODE descriptors can be built. The one used in most recent works [72,73], LODE (1, 1), which we refer to here simply as LODE, combines local and electrostatic representation as $|\chi_i^{(2)}\rangle = |\rho_i \otimes V_i\rangle$ [64,71]. The descriptor we call LOVV ($|\chi_i^{(2)}\rangle = |V_i \otimes V_i\rangle$) in this paper corresponds to LODE (0, 2) in the original nomenclature, meaning it has no density components and two potentiallike components. Its utility for long-range systems was not extensively investigated previously, and we demonstrate here its crucial role in moiré materials prediction. These descriptors exhibit distinct sensitivity ranges as quantified in Table II: SOAP is limited by its 6 Å cutoff, while LODE and LOVV maintain sensitivity up to 13 and 17 Å, respectively (see Sec. S5A of the Supplemental Material [47] for details). Performance comparisons across different materials and twist angles are presented in Sec. II, with physical discussion in Sec. III.

TABLE II. Descriptor representations and sensitivity ranges. Sensitivity range is defined as the interatomic distance at which descriptor changes drop below $10^{-2}$ Å$^{-1}$ upon atomic displacement (see Sec. S5A of the Supplemental Material [47]). The atomic density representation $\rho_i$ uses a 6 Å cutoff. $V_i$ denotes the electrostatic representation.

| Descriptor | $\lvert\chi_i^{(2)}\rangle$ | Sensitivity range |
|---|---|---|
| SOAP [69] | $\lvert\rho_i \otimes \rho_i\rangle$ | 6 Å |
| LODE [64,71] | $\lvert\rho_i \otimes V_i\rangle$ | 13 Å |
| LOVV [71] | $\lvert V_i \otimes V_i\rangle$ | 17 Å |

### C. Low-rank approximation for mitigating ill conditioning

The density-fitting auxiliary basis sets used in SALTED are intentionally overcomplete to ensure accurate electron density representation across diverse chemical environments. However, this overcompleteness leads to significant overlap between basis functions. This can result in an ill-conditioned overlap matrix $\mathbf{S}$, which appears in the loss function in Eq. (4). Large condition numbers $\kappa = \sigma_{\max}/\sigma_{\min}$, defined as the ratio between the maximum and minimum singular values (SVs), quantitatively indicate numerical instability in coefficient calculations due to near-null space directions (corresponding to small singular values) [74,75]. This introduces noise that degrades both training accuracy and generalization capability for DF-based ML methods.

For systems that display severe ill conditioning, we employ a low-rank approximation of the real-symmetric positive-definite overlap matrix via SV truncation with threshold $\delta$:

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^\top \approx \tilde{\mathbf{S}} = \mathbf{U}\tilde{\boldsymbol{\Sigma}}\mathbf{U}^\top, \tag{6}$$

where $\mathbf{U}$ contains the left singular vectors and $\tilde{\boldsymbol{\Sigma}}$ is the diagonal matrix of truncated singular values, retaining only those satisfying $\sigma_i > \delta$. The truncated overlap matrix $\tilde{\mathbf{S}}$ is then used in the loss function in Eq. (4) during model training, and also in density error evaluation in Eq. (7) during validation. This truncation serves two purposes: (1) implicit regularization by discarding near-null subspace directions associated with basis function redundancy rather than meaningful physical variations; and (2) numerical stability during training is improved by removing near-null space directions. Also, since we target extrapolating to moiré superlattices, the low-rank approximation improves generalization by preventing overfitting to noise-dominated components.

Considering the real, symmetric, and positive-definite nature of $\mathbf{S}$, eigenvalue decomposition is mathematically equivalent to singular value decomposition (SVD) for this application. In this work, we use SVD.

### D. Validation metrics

We quantified the prediction accuracy of the electron density using the integrated root mean square error (RMSE),

normalized by the residual density:

$$\Delta\rho = \left(\frac{\sum_k \int \mathrm{d}\mathbf{r}\left|\rho_k^{\mathrm{ML}}(\mathbf{r}) - \rho_k^{\mathrm{DF}}(\mathbf{r})\right|^2}{\sum_k \int \mathrm{d}\mathbf{r}\left|\rho_k^{\mathrm{DF}}(\mathbf{r}) - \rho_k^{\mathrm{sph}}(\mathbf{r})\right|^2}\right)^{1/2} = \left(\frac{\sum_k \left(\mathbf{c}_k^{\mathrm{ML}} - \mathbf{c}_k^{\mathrm{DF}}\right)^\top \mathbf{S}_k \left(\mathbf{c}_k^{\mathrm{ML}} - \mathbf{c}_k^{\mathrm{DF}}\right)}{\sum_k \left(\mathbf{c}_k^{\mathrm{DF}} - \mathbf{c}_k^{\mathrm{sph}}\right)^\top \mathbf{S}_k \left(\mathbf{c}_k^{\mathrm{DF}} - \mathbf{c}_k^{\mathrm{sph}}\right)}\right)^{1/2}, \quad (7)$$

where $k$ is the geometry index in the dataset, $\mathbf{S}_k$ is the DF auxiliary basis overlap matrix, and $\mathbf{c}_k^{\mathrm{sph}}$ represents spherical components of the DF coefficients. Normalization by the deviation from spherical approximation makes this metric extremely sensitive: For graphene, a 1% RMSE in Eq. (7) corresponds to less than 0.02% error if normalized by total density; $MoS_2$ has large residual densities, which suppresses its normalized density error in Table I (see Sec. S2D of the Supplemental Material [47] for details). Density fitting is expensive for large moiré structures, so twisted bilayer density error is reported within memory limits in Sec. S2D of the Supplemental Material [47]. The band structure error is reported for all test structures.

To evaluate subsequent electronic properties, we restart DFT calculation from the predicted density using the FHI-aims program package [76,77]. The SALTED-predicted density is used to initialize a DFT calculation, with just one Hamiltonian diagonalization performed to establish the electronic state and obtain energies. With this electronic state fixed, we evaluate subsequent electronic properties, including band structures as well as real-space electronic properties. This procedure avoids iterative convergence, while enabling direct comparison with fully converged DFT results.

Since moiré physics emerges from low-energy bands near the Fermi level, and these bands are most sensitive to long-range density features, we directly evaluated band structure prediction accuracy by MAE:

$$\Delta b(\{B\}) = \frac{1}{N_{\mathrm{k}} N_{\{B\}}} \sum_{\substack{i\in k \text{ path},\\ b\in\{B\}}} \left|\epsilon_b^{\mathrm{ML}}(\mathbf{k}_i) - \epsilon_b^{\mathrm{DFT}}(\mathbf{k}_i) + \delta_\epsilon\right|, \quad (8)$$

where $N_{\mathrm{k}}$ is the number of $k$ points ($\{\mathbf{k}_i\}$) in the $k$ path (see Sec. S1C of the Supplemental Material [47]), $\{B\}$ is a set of low-energy bands close to the Fermi level (4 valence + 4 conduction bands for graphene and hBN, and 6 + 6 for $TiS_2$, $ZrS_2$, and $MoS_2$), $N_{\{B\}}$ is the number of bands in $\{B\}$, $\epsilon_b^{\mathrm{ML}}(\mathbf{k}_i)$ and $\epsilon_b^{\mathrm{DFT}}(\mathbf{k}_i)$ are band energies from SALTED-predicted density and fully converged DFT, respectively, and $\delta_\epsilon$ aligns band structures from ML and DFT by gap centers. This metric simultaneously evaluates band gap and band dispersion accuracy in the low-energy regime governing moiré phenomena.

### E. Model hyperparameters and descriptor coverage

The optimal GPR regularization $\eta$ exhibits system-dependent behavior, related to the application of the low-rank approximation to the overlap matrix. For all materials, the validation density error decreases monotonically with decreasing $\eta$ [see Eq. (4)] until reaching a saturation value determined by the descriptor, which is expected since our loss function in Eq. (4) relies on the spatial integral of the squared error in the density. However, the error in the band structure prediction of the twisted bilayers shows contrasting dependence on $\eta$ across material classes, as shown in Fig. 6. For well-conditioned systems (graphene, hBN), the optimal $\eta$ for band prediction is larger than that for density minimization. Stronger regularization suppresses overfitting to local structural variations and residual noise from density fitting. For ill-conditioned systems ($TiS_2$, $ZrS_2$, and $MoS_2$), the optimal $\eta$ for band prediction is smaller than that for the density. GPR regularization here stabilizes against low-rank truncation artifacts rather than suppressing noise. These results suggest that a modification directly to the loss function targeting property optimization could improve the models further. Nevertheless, the model accuracy in practice is not highly sensitive to the precise value of $\eta$: For graphene and hBN, a single compromise $\eta$ between the density and band optima increases either metric by less than a factor of 2; for TMDCs, both metrics share the same small-$\eta$ plateau. The total energies error on the moiré test sets remains low regardless of whether $\eta$ is optimized for band structure or density (see Sec. S2B of the Supplemental Material [47] for details). The workflow does not require precise per-property hyperparameter tuning.

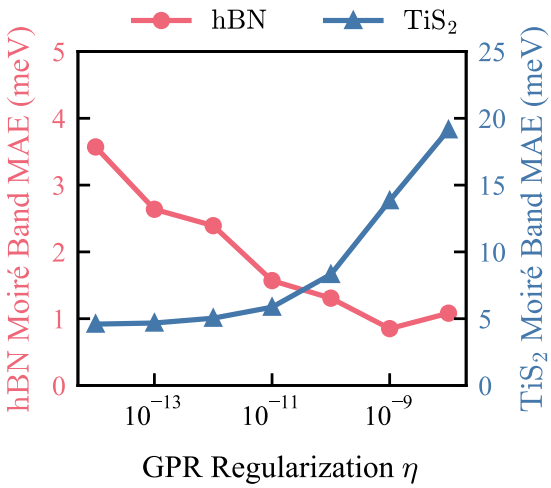


FIG. 6. Low-energy moiré band error varies by GPR regularization $\eta$ for hBN and $TiS_2$. As $\eta$ increases, low-energy moiré band error decreases for hBN but increases for $TiS_2$, showing contrary trends. The descriptors here are LOVV.

Gaussian process regression predicts via weighted similarity in descriptor space, meaning that accurate extrapolation from aligned to twisted structures relies on the descriptors used in the prediction of extrapolated structures being represented within the descriptors associated with the training data. We used dimensionality reduction via UMAP [78] to enable a qualitative descriptor coverage assessment.

As an example, the UMAP plots of the LOVV and LODE descriptors for $TiS_2$ are shown in Fig. 7. LODE descriptors show limited overlap between the training (aligned bilayers) and extrapolated (twisted bilayers) distributions, while the LOVV descriptors of extrapolated structures all lie within the distribution of the descriptors of the training structures. Equivalent plots for the other materials are provided in Sec. S5B of the Supplemental Material [47]. These results suggest qualitatively that short-range descriptors are less likely to cover the twisted bilayer descriptor space, leading to extrapolation

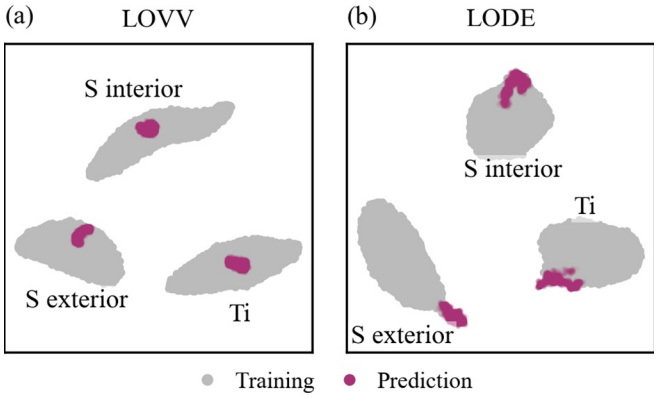


FIG. 7. Descriptor distribution dimension reduction by UMAP for $TiS_2$. For both (a) LOVV and (b) LODE descriptors, descriptors for each atom are automatically clustered according to atomic species (Ti and S) and geometric position (interior S atoms and exterior S atoms). Training (aligned bilayers) and test (twisted structures) atomic descriptors are colored in gray and purple, respectively. The prediction distribution can be well covered by the training distribution in LOVV, while in LODE, the prediction cluster of Ti atoms is not well covered by their training counterpart, and both interior and exterior S atom clusters lie on the edge of the training clusters.

failures, while LOVV's long-range nature ensures sufficient coverage for accurate predictions.

### F. Model training and conditioning

We trained SALTED models to predict the electron density and band structure of a range of 2D bilayer materials: graphene, hBN, $TiS_2$, $ZrS_2$, and $MoS_2$. We began by compiling a dataset of 512 aligned (displaced) bilayer structures of size $3 \times 3$ for each material, which were generated by randomly sampling in-plane displacements, interlayer spacings between the bilayers, and per-atom coordinate perturbations within each layer. The range of the displacements in each one of these coordinates is described in Sec. S1A of the Supplemental Material [47]. Each dataset was split into training and validation subsets of 400 and 112 structures, respectively.

The electron densities used to train the machine learning models and the reference band structures of the twisted bilayer structures were calculated using the PBE functional [79]. All of these calculations were performed using FHI-aims [76,77], using light or intermediate basis sets depending on DF accuracy in reproducing band structure as described in Sec. S1 of the Supplemental Material [47]. Detailed specifications of dataset generation, auxiliary basis set selection, hyperparameter optimization, and structural parameters are provided in Secs. S1 and S2 of the Supplemental Material [47]. Despite the shortcomings of PBE and other generalized gradient approximation functionals regarding the description of band gaps, we took a pragmatic choice of employing this functional to train the model because this is the *de facto* standard on previous studies of the materials reported here [6,56,80–82]. Training SALTED on data from hybrid functionals also works seamlessly and will be the topic of future work.

Figure 8 shows the distribution of the singular values of the overlap matrices for every structure in the training dataset of each material, along with the average condition numbers of these matrices. The condition numbers indicate particularly severe ill conditioning for the TMDC training datasets, likely due to the larger and more diffuse DF auxiliary basis sets required by heavier elements. As a result, the low-rank approximation to the overlap matrix described in Sec. IV C was applied to the TMDC training datasets, with $\delta = 10^{-6}$ used for $TiS_2$, and $\delta = 10^{-2}$ for both $MoS_2$ and $ZrS_2$.

*Workflow*: The overall workflow for predicting electron densities and band structures of 2D materials described in the preceding sections is summarized here:

(1) We begin by generating a training dataset, ensuring a range of intralayer and interlayer geometries are sampled.

(2) The electron densities of this training set are computed using DFT, and the corresponding DF coefficients are obtained using an auxiliary basis set that ensures density and band structure reproducibility.

(3) The condition number of the overlap matrix of these auxiliary basis functions is calculated for each structure in the dataset, and if the average condition number exceeds our empirical threshold ($\kappa > 10^{9.5}$), the low-rank approximation to the overlap matrices is applied.

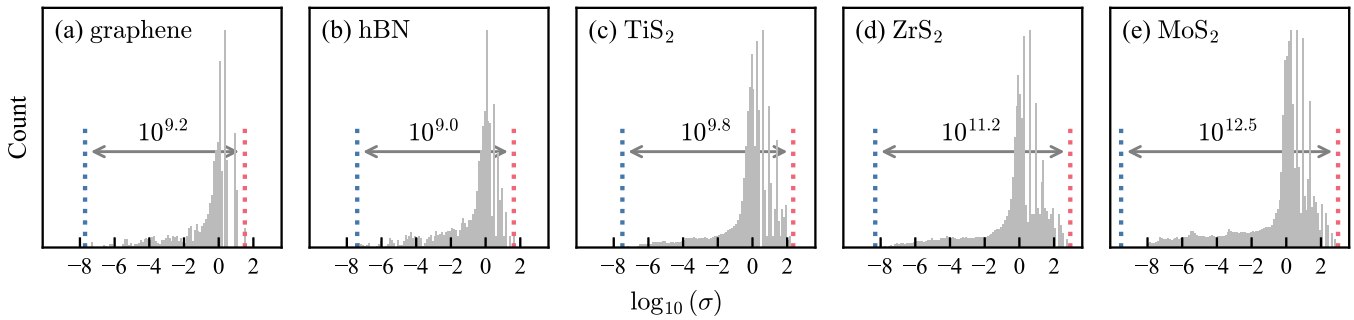


FIG. 8. Singular value distribution of overlap matrices. (a)–(e) Singular value decomposition results for DF basis overlap matrices across all training structures. The $y$ axis shows counts (arbitrary units), and the $x$ axis shows log-scale singular values. Vertical green/red lines indicate mean minimum/maximum singular values, with their separation representing the average condition number. TMDCs, especially $MoS_2$, exhibit significantly larger condition numbers than light-element systems.

(4) A descriptor is then selected to encode the training structures, with the appropriate choice depending on both the material of interest and the property being targeted.

(5) The SALTED model is then trained, optimizing the regularization hyperparameter $\eta$, and if appropriate the descriptor hyperparameters. The model is evaluated against both error in the predicted electron density and accuracy of the predicted band structure, since low errors in the former do not guarantee low errors in the latter.

To illustrate the practical computational gains of using the SALTED model, we take $ZrS_2$ as an example: Computing a twisted structure at $\theta = 2.281^\circ$ (3786 atoms per unit cell) would require about 11.5k CPU hours to converge the DFT electronic density (estimated by the scaling reported in Fig. S28 of the Supplemental Material [47]). In comparison, the total cost of $ZrS_2$ dataset generation and training of five SALTED models for $ZrS_2$ is around 6.7k CPU hours, attesting to the efficiency of using such models on the study of twisted bilayer systems.

## ACKNOWLEDGMENTS

The authors thank Andrea Grisafi for discussions and clarifications regarding different descriptors and the SALTED infrastructure. The authors also thank Lede Xian for initial discussions on twisted bilayer systems and Ángel Rubio for insightful input on the application aspects of this framework. Z.L. thanks Dante Kennes, Jingkai Quan, and Xinle Cheng for helpful discussions. Z.L. was partially funded by the Max Planck Graduate Center for Quantum Materials. Computational time was provided by the MPCDF. The work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project No. 555467911, CRC 1772/TP A06.

The authors declare no competing financial interest.

## DATA AVAILABILITY

Datasets and models are available on Zenodo [83].

[1] Y. Cao, V. Fatemi, S. Fang, K. Watanabe, T. Taniguchi, E. Kaxiras, and P. Jarillo-Herrero, Unconventional superconductivity in magic-angle graphene superlattices, Nature (London) **556**, 43 (2018).

[2] L. Xian, D. M. Kennes, N. Tancogne-Dejean, M. Altarelli, and A. Rubio, Multiflat bands and strong correlations in twisted bilayer boron nitride: Doping-induced correlated insulator and superconductor, Nano Lett. **19**, 4934 (2019).

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

[4] K. Yasuda, X. Wang, K. Watanabe, T. Taniguchi, and P. Jarillo-Herrero, Stacking-engineered ferroelectricity in bilayer boron nitride, Science **372**, 1458 (2021).

[5] D. M. Kennes, M. Claassen, L. Xian, A. Georges, A. J. Millis, J. Hone, C. R. Dean, D. N. Basov, A. N. Pasupathy, and A. Rubio, Moiré heterostructures as a condensed-matter quantum simulator, Nat. Phys. **17**, 155 (2021).

[6] M. Claassen, L. Xian, D. M. Kennes, and A. Rubio, Ultrastrong spin–orbit coupling and topological moiré engineering in twisted $ZrS_2$ bilayers, Nat. Commun. **13**, 4915 (2022).

[7] F. Wu, L. Li, Q. Xu, L. Liu, Y. Yuan, J. Zhao, Z. Huang, X. Zan, K. Watanabe, T. Taniguchi, D. Shi, L. Xian, W. Yang, L. Du, and G. Zhang, Coupled ferroelectricity and correlated states in a twisted quadrilayer $MoS_2$ moiré superlattice, Chin. Phys. Lett. **40**, 047303 (2023).

[8] C. S. Tsang, X. Zheng, T. Yang, Z. Yan, W. Han, L. W. Wong, H. Liu, S. Gao, K. H. Leung, C.-S. Lee, S. P. Lau, M. Yang, J. Zhao, and T. H. Ly, Polar and quasicrystal vortex observed in twisted-bilayer molybdenum disulfide, Science **386**, 198 (2024).

[9] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, Graphene bilayer with a twist: Electronic structure, Phys. Rev. Lett. **99**, 256802 (2007).

[10] R. Bistritzer and A. H. MacDonald, Moiré bands in twisted double-layer graphene, Proc. Natl. Acad. Sci. USA **108**, 12233 (2011).

[11] E. Suárez Morell, J. D. Correa, P. Vargas, M. Pacheco, and Z. Barticevic, Flat bands in slightly twisted bilayer graphene: Tight-binding calculations, Phys. Rev. B **82**, 121407 (2010).

[12] S. Pathak, T. Rakib, R. Hou, A. Nevidomskyy, E. Ertekin, H. T. Johnson, and L. K. Wagner, Accurate tight-binding model for twisted bilayer graphene describes topological flat bands without geometric relaxation, Phys. Rev. B **105**, 115141 (2022).

[13] C.-E. Ahn, W. Lee, K. Yananose, Y. Kim, and G. Y. Cho, Non-Abelian fractional quantum anomalous Hall states and first Landau level physics of the second moiré band of twisted bilayer $MoTe_2$, Phys. Rev. B **110**, L161109 (2024).

[14] C. Xu, N. Mao, T. Zeng, and Y. Zhang, Multiple Chern bands in twisted $MoTe_2$ and possible non-Abelian states, Phys. Rev. Lett. **134**, 066601 (2025).

[15] F. Guinea and N. R. Walet, Continuum models for twisted bilayer graphene: Effect of lattice deformation and hopping parameters, Phys. Rev. B **99**, 205134 (2019).

[16] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer, Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions, Nat. Commun. **10**, 5024 (2019).

[17] O. Unke, M. Bogojeski, M. Gastegger, M. Geiger, T. Smidt, and K.-R. Müller, SE(3)-equivariant prediction of molecular wavefunctions and electronic densities, Adv. Neural Inf. Process. Syst. **34**, 14434 (2021).

[18] H. Li, Z. Wang, N. Zou, M. Ye, R. Xu, X. Gong, W. Duan, and Y. Xu, Deep-learning density functional theory Hamiltonian for efficient *ab initio* electronic-structure calculation, Nat. Comput. Sci. **2**, 367 (2022).

[19] X. Gong, H. Li, N. Zou, R. Xu, W. Duan, and Y. Xu, General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian, Nat. Commun. **14**, 2848 (2023).

[20] Y. Zhong, H. Yu, M. Su, X. Gong, and H. Xiang, Transferable equivariant graph neural networks for the Hamiltonians of molecules and solids, npj Comput. Mater. **9**, 182 (2023).

[21] Z. Tang, H. Li, P. Lin, X. Gong, G. Jin, L. He, H. Jiang, X. Ren, W. Duan, and Y. Xu, A deep equivariant neural network approach for efficient hybrid density functional calculations, Nat. Commun. **15**, 8815 (2024).

[22] Y. Ma, H. Yu, Y. Zhong, S. Chen, X. Gong, and H. Xiang, Transferable machine learning approach for predicting electronic structures of charged defects, Appl. Phys. Lett. **126**, 044103 (2025).

[23] C. Zhang, Y. Zhong, Z.-G. Tao, X. Qin, H. Shang, Z. Lan, O. V. Prezhdo, X.-G. Gong, W. Chu, and H. Xiang, Advancing nonadiabatic molecular dynamics simulations in solids with E(3) equivariant deep neural Hamiltonians, Nat. Commun. **16**, 2033 (2025).

[24] Y. Zhong, R. Wang, X. Gong, and H. Xiang, A universal spin–orbit-coupled Hamiltonian model for accelerated quantum material discovery, Nat. Mach. Intell. **8**, 403 (2026).

[25] A. M. Lewis, A. Grisafi, M. Ceriotti, and M. Rossi, Learning electron densities in the condensed phase, J. Chem. Theory Comput. **17**, 7203 (2021).

[26] P. B. Jørgensen and A. Bhowmik, Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids, npj Comput. Mater. **8**, 183 (2022).

[27] A. Grisafi, A. M. Lewis, M. Rossi, and M. Ceriotti, Electronic-structure properties from atom-centered predictions of the electron density, J. Chem. Theory Comput. **19**, 4451 (2023).

[28] L. Fiedler, N. A. Modine, S. Schmerler, D. J. Vogel, G. A. Popoola, A. P. Thompson, S. Rajamanickam, and A. Cangi, Predicting electronic structures at any length scale with machine learning, npj Comput. Mater. **9**, 115 (2023).

[29] T. Lv, Z. Zhong, Y. Liang, F. Li, J. Huang, and R. Zheng, Deep charge: Deep learning model of electron density from a one-shot density functional theory calculation, Phys. Rev. B **108**, 235159 (2023).

[30] J. Rackers, L. Tecot, M. Geiger, and T. Smidt, A recipe for cracking the quantum scaling limit with machine learned electron densities, Mach. Learn.: Sci. Technol. **4**, 015027 (2023).

[31] B. Focassio, M. Domina, U. Patil, A. Fazzio, and S. Sanvito, Linear Jacobi-Legendre expansion of the charge density for machine learning-accelerated electronic structure calculations, npj Comput. Mater. **9**, 87 (2023).

[32] S. K. Achar, L. Bernasconi, and J. K. Johnson, Machine learning electron density prediction using weighted smooth overlap of atomic positions, Nanomaterials **13**, 1853 (2023).

[33] R.-G. Lee and Y.-H. Kim, Convolutional network learning of self-consistent electron density via grid-projected atomic fingerprints, npj Comput. Mater. **10**, 248 (2024).

[34] K. Ushenin, K. Khrabrov, A. Tsypin, A. Ber, E. Rumiantsev, and A. Kadurin, LAGNet: Better electron density prediction for LCAO-based data and drug-like substances, J. Cheminf. **17**, 65 (2025).

[35] C. Li, O. Sharir, S. Yuan, and G. K.-L. Chan, Image super-resolution inspired electron density prediction, Nat. Commun. **16**, 4811 (2025).

[36] T. Bao, R. Xu, H. Li, X. Gong, Z. Tang, J. Fu, W. Duan, and Y. Xu, Deep-learning database of density functional theory Hamiltonians for twisted materials, arXiv:2404.06449.

[37] W. Kohn, Density functional and density matrix method scaling linearly with the number of atoms, Phys. Rev. Lett. **76**, 3168 (1996).

[38] J. Behler, Four generations of high-dimensional neural network potentials, Chem. Rev. **121**, 10037 (2021).

[39] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, Phys. Rev. **136**, B864 (1964).

[40] X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, Resolution-of-identity approach to Hartree–Fock, hybrid density functionals, RPA, MP2 and *GW* with numeric atom-centered orbital basis functions, New J. Phys. **14**, 053020 (2012).

[41] A. C. Ihrig, J. Wieferink, I. Y. Zhang, M. Ropo, X. Ren, P. Rinke, M. Scheffler, and V. Blum, Accurate localized resolution of identity approach for linear-scaling hybrid density functionals and for many-body perturbation theory, New J. Phys. **17**, 093020 (2015).

[42] A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, Symmetry-adapted machine learning for tensorial properties of atomistic systems, Phys. Rev. Lett. **120**, 036002 (2018).

[43] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, Transferable machine-learning model of the electron density, ACS Cent. Sci. **5**, 57 (2019).

[44] A. M. Lewis, P. Lazzaroni, and M. Rossi, Predicting the electronic density response of condensed-phase systems to electric field perturbations, J. Chem. Phys. **159**, 014103 (2023).

[45] Z. Y. Zhu, Y. C. Cheng, and U. Schwingenschlögl, Giant spin-orbit-induced spin splitting in two-dimensional transition-metal dichalcogenide semiconductors, Phys. Rev. B **84**, 153402 (2011).

[46] W. P. Huhn and V. Blum, One-hundred-three compound band-structure benchmark of post-self-consistent spin-orbit coupling treatments in density functional theory, Phys. Rev. Mater. **1**, 033803 (2017).

[47] See Supplemental Material at http://link.aps.org/supplemental/10.1103/4575-9cmx for details of dataset generation, model optimization and validation, application analyses, computational benchmarks, and descriptor sensitivity analysis, which includes Refs. [84–98].

[48] K. Uchida, S. Furuya, J.-I. Iwata, and A. Oshiyama, Atomic corrugation and electron localization due to moiré patterns in twisted bilayer graphenes, Phys. Rev. B **90**, 155451 (2014).

[49] S. Carr, S. Fang, Z. Zhu, and E. Kaxiras, Exact continuum model for low-energy electronic states of twisted bilayer graphene, Phys. Rev. Res. **1**, 013001 (2019).

[50] G. Cantele, D. Alfè, F. Conte, V. Cataudella, D. Ninno, and P. Lucignano, Structural relaxation and low-energy properties of twisted bilayer graphene, Phys. Rev. Res. **2**, 043127 (2020).

[51] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csanyi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, Adv. Neural Inf. Process. Syst. **35**, 11423 (2022).

[52] A. Tkatchenko, R. A. DiStasio, Jr., R. Car, and M. Scheffler, Accurate and efficient method for many-body van der Waals interactions, Phys. Rev. Lett. **108**, 236402 (2012).

[53] M. J. Calderón and E. Bascones, Interactions in the 8-orbital model for twisted bilayer graphene, Phys. Rev. B **102**, 155149 (2020).

[54] M. Koshino, N. F. Q. Yuan, T. Koretsune, M. Ochi, K. Kuroki, and L. Fu, Maximally localized Wannier orbitals and the extended Hubbard model for twisted bilayer graphene, Phys. Rev. X **8**, 031087 (2018).

[55] D. Bennett, G. Chaudhary, R.-J. Slager, E. Bousquet, and P. Ghosez, Polar meron-antimeron networks in strained and twisted bilayers, Nat. Commun. **14**, 1629 (2023).

[56] D. S. Kim, R. C. Dominguez, R. Mayorga-Luna, D. Ye, J. Embley, T. Tan, Y. Ni, Z. Liu, M. Ford, F. Y. Gao, S. Arash, K. Watanabe, T. Taniguchi, S. Kim, C.-K. Shih, K. Lai, W. Yao, L. Yang, X. Li, and Y. Miyahara, Electrostatic moiré potential from twisted hexagonal boron nitride layers, Nat. Mater. **23**, 65 (2024).

[57] L. Gu, L. Zhang, S. Felsenfeld, B. Gao, R. Ma, S. Park, H. Jang, T. Taniguchi, K. Watanabe, and Y. Zhou, Quantum confining excitons with an electrostatic moiré superlattice, Phys. Rev. Lett. **135**, 026901 (2025).

[58] P. Zhao, C. Xiao, and W. Yao, Universal superlattice potential for 2D materials from twisted interface inside h-BN substrate, npj 2D Mater. Appl. **5**, 38 (2021).

[59] T. Vachaspati, *Kinks and Domain Walls: An Introduction to Classical and Quantum Solitons* (Cambridge University Press, Cambridge, 2006).

[60] J. S. Alden, A. W. Tsen, P. Y. Huang, R. Hovden, L. Brown, J. Park, D. A. Muller, and P. L. McEuen, Strain solitons and topological defects in bilayer graphene, Proc. Natl. Acad. Sci. USA **110**, 11256 (2013).

[61] G. X. Ni, H. Wang, B.-Y. Jiang, L. X. Chen, Y. Du, Z. Y. Sun, M. D. Goldflam, A. J. Frenzel, X. M. Xie, M. M. Fogler, and D. N. Basov, Soliton superlattices in twisted hexagonal boron nitride, Nat. Commun. **10**, 4360 (2019).

[62] C. R. Woods, P. Ares, H. Nevison-Andrews, M. J. Holwill, R. Fabregas, F. Guinea, A. K. Geim, K. S. Novoselov, N. R. Walet, and L. Fumagalli, Charge-polarized interfacial superlattices in marginally twisted hexagonal boron nitride, Nat. Commun. **12**, 347 (2021).

[63] Y. Zhang, T. Liu, and L. Fu, Electronic structures, charge transfer, and charge order in twisted transition metal dichalcogenide bilayers, Phys. Rev. B **103**, 155142 (2021).

[64] A. Grisafi, J. Nigam, and M. Ceriotti, Multi-scale approach for the prediction of atomic scale properties, Chem. Sci. **12**, 2078 (2021).

[65] J. Quan, N. Rybin, M. Scheffler, and C. Carbogno, Efficient band structure unfolding with atom-centered orbitals: General theory and application, Phys. Rev. B **113**, 085112 (2026).

[66] C. Chen *et al.*, Strong electron–phonon coupling in magic-angle twisted bilayer graphene, Nature (London) **636**, 342 (2024).

[67] C. Chen, W. Holtzmann, X.-W. Zhang, E. Anderson, S. He, Y. Zhao, W. Li, J. Liu, Y. Guo, C. Jozwiak, A. Bostwick, E. Rotenberg, K. Watanabe, T. Taniguchi, T. Cao, D. Xiao, X. Xu, and Y. Chen, Visualizing electronic structure of twisted bilayer $MoTe_2$ in devices, Commun. Phys. **9**, 62 (2026).

[68] R. Guo, H. Yu, L. Hong, S. Chen, X. Gong, and H. Xiang, Capturing long-range interactions with a reciprocal-space neural network, Phys. Rev. B **113**, 174101 (2026).

[69] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, Phys. Rev. B **87**, 184115 (2013).

[70] M. J. Willatt, F. Musil, and M. Ceriotti, Atom-density representations for machine learning, J. Chem. Phys. **150**, 154110 (2019).

[71] A. Grisafi and M. Ceriotti, Incorporating long-range physics in atomic-scale machine learning, J. Chem. Phys. **151**, 204105 (2019).

[72] A. Grisafi, A. Bussy, M. Salanne, and R. Vuilleumier, Predicting the charge density response in metal electrodes, Phys. Rev. Mater. **7**, 125403 (2023).

[73] M. Rossi, K. Rossi, A. M. Lewis, M. Salanne, and A. Grisafi, Learning the electrostatic response of the electron density through a symmetry-adapted vector field model, J. Phys. Chem. Lett. **16**, 2326 (2025).

[74] G. H. Golub and C. F. Van Loan, *Matrix Computations*, *Johns Hopkins Studies in the Mathematical Sciences*, 4th ed. (The Johns Hopkins University Press, Baltimore, 2013).

[75] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for Machine Learning* (Cambridge University Press, Cambridge, 2020).

[76] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, *Ab initio* molecular simulations with numeric atom-centered orbitals, Comput. Phys. Commun. **180**, 2175 (2009).

[77] J. W. Abbott *et al.*, Roadmap on advancements of the FHI-aims software package, Electron. Struct. (2026).

[78] L. McInnes, J. Healy, and J. Melville, UMAP: Uniform Manifold Approximation and Projection for dimension reduction, arXiv:1802.03426.

[79] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).

[80] S. Fang and E. Kaxiras, Electronic structure theory of weakly interacting bilayers, Phys. Rev. B **93**, 235153 (2016).

[81] G. Li, K. Yao, and G. Gao, Strain-induced enhancement of thermoelectric performance of $TiS_2$ monolayer based on first-principles phonon and electron band structures, Nanotechnology **29**, 015204 (2018).

[82] J. Xiao, M. Long, X. Li, Q. Zhang, H. Xu, and K. S. Chan, Effects of van der Waals interaction and electric field on the electronic structure of bilayer $MoS_2$, J. Phys.: Condens. Matter **26**, 405302 (2014).

[83] Z. Lou, A. M. Lewis, and M. Rossi, Datasets and models for the paper "Long-range machine learning of electron density for twisted bilayer moiré materials," Zenodo, 2026, https://doi.org/10.5281/zenodo.18556914.

[84] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, Continuum model of the twisted graphene bilayer, Phys. Rev. B **86**, 155449 (2012).

[85] J. He, K. Hummer, and C. Franchini, Stacking effects on the electronic and optical properties of bilayer transition metal dichalcogenides $MoS_2$, $MoSe_2$, $WS_2$, and $WSe_2$, Phys. Rev. B **89**, 075409 (2014).

[86] S. M. Gilbert, T. Pham, M. Dogan, S. Oh, B. Shevitski, G. Schumm, S. Liu, P. Ercius, S. Aloni, M. L. Cohen, and A. Zettl, Alternative stacking sequences in hexagonal boron nitride, 2D Mater. **6**, 021006 (2019).

[87] K. V. Zakharchenko, M. I. Katsnelson, and A. Fasolino, Finite temperature lattice properties of graphene beyond the quasiharmonic approximation, Phys. Rev. Lett. **102**, 046808 (2009).

[88] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, Structural relaxation made simple, Phys. Rev. Lett. **97**, 170201 (2006).

[89] A. Hjorth Larsen *et al.*, The atomic simulation environment—A Python library for working with atoms, J. Phys.: Condens. Matter **29**, 273002 (2017).

[90] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and

K. A. Persson, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, APL Mater. **1**, 011002 (2013).

[91] J. Hermann and A. Tkatchenko, Density functional model for van der Waals interactions: Unifying many-body atomic approaches with nonlocal functionals, Phys. Rev. Lett. **124**, 146401 (2020).

[92] J. T. Krogel, S. F. Yuk, P. R. C. Kent, and V. R. Cooper, Perspectives on van der Waals density functionals: The case of $TiS_2$, J. Phys. Chem. A **124**, 9867 (2020).

[93] Anisha, R. Kumar, S. Srivastava, and T. Kumar, Thermoelectric performance of 1T-$ZrS_2$ bilayer using stacking engineering, Phys. Scr. **99**, 015914 (2024).

[94] P. Tanwar, U. Paliwal, K. B. Joshi, and J. Kumar, First-principles study of structural, electronic and vibrational properties of bulk and monolayer $TiS_2$, J. Phys. Chem. Solids **179**, 111382 (2023).

[95] D. Zhao, K. Fang, and Z. Lian, Mechanical and vibrational behaviors of bilayer hexagonal boron nitride in different stacking modes, Sci. Rep. **14**, 10619 (2024).

[96] L. van der Maaten, Learning a parametric embedding by preserving local structure, Proc. Mach. Learn. Res. **5**, 384 (2009).

[97] J. Venna and S. Kaski, Local multidimensional scaling, Neural Netw. **19**, 889 (2006).

[98] J. Venna and S. Kaski, Neighborhood preservation in nonlinear projection methods: An experimental study, in *Artificial Neural Networks — ICANN 2001*, Vol. 2130, edited by G. Dorffner, H. Bischof, and K. Hornik (Springer, Berlin, Heidelberg, 2001), pp. 485–491.