

Supplementary Materials

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

Yang Zhong^{1,2}, Hongyu Yu^{1,2}, Xingao Gong^{1,2}, Hongjun Xiang^{1,2*}

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

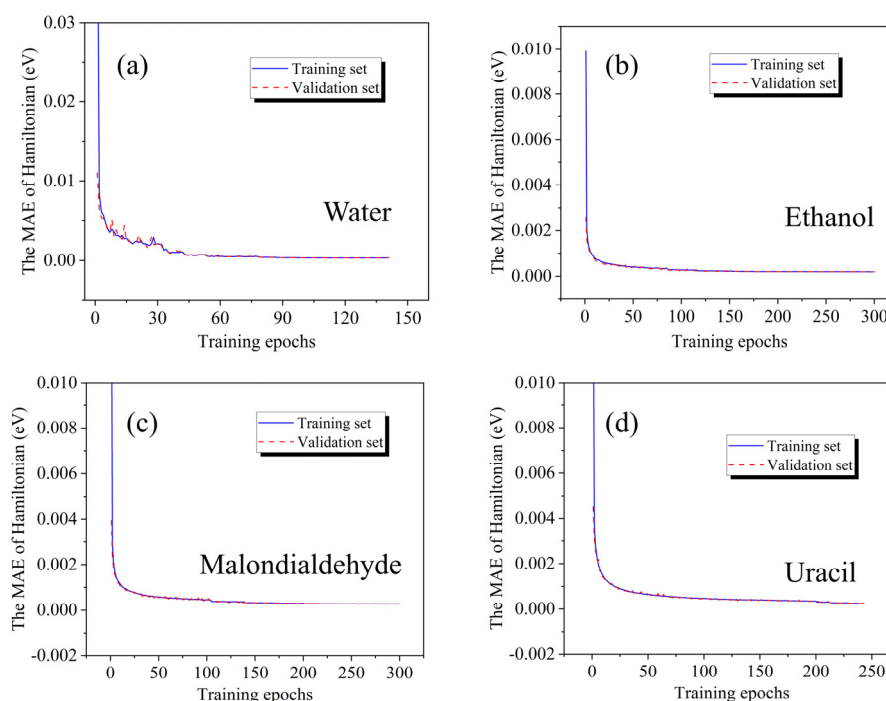
*E-mail: hxiang@fudan.edu.cn

Supplementary Table 1. The orbital energies (in eV) calculated by HamGNN and OpenMX near the band gap for four molecules randomly selected from the QM9 test set. The energy of the highest occupied molecular orbital (HOMO) is set to zero.

| Molecules | Methods | HOMO-2 | HOMO-1 | HOMO | HOMO+1 | HOMO+2 |
|---|---------|--------|--------|------|--------|--------|
| H ₉ C ₆ NO ₂ | HamGNN | -0.723 | -0.455 | 0.0 | 6.369 | 6.577 |
| | OpenMX | -0.728 | -0.450 | 0.0 | 6.350 | 6.555 |
| H ₈ C ₇ O ₂ | HamGNN | -0.602 | -0.506 | 0.0 | 5.633 | 6.042 |
| | OpenMX | -0.597 | -0.514 | 0.0 | 5.647 | 6.055 |
| H ₁₁ C ₇ NO | HamGNN | -1.905 | -0.268 | 0.0 | 3.689 | 6.279 |
| | OpenMX | -1.927 | -0.289 | 0.0 | 3.656 | 6.257 |
| H ₁₁ C ₆ NO | HamGNN | -1.617 | -1.532 | 0.0 | 6.005 | 6.076 |
| | OpenMX | -1.634 | -1.557 | 0.0 | 5.979 | 6.064 |

Supplementary Discussion 1: The analysis of the possibility of overfitting

To assess the overfitting possibility, we plot the error of HamGNN as a function of training epochs for both the training and validation sets of the molecules in Table 1. As shown in Supplementary Figure 1, the prediction errors of HamGNN on both the training and validation sets gradually decrease and eventually converge as the number of training epochs increases. The rate of error reduction for HamGNN on both the training and validation sets is highly consistent, indicating that the model does not suffer from any overfitting issues.



Supplementary Figure 1. The plot of the mean absolute errors (MAEs) of the predicted Hamiltonian matrices as a function of training epochs on the training and validation sets for water, ethanol, malondialdehyde, and uracil molecules from molecular dynamics trajectories.

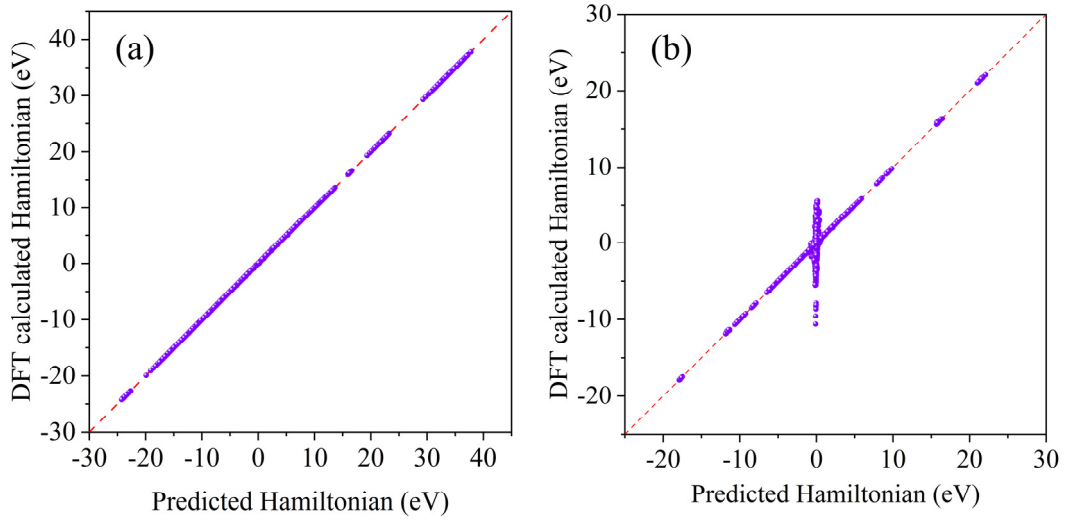
Supplementary Discussion 2: Limitations of SO(3) equivariant parametrized Hamiltonian

Based on symmetry analysis, each subblock of the tight-binding (TB) Hamiltonian matrix not only exhibits SO(3) rotational equivariance but also demonstrates a specific parity symmetry. The parameterized Hamiltonian that solely accounts for SO(3) rotational equivariance without explicitly considering parity symmetry lacks universality. PhiSNet is an equivariant network that has achieved high accuracy in predicting the Hamiltonian matrix of small molecules. However, as the parameterized Hamiltonian utilized by PhiSNet is constructed based on SO(3) irreducible representations, it does not explicitly account for the parity of the Hamiltonian matrix. Consequently, the parameterized Hamiltonian matrix constructed by PhiSNet may not accurately fit the Hamiltonian matrix of certain complex systems, such as periodic solids.

Supplementary Figure 2(a) shows the scatter plot of the Hamiltonian matrix predicted by PhiSNet versus those calculated by OpenMX on the test dataset of the water molecules from Table 1. As shown in Supplementary Figure 2(a), the scatter points are closely aligned with the diagonal line, indicating the high accuracy of PhiSNet in predicting Hamiltonian matrices for small molecules. To evaluate the accuracy of PhiSNet for periodic solids, we trained a PhiSNet model using Hamiltonian matrices computed by OpenMX for 225 graphene supercells, and subsequently evaluated its accuracy on an additional set of 45 graphene supercells. These graphene supercells were obtained through molecular dynamics (MD) simulations of a $6 \times 6 \times 1$ graphene supercell at 300K. Supplementary Figure 2(b) shows the scatter plot of the Hamiltonian matrix predicted by PhiSNet versus those calculated by OpenMX on the test dataset of graphene structures. Supplementary Figure 2(b) demonstrates that certain Hamiltonian matrix elements predicted by PhiSNet exhibit inaccuracies and consistently approach zero, resulting in a vertical band at the zero point of Supplementary Figure 2(b). This outcome is a result of PhiSNet's failure to consider the inherent parity symmetry in the Hamiltonian matrix.

PhiSNet uses the linear combinations of spherical harmonic tensors to represent each subblock of the Hamiltonian matrix: $H_{l_1 m_1, l_2 m_2} = \sum_{m_3=-l_3}^{l_3} C_{m_1, m_2, m_3}^{l_1, l_2, l_3} x_{m_3}^{l_3}$, where $C_{m_1, m_2, m_3}^{l_1, l_2, l_3}$ is the Clebsch-Gordan coefficients, $x_{m_3}^{l_3}$ is the spherical tensor of order l_3 learned by PhiSNet and satisfies SO(3) equivariance. The parity of $H_{l_1 m_1, l_2 m_2}$ is $(-1)^{l_1+l_2}$, which requires each irreducible spherical harmonic component $x_{m_3}^{l_3}$ to have the same parity with it. However, the absence of a parity constraint in each computational step during PhiSNet's forward propagation results in an uncertain assignment of parity for the learned $x_{m_3}^{l_3}$. This issue is mitigated to some extent in small molecules, where the dimension of the Hamiltonian matrix is relatively small, allowing the network parameters to be learned such that each $x_{m_3}^{l_3}$ approximately satisfies the parity symmetry constraints of the corresponding Hamiltonian block. However, when addressing periodic solid materials, the dimension of the Hamiltonian matrix significantly surpasses that of small molecules. Consequently, relying solely on training cannot guarantee

that all $\chi_{m_3}^{l_3}$ conform to the parity symmetry constraint for each sub-block of the solid system's Hamiltonian matrix. To satisfy the parity symmetry constraint of the Hamiltonian matrix, numerous spherical tensor $\chi_{m_3}^{l_3}$ becomes zero. The zero value can trivially satisfy any parity symmetry. This explains why PhiSNet predicts a substantial number of zeros in the Hamiltonian matrix for graphene supercells. Hence, it is necessary to explicitly consider the parity symmetry of the Hamiltonian matrix. This limitation significantly hinders the application of PhiSNet in solid materials.



Supplementary Figure 2. Comparison between the Hamiltonian matrix elements predicted by PhiSNet and those calculated by OpenMX on (a) the test dataset of water molecules and (b) graphene supercells.

Supplementary Discussion 3: The stability assessment of HamGNN

We trained 10 separate HamGNN models using the Hamiltonian matrices of 500 water molecules. We then evaluated the accuracy of each HamGNN model by testing it with the Hamiltonian matrices of another 500 water molecules. The MAEs obtained from these separate HamGNN models trained on the same water dataset are listed in Supplementary Table 2. Notably, the standard deviation (SD) of the MAEs from 10 repeated training results on the same water dataset is as low as 0.021 meV. This suggests that the deviation of each HamGNN model's MAE from the mean MAE is expected to be only 0.021 meV. Supplementary Table 3 lists the mean absolute errors (MAEs) of individual HamGNN models obtained from three repetitions of training on the Hamiltonian matrices of molecular datasets and solid datasets used in our study. Supplementary Table 3 reveals that HamGNN exhibits very small fluctuation in accuracy on both the molecular datasets and solid datasets, with standard deviations of MAE not exceeding 0.1 meV. These tests demonstrate the high stability of HamGNN.

Supplementary Table 2. Mean absolute errors (MAEs) (in meV) of separate HamGNN models from 10 repeated training results on the same water dataset. All the MAEs are evaluated using the same test sets.

| No. 1 | No. 2 | No. 3 | No. 4 | No. 5 | No. 6 | No. 7 | No. 8 | No. 9 | No. 10 | Average |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|---------|
| 0.748 | 0.756 | 0.743 | 0.751 | 0.743 | 0.776 | 0.754 | 0.786 | 0.805 | 0.737 | 0.760 |

Supplementary Table 3. Mean absolute errors (MAEs), average MAEs, and standard deviations (SD) of MAEs (in meV) of HamGNN from three repeated training results on the datasets of water, ethanol, malondialdehyde, uracil, carbon allotropes, silicon allotropes, and SiO₂ isomers.

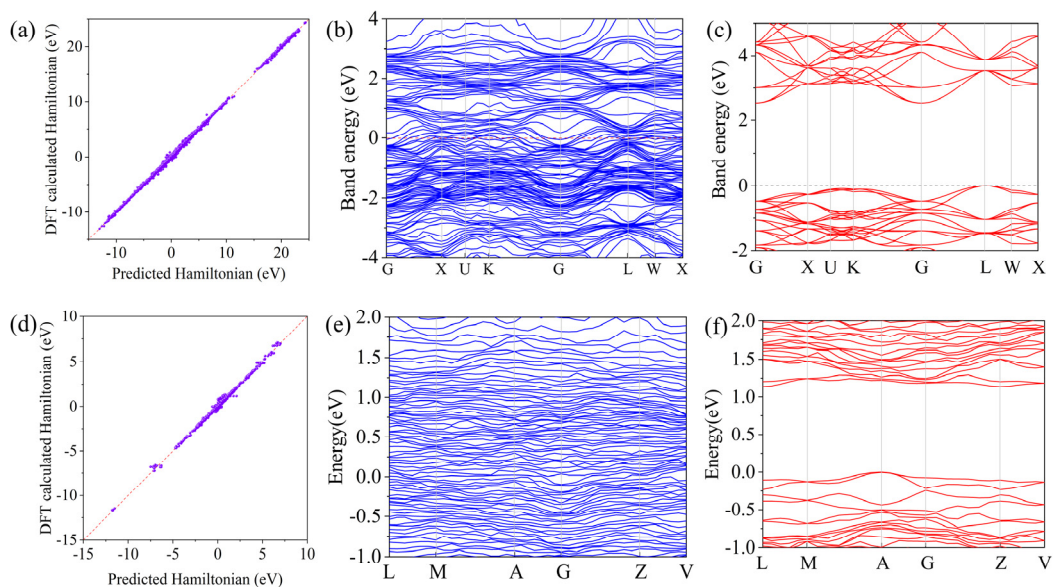
| Datasets | Training 1 | Training 2 | Training 3 | Average MAE | SD of MAEs |
|--------------------------|------------|------------|------------|-------------|------------|
| Ethanol | 0.193 | 0.204 | 0.188 | 0.195 | 0.00668 |
| Malondialdehyde | 0.283 | 0.261 | 0.267 | 0.270 | 0.00928 |
| Uracil | 0.250 | 0.302 | 0.239 | 0.264 | 0.0274 |
| Carbon allotropes | 1.550 | 1.692 | 1.541 | 1.594 | 0.0692 |
| Silicon allotropes | 2.013 | 2.134 | 2.088 | 2.078 | 0.0499 |
| SiO ₂ isomers | 2.242 | 2.216 | 2.180 | 2.213 | 0.0254 |

Supplementary Discussion 4: Transferability test for DeepH

DeepH rotates the Hamiltonian sub-blocks between two atoms in the laboratory coordinate system to matrices in the manually selected local coordinate system to deal with the equivariance issue of the Hamiltonian matrix. While a local coordinate system can be geometrically defined for two neighboring atoms based on their relative orientations with respect to the surrounding coordination atoms, defining the local coordinate system for long-range hopping between two atoms can prove very challenging. The maximum hopping range between two atoms in a crystal corresponds to the sum of the cutoff radii of the atomic orbital bases centered at those atoms. Taking the cutoff radii of hydrogen (H), sodium (Na), and potassium (K) atoms in OpenMX as examples, the cutoff radii for H, Na, and K atoms are 3.18 Å, 4.76 Å, and 5.291 Å, respectively. Consequently, the farthest hopping between two H atoms can extend up to 6.34 Å, while the farthest hopping between two K atoms can reach 10.58 Å, even though most chemical bonds typically do not exceed 3.5 Å. Establishing a well-defined local coordinate system for such long-range hopping between two atoms based on the relative orientations of chemical bonds becomes challenging. Furthermore, local coordinate systems are only applicable to a specific structure type in which they are defined. While DeepH exhibits favorable performance on perturbed structures with fixed atomic configurations, such as the graphene structures generated by molecular dynamics (MD), the accuracy of DeepH may degrade significantly when the training set contains crystals of various structure types.

We have conducted transferability tests for DeepH using the same carbon allotrope dataset and silicon allotrope dataset employed in the evaluation of HamGNN. The carbon allotrope dataset comprises 426 distinct carbon crystal structures, while the silicon allotrope dataset consists of 30 different silicon crystal structures. Each dataset was partitioned into training, validation, and test sets with ratios of 0.8, 0.1, and 0.1, respectively. On the test set of carbon allotropes, DeepH achieved a mean absolute error (MAE) of 4.41 meV for the predicted Hamiltonian matrix elements, approximately 2.8 times that of HamGNN. Supplementary Figure 3(a) illustrates the comparison between the DeepH-predicted and OpenMX-calculated Hamiltonian matrix elements. Additionally, we employed DeepH to predict the Hamiltonian matrix for the pentadiamond carbon structure (shown in Fig. 3(a)), which is not present in the training set. The resulting band structure, obtained by diagonalizing the predicted Hamiltonian matrix, is displayed in Supplementary Figure 3(b). Compared with the energy bands calculated by OpenMX (Supplementary Figure 3(c)), the energy bands predicted by DeepH do not agree with the OpenMX calculation results.

Supplementary Figure 3(d) illustrates the comparison between the Hamiltonian matrix elements predicted by DeepH and those obtained through DFT calculations. On the Si test dataset, DeepH exhibits a relatively high mean absolute error (MAE) of 11.7 meV for the predicted Hamiltonian matrix, which is approximately 5.8 times that of HamGNN. We used the silicon structure labeled MP-1199894 from the Materials Project to test the transferability of the DeepH model. A comparison of the energy bands predicted by DeepH (Supplementary Figure 3(e)) and the corresponding energy bands calculated by DFT (Supplementary Figure 3(f)) demonstrates the failure of DeepH to accurately reproduce the band structure for this crystal.

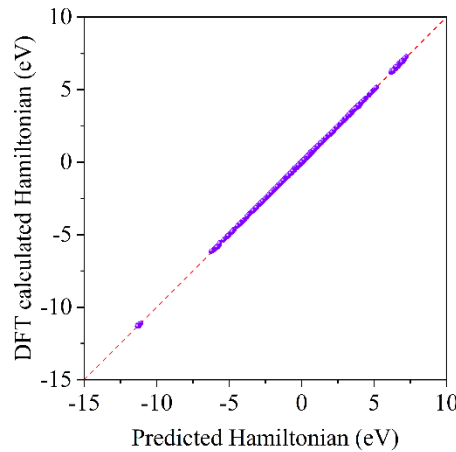


Supplementary Figure 3. Training results of DeepH on the training sets containing different structure types. (a) Comparison between the Hamiltonian matrix elements calculated by DeepH and OpenMX for the test set of carbon allotropes. (b) Predicted energy bands of pentadiamond by DeepH. (c) Energy bands of pentadiamond calculated by OpenMX. (d) Comparison between the Hamiltonian matrix elements calculated by DeepH and OpenMX for the test set of silicon allotropes. (e) Predicted energy bands of Si (MP-1199894) by DeepH. (f) Energy bands of Si (MP-1199894) calculated by OpenMX.

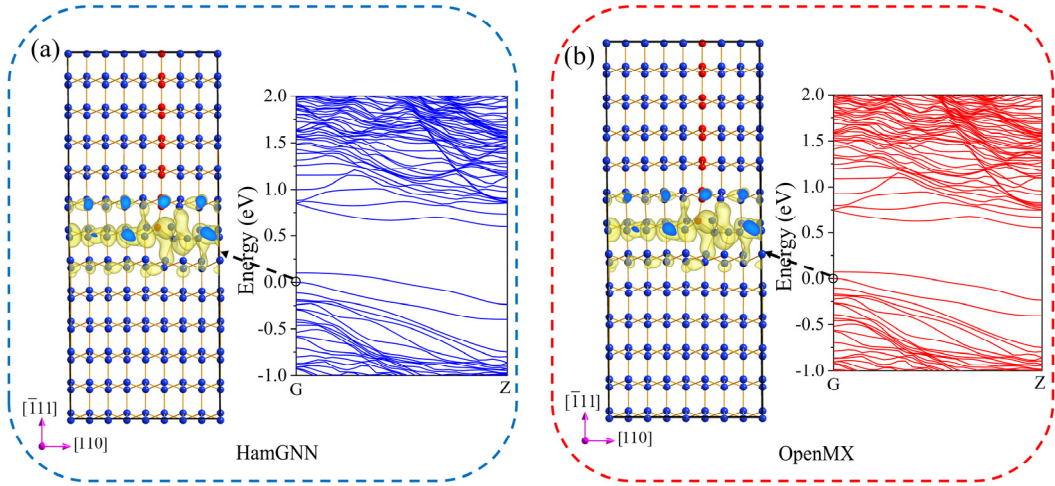
Supplementary Discussion 5: Comparison between HamGNN and DFT calculations on a small silicon supercell containing an edge dislocation

In the main text, we constructed a large supercell (containing 4284 atoms) with an isolated edge dislocation and predicted its band structure using HamGNN. To assess the accuracy of HamGNN and the impact of supercell size on the electronic structure of edge dislocations, we used both HamGNN and OpenMX to calculate the electronic structure of a small silicon supercell containing an edge dislocation. We constructed a supercell with 192 silicon atoms, which contains an edge dislocation with a Burgers vector $\mathbf{b} = \frac{1}{2}\langle 110 \rangle$. We optimized this silicon dislocation defect using GPUMD¹ with a force criterion of 0.01 Å. Supplementary Figure 4 illustrates a comparison between the Hamiltonian matrix elements calculated by HamGNN and DFT for the silicon dislocation model. The MAE of the predicted Hamiltonian matrix is only 1.03 meV. Despite being trained on only 30 perfect silicon allotropes, HamGNN demonstrates remarkable transferability by accurately predicting the Hamiltonian matrix for structures containing defects. The energy bands, obtained by diagonalizing the Hamiltonian matrix predicted by HamGNN, are shown in Supplementary Figure 5(a). The predicted valence band maximum (VBM) at the Gamma point is located on an occupied defect energy band introduced by the dislocation defect and is close to the internal bulk energy bands. The predicted VBM wave function is distributed along the glide direction of the dislocation, exhibiting distribution features similar to those of the bulk energy bands of the crystal. The accuracy of the energy bands and the VBM wavefunction predicted by HamGNN can be verified by the OpenMX calculation results shown in Supplementary Figure 5(b).

Unlike the distribution of the VBM wave function computed in the small supercell, the VBM wave function calculated in the large supercell is highly localized around the dislocation core (see Fig. 4). This test shows that the electronic structure of a dislocation defect in the small supercell is notably affected by its periodic image. Since the density of dislocations in real-world materials is very low, it is necessary to use sufficiently large dislocation models to achieve more realistic electronic structures for dislocation defects. Simulating dislocations in a large supercell using DFT methods is computationally expensive, while HamGNN can obtain the DFT-level electronic structures efficiently without self-consistent iterations.



Supplementary Figure 4. Comparison between the Hamiltonian matrix elements calculated by HamGNN and DFT for a small silicon supercell containing an edge dislocation.



Supplementary Figure 5. Comparison of the energy bands and VBM wavefunction distributions calculated by (a) HamGNN and (b) OpenMX for a small silicon supercell containing an edge dislocation.

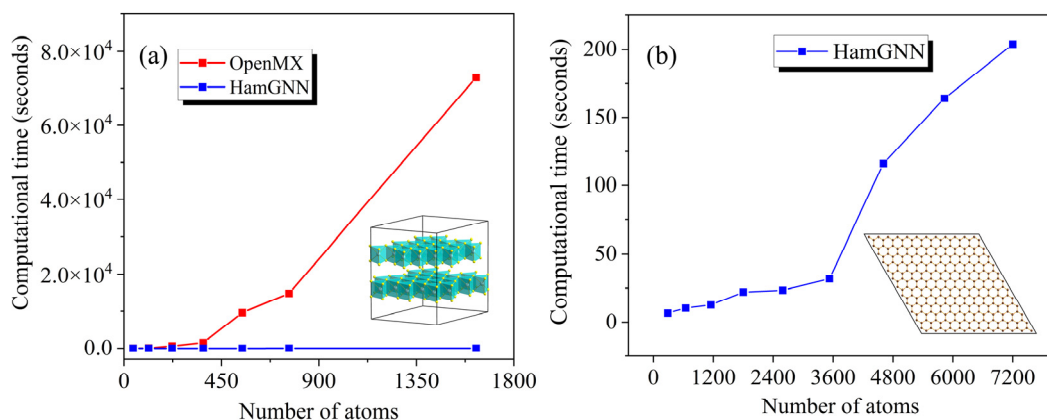
Supplementary Discussion 6: The computational efficiency of HamGNN

To compare the computational efficiency of HamGNN and DFT, we used HamGNN and OpenMX to calculate the Hamiltonian matrix for 7 MoS₂ bilayers with different twist angles. HamGNN and OpenMX performed the calculations on a node with 80 Intel(R) Xeon(R) Gold 6248 CPU cores. The computational time of HamGNN and OpenMX for each twist system is listed in Supplementary Table 4. The computational time of HamGNN on each system is much lower than the DFT methods. The curves of the computational time of DFT and HamGNN as a function of the system size are shown in Supplementary Figure 6(a). The computational time of OpenMX exhibits a rapid increase with system size, whereas that of HamGNN shows only a slow increase. The tests demonstrate that HamGNN achieves an exponential acceleration in comparison to the computational time of DFT. Generally, graph neural networks (GNNs) perform local aggregation operations on each node, such as weighted summation over neighboring nodes. The computational complexity of these operations is directly proportional to the number of neighbors of the node. Therefore, as the number of nodes or edges in the graph increases, the computational load also proportionally increases. Therefore, GNNs offer a highly scalable and efficient alternative to DFT for computing electronic Hamiltonians with nearly linear scalability.

To evaluate the computational scalability of HamGNN with system sizes, we utilized the HamGNN model that was trained on carbon allotropes to calculate Hamiltonian matrices for graphene supercells of varying dimensions. We used HamGNN to predict the Hamiltonian matrix for each graphene supercell on a computational node equipped with 80 Intel(R) Xeon(R) Gold 6248 CPU cores. The computational time for each supercell is presented in Supplementary Figure 6(b). Remarkably, HamGNN completes the prediction of the Hamiltonian matrix for a supercell containing up to 7200 atoms in just around 3 minutes, demonstrating an exceptional level of computational efficiency. The computational time of HamGNN exhibits an almost linear relationship with the system size until the number of atoms surpasses 3600. The location of the inflection point may be correlated with the hardware properties of the utilized computing system. Beyond this inflection point, the runtime of HamGNN exhibits a faster growth rate as the system size increases, while still maintaining an overall approximately linear trend.

Finally, we compare the computational cost of the HamGNN model with that of PhiSNet and DeepH models for two large graphene supercells with dimensions of $36 \times 36 \times 1$ and $48 \times 48 \times 1$. The computational time for the Hamiltonian matrices of the two graphene supercells using HamGNN, DeepH, PhiSNet, and OpenMX is presented in Supplementary Table 5. The computational time of OpenMX and DeepH is obtained from the supplementary information of ref. 2. The OpenMX calculations for the $48 \times 48 \times 1$ supercell of graphene are performed using 360 cores (10 compute nodes, each equipped with 2 Intel Xeon 6240 CPUs). The OpenMX calculations for the $36 \times 36 \times 1$ supercell of graphene are performed using 180 cores (5 compute nodes each equipped with 2 Intel Xeon 6240 CPUs). The DeepH calculations are carried out using 64 cores (one compute node equipped with 2 AMD EPYC 7542 CPUs). HamGNN and PhiSNet have performed the calculations on a computational node equipped with 80 Intel(R) Xeon(R) Gold 6248 CPU cores. Supplementary Table 5 shows that the machine learning models achieve an exponential speedup compared to DFT calculations. Among the three machine learning models, HamGNN has the lowest computation time for the large graphene supercells, while DeepH requires much more computational time than PhiSNet and HamGNN.

Although PhiSNet has a computational speed close to that of HamGNN, as mentioned earlier, it cannot accurately fit the Hamiltonian matrix of solids.



Supplementary Figure 6. Tests on the computational efficiency of HamGNN. (a) Comparison of the computational time between HamGNN and DFT on MoS₂ bilayers with varying twist angles. (b) The computational time of HamGNN on graphene supercells with different sizes.

Supplementary Table 4. The computational time (in seconds) of HamGNN and DFT on MoS₂ bilayers with varying twist angles.

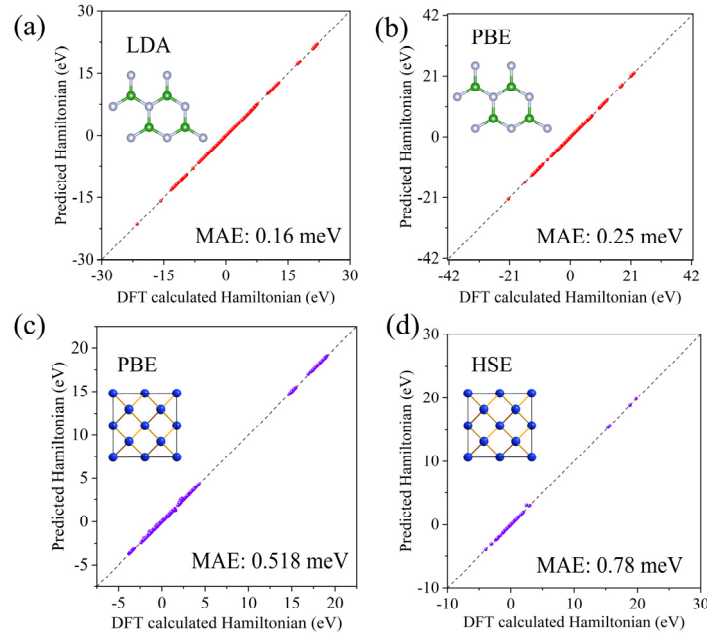
| Twist angle | Number of atoms | Computational time of OpenMX | Computational time of HamGNN |
|-------------|-----------------|------------------------------|------------------------------|
| 21.79° | 42 | 24.10 | 2.56 |
| 13.17° | 114 | 71.45 | 3.53 |
| 9.43° | 222 | 634.82 | 6.99 |
| 7.34° | 366 | 1536.62 | 9.57 |
| 6.01° | 546 | 9599.99 | 12.39 |
| 5.08° | 762 | 14816.50 | 14.10 |
| 3.5° | 1626 | 72922.25 | 20.63 |

Supplementary Table 5. Comparison of the computational time (in seconds) between DFT, DeepH, PhiSNet, and HamGNN.

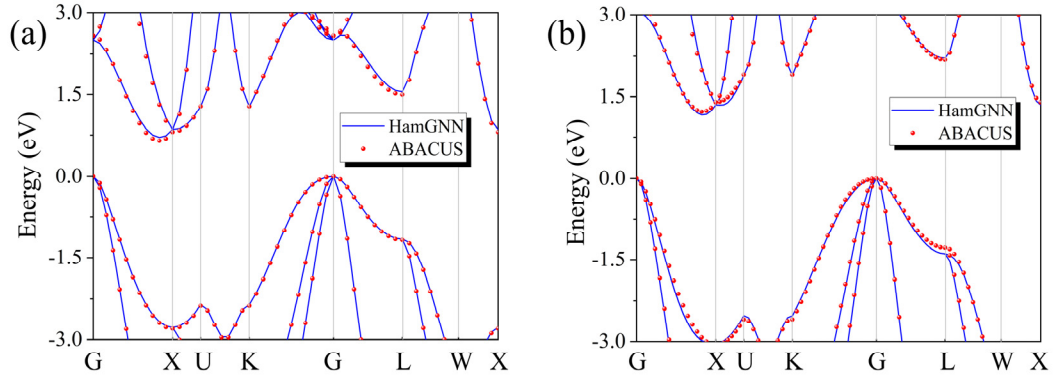
| Sizes of graphene supercells | OpenMX | DeepH | PhiSNet | HamGNN |
|------------------------------|------------|---------|---------|--------|
| 36×36×1 | 2,311,200 | 67,648 | 89.17 | 23.40 |
| 48×48×1 | 20,072,880 | 120,448 | 150.91 | 116.07 |

Supplementary Discussion 7: The impact of software and exchange-correlation functionals on the accuracy of HamGNN

The general parameterized Hamiltonian proposed in this work can be applied to all tight-binding Hamiltonian matrices based on atomic orbitals, which can be obtained through self-consistent iterations using software such as OpenMX^{3,4} or ABACUS^{5,6}. However, the method proposed in this paper can only be used to fit the tight-binding Hamiltonian matrix in real space and cannot be applied to fitting the Hamiltonian matrix in reciprocal space based on plane wave basis sets such as the Hamiltonian matrix generated by VASP^{7,8}, Abinit^{9,10}, etc. To test the accuracy of HamGNN on Hamiltonian matrices obtained by different software under various exchange-correlation functionals, we trained HamGNN using LDA (local density approximation), PBE (Perdew–Burke–Ernzerhof)¹¹, and HSE (Heyd-Scuseria-Ernzerhof)^{12,13} Hamiltonian matrices from OpenMX or ABACUS. We used OpenMX to calculate the Hamiltonian matrices under LDA and PBE functionals for 1000 monolayer BN structures generated by molecular dynamics. We used ABACUS to obtain PBE Hamiltonian matrices for 500 perturbed Si diamond structures and HSE Hamiltonian matrices for 200 perturbed Si diamond structures. The Hamiltonian matrices calculated under each exchange-correlation functional are divided into training, validation, and test sets in a ratio of 0.8:0.1:0.1, respectively. The training is carried out on the Hamiltonian data sets of each functional separately. The accuracy of the Hamiltonian matrices predicted by HamGNN on the test set of BN and diamond silicon is shown in Supplementary Figure 7. Supplementary Figure 7 demonstrates that HamGNN has powerful generalization ability and can achieve high accuracy in predicting the TB Hamiltonian matrices generated by different software under various exchange-correlation functionals.



Supplementary Figure 7. Scatter plots of (a) LDA and (b) PBE Hamiltonian matrix predicted by HamGNN versus those calculated by OpenMX for monolayer BN structures. Scatter plots of (c) PBE and (d) HSE Hamiltonian matrix predicted by HamGNN versus those calculated by ABACUS for Si diamond structures.



Supplementary Figure 8. (a) PBE and (b) HSE energy bands calculated by both HamGNN and ABACUS for the Si diamond structure.

We have predicted the PBE and HSE energy bands of diamond silicon by diagonalizing the PBE and HSE Hamiltonian matrices predicted by HamGNN. The PBE and HSE energy bands predicted by the HamGNN models are in good agreement with those of ABACUS, as shown in Supplementary Figure 8. This test shows the universality of the proposed parameterization method for the tight-binding Hamiltonian matrix. Our test also demonstrates the potential of utilizing HamGNN to accelerate computationally expensive HSE calculations.

Supplementary Table 6. The cutoff energy and K-point grid used by OpenMX in the calculation of the Hamiltonian matrices for each dataset.

| Datasets | Cutoff energy (Ry) | K points mesh |
|---|-----------------------|---------------|
| carbon allotropes | 180 | 6×6×6 |
| silicon allotropes | 150 | 5×5×5 |
| SiO ₂ isomers | 150 | 6×6×6 |
| bilayer MoS ₂ | 200 | 5×5×1 |
| Bi _x Se _y compounds | 300 | 5×5×5 |

Supplementary Table 7. The network and training parameters used for HamGNN on each dataset.

| Dataset | N_{batch} | N_{fea} | L_{max} | $r_{\text{cut}} (a_0)$ | N_k | N_{patience} | N_{stop} |
|-----------------------------------|--------------------|------------------|------------------|------------------------|-------|-----------------------|-------------------|
| QM9 | 5 | 64 | 4 | 15.0 | 1 | 10 | 30 |
| Molecules from MD trajectories | 2 | 128 | 4 | 15.0 | 0 | 10 | 30 |
| SACADA | 1 | 32 | 4 | 20.0 | 5 | 5 | 30 |
| Si | 1 | 32 | 4 | 20.0 | 5 | 5 | 30 |
| SiO ₂ | 1 | 32 | 4 | 20.0 | 5 | 5 | 30 |
| Bilayer MoS ₂ | 1 | 32 | 4 | 20.0 | 5 | 5 | 30 |
| Bi _x Se _y | 1 | 32 | 4 | 20.0 | 5 | 5 | 30 |

Supplementary References

1. Fan, Z., Chen, W., Vierimaa, V. & Harju, A. Efficient molecular dynamics simulations with many-body potentials on graphics processing units. *Comput. Phys. Comm.* **218**, 10-16 (2017).
2. Li, H. et al. Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation. *Nat. Comput. Sci.* **2**, 367-377 (2022).
3. Ozaki, T. Variationally optimized atomic orbitals for large-scale electronic structures. *Phys. Rev. B* **67**, 155108 (2003).
4. Ozaki, T. & Kino, H. Numerical atomic basis orbitals from H to Kr. *Phys. Rev. B* **69**, 195113 (2004).
5. Chen, M., Guo, G. C. & He, L. Systematically improvable optimized atomic basis sets for ab initio calculations. *J. Phys.: Condens. Matter* **22**, 445501 (2010).
6. Li, P. et al. Large-scale ab initio simulations based on systematically improvable atomic basis. *Comp. Mater. Sci.* **112**, 503-517 (2016).
7. 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).
8. Blochl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953-17979 (1994).
9. Torrent, M., Jollet, F., Bottin, F., Zerah, G. & Gonze, X. Implementation of the projector augmented-wave method in the ABINIT code: Application to the study of iron under

- pressure. *Comp. Mater. Sci.* **42**, 337-351 (2008).
10. Gonze, X. et al. ABINIT: First-principles approach to material and nanosystem properties. *Comp. Mater. Sci.* **180**, 2582-2615 (2009).
 11. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865-3868 (1996).
 12. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **118**, 8207-8215 (2003).
 13. Heyd, J., Scuseria, G. E. & Ernzerhof, M. Erratum: "Hybrid functionals based on a screened coulomb potential" [*J. Chem. Pys.* 118, 8207 (2003)]. *J. Chem. Phys.* **124**, 219906 (2006).