npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-025-01668-5

Bridging deep learning force fields and electronic structures with a physics-informed approach

Yubo Qi [1,2], Weiyi Gong[1] & Qimin Yan [1]

This work presents a physics-informed neural network approach bridging deep-learning force field and electronic structure simulations, illustrated through twisted two-dimensional large-scale material systems. The deep potential molecular dynamics model is adopted as the backbone, and the electronic structure simulation is integrated. Using Wannier functions as the basis, we categorize Wannier Hamiltonian elements based on physical principles to incorporate diverse information from a deep-learning force field model. This information-sharing mechanism streamlines the architecture of our dual-functional model, enhancing its efficiency and effectiveness. This Wannier-based dual-functional model for simulating electronic band and structural relaxation (WANDER) serves as a powerful tool to explore large-scale systems. By endowing a well-developed machine-learning force field with electronic structure simulation capabilities, the study marks a significant advancement in developing multimodal machine-learning-based computational methods that can achieve multiple functionalities traditionally exclusive to first-principles calculations. Moreover, utilizing Wannier functions as the basis lays the groundwork for predicting more physical quantities.

First-principles calculations based on density functional theory (DFT) have emerged as a powerful tool for predicting and calculating physical properties[1]. Their accurate descriptions of structural parameters, energy landscapes, and electronic structures, directly comparable with experimental measurements, have yielded deep physical insights[2]. However, as the system gets large, the computational cost of DFT calculation increases dramatically, making the simulations of many scientifically important solid-state systems, such as twisted 2D materials, heterostructures, and crystals with defects, technically challenging and even impossible. The recent developments in machine learning techniques demonstrate promising potential in addressing this challenge[3–17]. Specifically, machine-learning force fields have been shown to produce highly accurate atomic forces and crystalline energies[7–23]. In these force fields, the local atomic environment is digitized, and neural networks are employed to identify the relationship between the local atomic environment and atomic forces. In addition to the capabilities of predicting atomic forces and optimizing complex structures, machine learning techniques also demonstrate promising potential in electronic structure simulation. Recently, approaches (such as DeepH and HamGNN) based on message-passing graph neural networks have been developed to simulate the first-principles Hamiltonian and subsequently acquire the electronic structures[24–27]. Each of these works reproduces a primary functionality of DFT calculations with a significantly increased efficiency, representing a substantial step in developing machine learning-based computational methodologies.

The success of deep learning techniques in both force field and electronic structure prediction is rooted in physical principles. According to the "nearsightedness" model proposed by Prodan and Kohn, the change of potential at a distant position has little effect on local electronic properties[28]. This principle suggests that both atomic forces and electronic structures are inherently local properties. Deep learning has proven to be a powerful tool for identifying relationships between various quantities, including local structures and their corresponding physical properties. Moreover, this principle also implies that atomic forces and electronic structures can be predicted by similar models.

However, the developments in deep learning models for force fields and electronic structure simulations are imbalanced. While there has been remarkable progress in machine-learning force fields[7,18,19,23,29–33], there is still a notable scarcity of deep-learning models for solid-state electronic structure simulation, despite the success of models like DeepH, HamGNN, and their peers[24,27,34–38]. This observation prompts the question of whether it is feasible to bridge this gap by developing a method endowing existing machine learning force fields with the capability for electronic structure simulation. If so, the inquiry extends to identifying the guidelines for designing such a

[1]Department of Physics, Northeastern University, Boston, MA, USA. [2]Department of Physics, University of Alabama at Birmingham, Birmingham, AL, USA.
e-mail: q.yan@northeastern.edu

dual-functional model that not only ensures optimal efficiency and transferability but also attains high levels of accuracy.

Inspired by these challenges, we develop a physics-informed neural network approach for simulating both atomic and electronic structures. This approach uses the Wannier functions (generated from atomic orbitals) as the basis, and adopts the framework of a well-established and widely adopted machine learning force field model[7,29]. The physical quantities required for electronic band structure prediction, which are Wannier Hamiltonian elements, can be obtained using the well-developed and open-source package Wannier90 (see Supplementary Information Section 4 for technical details). Information learned from the force field model is used to facilitate the simulation of electronic band structures, and this information-sharing mechanism streamlines the architecture of our dual-functional model, enhancing its efficiency and effectiveness. To demonstrate the performance of this Wannier-based dual functional model for simulating electronic band and structural relaxation (WANDER), we take twisted MoS_2 bilayer systems as illustrative examples since the Moiré structures of 2D materials host a wide range of novel physical properties[39–41]. This work aims to bridge deep learning-based models for atomic-structure and electronic-structure simulations and provide insights on developing machine-learning-based computational methods offering multiple functionalities of first-principles calculations. Moreover, this research presents a feasible approach to augment machine learning models for electronic band structure simulations, potentially bridging the gap between the scarcity of machine learning-based electronic structure simulation approaches and the prosperity of machine learning-based force fields. Compared to peer models, WANDER is a multifunctional framework capable of predicting both atomic forces and electronic properties, with extensibility to additional physical quantities by using Wannier functions as the basis. Moreover, it employs physics-informed input representations, significantly reducing the number of free parameters and enhancing efficiency and transferability.

This work adopts Wannier functions as the basis, which are a complete set of real-space orthogonal functions acquired from the Fourier transforms of the Bloch functions[42]. Wannier functions have been widely adopted in the research of theoretical solid-state physics since they capture the essential physics of a material's electronic structure[43,44]. For example, the k-space Hamiltonian matrix elements, whose eigenvalues yield the electronic band structure, can be acquired through Wannier interpolation as

$$H(\mathbf{k})_{m\alpha,n\beta} = \sum_{\mathbf{a}} e^{i\mathbf{k}\cdot\mathbf{a}} \langle w_{m\alpha}(\mathbf{r})|H(\mathbf{r})|w_{n\beta}(\mathbf{r}+\mathbf{a})\rangle. \quad (1)$$

Here, $H(\mathbf{k})$ and $H(\mathbf{r})$ are the Hamiltonians in the reciprocal space and real space, respectively. **a**, $m(n)$, and $\alpha(\beta)$ are the indices of cells, atoms, and Wannier functions, respectively. In the WANDER, the Wannier Hamiltonian elements, which determine the electronic band structures, are divided into three categories (on-site interactions, intra-layer hopping integrals, and inter-layer hopping integrals) as suggested in a previous tight-binding model[45]. Our study shows that the input representation of each category should take different information from the deep learning force field model for improved performance in terms of accuracy, efficiency, and transferability. The selection of these representations is guided by the underlying physical mechanisms, aligning with the concept of physics-informed machine learning[46]. While our focus is primarily on atomic and electronic structure simulations, utilizing Wannier functions as the basis lays the groundwork for integrating additional functionalities into our approach. This expansion broadens the predictive capabilities of our model to encompass a diverse array of physical quantities calculable based on Wannier functions, including but not limited to spin Hall conductivity[47,48], shift current[49], transport properties[50], and electron-phonon coupling[51,52].

Results

Overview of Wannier functions

Wannier functions for a specific system are non-unique since a Bloch function $\phi_{n\mathbf{k}}$ can adopt a phase $e^{i\phi(\mathbf{k})}$ as an arbitrary function of $\mathbf{k}$. There are two widely adopted approaches for acquiring well-defined Wannier functions. The first approach is to minimize their spreads (through a process known as localization) to acquire maximally localized Wannier functions (MLWFs)[43,44]. MLWFs are exponentially decaying and widely adopted as the basis for constructing first-principles, tight-binding models[53,54]. However, MLWFs change with structural distortions dramatically, whose trend is challenging to track; for Chern insulators, it is even infeasible to construct exponentially localized Wannier functions[55]. An alternative approach, which can get rid of these challenges, is to use Wannierized atomic orbitals, without any localization steps, as the basis[56–58]. Wannier functions acquired through this method have larger spreads, leading to increased computational costs.

In this work, we adopt Wannier functions generated from atomic orbitals with finite localization as the basis, and the computational procedure is as follows. First, we calculate the MLWFs of the ground state MoS_2, which is a trivial insulator. Then, we employ atomic orbitals to approximate the MLWFs. These orbitals should be linearly independent and form a complete set (see Supplementary Information Section 1 for details). Finally, we use the atomic orbitals as the initial guess for projection, acquiring Wannier functions and minimizing their spreads for finite (40 in this study) iterations with the Wannier90 package[59,60]. The acquired "semi-localized" Wannier functions are used as the basis.

The essence of this approach closely resembles that of using Wannierized atomic orbitals as the basis. Using the atomic orbitals, which approximate the MLWFs of the ground state, as the initial projection will not introduce any error as long as the atomic orbitals form a complete set. However, the subsequent localization iterations may introduce errors. Here, we take the assumption that the semi-localized Wannier functions exhibit slight variation across different structures. In other words, the localization process for finite iterations decreases the spread of Wannier functions with slight shape changes (see Supplementary Information Section 2 for details). Our results, based on bilayer MoS_2, align with this assumption, demonstrating that using "semi-localized" Wannier functions as the basis has little impact on the model's accuracy.

Wannier functions have proven effective in simulating more complex systems, including metals and strongly correlated materials (by using more advanced functionals than GGA), which should also fall within the predictive scope of WANDER[44,61,62]. It is worth mentioning that localization with finite steps is completely optional and can be viewed as a trade-off between accuracy and efficiency. While it may introduce errors, the localization process also breaks the symmetry of the Wannier functions, making this approach unsuitable for investigating topological properties. To address these limitations, future work may focus on developing symmetry-adapted Wannier functions or utilizing Wannier functions constructed from atomic orbitals that preserve their symmetries[63,64]. However, these topics are beyond the scope of this study and will be pursued in future research.

Neural network scheme

The schematic plot of the architecture of WANDER is shown in Fig. 1. The first step is to train a force field model, as indicated by the blue rectangles in Fig. 1. Here, we adopt the deep potential molecular dynamics (DPMD) model introduced in Ref. 7,29. The local environment of a specific atom is transformed into input data by stacking the 4-dimensional vector $\left(\frac{1}{R}, \frac{R_x}{R^2}, \frac{R_y}{R^2}, \frac{R_z}{R^2}\right)$ contributed by each neighboring atom inside the cutoff radius R_C. Here, R is the length of the vector R pointing from the central atom to its neighbor. R_x, R_y, and R_z are the three components of $\boldsymbol{R}$. The atomic forces in Cartesian coordinates $\left(f_x, f_y, f_z\right)$ serve as the output. Two hidden layers, with 32 and 16 nodes, respectively, connect the input and output, forming a feed-forward network. Between the input and different hidden layers, data are transferred

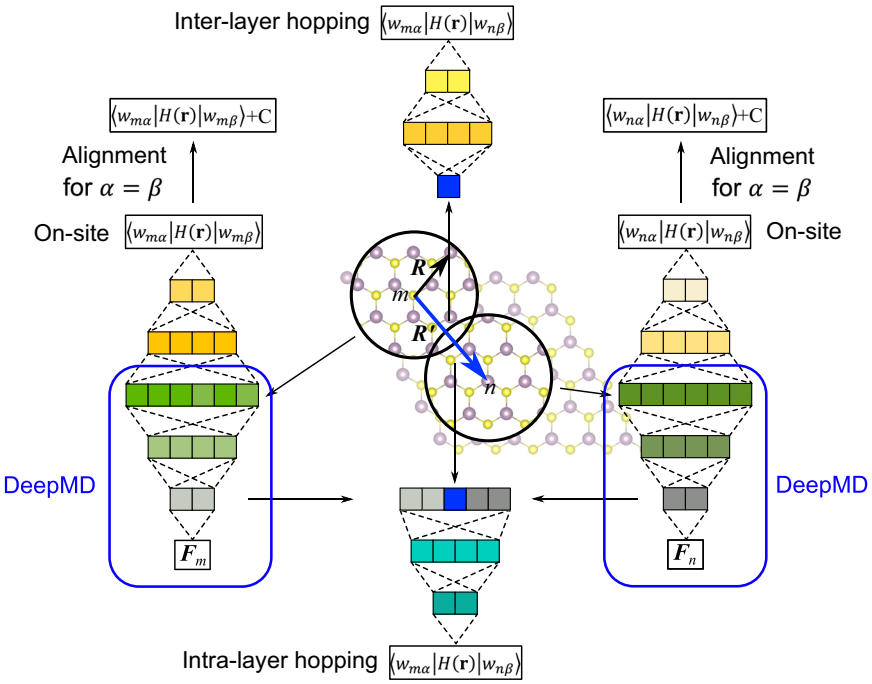

Fig. 1 | Schematic illustration of the architecture of the WANDER. First, an atomic force model is trained with the DPMD method, as indicated by the blue rectangles. The last hidden layer, referred to as RAER in this work, contains information about the local atomic environment. On-site interaction, intra-layer hopping integrals, and inter-layer hopping integrals are predicted with inputs for DPMD, RAERs plus relative-position vectors, and relative-position vectors, respectively.

through a linear transformation $\boldsymbol{d}_{k+1} = \boldsymbol{W}_k\boldsymbol{d}_k + \boldsymbol{b}_k$ followed by a rectified linear unit (ReLU) activation function. Here, $\boldsymbol{W}_k$ and $\boldsymbol{b}_k$ are learnable weights and biases at the kth step. Between the last hidden layer and the output layer, only a linear transformation is applied.

The Wannier Hamiltonian elements are classified into three categories as on-site interaction, intra-layer hopping integrals, and inter-layer hopping integrals. The underlying physics is that the local atomic environment affects the three types of interactions to varying extents. This classification is in accordance with the expression of the tight-binding Hamiltonian in ref. 45.

$$H_{\mathrm{TB}} = -\sum_{\langle m,n\rangle} t_{mn}^{\rho=0} c_m^{\dagger} c_n - \sum_{m,n} t_{mn}^{\rho=1} c_m^{\dagger} c_n + \sum_m \epsilon c_m^{\dagger} c_m, \tag{2}$$

where $\rho = 0$ and $\rho = 1$ stand for atoms i and j located at the same and different layers, respectively, and t_{mn} and ϵ are the coefficients for the hopping and on-site integrals.

The three categories of Wannier Hamiltonian elements (on-site interaction, intra-layer hopping integrals, and inter-layer hopping integrals) have different underlying physical mechanisms and should be predicted with different models, as shown in Fig. 1. Similar to an atomic force, an on-site interaction is a single-body term that is determined by the local atomic environment; predictions for atomic forces and on-site interactions share the same input. Hopping integrals are two-body interactions, and the relative positions of the two atoms should be incorporated into the input. The relative position is represented by a 4-dimensional vector $\left(\frac{1}{R'}, \frac{R'_x}{R'^2}, \frac{R'_y}{R'^2}, \frac{R'_z}{R'^2}\right)$, where R' is the length of the vector R' between the two atoms. To predict intra-layer hopping integrals, we extract the 16-dimensional vector of each atom from the last hidden layer in the DPMD model and stack it with the relative position vector as the input. Here, the last hidden layer in DPMD contains essential information about the local atomic environment and thus can be viewed as a reduced atomic environmental representation (RAER). The concept of RAER is related to the "reusing pretrained layers" approach, with the last hidden layer selected because our results show that it has a low dimensionality and contains the essential structural information. Later, we will show that this input containing RAER performs better than stacking the DPMD first-layer (input for DPMD) vector and the relative-position vector. The input for predicting the inter-layer hopping integral is the 4-dimensional relative-position vector. This is because there is no atom in the inter-layer gap affecting the Hamiltonian, making the hopping integral mainly depend on the relative positions of the two atoms. The deep learning architecture for predicting Wannier Hamiltonian elements is the same as that of DPMD, which comprises two hidden layers.

Dataset for training and vacuum level alignment

To build the dataset for training the force field model, we begin with the self-consistent field (SCF) calculations of 662 bilayer structures with 5 different biaxial strains spanning from − 2% to 8% (see Supplementary Fig. S3 for the distribution). In each structure, the top layer is translated with an arbitrary 3-dimensional vector. These structures will also serve as the dataset for training the electronic structure model.

To augment the dataset and capture the structures with slight distortions induced by the Moiré potential, we generate 1194 additional structures, each with the top layer translated by an arbitrary 3-dimensional vector, and optimize the structures. During the optimization, we anchor the four Mo atoms at the corners of each layer throughout the optimization (see Supplementary Information Section 3 for details). The first 18 steps of the structural relaxation trajectory (or the entire trajectory if relaxation terminated within 18 steps) were extracted for building the dataset. Together with the 662 structures with SCF calculations only, we have 22,052 configurations for training the force field. 80% of the data are used for training, and the remaining 20% is used for testing. In this study, we use only the training and testing datasets (without the validation dataset) because our simple model does not require extensive fine-tuning, aligning with prior work on DeepMD that achieved commendable performance.

To train models for predicting Wannier Hamiltonian elements (see Supplementary Information Section 4 for details), we employ the aforementioned subset of 662 configurations from the dataset containing 22,052 configurations for atomic-force prediction. 662 configurations give approximately 81,500,000 non-zero (with absolute magnitude larger than 1 meV) Wannier Hamiltonian elements. This quantity is substantial enough to facilitate the learning of the underlying rules. Details about the input data, output data, and neural network structure are in Supplementary Information Section 5.

To validate our model, we compare the DFT results of twisted bilayer MoS_2 structures. The twisted structures are undoubtedly not present in the

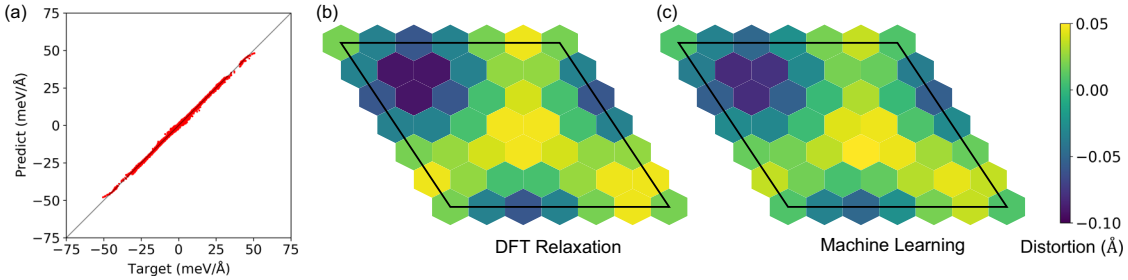

Fig. 2 | Performance of the WANDER for atomic-force predictions. a The parity plot for the atomic forces. Out-of-plane distortions induced by the Moiré potential given by **b** DFT calculation and **c** machine-learning model prediction. For a clear view, we only plot the distortions in the bottom layer.

training set composed of parallel non-twisted bilayer structures, serving as an outstanding demonstration of the transferability of this model. In twisted structures, one of the MoS_2 layers should be strained to be commensurate with periodic boundary conditions. For a large Moiré superlattice, the number of atoms in different layers can be different. In this case, a correction term should be added to the diagonal elements $< w_{m\alpha}|H|w_{m\alpha} >$ in the Wannier Hamiltonian to align the vacuum levels of the two layers.

In DFT calculations, the energy at the vacuum level E_{vl} is considered as zero energy[65]. E_{vl} is determined by the Hamiltonian H_{DFT}, which depends on the number of atoms in a supercell[66,67]. The removal or addition of atoms in a supercell invariably impacts the vacuum level of the entire system. In this model, the Wannier Hamiltonian elements are learned based on local atomic environments in those structures with equal numbers of atoms in each layer. To apply this mode to twisted bilayer systems, the vacuum level of the two layers with different numbers of atoms should be aligned by adding a constant C to the predicted diagonal terms $< w_{m\alpha}|H|w_{m\alpha} >$, as shown in Fig. 1. In other words, a diagonal term in the Wannier Hamiltonian depends not only on the neighboring atoms but also on the size of the molecular layer and the atoms outside the layer. This correction is necessary to ensure that the Wannier representation remains consistent with the DFT results.

Our calculations show that C for a layer depends on the strain ε and the ratio between the number of atoms N_{out} outside of the layer and the number of atoms N_{in} inside the layer. We build a shallow-level machine learning model (polynomial regression) to predict C with ε and $N_{\mathrm{out}}/N_{\mathrm{in}}$ as the input. The model performs quite well with an R-squared value of 0.9999999 (see Supplementary Information Section 6 for details).

Model performance for atomic forces prediction

Figure 2a shows the parity plot for atomic force predictions (see Supplementary Information Section 7 for the loss curves of models involved in this work). For the testing dataset, the model predictions match DFT results well with a 0.2 meV/Å mean absolute error (MAE). To demonstrate the transferability of this model, we consider a bilayer MoS_2 structure with 8.95° twisting angle and 201 atoms in a supercell. We relax the structure with a quasi-Newton method in which the atomic forces are predicted by the machine learning model, and compare the optimized structure with the DFT result. As shown in Fig. 2b and c, the out-of-plane distortions induced by the Moiré potential given by the DFT calculation and machine-learning model prediction match very well. Such a rumpling induced by the Moiré potential was also observed experimentally[68,69]. These results align with our expectations since the DPMD model has been proven successful and robust in numerous cases for atomic force prediction[30,31,70–76].

Model performance for electronic band predictions

We use the WANDER to predict the Wannier Hamiltonian elements of the 8.95°-twisted bilayer MoS_2, consisting of 201 atoms. We then compare these predictions with the results obtained from DFT calculations and Wannier90, illustrated in Fig. 3a. The MAE is 1.5 meV, indicating an exceptionally high level of machine learning prediction accuracy. The electronic band structure, generated through Wannier interpolation using the predicted Wannier Hamiltonian, especially the bands near Fermi levels, closely aligns with the DFT result, as depicted in Fig. 3b.

To further validate our model, we extend its application to predict the band structure of a 3.42°-twisted bilayer MoS_2 structure composed of 1308 atoms. This prediction is then compared with the results obtained from DFT calculations accelerated by GPUs. As illustrated in Fig. 3c, our predicted results exhibit a strong agreement with the DFT calculations, providing additional evidence for the robustness of our model. In Fig. 3d, we present a comparative analysis of the efficiency between our model and DFT calculations [using the Generalized Gradient Approximation (GGA) functional]. Our model can generate electronic band structures with an accuracy comparable to DFT calculations but with an acceleration in computational speed of 10^3–10^4 times.

Discussion

The selection of different inputs representing different categories of Wannier Hamilton elements is based on the underlying physical mechanisms and improves the performance of the model. The intra-layer hopping integrals primarily depend on the vector between two atoms. Most of the tight-binding models, which can simulate electronic band structures with reasonable accuracies, express the hopping coefficients as functions of interatomic distances only[77–80]. However, our work shows that the local atomic environments play a secondary but nonnegligible role. We consider a testing model in which only the relative-position vector is used as the input for predicting intra-layer hopping integrals. The performance of this testing model (as shown in Fig. 4b) is noticeably worse than the WANDER (as shown in Fig. 4a), even though they share the same hyperparameters. These results demonstrate that information about local atomic environments serves as an essential input for a higher-accuracy model.

We consider another testing model in which the inputs for the two atoms in the DPMD model are stacked with their relative-position vector as the input for intra-layer hopping integral prediction. However, even with the same hyperparameters, the performance of this testing model is inferior to our WANDER based on RAER, as shown in Fig. 4c. The rationale behind this lies in the reduced-dimensional nature of RAERs, which lowers the number of training parameters, facilitating the development of a model with a substantial weight on the relative position vector.

It is worth mentioning that the fundamental concept of RAER closely aligns with that of an autoencoder[81,82]. In an autoencoder, an input (such as the local atomic environment) undergoes a transformation into a code, and the original input is reconstructed through a decoding function. In this study, the last hidden layer (referred to as RAER) in the DPMD model serves as the code containing essential information about the local atomic environment. The utilization of RAER in this work serves to simplify the model, thereby enhancing the likelihood of achieving a model that effectively captures the underlying physics and, consequently, achieves a higher accuracy.

In predicting inter-layer hopping integrals, we employ relative-position vectors as inputs. This choice is motivated by the absence of atoms in the inter-layer gap affecting the Hamiltonian, and the hopping integral mostly depends on the relative positions of the atomic pair[44]. Introducing local

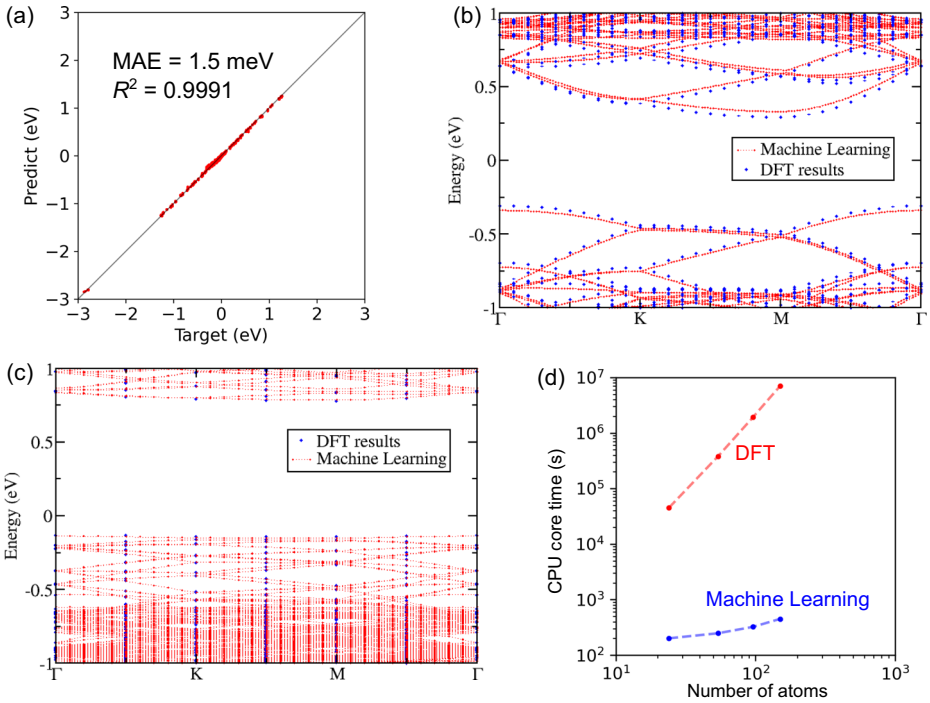

Fig. 3 | Performance of the WANDER for electronic band predictions. a The parity plot for Wannier Hamiltonian elements for the 8.95°-twisted bilayer MoS_2. Comparisons between the electronic band structures given by the WANDER and DFT calculations for **b** an 8.95°-twisted bilayer MoS_2 structure with 201 atoms and **c** a 3.42°-twisted bilayer MoS_2 structure with 1308 atoms. **d** Computational costs of DFT and WANDER for calculating the electronic bands of structures with different numbers of atoms.

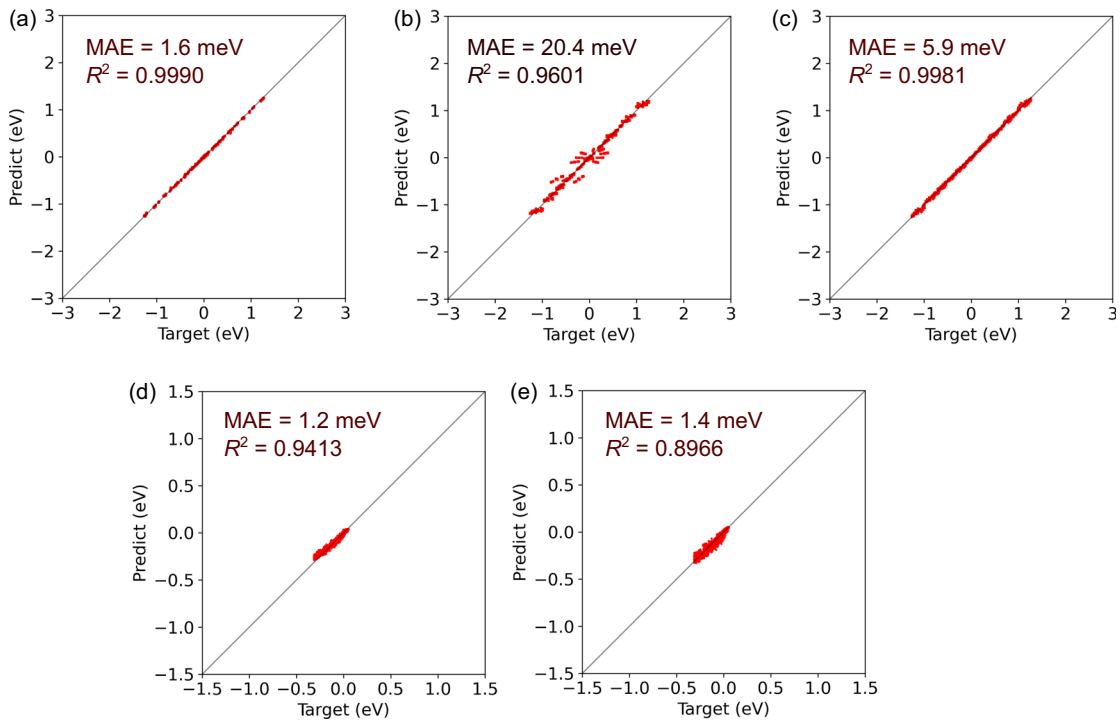

Fig. 4 | Comparisons of the performances of models with different inputs in predicting Hamiltonian elements. The parity plots for different models predicting intra-layer Hamiltonian elements of structures in the testing dataset. The inputs are **a** RAERs plus relative-position vectors (adopted in the model shown in Fig. 1), **b** relative-position vectors only, and **c** inputs in DPMD plus relative-position vectors. The parity plots for different models predicting inter-layer Hamiltonian elements of the 8.95°-twisted bilayer MoS_2 structure with 201 atoms. Here, we focus on transferability so that we consider a twisted structure that is much different from the structures in the training dataset. The inputs are **d** relative-position vectors only (adopted in the model shown in Fig. 1), and **e** RAERs plus relative-position vectors.

atomic environment information into the input could complicate the model and hamper its transferability; generally speaking, models with fewer parameters often exhibit better transferability in deep learning[83,84]. Twisted structures are not included in the training dataset due to their high computational costs. Consequently, the transferability of the model becomes crucial. Figure 4d and e illustrate a performance comparison on an 8.95°-twisted bilayer MoS_2 structure between our WANDER and a testing model using RAERs plus relative-position vectors as inputs. The WANDER exhibits significantly higher accuracy, as evidenced by an R-squared value of 0.9413 compared to 0.8966 observed in the testing model.

By carefully choosing input parameters for distinct categories of Wannier Hamiltonian elements, the WANDER is able to simulate atomic and electronic structures simultaneously with accuracy comparable to that of DFT but at a significantly lower computational cost. This dual-functional model shows great promise for applications in exploring interactions between structures and electronic behaviors, including phenomena like electron-phonon coupling and structural change-induced metal-insulator transitions, especially in large-scale systems.

As a departure from the DeepH, HamGNN, and their peers[24,24–27,27,34–38], which are pioneering works in machine-learning electronic structure simulations, the WANDER offers an alternative for predicting the electronic band structures of scientifically significant systems with low periodicities, such as twisted bilayer and multilayer 2D materials. Moreover, a key emphasis of this work is showcasing how existing force fields can be endowed with the capability for electronic structure simulations. We chose the DPMD model as an illustrative example of this successful integration. However, the fundamental principles of this method, which involve utilizing the local atomic environment for on-site interaction prediction, employing a reduced atomic environment representation along with a relative-position vector for intra-layer hopping prediction, and using a relative-position vector for inter-layer hopping prediction, are rooted in physical laws. These design principles are expected to be applicable to any machine-learning force field model. For example, the electronic band simulation model can be integrated into the state-of-the-art graph neural network (GNN) force fields, by incorporating the length and direction of the interatomic vector during the graph representation and embedding processes. The final embedding layer can serve as the RAER. The computational cost of the new dual-functional model consists of two components: the force field part, which inherits the cost of the adopted force field model, and the electronic structure part, which remains the same as our WANDER model.

Another significant aspect of the WANDER is its small number of training parameters (62,411 parameters; refer to Supplementary Information Section 5 for details). In WANDER, we utilize the RAER obtained from the force field model to construct input representations for predicting Wannier Hamiltonian elements. RAER encapsulates essential information about the local atomic environment while maintaining a lower dimensionality. Furthermore, WANDER is a physics-informed model that utilizes the relative position vector as input for predicting inter-layer terms. These characteristics contribute to a reduction in the number of training parameters, simplifying the model architecture, facilitating training, and improving transferability.

It is worth mentioning that in this work, we use twisted bilayer MoS_2 structures as illustrative examples. However, the WANDER method we developed is expected to be applicable to a wide range of materials, including molecules and three-dimensional (3D) bulk crystals, by constructing datasets with DFT calculations and retraining. For 3D structures, only the on-site and intra-layer hopping terms can be included, while the inter-layer hopping terms are omitted, since there is generally no weak Van der Waals interactions in 3D materials. This broad applicability is due to the fact that our model is based on DPMD and Wannier functions, whose transferability has been proven robust across different systems.

This model presents opportunities for enhancement across several dimensions. While it exhibits satisfactory transferability in the context of twisted bilayer structures of MoS_2, its performance on other unseen structures remains unpredictable, which is a common challenge in many machine-learning models. For example, achieving high model accuracy across a broader range of twist angles and strains requires augmenting the dataset. Addressing this issue can be approached through the following strategies. Firstly, we can enhance the model by simplifying it, which can be achieved by reducing the number of free parameters[83,84]. In this study, we designate the last hidden layer in the DPMD model as the RAER. A more in-depth exploration, such as employing the autoencoder method[81], can be conducted to determine the minimal dimensionality required for the RAER. Furthermore, expanding the dataset to include a broader range of structure categories can mitigate the challenge of encountering unpredictable structures. This approach necessitates the development of high-throughput computational methods for generating large amounts and standardized Wannier Hamiltonians with well-developed workflow[54], serving as the training dataset for creating a universal model[85]. Moreover, our work uses generalized gradient approximation (GGA) functionals, which have limitations in accurately describing strongly correlated d orbitals. To improve the accuracy of electronic structure simulations, we can employ advanced methodologies such as the DFT+U method, meta-GGA functionals, and GW, as well as consider the inclusion of spin-orbit coupling. Subsequent Wannierization based on these methods would yield even higher precision. Additionally, we notice that energy bands far from the Fermi level may suffer from accumulated errors. Although this generally has only a minor effect on the most important physical properties, which are typically governed by states near the Fermi level, addressing this issue will be part of our future work.

This work also serves as inspiration for additional investigations in various directions. For instance, this work adopts Wannier functions, widely utilized in solid-state physics, as the basis. Theoretically speaking, any physical quantities calculable based on Wannier functions, including but not limited to spin Hall conductivity[47,48], transport properties[50], electron-phonon coupling[51,52], and non-equilibrium processes shift current[49] can be integrated into the predictive scope of this model. Since Wannier functions are defined in real space, extending their use to predict electron densities could be a promising future direction. Moreover, this work adopts the input representation of DeepMD, which can be further extended to a GNN-based representation for an efficient description of the local atomic environment. This work aims to pave the path toward developing multifunctional machine-learning models for simulating a wide range of physical quantities.

Methods

Data preparation and computational details

To prepare the dataset, we carry out DFT calculations on 4×4 bilayer MoS_2 supercells using the QUANTUM–ESPRESSO package[86]. The functionals used are generalized gradient approximation[87] and a $2 \times 2 \times 1$ Monkhorst–Pack k-point mesh is used to sample the Brillouin zone[88]. The kinetic energy cutoff for wavefunctions is 50 Ry. The Van der Waals interaction is simulated with the DFT-D method[89]. The Wannier Hamiltonians are calculated with the Wannier90 package[59,60]. More computational details can be found in Supplementary Information Sections 3 and 4.

Structural relaxation with the DPMD model

When generating the input data for the DPMD model, R_C is selected as 5.2 Å. After acquiring the DPMD model, we optimize the twisted structure with a quasi-Newton method, in which

$$\boldsymbol{x}_{i,n+1} = \boldsymbol{x}_{i,n} + \alpha \boldsymbol{f}_{i,n}. \tag{3}$$

Here, $\boldsymbol{x}_{i,n}$ is the position of atom i at the nth step. $\boldsymbol{f}_{i,n}$ is the force on atom i at the nth step. α is set as 0.529 Å^2/Ry.

Details about the deep learning model

The model is constructed using the PyTorch Python library[90]. Before progressing into the hidden layers, batch normalization is applied to the input to accelerate deep network training by reducing internal covariate shifts. When training the models to predict atomic forces and intra-layer hopping integrals, the two hidden layers consist of 64 and 32 nodes, respectively. For predicting the Wannier Hamiltonian elements, different neural networks are trained for different kinds of atomic pairs (Mo-Mo, Mo-S, S-Mo, or S-S) and different inter-atomic distance ranges (long-range interaction for $R >$ 6 Å and short-range interaction for $R <$ 6 Å). There are 62411 training parameters in the model predicting the Wannier Hamiltonian (see Supplementary Information Section 5 for details). The Adam optimizer is used in this work.

Data availability

The datasets for training the models in this work are available at the Digital Repository Service of Northeastern University (http://hdl.handle.net/2047/D20630194).

Code availability

The code for this work is available on GitHub (http://github.com/yuboqiuab/multifunctional).

Received: 10 May 2024; Accepted: 12 May 2025;
Published online: 12 June 2025

References

1. Kohn, W., Becke, A. D. & Parr, R. G. Density functional theory of electronic structure. *J. Phys. Chem.* **100**, 12974–12980 (1996).
2. Neugebauer, J. & Hickel, T. Density functional theory in materials science. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 438–448 (2013).
3. Carleo, G. et al. Machine learning and the physical sciences. *Rev. Mod. Phys.* **91**, 045002 (2019).
4. Schmidt, J., Marques, M. R., Botti, S. & Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **5**, 83 (2019).
5. Behler, J. Four generations of high-dimensional neural network potentials. *Chem. Rev.* **121**, 10037–10072 (2021).
6. Unke, O. T. et al. Machine learning force fields. *Chem. Rev.* **121**, 10142–10186 (2021).
7. Zhang, L., Han, J., Wang, H., Car, R. & Weinan, E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
8. Artrith, N., Morawietz, T. & Behler, J. High-dimensional neural-network potentials for multicomponent systems: Applications to zinc oxide. *Phys. Rev. B* **83**, 153101 (2011).
9. Sosso, G. C., Miceli, G., Caravati, S., Behler, J. & Bernasconi, M. Neural network interatomic potential for the phase change material gete. *Phys. Rev. B* **85**, 174103 (2012).
10. Deringer, V. L., Caro, M. A. & Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Adv. Mater.* **31**, 1902765 (2019).
11. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
12. Shapeev, A. V. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).
13. Thompson, A. P., Swiler, L. P., Trott, C. R., Foiles, S. M. & Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).
14. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
15. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
16. Unke, O. T. & Meuwly, M. Physnet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
17. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. Schnet–a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
18. Choudhary, K. et al. Unified graph neural network force-field for the periodic table: solid state applications. *Digit. Discov.* **2**, 346–355 (2023).
19. Park, C. W. et al. Accurate and scalable graph neural network force field and molecular dynamics with direct force architecture. *Npj Comput. Mater.* **7**, 73 (2021).
20. Haghighatlari, M. et al. Newtonnet: A newtonian message passing network for deep learning of interatomic potentials and forces. *Digit. Discov.* **1**, 333–343 (2022).
21. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
22. Husic, B. E. et al. Coarse graining molecular dynamics with graph neural networks. *J. Chem. Phys.* **153**, 194101 (2020).
23. Xiao, H. et al. An invertible, invariant crystal representation for inverse design of solid-state materials using generative deep learning. *Nat. Commun.* **14**, 7027 (2023).
24. Li, H. et al. Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.* **2**, 367–377 (2022).
25. Li, H. et al. Deep-learning electronic-structure calculation of magnetic superstructures. *Nat. Comput. Sci.* **3**, 321–327 (2023).
26. Gong, X. et al. General framework for e(3)-equivariant neural network representation of density functional theory hamiltonian. *Nat. Commun.* **14**, 2848 (2023).
27. Zhong, Y., Yu, H., Su, M., Gong, X. & Xiang, H. Transferable equivariant graph neural networks for the hamiltonians of molecules and solids. *npj Comput. Mater.* **9**, 182 (2023).
28. Prodan, E. & Kohn, W. Nearsightedness of electronic matter. *Proc. Natl Acad. Sci.* **102**, 11635–11638 (2005).
29. Wang, H., Zhang, L., Han, J. & Weinan, E. Deepmd-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **228**, 178–184 (2018).
30. Wu, J., Yang, J., Ma, L., Zhang, L. & Liu, S. Modular development of deep potential for complex solid solutions. *Phys. Rev. B* **107**, 144102 (2023).
31. Wu, J., Zhang, Y., Zhang, L. & Liu, S. Deep learning of accurate force field of ferroelectric hfo_2. *Phys. Rev. B* **103**, 024108 (2021).
32. Xu, N., Shi, Y., He, Y. & Shao, Q. A deep-learning potential for crystalline and amorphous Li–Si alloys. *J. Phys. Chem. C.* **124**, 16278–16288 (2020).
33. Ouyang, X. et al. Quantum-accurate modeling of ferroelectric phase transition in perovskites from message-passing neural networks. *J. Phys. Chem. C.* **127**, 20890–20902 (2023).
34. Wang, Z. et al. Graph representation-based machine learning framework for predicting electronic band structures of quantum-confined nanostructures. *Sci. China Mater.* **65**, 3157–3170 (2022).
35. Knøsgaard, N. R. & Thygesen, K. S. Representing individual electronic states for machine learning GW band structures of 2d materials. *Nat. Commun.* **13**, 468 (2022).
36. Wang, Z. et al. Machine learning method for tight-binding hamiltonian parameterization from ab-initio band structure. *npj Comput. Mater.* **7**, 11 (2021).
37. Gong, W. et al. Graph transformer networks for accurate band structure prediction: An end-to-end approach. *arXiv preprint arXiv:2411.16483* (2024).
38. Fu, L., Wu, Y., Shang, H. & Yang, J. Transformer-based neural-network quantum state method for electronic band structures of real solids. *J. Chem. Theory Comput.* **20**, 6218–6226 (2024).
39. Carr, S., Fang, S. & Kaxiras, E. Electronic-structure methods for twisted moiré layers. *Nat. Rev. Mater.* **5**, 748–763 (2020).
40. Andrei, E. Y. & MacDonald, A. H. Graphene bilayers with a twist. *Nat. Mater.* **19**, 1265–1275 (2020).
41. Bistritzer, R. & MacDonald, A. H. Moiré bands in twisted double-layer graphene. *Proc. Natl Acad. Sci.* **108**, 12233–12237 (2011).
42. Kohn, W. Analytic properties of Bloch waves and Wannier functions. *Phys. Rev.* **115**, 809 (1959).
43. Souza, I., Marzari, N. & Vanderbilt, D. Maximally localized Wannier functions for entangled energy bands. *Phys. Rev. B* **65**, 035109 (2001).

44. Marzari, N., Mostofi, A. A., Yates, J. R., Souza, I. & Vanderbilt, D. Maximally localized wannier functions: Theory and applications. *Rev. Mod. Phys.* **84**, 1419 (2012).
45. Wang, H., Ma, S., Zhang, S. & Lei, D. Intrinsic superflat bands in general twisted bilayer systems. *Light Sci. Appl.* **11**, 159 (2022).
46. Karniadakis, G. E. et al. Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
47. Wang, X., Yates, J. R., Souza, I. & Vanderbilt, D. Ab initio calculation of the anomalous hall conductivity by wannier interpolation. *Phys. Rev. B* **74**, 195118 (2006).
48. Zhou, J., Qiao, J., Bournel, A. & Zhao, W. Intrinsic spin hall conductivity of the semimetals $MoTe_2$ and WTe_2. *Phys. Rev. B* **99**, 060408 (2019).
49. Ibañez-Azpiroz, J., Tsirkin, S. S. & Souza, I. Ab initio calculation of the shift photocurrent by Wannier interpolation. *Phys. Rev. B* **97**, 245143 (2018).
50. Pizzi, G., Volja, D., Kozinsky, B., Fornari, M. & Marzari, N. BoltzWann: A code for the evaluation of thermoelectric and electronic transport properties with a maximally-localized Wannier functions basis. *Comput. Phys. Commun.* **185**, 422–429 (2014).
51. Giustino, F., Cohen, M. L. & Louie, S. G. Electron-phonon interaction using Wannier functions. *Phys. Rev. B* **76**, 165108 (2007).
52. Poncé, S., Margine, E. R., Verdi, C. & Giustino, F. Epw: Electron–phonon coupling, transport and superconducting properties using maximally localized wannier functions. *Comput. Phys. Commun.* **209**, 116–133 (2016).
53. Koshino, M. et al. Maximally localized Wannier orbitals and the extended Hubbard model for twisted bilayer graphene. *Phys. Rev. X* **8**, 031087 (2018).
54. Garrity, K. F. & Choudhary, K. Database of wannier tight-binding hamiltonians using high-throughput density functional theory. *Sci. Data* **8**, 106 (2021).
55. Brouder, C., Panati, G., Calandra, M., Mourougane, C. & Marzari, N. Exponential localization of Wannier functions in insulators. *Phys. Rev. Lett.* **98**, 046402 (2007).
56. Zhang, W., Yu, R., Zhang, H.-J., Dai, X. & Fang, Z. First-principles studies of the three-dimensional strong topological insulators Bi_2Te_3, Bi_2Se_3 and Sb_2Te_3. *N. J. Phys.* **12**, 065013 (2010).
57. Shi, W. et al. A charge-density-wave topological semimetal. *Nat. Phys.* **17**, 381–387 (2021).
58. Ponraj, J. S. et al. Photonics and optoelectronics of two-dimensional materials beyond graphene. *Nanotechnol* **27**, 462001 (2016).
59. Mostofi, A. A. et al. An updated version of wannier90: A tool for obtaining maximally-localised Wannier functions. *Comput. Phys. Commun.* **185**, 2309–2310 (2014).
60. Pizzi, G. et al. Wannier90 as a community code: new features and applications. *J. Condens. Matter Phys.* **32**, 165902 (2020).
61. Lechermann, F. et al. Dynamical mean-field theory using Wannier functions: A flexible route to electronic structure calculations of strongly correlated materials. *Phys. Rev. B—Condens. Matter Mater. Phys.* **74**, 125120 (2006).
62. Sporkmann, B. & Bross, H. Calculation of Wannier functions for fcc transition metals by Fourier transformation of Bloch functions. *Phys. Rev. B* **49**, 10869 (1994).
63. Sakuma, R. Symmetry-adapted Wannier functions in the maximal localization procedure. *Phys. Rev. B* **87**, 235109 (2013).
64. Lei, B.-H., Pan, S., Yang, Z., Cao, C. & Singh, D. J. Second harmonic generation susceptibilities from symmetry adapted wannier functions. *Phys. Rev. Lett.* **125**, 187402 (2020).
65. Hinuma, Y., Grüneis, A., Kresse, G. & Oba, F. Band alignment of semiconductors from density-functional theory and many-body perturbation theory. *Phys. Rev. B* **90**, 155405 (2014).
66. Choe, D.-H., West, D. & Zhang, S. Band alignment and the built-in potential of solids. *Phys. Rev. Lett.* **121**, 196802 (2018).
67. Zhang, J., Xie, W., Zhao, J. & Zhang, S. Band alignment of two-dimensional lateral heterostructures. *2D Mater.* **4**, 015038 (2016).
68. Kim, D. S. et al. Electrostatic moiré potential from twisted hexagonal boron nitride layers. *Nat. Mater.* **23**, 65–70 (2024).
69. Tilak, N., Li, G., Taniguchi, T., Watanabe, K. & Andrei, E. Y. Moiré potential, lattice relaxation, and layer polarization in marginally twisted MoS_2 bilayers. *Nano Lett.* **23**, 73–81 (2022).
70. Yang, M., Raucci, U. & Parrinello, M. Reactant-induced dynamics of lithium imide surfaces during the ammonia decomposition process. *Nat. Catal.* **6**, 829–836 (2023).
71. Niu, H., Bonati, L., Piaggi, P. M. & Parrinello, M. Ab initio phase diagram and nucleation of gallium. *Nat. Commun.* **11**, 2654 (2020).
72. Rodriguez, A., Lam, S. & Hu, M. Thermodynamic and transport properties of lif and flibe molten salts with deep learning potentials. *ACS Appl. Mater. Interfaces* **13**, 55367–55379 (2021).
73. Piaggi, P. M., Weis, J., Panagiotopoulos, A. Z., Debenedetti, P. G. & Car, R. Homogeneous ice nucleation in an ab initio machine-learning model of water. *Proc. Natl Acad. Sci.* **119**, e2207294119 (2022).
74. Chen, T. et al. Modeling the high-pressure solid and liquid phases of tin from deep potentials with ab initio accuracy. *Phys. Rev. Mater.* **7**, 053603 (2023).
75. Liu, R. et al. Structural and dynamic properties of solvated hydroxide and hydronium ions in water from ab initio modeling. *J. Chem. Phys.* **157**, 024503 (2022).
76. Wu, J. et al. Accurate force field of two-dimensional ferroelectrics from deep learning. *Phys. Rev. B* **104**, 174107 (2021).
77. Liu, G.-B., Shan, W.-Y., Yao, Y., Yao, W. & Xiao, D. Three-band tight-binding model for monolayers of group-vib transition metal dichalcogenides. *Phys. Rev. B* **88**, 085433 (2013).
78. Lenosky, T. J. et al. Highly optimized tight-binding model of silicon. *Phys. Rev. B* **55**, 1528 (1997).
79. Tang, M. S., Wang, C. Z., Chan, C. T. & Ho, K. M. Environment-dependent tight-binding potential model. *Phys. Rev. B* **53**, 979 (1996).
80. Ribeiro, R. M. & Peres, N. M. Stability of boron nitride bilayers: Ground-state energies, interlayer distances, and tight-binding description. *Phys. Rev. B* **83**, 235312 (2011).
81. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *science* **313**, 504–507 (2006).
82. Xie, T., Fu, X., Ganea, O.-E., Barzilay, R. & Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197* (2021).
83. Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. Understanding deep learning (still) requires rethinking generalization. *Commun. ACM* **64**, 107–115 (2021).
84. Moradi, R., Berangi, R. & Minaei, B. A survey of regularization strategies for deep models. *Artif. Intell. Rev.* **53**, 3947–3986 (2020).
85. Vitale, V. et al. Automated high-throughput wannierisation. *Npj Comput. Mater.* **6**, 66 (2020).
86. Giannozzi, P. et al. Quantum espresso: A modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter* **21**, 395502–20 (2009).
87. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
88. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).
89. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **27**, 1787–1799 (2006).
90. Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703* (2019).

Acknowledgements

Y.Q. and Q.Y. were supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award No. DE-SC0023664. W.G. was

supported by the U.S. National Science Foundation under grant No. DMR-2323469. The research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award BES-ERCAP0029544. We thank Fei Xue for the valuable discussions.

Author contributions

Q.Y. conceived and coordinated the research project. Y.Q. and W.G. designed and implemented the deep learning codes with Pytorch. Q.Y., Y.Q., and W.G. prepared the manuscript. All authors participated in discussing and analyzing the data and editing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-025-01668-5.

Correspondence and requests for materials should be addressed to Qimin Yan.