

ARTICLE OPEN



Transferable equivariant graph neural networks for the Hamiltonians of molecules and solids

Yang Zhong^{1,2}, Hongyu Yu^{1,2}, Mao Su^{1,2}, Xingao Gong^{1,2} and Hongjun Xiang^{1,2}✉

This work presents an E(3) equivariant graph neural network called HamGNN, which can fit the electronic Hamiltonian matrix of molecules and solids by a complete data-driven method. Unlike invariant models that achieve equivariance approximately through data augmentation, HamGNN employs E(3) equivariant convolutions to construct the Hamiltonian matrix, ensuring strict adherence to all equivariant constraints inherent in the physical system. In contrast to previous models with limited transferability, HamGNN demonstrates exceptional accuracy on various datasets, including QM9 molecular datasets, carbon allotropes, silicon allotropes, SiO₂ isomers, and Bi_xSe_y compounds. The trained HamGNN models exhibit accurate predictions of electronic structures for large crystals beyond the training set, including the Moiré twisted bilayer MoS₂ and silicon supercells with dislocation defects, showcasing remarkable transferability and generalization capabilities. The HamGNN model, trained on small systems, can serve as an efficient alternative to density functional theory (DFT) for accurately computing the electronic structures of large systems.

npj Computational Materials (2023)9:182; <https://doi.org/10.1038/s41524-023-01130-4>

INTRODUCTION

Nowadays machine learning (ML) has a wide range of applications in molecular and materials science, including the direct prediction of various properties of materials^{1–3}, the construction of machine learning interatomic potentials (MLIPs) with quantum mechanical precision^{4–7}, the high-throughput generation of molecular and crystal structures^{8–10}, and the construction of more precise exchange-correlation functionals^{11,12}. However, the determination of materials' electronic structure still heavily relies on calculations based on density functional theory (DFT). Unfortunately, these methods require a significant amount of time to get the electronic Hamiltonian matrices through self-consistent iterations and exhibit poor scalability with respect to system sizes. Semi-empirical tight-binding (TB) approximations¹³, such as the Slater-Koster method¹⁴, can reduce a lot of computation compared to the DFT methods. However, this approach often directly uses the existing or manually fine-tuned TB parameters and thus cannot accurately reproduce the electronic structure of general systems. Developing truly transferable, fully data-driven TB models that can be applied across various materials, geometries, and boundary conditions can reconcile accuracy with speed but is rather challenging.

Hegde and Brown first used kernel ridge regression (KRR) to learn semi-empirical tight-binding Hamiltonian matrices¹⁵. They successfully fitted the Hamiltonian matrix of the Cu system containing only rotation invariant *s* orbitals and the diamond system consisting only of *s* and *p* orbitals. Similarly, Wang et al. designed a neural network model to obtain semi-empirical TB parameters by fitting the band structures calculated by DFT¹⁶. An important feature of the Hamiltonian matrix is that its components transform equivariantly with the rotation of the coordinate system. However, none of these approaches deals with the rotational equivariance of the Hamiltonian matrix. Moreover, the two methods are limited to fitting the empirical model Hamiltonian matrices rather than the true ab initio tight-binding Hamiltonian matrices generated by the self-consistent iteration of ab initio tight-binding methods such as OpenMX^{17,18} and Siesta^{19,20}.

Because of the spherical harmonic part of the atomic basis functions, the TB Hamiltonian matrix must satisfy two fundamental constraints: rotational equivariance and parity symmetry. Due to the presence of periodic boundary conditions, the tight-binding Hamiltonian matrix for solids also exhibits translational invariance. The fundamental symmetry constraints satisfied by the tight-binding Hamiltonian matrix belong to the E(3) group. The notation E(3) denotes the Euclidean group in three-dimensional (3D) space, i.e., the group of rotations, translations, and inversion in 3D space. If a mapping $L(\mathbf{q})$ is E(3) equivariant, then the following equivariant condition is satisfied for any rotation or inversion operation $\hat{g} \in O(3)$ and translation operation: $L(\hat{g} \circ \mathbf{q} + \mathbf{t}) = \hat{g} \circ L(\mathbf{q})$, where \mathbf{t} is the translation vector. In other words, the mapping $L(\mathbf{q})$ remains invariant under translation while being equivariant under rotation and inversion. The tight-binding Hamiltonian matrix $H(\{\mathbf{r}_i\})$ can be viewed as a function of the atomic positions $\{\mathbf{r}_i\}$ of the system. Here we take the rotational equivariance of the Hamiltonian matrix as an example. When a rotation $\hat{Q} \in SO(3)$ is applied to the system, the equivariance between the rotated and original Hamiltonian matrices can be expressed as $H(\{\mathbf{r}'_i\}) = D(\hat{Q})H(\{\mathbf{r}_i\})D(\hat{Q})^\dagger$, where $\{\mathbf{r}'_i\}$ is the set of the rotated atomic positions, and D is the Wigner D matrix^{21,22}. It is crucial for the mapping of atomic coordinates to the Hamiltonian matrix to satisfy equivariance. If equivariance is not achieved in the mapping, the Hamiltonian matrix may undergo an unphysical transformation, resulting in $H(\{\mathbf{r}'_i\}) \neq D(\hat{Q})H(\{\mathbf{r}_i\})D(\hat{Q})^\dagger$ when the system undergoes any rotation. Non-equivariant models necessitate a substantial amount of training data to approximately achieve equivariance by learning how to predict the Hamiltonian matrix in all possible directions of the system. Non-equivariant models often struggle with transferability, especially when encountering structures or configurations that differ significantly from the training set. As a consequence, they may lack generalization capabilities when applied to new or complex materials. In contrast, equivariant models possess a more robust ability for complex structures due to their ability to capture and preserve symmetries.

¹Key Laboratory of Computational Physical Sciences (Ministry of Education), Institute of Computational Physical Sciences, State Key Laboratory of Surface Physics, and Department of Physics, Fudan University, Shanghai 200433, China. ²Shanghai Qi Zhi Institute, Shanghai 200030, China. ✉email: hxiang@fudan.edu.cn

Zhang et al.²³ and Nigam et al.²⁴ proposed a method to predict the ab-initio TB Hamiltonians of small molecules and simple solid systems by constructing an equivariant kernel in Gaussian Process Regression (GPR) to parameterize the Hamiltonian matrices. Due to the fixed kernel and representation used in GPR^{25–28}, its training accuracy and multi-element generalization ability are typically inferior to those of deep neural networks like message-passing neural networks (MPNNs)^{29–34} when a sufficient number of training samples are available. Hence, the development of graph neural networks (GNNs) with the capability to predict the Hamiltonian matrix of both general periodic and aperiodic systems emerges as the most favorable approach.

However, traditional GNNs are limited to predicting rotation-invariant scalars such as energy, band gap, and the like. To predict equivariant directional properties, GNNs need to encode the directional information of the system suitably. To make the predicted Hamiltonian matrices satisfy the rotational equivariance, Schütt et al. designed the SchNorb neural network architecture by embedding the direction information of the bonds into the message-passing function³⁵. This network constructs the TB Hamiltonian of molecules from the directional edge features of atom pairs. However, SchNorb needs to learn the rotational equivariance of the Hamiltonian matrix through data augmentation, which greatly increases the amounts of training data and redundant parameters of the network. Unke et al. proposed the PhisNet model³⁶, which achieved SE(3) equivariant parameterization of TB Hamiltonian matrix using GNN based on SO(3) representations. Their model demonstrated high accuracy on the Hamiltonian matrices of small molecules such as water and ethanol. However, it should be noted that the PhisNet model cannot be considered a universal representation of Hamiltonian matrices due to its neglect of parity symmetry, which can lead to significant issues in predicting periodic systems.

Recently, Li et al. proposed a GNN model named DeepH to predict the TB Hamiltonian matrix by constructing local coordinate systems in a crystal³⁷. DeepH successfully predicted the TB Hamiltonian matrices of some simple periodic systems such as graphene and carbon nanotubes. Their original intention of introducing the local coordinate system is to solve the rotation equivariance problem of the Hamiltonian, but DeepH still embeds the local directional information of interacting atom pairs in the invariant message passing function, which will undoubtedly increase the number of redundant parameters of the network but may require less data augmentation than a fully invariant model without local coordinate systems. In addition, the hopping distance between two interacting atoms far exceeds the typical lengths of chemical bonds. Taking the smallest hydrogen atom as an example, the cutoff radius of the numerical atomic orbital of the hydrogen atom used by OpenMX is 6 Bohr, so the furthest hopping between any two atomic bases used by OpenMX in periodic systems can exceed at least 12 Bohr (~6.4 Å), a distance that even exceeds the lattice parameters of certain crystals. Therefore, it is difficult to describe such long hopping in a well-defined local coordinate system.

In this work, we developed a general framework to parameterize the Hamiltonian matrix by decomposing each block of the Hamiltonian matrix into a vector coupling of equivariant irreducible spherical tensors³⁸ (ISTs) with appropriate parity symmetry. The parametrized Hamiltonian matrix strictly adheres to rotational equivariance and parity symmetry, and can be further extended to a parameterized Hamiltonian that satisfies SU(2) and time-reversal equivariance to accurately fit the Hamiltonian matrix with spin-orbit coupling (SOC) effects. Based on this universal parametrized Hamiltonian, we designed the E(3) equivariant HamGNN model for predicting the TB Hamiltonian matrices of molecules and solids. HamGNN has demonstrated its universality for the Hamiltonian matrix, showcasing exceptional accuracy across various datasets, including QM9 molecular datasets, carbon allotropes, silicon allotropes, SiO₂ isomers, and Bi_xSe_y compounds.

The trained HamGNN model can predict the Hamiltonian matrices, energy bands, and wavefunctions of the structures not present in the training set. The powerful fitting and generalization ability exhibited by HamGNN enables researchers to efficiently compute the electronic structures of large-scale crystal systems that were previously deemed challenging or inaccessible.

RESULTS

E(3) equivariant parametrized Hamiltonian

The core of the electronic structure problem in DFT is to solve the Kohn-Sham equation for electrons. If the Kohn-Sham Hamiltonian is represented by numerical atomic orbitals centered on each atom, such as those defined in packages like OpenMX and Siesta, then the Kohn-Sham equation can be formulated as a generalized eigenvalue problem:

$$H^{(\mathbf{k})} \psi_{n\mathbf{k}} = \epsilon_{n\mathbf{k}} S^{(\mathbf{k})} \psi_{n\mathbf{k}} \quad (1)$$

where $H_{n_i l_i m_i, n_j l_j m_j}^{(\mathbf{k})} = \sum_{n_c} e^{i\mathbf{k} \cdot \mathbf{R}_{n_c}} H_{n_i l_i m_i, n_j l_j m_j}^{(\mathbf{R}_{n_c})}$ and $S_{n_i l_i m_i, n_j l_j m_j}^{(\mathbf{k})} = \sum_{n_c} e^{i\mathbf{k} \cdot \mathbf{R}_{n_c}} S_{n_i l_i m_i, n_j l_j m_j}^{(\mathbf{R}_{n_c})}$ are the Kohn-Sham Hamiltonian and overlap matrices at the point \mathbf{k} in the reciprocal space. $H_{n_i l_i m_i, n_j l_j m_j}^{(\mathbf{k})}$ and $S_{n_i l_i m_i, n_j l_j m_j}^{(\mathbf{k})}$ are obtained by Fourier transform of the real-space TB Hamiltonian matrix $H_{n_i l_i m_i, n_j l_j m_j}^{(\mathbf{R}_{n_c})} = \langle \phi_{n_i l_i m_i}(\mathbf{r} - \boldsymbol{\tau}_i) | \hat{H} | \phi_{n_j l_j m_j}(\mathbf{r} - \boldsymbol{\tau}_j - \mathbf{R}_{n_c}) \rangle$ and overlap matrix $S_{n_i l_i m_i, n_j l_j m_j}^{(\mathbf{R}_{n_c})} = \langle \phi_{n_i l_i m_i}(\mathbf{r} - \boldsymbol{\tau}_i) | \phi_{n_j l_j m_j}(\mathbf{r} - \boldsymbol{\tau}_j - \mathbf{R}_{n_c}) \rangle$, respectively, in the basis of atomic orbitals $\phi_{n_i l_i m_i}$ at the site $\boldsymbol{\tau}_i$ and $\phi_{n_j l_j m_j}$ at the site $\boldsymbol{\tau}_j + \mathbf{R}_{n_c}$, where \mathbf{R}_{n_c} is the shift vector of the n_c -th periodic image cell. Therefore, once we have obtained the Hamiltonian matrix and overlap matrix in real space, we can further solve the electronic structure in the whole reciprocal space.

Due to the spherical symmetry of the atomic potential, the atomic orbital bases as its eigenfunctions not only satisfy the rotational equivariance under the operation $Q \in SO(3)$ but also has a certain parity symmetry under the inversion operation $g \in \{E, I\}$. Under a rotary-inversion operation $gQ \in O(3)$, the TB Hamiltonian matrix element $H_{n_i l_i m_i, n_j l_j m_j}$ in real space becomes (we omit the notation \mathbf{R}_{n_c} for convenience in the following discussion):

$$H'_{n_i l_i m_i, n_j l_j m_j} = \langle gQ \phi_{n_i l_i m_i} | \hat{H} | gQ \phi_{n_j l_j m_j} \rangle \quad (2)$$

The irreducible representation of gQ is $\sigma_p(g) \otimes D(Q)$, where $D(Q)$ is the Wigner D matrix and $\sigma_p(g)$ is the scalar irreducible representation of the inversion operation, which is defined as follows

$$\sigma_p(g) = \begin{cases} 1, & g = E \\ p, & g = I \end{cases} \quad (3)$$

Substitute the irreducible representation of gQ into Eq. (2), we can get

$$H'_{n_i l_i m_i, n_j l_j m_j} = \sigma_{p_i p_j}(g) \sum_{\mu_i \mu_j} D_{m_i \mu_i}^{(l_i)}(Q) D_{m_j \mu_j}^{(l_j)}(Q) H_{n_i l_i \mu_i, n_j l_j \mu_j} \quad (4)$$

where $\sigma_{p_i p_j}(g) = \sigma_{p_i}(g) \sigma_{p_j}(g)$. We further write the right-hand side of Eq. (4) in the form of matrix-vector multiplication:

$$H'_{n_i l_i m_i, n_j l_j m_j} = \sigma_{p_i p_j}(g) \sum_{\mu_i \mu_j} [D^{(l_i)}(Q) \otimes D^{(l_j)}(Q)]_{m_i \mu_i, m_j \mu_j} H_{n_i l_i \mu_i, n_j l_j \mu_j} \quad (5)$$

It can be seen from the above equation that each sub-block $H_{n_i l_i \mu_i, n_j l_j \mu_j}$ ($|\mu_i| \leq l_i$, $|\mu_j| \leq l_j$) of the TB Hamiltonian matrix based on atomic orbitals can be regarded as a spherical tensor^{22,38} $\mathbf{T}_{\boldsymbol{\mu}, p_i p_j}^{n_i l_i, n_j l_j} \equiv H_{n_i l_i \mu_i, n_j l_j \mu_j}$ with the parity $p_i p_j$, which is rotationally equivariant according to the generalized Wigner D matrix $\mathbf{D}_{\boldsymbol{\mu} \boldsymbol{m}}^{\boldsymbol{l}}(Q) = D_{\mu_i m_i}^{l_i}(Q) D_{\mu_j m_j}^{l_j}(Q)$, where $\boldsymbol{l} \equiv (l_i, l_j)$, $\boldsymbol{\mu} \equiv (\mu_i, \mu_j)$, $\boldsymbol{m} \equiv (m_i, m_j)$.

According to the angular momentum theory^{21,22}, $D^l(Q) \otimes D^j(Q)$ is a reducible representation and can be further decomposed into the direct sum of several irreducible Wigner D matrices:

$$D^l \otimes D^j = D^{|l-j|} \oplus D^{|l-j|+1} \oplus \dots \oplus D^{l+j} \quad (6)$$

Combining the parity of the Hamiltonian matrix block (n_l, l_l, n_j, l_j) , we can get

$$\sigma_{p_l, p_j}(g) D^l(Q) \otimes D^j(Q) = \sum_{L=|l_l-l_j|}^{l_l+l_j} \oplus \sigma_{p_l, p_j}(g) D^L(Q) \quad (7)$$

According to Eq. (7), $\mathbf{T}_{\mu, p_l, p_j}^{n_l, l_l, n_j, l_j}$ is reducible and the coupled irreducible spherical tensor $T_{L, p_l, p_j, m}^{n_l, l_l, n_j, l_j}$ in each order $L = |l_l - l_j|, \dots, l_l + l_j$ can be obtained by the vector coupling of $\mathbf{T}_{\mu, p_l, p_j}^{n_l, l_l, n_j, l_j}$:

$$T_{L, p_l, p_j, m}^{n_l, l_l, n_j, l_j} = \sum_{\mu_l = -l_l}^{l_l} \sum_{\mu_j = -l_j}^{l_j} C_{m \mu_l \mu_j}^{L l_l l_j} \mathbf{T}_{\mu, p_l, p_j}^{n_l, l_l, n_j, l_j} \quad (8)$$

where $C_{m \mu_l \mu_j}^{L l_l l_j}$ is the vector coupling coefficient, namely the Clebsch-Gordan coefficient. Each IST $T_{L, p_l, p_j, m}^{n_l, l_l, n_j, l_j}$ has the parity symmetry of p_l, p_j and satisfies the rotational equivariance of order L . By inverse linear transformation of Eq. (8), $\mathbf{T}_{\mu, p_l, p_j}^{n_l, l_l, n_j, l_j}$ can be constructed from ISTs $T_{L, p_l, p_j, m}^{n_l, l_l, n_j, l_j}$:

$$\mathbf{T}_{\mu, p_l, p_j}^{n_l, l_l, n_j, l_j} = \sum_{L=|l_l-l_j|}^{l_l+l_j} \sum_{m=-L}^L C_{m \mu_l \mu_j}^{L l_l l_j} T_{L, p_l, p_j, m}^{n_l, l_l, n_j, l_j} \quad (9)$$

Therefore, as long as we find all ISTs corresponding to each block of the Hamiltonian matrix, we can construct the entire Hamiltonian matrix in a block-wise manner through Eq. (9). We construct two O(3) equivariant vectors $\mathbf{\Omega}_i^{on}$ and $\mathbf{\Omega}_{ij}^{off}$ by the direct summation of all the ISTs required by the on-site ($i=j$) Hamiltonian and the off-site ($i \neq j$) Hamiltonian respectively:

$$\mathbf{\Omega}_i^{on} = \sum_{n_l} \sum_{n'_l} \sum_{L=|l_l-l'_l|}^{l_l+l'_l} \oplus [T_{L, p_l, p'_l, m}^{n_l, l_l, n'_l, l'_l}]_{-L \leq m \leq L} \quad (10)$$

$$\mathbf{\Omega}_{ij}^{off} = \sum_{n_l} \sum_{n_j} \sum_{L=|l_l-l_j|}^{l_l+l_j} \oplus [T_{L, p_l, p_j, m}^{n_l, l_l, n_j, l_j}]_{-L \leq m \leq L} \quad (11)$$

The prediction of the Hamiltonian matrix is transformed into the prediction of $\mathbf{\Omega}_i^{on}$ and $\mathbf{\Omega}_{ij}^{off}$, which can be obtained by mapping from the equivariant features of the nodes and the pair interactions, respectively. The final parameterized Hamiltonian can be expressed as:

$$\tilde{H}_{n_l, l_l, m_l, n_j, l_j, m_j} = \begin{cases} \tilde{H}_{n_l, l_l, m_l, n'_l, l'_l, m'_l}^{on} = \sum_{L=|l_l-l'_l|}^{l_l+l'_l} \sum_{m=-L}^L C_{m, m'_l, m}^{L l_l l'_l} T_{L, p_l, p'_l, m}^{n_l, l_l, n'_l, l'_l} & i=j \\ \tilde{H}_{n_l, l_l, m_l, n_j, l_j, m_j}^{off} = \sum_{L=|l_l-l_j|}^{l_l+l_j} \sum_{m=-L}^L C_{m, m_j, m}^{L l_l l_j} T_{L, p_l, p_j, m}^{n_l, l_l, n_j, l_j} & i \neq j \end{cases} \quad (12)$$

The above formula is O(3) equivariant. Since GNN naturally has translational symmetry, the parameterized Hamiltonian represented by $\mathbf{\Omega}_i^{on}$ and $\mathbf{\Omega}_{ij}^{off}$ obtained from GNN has E(3) equivariance.

When the spin-orbit coupling (SOC) effects are considered, the real-space Hamiltonian matrices are complex-valued and can be divided into four sub-blocks $\hat{H}_{s_l, s_j}(s_l, s_j = \uparrow \text{ or } \downarrow)$ by the spin degree. In this case, the complete Hamiltonian matrices satisfy the

following SU(2) rotational equivariance:

$$\langle gQ(\phi_{n_l, l_l, m_l, s_l}) | \hat{H} | gQ(\phi_{n_j, l_j, m_j, s_j}) \rangle = \sigma_{p_l, p_j}(g) \sum_{\mu_l = -l_l}^{l_l} \sum_{\mu_j = -l_j}^{l_j} \sum_{s_l = -1/2}^{1/2} \sum_{s_j = -1/2}^{1/2} (D_{m_l \mu_l}^{(l_l)}(Q) D_{m_j \mu_j}^{(l_j)}(Q) D_{s_l s_j}^{(1/2)*}(Q) D_{s_l s_j}^{(1/2)}(Q) \langle \phi_{n_l, l_l, \mu_l, s_l} | \hat{H} | \phi_{n_j, l_j, \mu_j, s_j} \rangle) \quad (13)$$

Although each subblock \hat{H}_{s_l, s_j} satisfies the O(3) rotational equivariance, they are coupled to each other under the rotational operations. Therefore, the four subblocks predicted independently with the O(3) equivariant parameterized Hamiltonian cannot be used to construct the complete SU(2) equivariant Hamiltonian with the SOC effect. In addition, the real and imaginary parts of the SU(2) equivariant Hamiltonian matrices are also coupled during rotation, so the complete Hamiltonian cannot be constructed by using the independently predicted real and imaginary parts. These methods not only rely on a large number of fitting parameters but also can not make the constructed Hamiltonian matrices strictly meet the SU(2) equivariance. To ensure that the SOC effect learned by the network complies with the physical rules and SU(2) equivariance, we explicitly express the complete Hamiltonian as the sum of the spin-less part and the SOC part:

$$\hat{H} = \hat{H} \otimes I_2 + \hat{H}_{SOC} \quad (14)$$

where $\hat{H}_{SOC} = \frac{1}{2} \xi(r) \hat{L} \cdot \hat{\sigma} = \frac{1}{2} \xi(r) (\hat{L}_x \hat{\sigma}_x + \hat{L}_y \hat{\sigma}_y + \hat{L}_z \hat{\sigma}_z)$, which satisfies SU(2) rotational equivariance and time-reversal equivariance. $\xi(r)$ is an invariant coefficient describing the strength of the SOC effects³⁹. According to the above equation, we can get the following parameterized Hamiltonian matrices with SOC effect in the atomic orbitals^{39,40}:

$$\tilde{H}_{n_l, l_l, m_l, s_l, n_j, l_j, m_j, s_j}^{SOC} = \begin{cases} \tilde{H}_{n_l, l_l, m_l, n_j, l_j, m_j} + \frac{1}{2} \xi_{n_l, l_l, n_j, l_j} \langle \phi_{n_l, l_l, m_l} | \hat{L}_z | \phi_{n_j, l_j, m_j} \rangle s_l = \uparrow, s_j = \uparrow \\ \frac{1}{2} \xi_{n_l, l_l, n_j, l_j} (\langle \phi_{n_l, l_l, m_l} | \hat{L}_x | \phi_{n_j, l_j, m_j} \rangle - i \langle \phi_{n_l, l_l, m_l} | \hat{L}_y | \phi_{n_j, l_j, m_j} \rangle) s_l = \uparrow, s_j = \downarrow \\ \frac{1}{2} \xi_{n_l, l_l, n_j, l_j} (\langle \phi_{n_l, l_l, m_l} | \hat{L}_x | \phi_{n_j, l_j, m_j} \rangle + i \langle \phi_{n_l, l_l, m_l} | \hat{L}_y | \phi_{n_j, l_j, m_j} \rangle) s_l = \downarrow, s_j = \uparrow \\ \tilde{H}_{n_l, l_l, m_l, n_j, l_j, m_j} - \frac{1}{2} \xi_{n_l, l_l, n_j, l_j} \langle \phi_{n_l, l_l, m_l} | \hat{L}_z | \phi_{n_j, l_j, m_j} \rangle s_l = \downarrow, s_j = \downarrow \end{cases} \quad (15)$$

Since the matrix representation of the angular momentum operator $\langle \phi_{n_l, l_l, m_l} | \hat{L}_a | \phi_{n_j, l_j, m_j} \rangle = \langle \phi_{n_l, l_l, m_l}(\mathbf{r}_l) | \hat{L}_a | \phi_{n_j, l_j, m_j}(\mathbf{r}_j - \mathbf{r}_{jl}) \rangle$ under the atomic orbital basis can be directly calculated analytically, the only learnable parameters in Eq. (15) are $\tilde{H}_{n_l, l_l, m_l, n_j, l_j, m_j}$ and ξ_{n_l, l_l, n_j, l_j} . $\tilde{H}_{n_l, l_l, m_l, n_j, l_j, m_j}$ can be directly expressed by Eq. (12), and ξ_{n_l, l_l, n_j, l_j} is an invariant scalar coefficient that can be mapped from the features of atom pairs ij .

Network Implementation

As can be seen from the above discussion, each Hamiltonian matrix block satisfies the rotational equivariance and has a definite parity under the inversion operation, so we designed an E(3) equivariant HamGNN model based on MPNN to fit TB Hamiltonian matrix. This framework directly captures the electronic structure without expensive self-consistent iterations by constructing local equivariant representations of each atomic orbit. The network architecture of HamGNN is shown in Fig. 1a. HamGNN can achieve a direct mapping from atomic species $\{Z\}$ and positions $\{\mathbf{r}\}$ to TB Hamiltonian matrix.

Based on the equivariant parameterized Hamiltonian matrix introduced in this work, the prediction of the Hamiltonian matrix is converted to predicting two reducible equivariant tensors: $\mathbf{\Omega}_i^{on}$ for on-site Hamiltonians and $\mathbf{\Omega}_{ij}^{off}$ for off-site Hamiltonians. These tensors satisfy the rotation and inversion equivariance of the O(3) group, therefore we need to construct them in the representation of the O(3) group. HamGNN utilizes equivariant atomic features that are constructed by the direct sum of O(3) representations with different rotation orders l . The O(3) representations of rotation order l can characterize atomic orbitals with angular quantum number l , as they possess the same rotational equivariance. The atomic features are

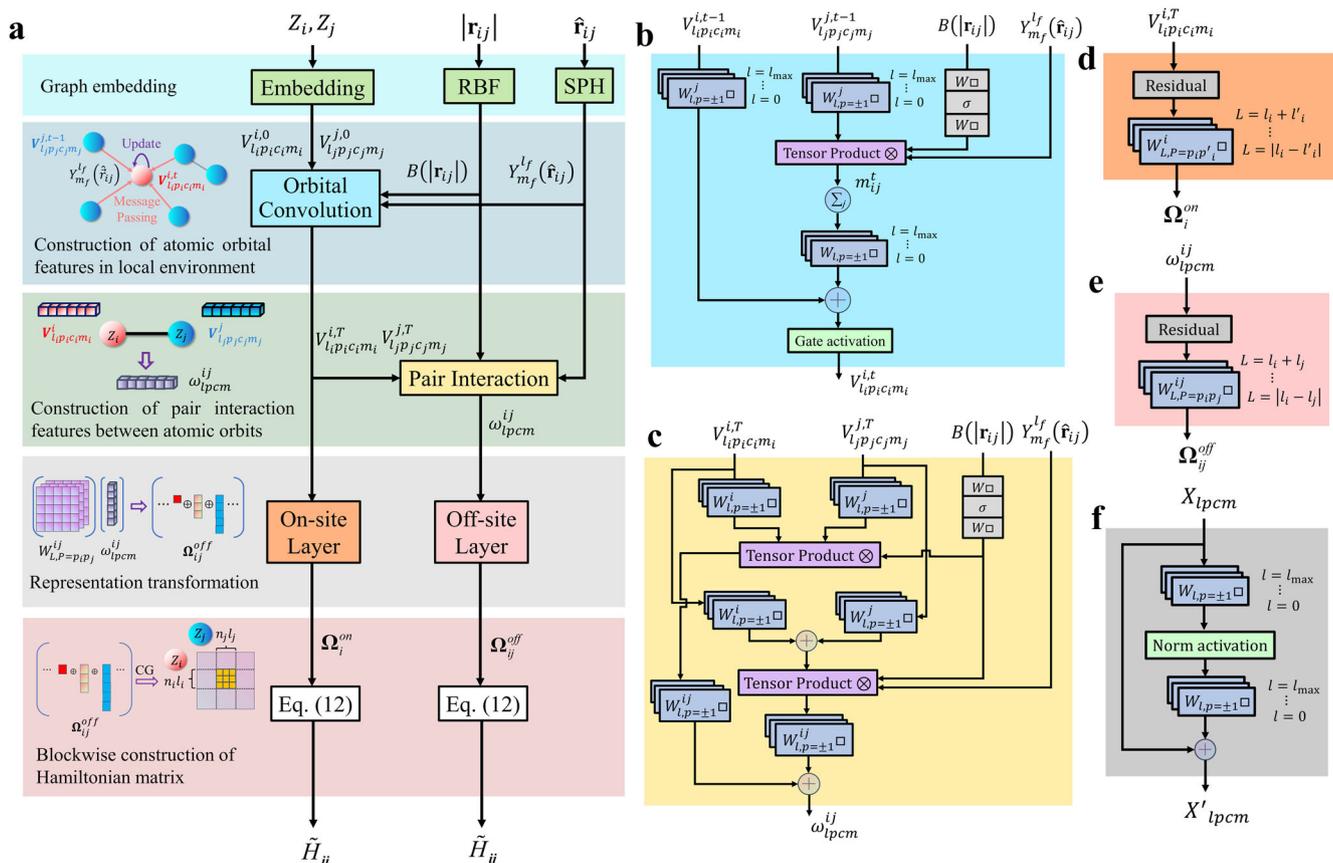


Fig. 1 HamGNN architecture and the illustration of its subnetworks. **a** The overall architecture of HamGNN. This neural network architecture predicts the Hamiltonian matrix through five steps. The prediction starts from the initial graph embedding of the species, interatomic distances, and interatomic directions of molecules and crystals. The atomic orbital features with angular momentum l in the local environment are included in the l -order components of the E(3) equivariant atom features and are refined through T orbital convolution blocks. In the third step, pair interaction features $\omega_{l_i p_i c_i m_i}^{ij}$ of atomic orbitals are constructed by pair interaction blocks. In the fourth step, the IST representations of on-site and off-site Hamiltonian matrices are constructed by passing the features of atomic orbitals and pair interactions through the on-site layer and off-site layer, respectively. The final step is to construct the on-site and off-site Hamiltonian matrices block-by-block via the parameterized Hamiltonian given by Eq. (12). **b** Orbital convolution block. The equivariant atomic features that include the features of atomic orbitals of each angular momentum l are refined by the equivariant message passing and update functions. **c** Pair interaction block. This block is used to construct the pair interaction features between the orbitals of two adjacent atoms by equivariant tensor product. **d** On-site layer. The equivariant features of atoms are transformed into ISTs of on-site blocks Ω_i^{on} by the on-site layer. **e** Off-site layer. The pair interaction features between atomic orbitals are transformed into ISTs of the off-site block Ω_{ij}^{off} by the off-site layer. **f** Residual block. The residual block is used in the on-site layer and off-site layer to perform a nonlinear equivariant transformation of input features.

refined through orbital convolution layers (shown in Fig. 2b), which update the atomic features by aggregating equivariant messages constructed from local chemical environments. After refined by T orbital convolution layers, the atomic features are used to construct equivariant pair interaction features via the pair interaction layer (shown in Fig. 2c). Ω_i^{on} and Ω_{ij}^{off} are obtained by representation transformation of atomic features and pair interaction features, respectively. The Hamiltonian matrix is finally constructed using Ω_i^{on} and Ω_{ij}^{off} via Eq. (12).

HamGNN first encodes elements, inter-atomic distances, and relative directions as initial graph embeddings. The atomic numbers Z_i are encoded as one-hot vectors and are subsequently transformed into initial atomic features through a multilayer perceptron (MLP) layer. The distance between atom i and its neighboring atom j within the cutoff radius r_c is expanded using the Bessel basis function⁴¹:

$$B(|\mathbf{r}_{ij}|) = \sqrt{\frac{2}{r_c}} \frac{\sin(n\pi|\mathbf{r}_{ij}|/r_c)}{|\mathbf{r}_{ij}|} f_c(|\mathbf{r}_{ij}|) \quad (16)$$

where f_c is the cosine cutoff function⁴², which guarantees physical continuity for the neighbor atoms close to the cutoff sphere. The list

of atomic neighbors is determined by the cutoff radius of each atom's orbital basis. The interatomic distance is expanded by a set of Bessel functions with $n = [1, 2, \dots, N_b]$, where N_b is the number of Bessel basis functions. In this work, N_b is set to 8. The directional information between atom i and atom j is embedded in a set of real spherical harmonics $\{Y_{m_l}^l(\hat{\mathbf{r}}_{ij})\}$, which will be used to construct the rotation-equivariant filter⁴³ in the equivariant convolution functions.

The atomic feature tensor $V = V_{l_0} \oplus V_{l_1} \oplus \dots \oplus V_{l_{max}}$ in HamGNN is represented as a direct sum of different irreducible representations of the O(3) group up to a maximum rotation order l_{max} . The features corresponding to each order can characterize the atomic orbitals with distinct angular quantum numbers. If the input structure is rotated, the atomic features will be transformed by the direct sum of the Wigner D matrix, which is represented as a block diagonal matrix $D = D_{l_0} \oplus D_{l_1} \oplus \dots \oplus D_{l_{max}}$. $V_{l_i p_i c_i m_i}^{i, t}$ denotes the tensor element of the equivariant features of atom i in the orbital convolution layer t , where $l_i \leq l_{max}$ is the rotation order of the O(3) irreducible representation, $p_i \in \{1, -1\}$ denotes the parity of the equivariant components of the order l_i , $-l_i \leq m_i \leq l_i$ is the index of each projection of the equivariant representation, c_i is the channel index (the dimension of the features). We use T

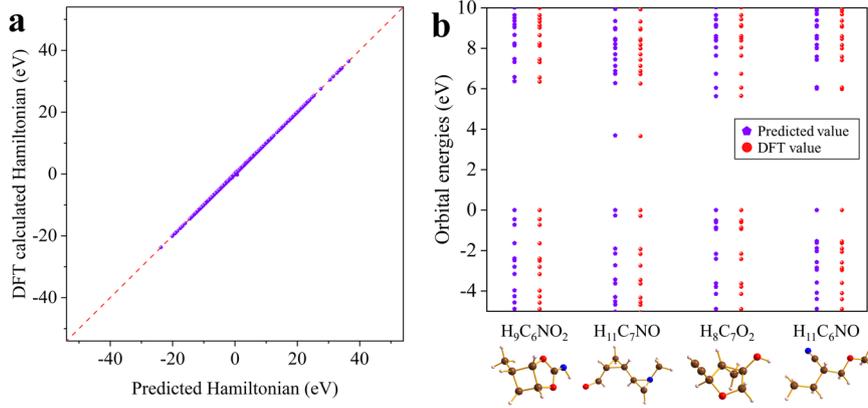


Fig. 2 Application of HamGNN on molecules in the QM9 dataset. **a** The comparison between the Hamiltonian matrix elements predicted by HamGNN and those calculated by OpenMX on the QM9 test set. **b** Comparison of the orbital energies predicted by HamGNN and those calculated by OpenMX for four molecules randomly selected from the QM9 test set.

orbital convolution layers to construct the equivariant features of the atomic orbitals in the local environment. In each orbital convolution layer, the input atomic feature tensor is updated by aggregating the equivariant messages from neighboring atoms. Each equivariant message is generated through the coupling of the feature tensor of neighboring atom j and the convolution filters in an equivariant manner. The rotation equivariant convolution filters are constrained to be a product of learnable radial functions and spherical harmonic functions:

$$F_m^l(\mathbf{r}_{ij}) = S(|\mathbf{r}_{ij}|)Y_{m_l}^{l_l}(\hat{\mathbf{r}}_{ij}) \quad (17)$$

where the radial function $S(|\mathbf{r}_{ij}|) = MLP[B(|\mathbf{r}_{ij}|)]$ is a mapping that transforms the radial basis $B(|\mathbf{r}_{ij}|)$ into the rotation-invariant scalar weights through a multilayer perceptron. The tensor product between the feature tensor of the neighboring atom j and the convolution filter is used to generate the equivariant messages sent to atom i . A tensor product of representations is a mathematical operation for combining two given representations $x^{(l_1)} \in l_1$ and $y^{(l_2)} \in l_2$ to form another equivariant feature $x^{(l_1)} \otimes y^{(l_2)}$. The tensor product $x^{(l_1)} \otimes y^{(l_2)}$ is reducible and can be expanded into a direct sum of irreducible representations. The value for the irreducible representation of rotation degree $l \in \{|l_1 - l_2|, \dots, l_1 + l_2\}$ in the direct sum representation of the tensor product $x^{(l_1)} \otimes y^{(l_2)}$ is given by

$$(x \otimes y)_m^l = \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} C_{m, m_1, m_2}^{l, l_1, l_2} x_{m_1}^{l_1} y_{m_2}^{l_2} \quad (18)$$

where $C_{m, m_1, m_2}^{l, l_1, l_2}$ are Clebsch-Gordan coefficients. Finally, the atomic features in each orbital convolution layer are updated by aggregating the equivariant messages of neighboring atoms using the following formulas:

$$m_{l_i, p_i, c, m_i}^{j, t} = \sum_{m_f=-l_f}^{l_f} \sum_{m_j=-l_j}^{l_j} C_{m_i, m_f, m_j}^{l_i, l_f, l_j} S_{l_i, p_i, c}^{l_i, p_i, c}(|\mathbf{r}_{ij}|) Y_{m_f}^{l_f}(\hat{\mathbf{r}}_{ij}) V_{l_j, p_j, c, m_j}^{j, t-1} \quad (19)$$

$$V_{l_i, p_i, c, m_i}^{i, t} = V_{l_i, p_i, c, m_i}^{i, t-1} + \sum_{j \in N(i)} m_{l_i, p_i, c, m_i}^{j, t} \quad (20)$$

The invariant scalar features of the interatomic distances are used to scale the equivariant output of each rotation order. $S_{l_i, p_i, c}^{l_i, p_i, c}(|\mathbf{r}_{ij}|)$ is a learnable scalar weight for each filter-feature tensor product path $(l_f, p_f) \otimes (l_j, p_j, c) \rightarrow (l_i, p_i, c)$. To respect the parity equivariance, each filter-feature tensor product path $(l_f, p_f) \otimes (l_j, p_j, c) \rightarrow (l_i, p_i, c)$ satisfies the parity selection rule: $p_i = p_f p_j$. Equation (20) is the update function, which aggregates equivariant messages from neighboring atoms to update the input atomic features. The updated atomic features are passed to a nonlinear gate activation

function⁴⁴, which scales the input features equivariantly with the invariant field ($l \neq 0$) of the input features as the gate.

The pair interaction features, which are utilized to construct the off-site Hamiltonian, are generated by the pair interaction layer. The pair interaction features are the sum of two parts. The first part is the tensor product between the features of two interacting atoms i and j : $(l_i, p_i, c) \otimes (l_j, p_j, c) \rightarrow (l, p, c)$, where l and p denote the rotation order and the parity of the pair interaction features, respectively. The second part is the tensor product between the mixed features of atom pairs ij and the convolution filters: $(l', p', c) \otimes (l_f, p_f) \rightarrow (l, p, c)$, where l' and p' denote the rotation order and the parity of the mixed features of atom pairs ij , respectively. The pair interaction features are finally constructed by the following equation:

$$\omega_{l, p, c, m}^{ij} = \sum_{m_i=-l_i}^{l_i} \sum_{m_j=-l_j}^{l_j} C_{m_i, m_j, m}^{l_i, l_j, l} S_{l_i, p_i, c}^{l_i, p_i, c}(|\mathbf{r}_{ij}|) \left(\sum_{c'} W_{l_i, p_i, c, c'}^i V_{l_i, p_i, c, m_i}^{i, T} \right) \left(\sum_{c'} W_{l_j, p_j, c, c'}^j V_{l_j, p_j, c, m_j}^{j, T} \right) + \sum_{m_f=-l_f}^{l_f} \sum_{m'=-l'}^{l'} C_{m', m_f, m}^{l', l_f, l} S_{l', p', c}^{l', p', c}(|\mathbf{r}_{ij}|) Y_{m_f}^{l_f}(\hat{\mathbf{r}}_{ij}) V_{l', p', c, m'}^{j, j} \quad (21)$$

where $W_{l_i, p_i, c, c'}^i$ and $W_{l_j, p_j, c, c'}^j$ are learnable weight matrices used to linearly couple the equivariant features from different channels, $V_{l_i, p_i, c, m_i}^{j, j} = \sum_{c'} W_{l_i, p_i, c, c'}^i V_{l_i, p_i, c, m_i}^{i, T} + \sum_{c'} W_{l_i, p_i, c, c'}^j V_{l_i, p_i, c, m_i}^{j, T}$ is the mixed feature tensor of the atomic features $V_{l_i, p_i, c, m_i}^{i, T}$ and $V_{l_j, p_j, c, m_j}^{j, T}$.

The on-site layer and off-site layer are used to convert the node features $V_{l_i, p_i, c, m_i}^{i, T}$ and pair interaction features $\omega_{l, p, c, m}^{ij}$ into the direct sums Ω_i^{on} and Ω_{ij}^{off} of the ISTs required to construct on-site and off-site Hamiltonian blocks, respectively. We add shortcut connections in the on-site layer and off-site layer and use a norm activation function that scales the modulus of the irreducible representations of each order nonlinearly to increase the non-linear fitting ability of the network. In the last step, the network uses the ISTs in Ω_i^{on} and Ω_{ij}^{off} to construct the on-site and off-site Hamiltonian blocks through Eq. (12). The final predicted Hermitian Hamiltonian is obtained by the following symmetrization:

$$H_{n_i l_i m_i, n_j l_j m_j} = \begin{cases} H_{n_i l_i m_i, n_j l_j m_j}^{on} = (\tilde{H}_{n_i l_i m_i, n_j l_j m_j}^{on} + \tilde{H}_{n_j l_j m_j, n_i l_i m_i}^{on*}) / 2i = j \\ H_{n_i l_i m_i, n_j l_j m_j}^{off} = (\tilde{H}_{n_i l_i m_i, n_j l_j m_j}^{off} + \tilde{H}_{n_j l_j m_j, n_i l_i m_i}^{off*}) / 2i \neq j \end{cases} \quad (22)$$

Tests and applications

To assess the accuracy and transferability of HamGNN, we trained and tested HamGNN on the Hamiltonian matrices and electronic

Table 1. Comparison of the MAEs of the Hamiltonian matrices predicted by HamGNN, DeepH, and PhiSNet for the molecule structures from molecular dynamics trajectories.

Dataset	PhiSNet (meV)	DeepH (meV)	HamGNN (meV)
Water	0.808	1.048	0.760
Ethanol	0.288	0.601	0.195
Malondialdehyde	0.340	0.547	0.270
Uracil	0.487	0.470	0.264

The lowest MAE among the considered models for each subsystem is displayed in bold.

structures for the periodic and aperiodic systems, including various molecules, periodical solids, a nanoscale dislocation defect, and a Moiré superlattice. Previously reported models such as SchNorb³⁵, PhiSNet³⁶, and DeepH⁴⁵ are trained and tested on the slightly perturbed structures from only one configuration each time. Since HamGNN is based on the universal parameterized Hamiltonian proposed in this work, our model can be trained and tested on structures with the same atomic species but different configurations in the same way as MLIPs.

Molecules. The QM9 dataset^{46,47} contains 134k stable small organic molecules made up of CHONF. These small organic molecules are important candidates for drug discovery. The development of general ML models for rapid screening of the electronic structure properties of drug molecules is beneficial for understanding the mechanisms of drugs and shortening the cycle of drug development. We calculated the real-space TB Hamiltonian matrices using OpenMX for 10,000 randomly selected molecules from the QM9 dataset. We divided the whole dataset into the training, validation, and test set with a ratio of 0.8: 0.1: 0.1. As shown in Fig. 2a, the predicted values exhibit a high degree of agreement with the DFT-calculated values of the Hamiltonian matrices for various configurations in the test set. The mean absolute error (MAE) of the Hamiltonian matrix element predicted by the trained HamGNN model on the test set is only 1.49 meV. Please note that all MAEs mentioned in our manuscript refer to the mean absolute errors between predicted values and ground truth (DFT values), unless otherwise stated. Each of the orbital energy calculated by diagonalizing the predicted Hamiltonian matrix coincides almost exactly with each DFT-calculated orbital energy (see Fig. 2b), showing high precision and transferability. The orbital energies calculated by HamGNN and OpenMX for the four molecules shown in Fig. 2b are listed in Supplementary Table 1.

We also trained HamGNN on the Hamiltonians of several specific small molecules generated by ab initio molecular dynamics (MD) and compared the accuracy of HamGNN with two recently reported models, PhiSNet³⁶ and DeepH⁴⁵. The Hamiltonian matrices of these molecules were calculated by OpenMX and divided into the training, validation, and test sets in the same way as PhiSNet in ref. ³⁶. The MAEs of the Hamiltonian matrices predicted by DeepH are from ref. ⁴⁵. As can be seen from Table 1, HamGNN achieves the highest accuracy among the three models. The error of HamGNN as a function of training epochs on both the training and validation sets is highly consistent (see Supplementary Discussion 1). The prediction error of DeepH is higher than that of PhiSNet and HamGNN because the local coordinate system used by DeepH is not strictly equivariant. Although SE(3) equivariant PhiSNet shows high accuracy in predicting the molecules, it is not a universal equivariant model because it does not satisfy the parity symmetry of the Hamiltonian matrix strictly. Our tests have revealed that neglecting parity symmetry can significantly impair its fitting ability for solid materials (see Supplementary Discussion 2).

Periodic solids. We have collected 426 carbon allotropes from the Samara Carbon Allotrope Database (SACADA)⁴⁸ and 30 silicon allotropes from Materials Project⁴⁹, each containing no more than 60 atoms in its unit cell. We have also collected 221 SiO₂ isomers from the Materials Project, each containing no more than 80 atoms in its unit cell. We have perturbed each of the collected SiO₂ isomers three times, resulting in 663 perturbed SiO₂ structures with a random atomic displacement of 0.1 Å. We performed DFT calculations using OpenMX to obtain the ab initio TB Hamiltonian matrices for these carbon allotropes, silicon allotropes, and perturbed SiO₂ isomers. The Hamiltonian matrices in each dataset were divided into training, validation, and test sets with a ratio of 0.8:0.1:0.1. Three separate HamGNN models were trained on the carbon allotropes, silicon allotropes, and perturbed SiO₂ isomers, respectively.

The MAEs of the Hamiltonian matrix predicted by HamGNN for the structures in the test set of carbon allotropes, silicon allotropes, and SiO₂ isomers are 1.55 meV, 2.01 meV, and 2.29 meV, respectively. There are tiny variations in model accuracy for the HamGNN models trained with different initial random network weights (see Supplementary Discussion 3), indicating the high stability of HamGNN. The MAE of HamGNN on the carbon allotropes is even lower than the error (2.0 meV) of DeepH on the training dataset of only the graphene structures⁴⁵. Most importantly, the HamGNN model trained on the carbon allotropes is transferable and can fit the Hamiltonian matrices of the carbon allotropes with arbitrary sizes and configurations beyond the training set. Achieving transferable predictions for the DeepH model that utilizes local coordinate systems is difficult, as discussed in Supplementary Discussion 4.

We used pentadiamond⁵⁰, Moiré twisted bilayer graphene (TBG), Si (MP-1199894), and SiO₂ (MP-1257168) to test the accuracy and transferability of the HamGNN models trained on the three datasets. The test structures are shown in Fig. 3a-d. Pentadiamond is a three-dimensional carbon foam constructed from carbon pentagons and contains 88 carbon atoms in the unit cell⁵⁰. There are 1084 carbon atoms in the twisted bilayer graphene with a Moiré angle $\theta \approx 3.48^\circ$. The interlayer spacing of TBG is about 4.0 Å. The carbon atoms within each layer are connected by strong covalent bonds, while the layers are stacked by the weak van der Waals forces. The Si structure labeled MP-1199894 contains 82 atoms in the unit cell and crystallizes in the monoclinic C2/c space group. The SiO₂ structure labeled MP-1257168 is characterized by the complex porous structures built by SiO₄ tetrahedra and has 180 atoms in the unit cell. The MAE of the Hamiltonian matrix elements predicted by HamGNN for the pentadiamond, TBG, Si (MP-1204046), and SiO₂ (MP-667371) is only 1.54 meV, 3.23 meV, 1.39 meV, and 1.67 meV, respectively. The high accuracy can be seen in the scatter plots of the predicted Hamiltonian matrices versus the DFT calculated Hamiltonian matrices shown in Fig. 3e-h. Although our carbon allotropic dataset does not contain any bilayer graphene structures, HamGNN automatically learns van der Waals interactions from this dataset and successfully predicts the Hamiltonian matrix of TBG. The energy bands obtained by diagonalizing the predicted Hamiltonians closely align with those from the DFT calculations, as shown in Fig. 3i-l.

Dislocation defect. In general, point defects only induce local structural distortions, so the supercell used for point defect simulation does not need a very large size and can be calculated directly using the DFT method^{51,52}. However, as one-dimensional line defects, edge dislocations induce elastic stress fields in their surroundings⁵³⁻⁵⁵, resulting in significantly greater lattice distortions compared to point defects. Simulating an isolated edge dislocation necessitates the utilization of a large supercell to mitigate the strong elastic interactions between the dislocation and its periodic image. The high density of dislocations in small

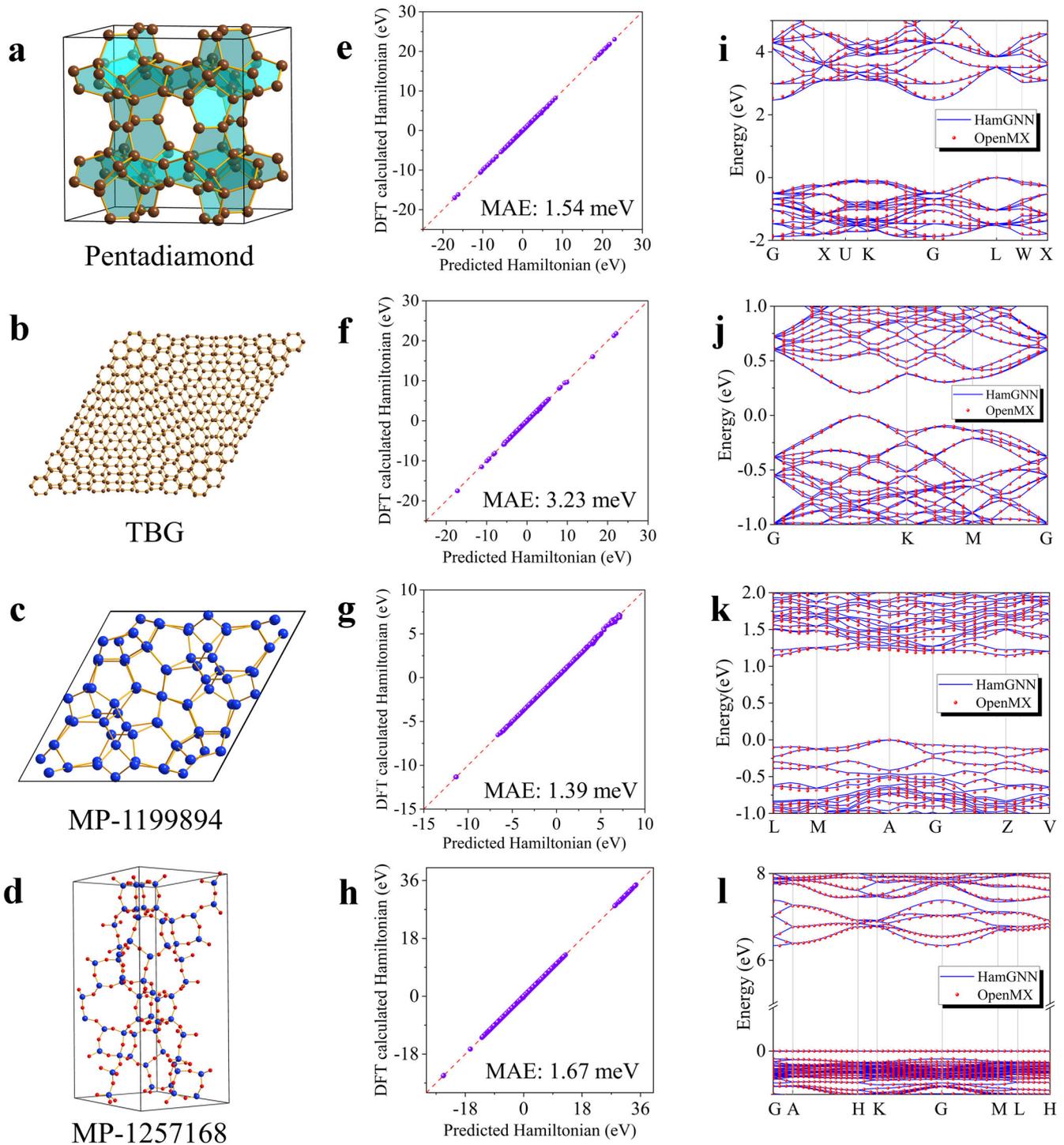


Fig. 3 The prediction of HamGNN on several periodic solids that are not present in the training sets. **a–d** Crystal structures of pentadiamond, Moiré twisted bilayer graphene (TBG), Si (MP-1199894), and SiO₂ (MP-1257168). **e–h** Comparison of the Hamiltonian matrix elements predicted by HamGNN and those calculated by OpenMX for pentadiamond, TBG, Si (MP-1199894), and SiO₂ (MP-1257168). **i–l** Comparison of the energy bands predicted by HamGNN (solid line) and those (dashed line) calculated by OpenMX for pentadiamond, TBG, Si (MP-1199894), and SiO₂ (MP-1257168).

supercells may lead to unrealistic lattice distortion and inaccurate electronic structures. However, the computational complexity of the large supercells required to simulate the edge dislocations is a great challenge for DFT methods. HamGNN enables the direct mapping from structure to the Hamiltonian matrix, providing the possibility to compute electronic structures of dislocation defects in large supercells.

In this work, we take an isolated edge dislocation in a large silicon supercell as an example to demonstrate the high efficiency of the HamGNN model to simulate nanoscale defects in large systems. Crystalline silicon has a diamond-type crystal structure, whose most favorable slip system belongs to the type $\frac{1}{2}\{110\}\{111\}$. Taking the $\{111\}$ plane of silicon as the slip plane, we constructed an isolated edge dislocation with Burgers vector $1/$

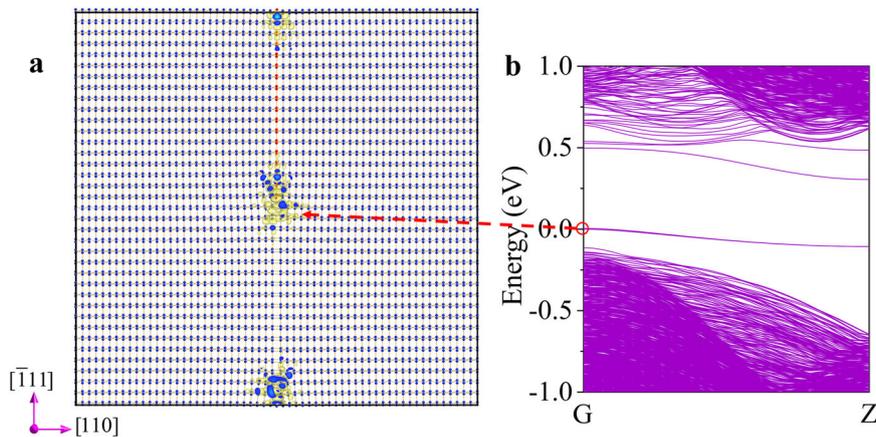


Fig. 4 The electronic structure prediction on silicon dislocation. **a** Atomic structure and charge density of the valence band maximum (VBM) of an isolated dislocation in silicon. The atoms on the dislocation plane are colored red. **b** The energy bands of the dislocation model along G(0.0, 0.0, 0.0) to Z(0.0, 0.0, 0.5).

$2 < 110 \rangle$, as shown in Fig. 4a. There are 4284 atoms in this supercell containing an isolated edge dislocation. We utilized the HamGNN model, which was trained on Si allotropes, to predict the Hamiltonian matrix of this supercell. Subsequently, we computed its band structure and charge density of valence band maximum (VBM) using the predicted Hamiltonian matrix.

The defect energy bands introduced by the dislocation defect appear within the band gap, as shown in Fig. 4b, which significantly narrows the band gap of bulk Si. The predicted valence band maximum (VBM) at the Gamma point is located on an occupied defect energy band. The VBM wave function shown in Fig. 4a is mainly distributed in the dislocation core, indicating that this occupied band is caused by the hanging bonds and structural distortion in the core. The predicted electronic structure in the dislocation core is analogous to that obtained from the DFT simulation for a non-periodical dislocation model containing only 358 atoms⁵⁶. However, the dangling bonds at the boundaries of the non-periodical supercell are saturated by hydrogens⁵⁶, which may introduce unrealistic electronic states from hydrogens. We use HamGNN to directly compute the electronic structures of dislocations in large periodic supercells, thus overcoming the potential issues arising from the use of small or non-periodic supercells. With the ability to establish a shortcut from structure to ab initio Hamiltonian matrix, HamGNN enables direct calculation of electronic structures for large supercells without the costly SCF iterations. HamGNN demonstrated exceptional speed and efficiency, taking only 36 seconds to calculate the Hamiltonian matrix of the silicon supercell containing 4284 atoms on a node equipped with 80 Intel(R) Xeon(R) Gold 6248 CPU cores. To verify the reliability of the trained HamGNN model in predicting the electronic structures for silicon dislocation defects, we constructed a small silicon supercell containing an edge dislocation defect and compared the Hamiltonian matrix, energy bands, and the wave function predicted by HamGNN with those calculated by OpenMX (see Supplementary Discussion 5).

Moiré superlattice of bilayer MoS₂. MoS₂ is a 2D transition metal dichalcogenide (TMD) that has attracted much attention because it is an excellent semiconductor with a wide range of applications in the field of electronics and optoelectronics^{57–60}. Different from monolayer or untwisted bilayer MoS₂, the twisted bilayer MoS₂ with Moiré angles has been found to have flat bands and shear solitons^{57,61–63}, which may lead to some interesting physical phenomena, such as superconducting states, quantum Hall insulators, Mott-insulating phases. However, the electronic structure calculation of Moiré twisted bilayer MoS₂ by DFT is relatively expensive due to its large size. To demonstrate the accuracy and

efficiency of HamGNN in replacing DFT for calculating the electronic structure of Moiré twisted two-dimensional materials, we conducted a quantitative comparison between the HamGNN prediction and the DFT calculation for twisted bilayer MoS₂ with a Moiré angle of 3.5°. Before predicting the electronic structure of the Moiré twisted bilayer MoS₂ superlattice, we trained a HamGNN model using a dataset consisting of 500 untwisted bilayer MoS₂ structures, each containing 54 atoms. Each MoS₂ bilayer structure in the dataset has a random interlayer sliding distance of up to 2 angstroms along a random direction. The layer spacing of each MoS₂ bilayer structure in the dataset was randomly shifted by a maximum of 0.5 angstroms. We used OpenMX to calculate the Hamiltonian matrices for the untwisted MoS₂ bilayer structures, and divided the dataset into training, validation, and test sets with a ratio of 0.8:0.1:0.1. The MAE of the trained HamGNN on the test set is only 0.82 meV.

We used OpenMX and the trained HamGNN model to calculate the Hamiltonian matrix of the twisted bilayer MoS₂ with a Moiré angle of 3.5°, which contains 1626 atoms in a unit cell. HamGNN and OpenMX performed the calculations on a node with 80 Intel(R) Xeon(R) Gold 6248 CPU cores. HamGNN only took about 21 s, while OpenMX required approximately 73,000 s to complete this task. This indicates that HamGNN significantly improves the efficiency of electronic structure calculations. The MAE between the Hamiltonian matrix predicted by HamGNN and that calculated by OpenMX is only 0.89 meV. The energy bands, obtained by diagonalizing the Hamiltonian matrices from HamGNN and OpenMX, are shown in Fig. 5a. The predicted bands show a high level of agreement with those computed by OpenMX. Besides, the appearance of flat bands at the valence band edge and the Dirac cone at the K point agrees well with the energy bands calculated by VASP in ref. ⁶¹. The spatial distribution of the predicted VBM wave function, as illustrated in Fig. 5b, exhibits a high degree of localization around the Moiré patterns, which is consistent with the computational results obtained from VASP calculations under LDA and PBE functionals⁶¹.

Through the quantitative comparisons, we demonstrate that HamGNN models trained on small systems can serve as an efficient alternative to DFT for accurately calculating the electronic structure of large systems. After being trained on the Hamiltonian matrices of small untwisted bilayer structures, the HamGNN model can effectively predict the electronic structure of small-angle twisted materials, which is computationally expensive for DFT calculations. The high accuracy and efficiency of HamGNN demonstrate its immense potential in accelerating the electronic structure calculations for small-angle twisted bilayers of various transition metal dichalcogenides (TMDs). In Supplementary

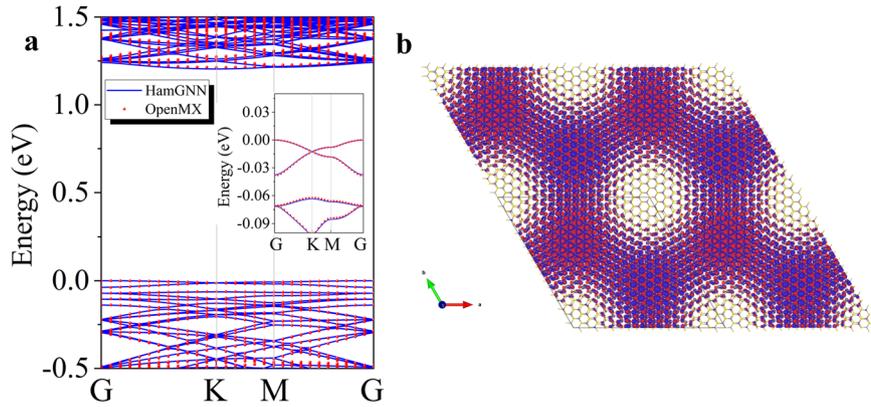


Fig. 5 The electronic structure prediction for the twisted bilayer MoS_2 with a Moiré angle of 3.5° . **a** Comparison of the energy bands predicted by HamGNN and those calculated by OpenMX for the Moiré twisted bilayer MoS_2 . **b** The spatial distribution of VBM wave function in the Moiré twisted bilayer MoS_2 .

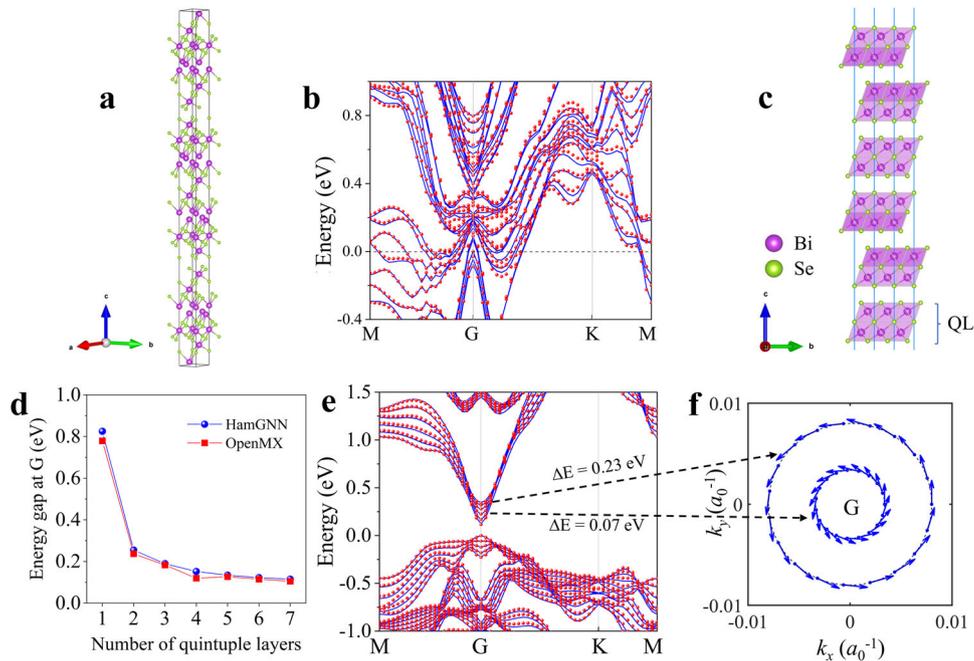


Fig. 6 The electronic structure prediction on the Bi_xSe_y quantum materials. **a** The crystal structure of Bi_8Se_7 . **b** Comparison of the energy bands predicted by HamGNN (solid line) with those calculated by DFT (dashed line) for Bi_8Se_7 . **c** Schematic representation of the layered crystal structure of Bi_2Se_3 . **d** Comparison between the energy gaps at G point predicted by HamGNN with those calculated by DFT for the Bi_2Se_3 slab models with various quintuple layers (QL). **e** Comparison of the energy bands predicted by HamGNN (solid line) with those calculated by DFT (dashed line) for Bi_2Se_3 slab with 6 QLs. **f** The predicted spin textures of two unoccupied states near the conduction band minimum (CBM).

Discussion 6, we have discussed the computational scalability of HamGNN with respect to system sizes and also compared its computational efficiency with OpenMX for bilayer MoS_2 slabs of varying twist angles.

Bi_xSe_y quantum materials. Bi and Se have multiple chemical valences and can form a set of binary compounds Bi_xSe_y with various stoichiometric ratios^{64,65}. Bi is a heavy element whose d electrons have strong SOC effects. A total of 19 Bi_xSe_y compounds can be found on Materials Project⁴⁹. The compound Bi_8Se_7 (id: MP-680214) shown in Fig. 6a, which contains 45 atoms in the unit cell, was used to test the transferability and accuracy of HamGNN. The remaining 18 Bi_xSe_y compounds, which contain no more than 40 atoms in the unit cell, were used to generate the training set for the network. To increase the size of the training set, we applied a

random perturbation up to 0.02 \AA to the atoms of each Bi_xSe_y structure to generate 50 perturbed structures and obtained 900 structures in total. These structures were randomly divided into the training, validation, and test sets with a ratio of 0.8: 0.1: 0.1. The MAE of the real part of the SOC Hamiltonian predicted by the trained model for Bi_8Se_7 is 1.29 meV, and the MAE of the imaginary part of the SOC Hamiltonian is only 5.0×10^{-7} meV. As the SOC effect primarily manifests in the imaginary component of the Hamiltonian matrix, such a low MAE for the imaginary part indicates that our parameterized SOC Hamiltonian has a strong fitting ability across various structures. As shown in Fig. 6b, the predicted energy bands of Bi_8Se_7 are in good agreement with those calculated by OpenMX. Although the training set does not contain any compounds with a stoichiometric ratio of Bi_8Se_7 , the HamGNN model still accurately predicted the SOC Hamiltonian

matrix and energy bands of this structure, demonstrating its high transferability.

Bi_2Se_3 is a widely recognized 3D topological insulator material, which serves as an exceptional platform for exploring quantum phenomena related to the effects of spin-orbit coupling (SOC)^{66–70}. Bulk Bi_2Se_3 is stacked by quintuple layers (QLs) through the van der Waals (vdW) interaction, as is shown in Fig. 6c. Each QL layer is composed of five atomic layers of Se-Bi-Se-Bi-Se combined by strong covalent bonds. As shown in Fig. 6d, the G-point band gaps predicted by HamGNN exhibit a high degree of agreement with those obtained from DFT calculations for Bi_2Se_3 slabs comprising 1 to 7 QLs. As shown in Fig. 6e, the addition of more QL layers results in a gradual decrease of the G-point band gap and a tendency towards linear band dispersion at the G point, ultimately forming a Dirac cone. The spin textures of the lowest unoccupied states, located 0.07 eV and 0.23 eV above the conduction band minimum (CBM), were computed using the Hamiltonian matrix predicted by HamGNN, as illustrated in Fig. 6f. The predicted spin textures are in good agreement with the features of the Dirac cone, which is a topological surface state protected by time-reversal symmetry and characterized by spin-momentum locking.

DISCUSSION

DFT methods are now widely used to calculate various properties of molecules and materials. However, successful implementation of DFT calculations on large systems remains infrequent due to the significant computational resources and running time required. A typical DFT calculation often requires tens to hundreds of self-consistent iterations to obtain the final Hamiltonian and wave function, with the diagonalization of the Hamiltonian matrix performed on a dense k-point grid at each iteration step. This process takes up most of the running time of DFT calculations and can not be circumvented. The emergence of deep learning in recent years has enabled efficient atomic simulations with DFT accuracy, as evidenced by the widespread use of machine learning interatomic potentials (MLIPs) that offer quantum mechanical precision to accelerate long-time molecular dynamics simulations for large systems. As potential energy is just an invariant scalar, the implementation of transferable MLIP models is relatively straightforward. However, due to the rotational equivariance and parity symmetry of the Hamiltonian matrix, developing a transferrable model for directly predicting the Hamiltonian matrix is highly challenging.

In this work, we propose an analytical E(3) equivariant parameterized Hamiltonian matrix that explicitly takes into account rotation equivariance and parity symmetry. Furthermore, we extend it to a parameterized Hamiltonian matrix satisfying SU(2) and time-reversal equivariance to fit the Hamiltonian matrix with SOC effects. Based on this parameterized Hamiltonian matrix, we develop an E(3) equivariant deep neural network called HamGNN to fit the Hamiltonian matrix of various molecules and solids. Previously reported models were trained and tested on the datasets consisting of slightly perturbed molecules or crystals from the same configuration. To demonstrate the accuracy and transferability of our parameterized Hamiltonian matrix, we used the trained HamGNN model to predict the electronic structures of the molecules, periodic solids, the silicon dislocation defect, Moiré twisted bilayer MoS_2 , and Bi_xSe_y quantum materials. The results of actual tests demonstrate that our model exhibits a high level of accuracy in comparison to DFT, while also demonstrating a significant degree of transferability similar to that of MLIPs. These performances provide a crucial foundation for the extensive implementation of electronic structure methods in machine learning. Since our model can establish a direct mapping from the structure to the self-consistent Hamiltonian matrix without the time-consuming self-consistent iterations in DFT, it can be used to

accelerate the electronic structure calculation of large systems and other costly advanced calculations, such as the electron-phonon coupling matrix via the automatic differentiation ability of the neural network.

METHODS

Hamiltonian datasets and DFT calculation details

QM9 structure set and the molecules in Table 1 are available from <http://quantum-machine.org/datasets/>, SACADA structure set is available from https://www.sacada.info/sacada_3D.php, and the structures of Si allotropes, SiO_2 isomers, and Bi_xSe_y crystals are downloaded from the Materials Project site⁴⁹. To prepare the training set of untwisted bilayer MoS_2 , a random perturbation of up to 0.02 Å is applied to each atom, and a slip of up to 2 Å is performed in a random direction within the XY plane. The layer spacing of the bilayer MoS_2 was randomly shifted by a maximum of 0.5 angstroms. We performed DFT calculations on the structures in the above datasets to obtain TB Hamiltonian matrices via OpenMX¹⁷, a software package for nano-scale material simulations based on norm-conserving pseudopotentials and pseudo-atomic localized basis functions. The PBE (Perdew-Burke-Ernzerhof)⁷¹ functional is employed for all OpenMX calculations presented in this study. H6.0-s2p1, C6.0-s2p2d1, N6.0-s2p2d1, O6.0-s2p2d1, F6.0-s2p2d1, Si7.0-s2p2d1, Mo7.0-s3p2d2, Bi8.0-s3p2d2, and Se7.0-s3p2d2 pseudoatomic orbitals (PAOs) were used as the basis for the calculations. The truncation radius of the atomic orbits of H, C, N, O, and F is 6.0 Bohr, the truncation radius of the atomic orbits of Si, Mo, and Se is 7.0 Bohr, the truncation radius of the atomic orbits of Bi is 8.0 Bohr. The cutoff energy and K-point grid used by OpenMX in the calculation of the Hamiltonian matrices for each dataset are listed in Supplementary Table 6. The Si dislocation model was built by AtomsK⁷² and relaxed by GPUMD⁷³ with a force criterion of 0.1 eV · Å⁻¹. The VASP^{74,75} results of Moiré superlattice of bilayer MoS_2 , which serve to validate the accuracy of our HamGNN predictions, are referenced from Naik et al.'s work⁶¹. They utilized LDA and PBE functionals to calculate the electronic structure of twisted bilayer MoS_2 with a Moiré angle of 3.5°. For the OpenMX calculation on the Moiré twisted bilayer MoS_2 , we used a cutoff energy of 150 Ry and a 4 × 4 × 1 k-point mesh to sample the Brillouin zone.

Hamiltonian construction details

The majority of computational time in DFT calculations is dedicated to self-consistent iterations aimed at obtaining the self-consistent charge density ρ , which is used to determine the final Hamiltonian matrix and wave functions. The Kohn-Sham Hamiltonian can be written as $H = T + V_{\text{eff}}$, where $V_{\text{eff}} = \sum_l V_{\text{ec}}(\mathbf{r} - \mathbf{r}_l) + V_{\text{scf}}[\rho]$. $V_{\text{ec}}(\mathbf{r} - \mathbf{r}_l)$ is the Coulomb potential of the core charges of each atom l , $V_{\text{scf}}[\rho] = \frac{\delta E_{\text{ee}}}{\delta \rho} + \frac{\delta E_{\text{xc}}}{\delta \rho}$ denotes the sum of the potentials that arise from the electron-electron interaction energy $E_{\text{ee}}[\rho]$ and the exchange-correlation energy $E_{\text{xc}}[\rho]$. The matrix elements of the kinetic energy operator $\langle \phi_{n_l, l, m_l}(\mathbf{r} - \mathbf{r}_l) | \hat{T} | \phi_{n_j, j, m_j}(\mathbf{r} - \mathbf{r}_j - \mathbf{R}_{n_c}) \rangle$ and the Coulomb potential of the atomic cores $\langle \phi_{n_l, l, m_l}(\mathbf{r} - \mathbf{r}_l) | \sum_l V_{\text{ec}}(\mathbf{r} - \mathbf{r}_l) | \phi_{n_j, j, m_j}(\mathbf{r} - \mathbf{r}_j - \mathbf{R}_{n_c}) \rangle$ are not functionals of the self-consistent charge density and can be calculated analytically without SCF iterations. The complete Hamiltonian can be obtained by the sum of the SCF-independent matrices calculated analytically from the atomic coordinates and the SCF-dependent matrix $V_{\text{scf}}[\rho]$ that can be fitted by the equivariant neural networks. The general parameterized Hamiltonian proposed in this work can be applied to all tight-binding Hamiltonian matrices calculated using various exchange-correlation functionals (see Supplementary Discussion 7).

Network and training details

The Pytorch-1.11.0⁷⁶, PyG-2.0.4⁷⁷, Pymatgen⁷⁸, Nequip-0.5.3³³, and e3nn-0.5.0⁷⁹ libraries are used to implement HamGNN. All models used in this work have five orbital convolution layers, a pair interaction layer, an on-site layer, and an off-site layer. The edge distances between each atom and its neighbors are expanded by eight Bessel function bases. The spherical harmonic functions with a maximum degree $L_{\max} = 4$ are used to embed the directions of the edges. The $O(3)$ representations used for the atom features and pair interaction features have N_{fea} channels with a maximum degree L_{\max} and parities $p = \pm 1$. The number of feature channels N_{fea} and the maximum degree L_{\max} for each dataset are listed in Supplementary Table 7. A two-layer MLP with 64 neurons is used to map the invariant edge embeddings to the weights of each tensor product path in Eqs. (19) and (21). Shifted softplus³¹ function is used as the activation function in the MLP. The gate activation function scales the input features $(\oplus_i u_{p_i}^{(0)}) \oplus (\oplus_j v_{p'_j}^{(0)}) \oplus (\oplus_j w_{p'_j}^{(j>0)})$ with its invariant field $(\oplus_i u_{p_i}^{(0)})$ and $(\oplus_j v_{p'_j}^{(0)})$ as the gate.

The output equivariant features of gate nonlinearity are $[\oplus_i \phi_{p_i}^1(u_{p_i}^{(0)})] \oplus [\oplus_j \phi_{p'_j}^2(v_{p'_j}^{(0)})w_{p'_j}^{(j)}]$, where $\phi_{p_i}^1$ and $\phi_{p'_j}^2$ are the activation functions that vary with the parity of the scalar input, defined as follows³³:

$$\phi_{p_i}^1(x) = \begin{cases} \text{ssp}(x) = \ln(0.5 \times e^x + 0.5) & p_i = 1 \\ \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} & p_i = -1 \end{cases} \quad (23)$$

$$\phi_{p'_j}^2(x) = \begin{cases} \text{ssp}(x) = \ln(0.5 \times e^x + 0.5) & p'_j = 1 \\ |x| & p'_j = -1 \end{cases} \quad (24)$$

$\phi_{p_i}^1(x)$ has the same parity as the input scalar x , while $\phi_{p'_j}^2(x)$ always has even parity. This ensures that the parity of the output features of the Gate activation function is equivariant.

To increase transferability and avoid overfitting, we include the error of the calculated energy bands as a regularization term in the loss function:

$$L = \|\tilde{H} - H\| + \frac{\lambda}{N_{\text{orb}} \times N_k} \sum_{k=1}^{N_k} \sum_{n=1}^{N_{\text{orb}}} \|\tilde{\epsilon}_{nk} - \epsilon_{nk}\| \quad (25)$$

where the variables marked with a tilde refer to the corresponding predictions and λ denotes the loss weight of the band energy error. λ equals 0.001 in our training. When the training of the network has not converged, the error of the predicted Hamiltonian is large, resulting in poor or even divergent prediction values of the energy bands. Adding the band loss value at the beginning of training may cause the total loss value to diverge. Therefore, we train the network in two steps. First, only the mean absolute error of Hamiltonian matrices is used as the loss value to train the network until the network weights converge. The parameters were optimized with AdamW^{80,81} optimizer using an initial learning rate of 10^{-3} . Then the mean absolute error of each band calculated at N_k random points in the reciprocal space is added to the loss function and starts the training at an initial learning rate of 10^{-4} . When the accuracy of the model on the validation set is not improved after successive N_{patience} epochs, the learning rate will be reduced by a factor of 0.5. When the accuracy of the model on the validation set is not improved after successive N_{stop} epochs or the learning rate is lower than 10^{-6} , the training will be stopped and the model that has the best accuracy on the validation set will be used on the test set. The values of some key network and training parameters on each dataset are listed in Supplementary Table 7. All models were trained on a single NVIDIA A100 GPU.

DATA AVAILABILITY

The pre-trained models and the test examples are available on Zenodo (<https://doi.org/10.5281/zenodo.8147631>). The training datasets are available on Zenodo (<https://doi.org/10.5281/zenodo.8157128>).

CODE AVAILABILITY

The HamGNN code used in the current study is available on GitHub (<https://github.com/QuantumLab-ZY/HamGNN>).

Received: 24 March 2023; Accepted: 14 September 2023;

Published online: 06 October 2023

REFERENCES

- Choudhary, K. et al. High-throughput density functional perturbation theory and machine learning predictions of infrared, piezoelectric, and dielectric responses. *npj Comput. Mater.* **6**, 64 (2020).
- Liu, Y. T., Zhou, Q. & Cui, G. L. Machine learning boosting the development of advanced lithium batteries. *Small Methods* **5**, 2100442 (2021).
- Zhang, N. et al. Machine learning in screening high performance electrocatalysts for CO₂ reduction. *Small Methods* **5**, 2100987 (2021).
- Manzhos, S. & Carrington, T. Jr Neural network potential energy surfaces for small molecules and reactions. *Chem. Rev.* **121**, 10187–10217 (2021).
- Choudhary, K. et al. Unified graph neural network force-field for the periodic table: solid state applications. *Digit. Discov.* **2**, 346–355 (2023).
- Cheng, Z. et al. Building quantum mechanics quality force fields of proteins with the generalized energy-based fragmentation approach and machine learning. *Phys. Chem. Chem. Phys.* **24**, 1326–1337 (2022).
- Unke, O. T. et al. SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.* **12**, 7273 (2021).
- Cheng, G. J., Gong, X. G. & Yin, W. J. Crystal structure prediction by combining graph network and optimization algorithm. *Nat. Commun.* **13**, 1492 (2022).
- Ganea, O. et al. GeoMol: torsional geometric generation of molecular 3D conformer ensembles. *Adv. Neural Inf. Process. Syst.* **34**, 13757–13769 (2021).
- Xie, T., Fu, X., Ganea, O.-E., Barzilay, R. & Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations (ICLR)*, 2019.
- Kirkpatrick, J. et al. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science* **374**, 1385–1389 (2021).
- Nagai, R., Akashi, R. & Sugino, O. Machine-learning-based exchange correlation functional with physical asymptotic constraints. *Phys. Rev. Res.* **4**, 013106 (2022).
- Goringe, C. M., Bowler, D. R. & Hernandez, E. Tight-binding modelling of materials. *Rep. Prog. Phys.* **60**, 1447–1512 (1997).
- Papaconstantopoulos, D. A. & Mehl, M. J. The Slater-Koster tight-binding method: a computationally efficient and accurate approach. *J. Phys. Condens. Matter* **15**, R413–R440 (2003).
- Hegde, G. & Bowen, R. C. Machine-learned approximations to density functional theory Hamiltonians. *Sci. Rep.* **7**, 42669 (2017).
- Wang, Z. F. et al. Machine learning method for tight-binding Hamiltonian parameterization from ab-initio band structure. *npj Comput. Mater.* **7**, 11 (2021).
- Ozaki, T. Variationally optimized atomic orbitals for large-scale electronic structures. *Phys. Rev. B* **67**, 155108 (2003).
- Ozaki, T. & Kino, H. Numerical atomic basis orbitals from H to Kr. *Phys. Rev. B* **69**, 195113 (2004).
- Artacho, E. et al. The SIESTA method; developments and applicability. *J. Phys.* **20**, 064208 (2008).
- Garcia, A. et al. Siesta: recent developments and applications. *J. Chem. Phys.* **152**, 204108 (2020).
- Morrison, M. A. & Parker, G. A. A guide to rotations in quantum-mechanics. *Aust. J. Phys.* **40**, 465–497 (1987).
- Weinert, U. Spherical tensor representation. *Arch. Ration. Mech.* **74**, 165–196 (1980).
- Zhang, L. et al. Equivariant analytical mapping of first principles Hamiltonians to accurate and transferable materials models. *npj Comput. Mater.* **8**, 158 (2022).
- Nigam, J., Willatt, M. J. & Ceriotti, M. Equivariant representations for molecular Hamiltonians and N-center atomic-scale properties. *J. Chem. Phys.* **156**, 014115 (2022).
- Unke, O. T. & Meuwly, M. A reactive, scalable, and transferable model for molecular energies from a neural network approach based on local information. *J. Chem. Phys.* **148**, 241708 (2018).

26. Christensen, A. S., Bratholm, L. A., Faber, F. A. & Anatole von Lilienfeld, O. FCHL revisited: Faster and more accurate quantum machine learning. *J. Chem. Phys.* **152**, 044107 (2020).
27. Faber, F. A., Christensen, A. S., Huang, B. & von Lilienfeld, O. A. Alchemical and structural distribution based representation for universal quantum machine learning. *J. Chem. Phys.* **148**, 241717 (2018).
28. Bartok, A. P., Kondor, R. & Csanyi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
29. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. *Proc. 34th Int. Conf. Mach. Learn.* **70**, 1263–1272 (2017).
30. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
31. Schutt, K. T., Sauceda, H. E., Kindermans, P. J., Tkatchenko, A. & Muller, K. R. SchNet - A deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
32. Choudhary, K. & DeCost, B. Atomistic line graph neural network for improved materials property predictions. *npj Comput. Mater.* **7**, 185 (2021).
33. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
34. Wang, Z. et al. Heterogeneous relational message passing networks for molecular dynamics simulations. *npj Comput. Mater.* **8**, 53 (2022).
35. Schutt, K. T., Gastegger, M., Tkatchenko, A., Muller, K. R. & Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **10**, 5024 (2019).
36. Unke, O. et al. SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. *Adv. Neural Inf. Process. Syst.* **34**, 14434–14447 (2021).
37. Li, H. C., Collins, C., Tanha, M., Gordon, G. J. & Yaron, D. J. A density functional tight binding layer for deep learning of chemical Hamiltonians. *J. Chem. Theory Comput.* **14**, 5764–5776 (2018).
38. Grisafi, A., Wilkins, D. M., Csanyi, G. & Ceriotti, M. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Phys. Rev. Lett.* **120**, 036002 (2018).
39. Jones, M. D. & Albers, R. C. Spin-orbit coupling in an f-electron tight-binding model: Electronic properties of Th, U, and Pu. *Phys. Rev. B* **79**, 045107 (2009).
40. Hemstreet, L. A., Fong, C. Y. & Nelson, J. S. First-principles calculations of spin-orbit splittings in solids using nonlocal separable pseudopotentials. *Phys. Rev. B: Condens. Matter* **47**, 4238–4243 (1993).
41. Klicpera, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations (ICLR)*, (2019).
42. 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).
43. Thomas, N. et al. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. Preprint at <https://doi.org/10.48550/arXiv.1802.08219> (2018).
44. Weiler, M., Geiger, M., Welling, M., Boomsma, W. & Cohen, T. S. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. *Adv. Neural Inf. Process. Syst.* **31**, 10381–10392 (2018).
45. Li, H. et al. Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.* **2**, 367–377 (2022).
46. Ruddigkeit, L., van Deursen, R., Blum, L. C. & Reymond, J. L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **52**, 2864–2875 (2012).
47. Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022 (2014).
48. Hoffmann, R., Kabanov, A. A., Golov, A. A. & Proserpio, D. M. Homo citans and carbon allotropes: for an ethics of citation. *Angew. Chem. Int. Ed.* **55**, 10962–10976 (2016).
49. Jain, A. et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
50. Fujii, Y., Maruyama, M., Cuong, N. T. & Okada, S. Pentadiamond: a hard carbon allotrope of a pentagonal network of sp^2 and sp^3 C atoms. *Phys. Rev. Lett.* **125**, 016001 (2020).
51. Puska, M. J. Point defects in silicon, first-principles calculations. *Comput. Mater. Sci.* **17**, 365–373 (2000).
52. McCluskey, M. D. & Janotti, A. Defects in semiconductors. *J. Appl. Phys.* **127**, 190401 (2020).
53. Blumenau, A. T. et al. Dislocations in diamond: Dissociation into partials and their glide motion. *Phys. Rev. B* **68**, 014115 (2003).
54. Li, Z. & Picu, R. C. Shuffle-glide dislocation transformation in Si. *J. Appl. Phys.* **113**, 083519 (2013).
55. Das, S. & Gavini, V. Electronic structure study of screw dislocation core energetics in Aluminum and core energetics informed forces in a dislocation aggregate. *J. Mech. Phys. Solids* **104**, 115–143 (2017).
56. Wang, J. W., Xu, W. W., Wang, R., Laref, A. & Wu, X. Z. Structural and electronic properties of 90 degrees dislocations in silicon nanorods: A first-principles calculation. *Comput. Mater. Sci.* **149**, 243–249 (2018).
57. Wu, F., Lovorn, T., Tutuc, E., Martin, I. & MacDonald, A. H. Topological insulators in twisted transition metal dichalcogenide homobilayers. *Phys. Rev. Lett.* **122**, 086402 (2019).
58. Li, X. & Zhu, H. Two-dimensional MoS_2 : Properties, preparation, and applications. *J. Mater.* **1**, 33–44 (2015).
59. Conley, H. J. et al. Bandgap engineering of strained monolayer and bilayer MoS_2 . *Nano. Lett.* **13**, 3626–3630 (2013).
60. Xu, K. et al. The role of Anderson's rule in determining electronic, optical and transport properties of transition metal dichalcogenide heterostructures. *Phys. Chem. Chem. Phys.* **20**, 30351–30364 (2018).
61. Naik, M. H. & Jain, M. Ultraflatbands and shear solitons in Moiré patterns of twisted bilayer transition metal dichalcogenides. *Phys. Rev. Lett.* **121**, 266401 (2018).
62. Devakul, T., Crepel, V., Zhang, Y. & Fu, L. Magic in twisted transition metal dichalcogenide bilayers. *Nat. Commun.* **12**, 6730 (2021).
63. Liu, Y. et al. Moiré superlattices and related Moiré excitons in twisted van der Waals heterostructures. *Chem. Soc. Rev.* **50**, 6401–6422 (2021).
64. Majhi, K. et al. Emergence of a weak topological insulator from the Bi_2Se_3 family. *Appl. Phys. Lett.* **110**, 162102 (2017).
65. Goncalves, P. H. R. et al. Formation of Bi_2Se_3 phases upon annealing of the topological insulator Bi_2Te_3 : Stabilization of in-depth bismuth bilayers. *J. Phys. Chem. Lett.* **9**, 954–960 (2018).
66. Yazyev, O. V., Moore, J. E. & Louie, S. G. Spin polarization and transport of surface states in the topological insulators Bi_2Se_3 and Bi_2Te_3 from first principles. *Phys. Rev. Lett.* **105**, 266806 (2010).
67. Zhang, Y. et al. Crossover of the three-dimensional topological insulator Bi_2Te_3 to the two-dimensional limit. *Nat. Phys.* **6**, 584–588 (2010).
68. Crowley, J. M., Tahir-Kheli, J. & Goddard, W. A. 3rd accurate ab initio quantum mechanics simulations of Bi_2Se_3 and Bi_2Te_3 topological insulator surfaces. *J. Phys. Chem. Lett.* **6**, 3792–3796 (2015).
69. Mazumder, K. & Shirage, P. M. A brief review of Bi_2Se_3 based topological insulator: From fundamentals to applications. *J. Alloy. Compd.* **888**, 161492 (2021).
70. Shirali, K., Shelton, W. A. & Vekhter, I. Importance of van der Waals interactions for ab initio studies of topological insulators. *J. Phys. Condens. Matter* **33**, 035702 (2020).
71. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
72. Hirel, P. AtomsK: A tool for manipulating and converting atomic data files. *Comput. Phys. Comm.* **197**, 212–219 (2015).
73. Fan, Z. et al. GPUMD: A package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations. *J. Chem. Phys.* **157**, 114801 (2022).
74. Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comp. Mater. Sci.* **6**, 15–50 (1996).
75. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
76. Paszke, A. et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32**, 8026–8037 (2019).
77. Fey, M. & Lenssen, J. E. Fast graph representation learning with pytorch geometric. In *Workshop of International Conference on Learning Representations (ICLR)*, (2019).
78. Ong, S. P. et al. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comp. Mater. Sci.* **68**, 314–319 (2013).
79. Geiger M. et al. *e3nn/e3nn: 2022-04-13 (0.5.0)*. Zenodo <https://doi.org/10.5281/zenodo.6459381> (2022).
80. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations* (2021).
81. Reddi, S. J., Kale, S. & Kumar, S. On the convergence of adam and beyond. Preprint at <https://doi.org/10.48550/arXiv.1904.09237> (2019).

ACKNOWLEDGEMENTS

We thank Dr. Hongli Guo for providing the structure of the Moiré superlattice of bilayer MoS_2 . We acknowledge financial support from the Ministry of Science and Technology of the People's Republic of China (No. 2022YFA1402901), NSFC (grants No. 11825403, 11991061, 12188101), and the Guangdong Major Project of the Basic and Applied Basic Research (Future functional materials under extreme conditions–2021B0301030005).

AUTHOR CONTRIBUTIONS

H.J.X. and X.G.G. supervised the project. Y.Z. and H.J.X. proposed the parameterization framework of Hamiltonian. Y.Z. coded the framework and carried out the experiments with the help of H.Y.Y. and M.S. Y.Z. and H.J.X. prepared the manuscript. All authors discussed the results and commented on 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/s41524-023-01130-4>.

Correspondence and requests for materials should be addressed to Hongjun Xiang.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023