Equivariant Electronic Hamiltonian Prediction with Many-Body Message Passing

Chen Qian,[1] Valdas Vitartas,[1, 2] James R. Kermode,[2] and Reinhard J. Maurer[1, 3, *]

[1] *Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom*
[2] *Warwick Centre for Predictive Modelling, School of Engineering,*
University of Warwick, Coventry, CV4 7AL, United Kingdom
[3] *Department of Physics, University of Warwick, Coventry, CV4 7AL, United Kingdom*

Machine learning surrogates of Kohn-Sham Density Functional Theory Hamiltonians offer a powerful tool to accelerate the prediction of electronic properties of materials, such as electronic band structures and densities-of-states. For large-scale applications, an ideal model would exhibit high generalization ability and computational efficiency. Here, we introduce the MACE-H graph neural network, which combines high body-order message passing with a node-order expansion to efficiently obtain all relevant $O(3)$ irreducible representations. The model achieves high accuracy and computational efficiency and captures the full local chemical environment features of, currently, up to f orbital matrix interaction blocks. We demonstrate the model's accuracy and transferability on several open materials benchmark datasets of two-dimensional materials and a new dataset for bulk gold, achieving sub-meV prediction errors on matrix elements and eigenvalues across all systems. We further analyse the interplay of high body order message passing and locality that makes this model a good candidate for high-throughput material screening.

I. INTRODUCTION

Machine learning and deep learning have become central tools in computational materials science, offering new ways to accelerate simulations and discover materials [1]. A key remaining challenge is how to scale electronic structure calculations to larger time and length scales without sacrificing accuracy. Traditional first-principles methods such as Kohn–Sham density functional theory (KS-DFT) are limited by their steep computational scaling, making them impractical for large or complex systems. Decoupling the accuracy of electronic structure methods from their numerical bottlenecks is essential for enabling predictive simulations across broader domains. Machine learning methods can achieve ab initio accuracy with high computational efficiency, thereby extending applicability to systems with sizes beyond the limits of current approaches.

Machine learning interatomic potentials (MLIPs) have made significant progress in this direction [2], providing efficient and accurate models for atomic interactions [3–10]. These models are typically trained on ab initio energy and force data and have demonstrated remarkable transferability and fidelity. These advances have spurred the development of universal potentials, such as MACE-MP-0 [11] and GNoME [12]. Despite the success of MLIPs in accelerating atomistic simulations, these models fundamentally omit explicit electronic degrees of freedom. As a result, they are limited to predicting structural and thermodynamic properties, and cannot access electronic observables such as band structures, charge transport, or optical responses.

Historically, efforts to make KS-DFT more efficient have involved simplifying the treatment of electronic interactions, either by neglecting certain integrals or by imposing strong approximations on the range or body order of the effective potential. These simplifications underpin semi-empirical methods, which trade physical fidelity for computational speed. [13–15]

An alternative route to increased efficiency has been the development of linear-scaling electronic structure methods. These approaches exploit the locality of atomic orbital basis sets and the sparsity of real-space Hamiltonians. By leveraging the compactness of atom-centered representations, such methods can scale to larger systems while retaining a representation of the electronic structure. [16, 17] However, their accuracy and transferability remain constrained by the rigidity of their functional forms. These methods rely on the sparsity of the real-space Kohn-Sham (KS) Hamiltonian matrix for their efficiency, as illustrated in Fig. 1.

Broadly, ML surrogate models explored to date fall into two categories: tight-binding (TB) methods and direct KS Hamiltonian prediction. Machine learning tight-binding methods predict tight-binding parameters and transform these parameters into matrix elements by Slater-Koster formulas. Due to the scalar nature of Slater-Koster parameters, these models can be easily trained on band structures or electronic density using well-established ML methods like Gaussian process regression (GPR), multilayer perceptrons (MLPs) [18–20], and graph neural networks [21]. However, the accuracy of tight-binding methods suffers due to limited body order and parameter approximations, limiting applicability to systems dominated by delocalized interactions.

The second category of approaches uses parametric surrogate models for KS Hamiltonians directly. The first such framework was SchNOrb [22], which demonstrated the feasibility of predicting molecular Hamiltonians using deep tensor networks. Further examples including SchNet [23], CGCNN [24], [25–27] have constructed mappings from geometric information to scalar properties with high accuracy. While successful for small molecules, these

[*] r.maurer@warwick.ac.uk

models typically rely on low-order interactions and are limited in their ability to generalize to extended systems or complex environments.

Our approach in this work builds on recent advances in E(3)-equivariant neural networks, which leverage irreducible representations and tensor product operations to naturally encode rotational symmetries in atomic systems. Equivariant frameworks, such as e3nn and E3x [28, 29], can intrinsically handle the symmetry properties of Hamiltonian matrices involving orbitals of arbitrary angular momentum. Compared to invariant GNNs that rely on local coordinate systems—such as SchNOrb [22] or DeepH [30], irreps-based equivariant models like PhiSNet [31], HamGNN [32], NICE [33], and DeepH-E3 [34] offer enhanced geometric awareness and have demonstrated superior accuracy in electronic structure prediction.

While these models have successfully introduced equivariance into Hamiltonian learning, they typically rely only on two-body interactions. In contrast, our model incorporates high-body-order message passing into the learning of electronic Hamiltonians, building on the success of high-body messages seen in MLIPs and in successful linear models for Hamiltonians using ACE features [3, 35, 36]. By doing so, we capture complex many-body correlations that are essential for modeling extended systems and intricate local environments, such as those found in twisted bilayers and defected bulk materials. This combination of equivariant architecture and high-order interaction modeling enables accurate, transferable predictions of electronic structure across diverse materials classes.

Specifically, we introduce the MACE-H model, which incorporates the efficient many-body order message-passing scheme from MACE [9] into a graph neural network representation of KS DFT Hamiltonians. Taking advantage of the increased body order in the message passing phase, the model achieves high prediction accuracy and computational and data efficiency. MACE-H achieves high expressiveness and is compatible with spin-orbit coupling and multi-element systems. The performance of MACE-H was benchmarked across various datasets against the two-body equivariant message-passing GNN DeepH-E3. The datasets include 2D shifted and twisted bilayers and 3D bulk materials using metrics based on matrix elements, band structures, and density of states, such as the eigenvalue error and electronic entropy error. The latter tests show that high accuracy in matrix elements persists in downstream property predictions based on the eigenvalues. Further tests reveal the interplay between the many-body expansion of messages and the range of interactions in the model. In addition to the many-body message-passing scheme, we also employed a shift and scale operation that increases accuracy, stability and accelerates convergence. Finally, we found that the model prediction error can be estimated by the Hermiticity of the model output, which is proposed as a possible uncertainty quantification measure during active learning. While the model is designed for the prediction of KS DFT Hamiltonians, it can be more generally employed for the prediction of quantum operators in local basis representation, which will benefit high-throughput materials discovery and coupled electron-nuclear dynamics simulations.

II. RESULTS

A. The MACE-H model

The MACE-H graph neural network includes three different modules: 1) the node-wise (atom-wise) MACE message passing, 2) the node degree expansion, and 3) the edge-wise message updating. It will be shown below that the many-body expansion increases data efficiency and prediction accuracy while maintaining high expressiveness in the message passing phase. The overall workflow is illustrated in Fig. 2. The specific implementation of the three modules is explained in the following sections.

1. MACE message passing

The node-wise message passing, used to obtain the atom-wise feature representation, is implemented by the MACE block. Similar to the implementation of the MACE interatomic potential [9], the message-passing phase of atom i produces an intermediate atom feature $A_i^{(t)}$ with 2-body interaction. For this, a message function, $\tilde{A}^{(t)}_{i,kl_3m_3}$, is produced by aggregation of the atomic neighborhood $\mathcal{N}(i)$ at layer t:

$$\tilde{A}^{(t)}_{i,kl_3m_3} = \sum_{j\in\mathcal{N}(i)} \sum_{l_1m_1,l_2m_2}^{l_3m_3} R^{(t)}_{kl_1l_2l_3}(r_{ji}) C^{l_3m_3}_{l_1m_1,l_2m_2} Y^{m_1}_{l_1}(\hat{r}_{ji}) \cdot \sum_{\tilde{k}} W^{(t)}_{k\tilde{k}l_2} h^{(t-1)}_{j,\tilde{k}l_2m_2}\,, \tag{1}$$

where $R^{(t)}_{kl_1l_2l_3}$ is a learnable weight of each Clebsch-Gordan (CG) tensor product path, which is obtained by a multilayer perceptron (MLP) taking the interatomic distance r_{ji} as input, $C^{l_3m_3}_{l_1m_1,l_2m_2}$ is the CG coefficient to retain the equivariance of the CG tensor products, $Y^{m_1}_{l_1}(\hat{r}_{ji})$ is a spherical harmonic function of the interatomic vector $\hat{r}_{ji}$, where the hat indicates a unit vector. The learnable weights $W^{(t)}_{k\tilde{k}l_2}$ linearly transform hidden states $h^{(t-1)}_{j,\tilde{k}l_2m_2}$ from a previous layer. Eq. 1 is implemented via a geometric tensor product (Geo TP) following the same formalism as NequIP [7], which has fewer paths compared to the fully connected tensor product (FC TP) and thus avoids overparameterization and computational overhead. This implementation is equivalent to the depth-wise tensor product in Equiformer [37, 38]. As a special case, in the first layer, Eq. 1 is simplified with $h^{(0)}_{j,\tilde{k}l_2m_2}$ being a one-hot embedding of the element Z_j.

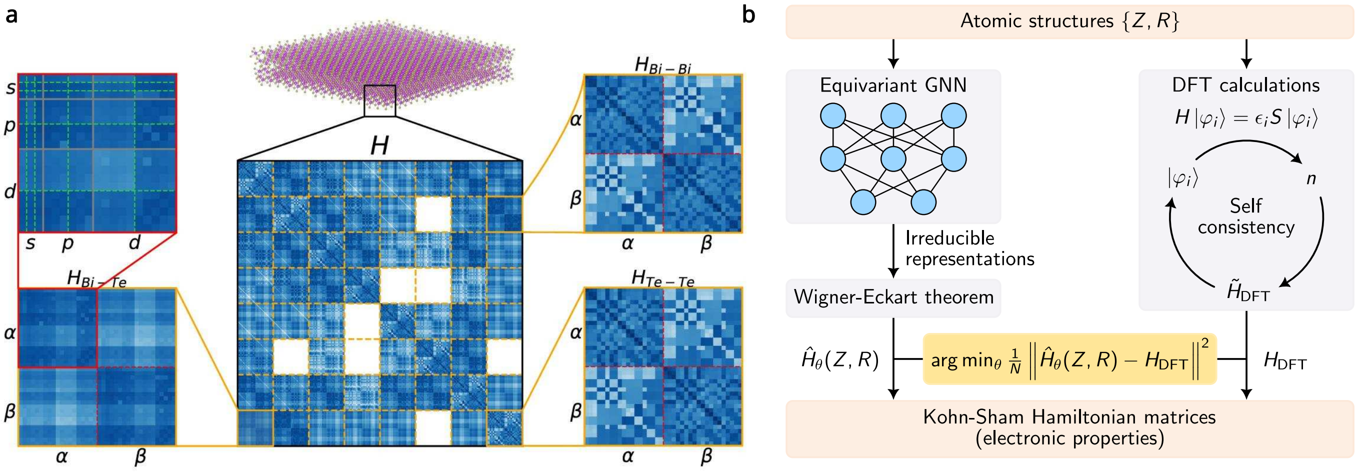

a
s
p
d
s
p
d
H_{Bi-Te}
α
β
α
β
H
H_{Bi-Bi}
α
β
α
β
H_{Te-Te}
α
β
α
β
b
Atomic structures $\{Z, R\}$
Equivariant GNN
DFT calculations
$H\left|\varphi_i\right\rangle = \epsilon_i S \left|\varphi_i\right\rangle$
$\left|\varphi_i\right\rangle$
n
Self consistency
$\tilde{H}_{\mathrm{DFT}}$
Irreducible representations
Wigner-Eckart theorem
$\hat{H}_\theta(Z,R)$
$\arg\min_\theta \frac{1}{N}\left\|\hat{H}_\theta(Z,R) - H_{\mathrm{DFT}}\right\|^2$
H_{DFT}
Kohn-Sham Hamiltonian matrices (electronic properties)

FIG. 1. **Illustration of the Hamiltonian Matrix Prediction**. **a**. The composition of Hamiltonian matrices for a twisted Bi_2Te_3 bilayer with spin-orbit coupling, the matrix can be decomposed into atom-pair-wise blocks, which consist of spin-polarized orbital-pair subblocks, **b**. The comparison of Hamiltonian matrix calculations via neural network and self-consistent field density functional theory. The training of the neural network is supervised by the labels generated from the density functional theory.

In interatomic potentials, only the energies as scalar outputs ($l = 0$) are the final prediction results. For Hamiltonian predictions, we require irreps with higher degrees ($l > 0$) to construct Hamiltonian subblocks with high numerical stability. To facilitate model convergence and numerical stability of the gradient, MACE-H directly adopts summation as the aggregation scheme without division by the average node degree (as used in MACE), and also employs the E(3) layer normalization operation in DeepH-E3 [34]:

$$\begin{aligned} A^{(t)}_{i,kl_3m_3} &= \mathrm{E3LayerNorm}\left(\tilde{A}^{(t)}_{i,kl_3m_3}\right) \\ &= W^{(t)}_{kl_3}\frac{\tilde{A}^{(t)}_{i,kl_3m_3} - \mu^{(t)}_{kl_3}}{\sigma^{(t)}_{l_3} + \varepsilon} + b^{(t)}_{kl_3}\,, \end{aligned} \tag{2}$$

where W_{kl_3} and b_{kl_3} are learnable affine parameters, but b_{kl_3} is only used for scalars. The mean values $\mu^{(t)}_{kl_3} = \frac{1}{Nn}\sum_{i=1}^{N}\sum_{k=1}^{n}\tilde{A}^{(t)}_{i,kl_3m_3}$ are calculated separately for each irrep, l_3, by averaging across the N nodes in the same graph and n channels. The standard deviation $\sigma^{(t)}_{l_3} = \sqrt{\frac{1}{Nn}\sum_{i=1}^{N}\sum_{k=1}^{n}\sum_{m_3=-l_3}^{l_3}\left\|\tilde{A}^{(t)}_{i,kl_3m_3} - \mu^{(t)}_{kl_3}\right\|^2}$ is only active for scalars in MACE-H. ε is a small number in the denominator to guarantee numerical stability.

Higher-order $(\nu+1)$-body information is introduced with the many-body expansion operation on the node-wise messages, where messages, labelled with $\eta_\nu = (l_1, \ldots, l_\nu)$, with a given correlation order ν contain irreps up to l_ν:

$$\begin{gathered} B^{(t)}_{i,\eta_\nu kLM} = \sum_{\boldsymbol{lm}}\mathcal{C}^{LM}_{\eta_\nu,\boldsymbol{lm}}\prod_{\xi=1}^{\nu}\sum_{\tilde{k}}W^{(t)}_{k\tilde{k}l_\xi}A^{(t)}_{i,\tilde{k}l_\xi m_\xi}\,, \\ \boldsymbol{lm} = (l_1m_1, \ldots, l_\nu m_\nu)\,, \end{gathered} \tag{3}$$

$$m^{(t)}_{i,kLM} = \sum_{\nu}\sum_{\eta_\nu}W^{(t)}_{Z_ikL,\eta_\nu}B^{(t)}_{i,\eta_\nu kLM}\,, \tag{4}$$

where the $W^{(t)}_{k\tilde{k}l_\xi}$ is the learnable weight for the linear transformation, the generalized CG coefficients are calculated based on the standard CG coefficients and the related indices to symmetrize the many-body features $B^{(t)}_{i,\eta_\nu kLM}$. $W^{(t)}_{Z_ikL,\eta_\nu}$ is the learnable element-dependent weight to perform the linear combination of $B^{(t)}_{i,\eta_\nu kLM}$ and obtain the feature $m^{(t+1)}_{i,kLM}$. The loop tensor contraction algorithm as proposed in MACE is employed for an efficient implementation of the many-body expansion [9].

The hidden state of each layer $h^{(t)}_{i,kLM}$ is obtained by the residual connection of the layer input, i.e., the last layer hidden state $h^{(t-1)}_{i,kLM}$, and corresponding linear transformations over the channels:

$$h^{(t)}_{i,kLM} = \sum_{\tilde{k}}W^{(t)}_{Lk\tilde{k}}m^{(t)}_{i,\tilde{k}LM} + \sum_{\tilde{k}}W^{(t)}_{Z_iLk\tilde{k}}h^{(t-1)}_{i,\tilde{k}LM}\,. \tag{5}$$

2. *Node degree expansion*

One significant characteristic of the KS Hamiltonian matrix prediction task is the presence of irreps with high angular momentum. For instance, a $f-f$ matrix subblock with spin orbit coupling will require irreps with maximum degree $l_{\max} = 9$, which is intractable for the CG tensor product in the implementation of the message function and many-body expansion. Moreover, as the generalized CG coefficients are pre-calculated with high sparsity, the size increases exponentially with the increase in angular momentum, which requires excessive floating-point operations and GPU memory. As a result, the many-body expansion operation is not straightforwardly applicable to KS Hamiltonian prediction without adaptation. To this end, we introduce the node degree expansion block

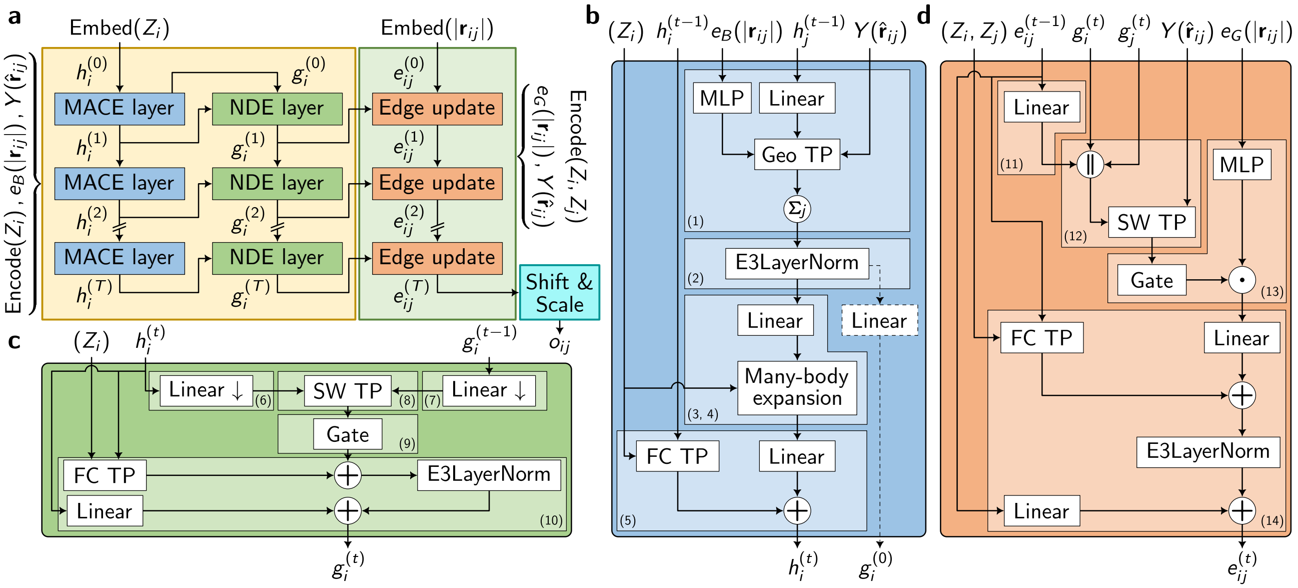


FIG. 2. **Graph Neural Network Architecture of MACE-H. a.** The overall workflow of module blocks. The node-wise message passing operations (as denoted in the yellow rectangle) consist of L message neural network layers. Encode(Z_i) and Encode(Z_i, Z_j) are one-hot-encoding of the element Z of atom i and the element pair $Z_i - Z_j$ between atom i and j, respectively. The initial node feature in MACE block and edge feature in the edge-update block in the first layer are the element embedding Embed (Z_i) and distance embedding Embed $(|\boldsymbol{r}_{ij}|)$. e_B $(|\boldsymbol{r}_{ij}|)$ and e_G $(|\boldsymbol{r}_{ij}|)$ stands for Bessel and Gaussian radial basis set, respectively. **b.** The MACE layer is used to aggregate the atom-level chemical environment feature representation, which encodes atom embedding, edge length, and orientation information to higher-body features. **c.** The node degree expansion (NDE) layer elevates the node feature degree in order to be compatible with the edge-wise irreducible representations (irreps) corresponding to the Hamiltonian subblocks. **d.** The edge update block converts the node-wise features and geometry information into the edge-wise features corresponding to the matrix block. This is identical to the edge update block in DeepH-E3 [34]. All abbreviations are defined in the main text.

to increase the degree of features with affordable node-wise operations, which bridge the angular momentum mismatch between edge-wise high-degree irreps and node-wise hidden states created by the many-body expansion.

The core implementation is the tensor product between two irreps with one being the hidden states $h_{i,kLM}^{(t)}$ from the same layer MACE block and the other being the $\tilde{A}_{i,kl_3m_3}^{(1)}$ for the first layer and the last node degree expansion layer outputs $g_{i,klm}^{(t-1)}$ for number of layers $t \geq 2$. To obtain the higher degree intermediate feature $E_{i,kL_3M_3}^{(t)}$, a separate-weight tensor product (SW TP) acts on the two inputs that are linearly transformed to achieve down-sampling in the number of channels:

$$\bar{h}_{i,k_1L_1M_1}^{(t)} = \sum_{\tilde{k}_1} W_{L_1k_1\tilde{k}_1}^{(t)} h_{i,\tilde{k}_1L_1M_1}^{(t)} , \tag{6}$$

$$\bar{g}_{i,k_2L_2M_2}^{(t-1)} = \sum_{\tilde{k}_2} W_{L_2k_2\tilde{k}_2}^{(t)} g_{i,\tilde{k}_2L_2M_2}^{(t-1)} , \tag{7}$$

$$E_{i,kL_3M_3}^{(t)} = \sum_{L_1M_1,L_2M_2}^{L_3M_3} \sum_{k_1,k_2} W_{L_1,kk_1}^{(t)} W_{L_2,kk_2}^{(t)} C_{L_1M_1,L_2M_2}^{L_3M_3} \bar{h}_{i,k_1L_1M_1}^{(t)} \bar{g}_{i,k_2L_2M_2}^{(t-1)} . \tag{8}$$

For higher nonlinearity, we adopted the gate activation for $E_{i,kL_3M_3}^{(t)}$:

$$\begin{aligned} F_i^{(t)} &= \text{Gate}(E_i^{(t)}) \qquad (9) \\ &= \left(\bigoplus_{L_1=0} \phi\left(E_{\dot{k}L_1}^{(t)}\right)\right) \oplus \left(\bigoplus_{L_2=0,\, L_3\geq 1} \varphi\left(E_{\ddot{k}L_2}^{(t)}\right) E_{\hat{k}L_3}^{(t)} \right) , \end{aligned}$$

where $\oplus$ denotes the direct sum of irreps, $E^{(t)}_{\tilde{k}L_1}$ corresponds to the scalar components of the outputs, $E^{(t)}_{\tilde{k}L_2}$ is the gating scalar, $E^{(t)}_{\tilde{k}L_3}$ is the nonscalar component of the irreps, ϕ and φ are activation functions. ϕ is the SiLU and tanh activation function for even and odd parity scalars, respectively. For φ, the sigmoid and tanh activation functions are used for gating scalars with even and odd parity, respectively.

The output $g^{(t)}_{i,kLM}$ is obtained through linear transformation and residual connection to the hidden states $g^{(t)}_{i,kLM}$:

$$g^{(t)}_{i,kLM} = \text{E3LayerNorm}\left(F^{(t)}_{i,kLM} + \sum_{\tilde{k}} W^{(t)}_{Z_i Lk\tilde{k}} h^{(t)}_{i,\tilde{k}LM}\right) + \sum_{\tilde{k}} W^{(t)}_{Lk\tilde{k}} h^{(t)}_{i,\tilde{k}LM} \,. \tag{10}$$

In this equation, NDE inputs $h^{(t)}_{i,\tilde{k}l_3m_3}$ with lower degree are padded with zeros to a higher degree LM as they have lower maximum angular momentum than the L and M created through the node degree expansion.

3. Edge update

In MACE-H, we directly implement the edge update block previously proposed as part of the DeepH-E3 model [34]. The edge-wise feature $e^{(t)}_{ij}$ in each layer is updated by the corresponding node features, edge vector, and last layer output. Intermediate edge features, $S^{(t)}_{ij}$, are created by linear transformation of the last layer edge update output $e^{(t-1)}_{ij}$. These are concatenated with the related node features $g^{(t)}_i$ and $g^{(t)}_j$ and updated features $U^{(t)}_{ij,kl_3m_3}$ are calculated with a separate-weight CG tensor product with corresponding spherical harmonics $Y(\hat{r}_{ji})$:

$$S^{(t)}_{ij,klm} = \sum_{\tilde{k}} W^{(t)}_{lk\tilde{k}} e^{(t-1)}_{ij,\tilde{k}lm} \,, \tag{11}$$

$$U^{(t)}_{ij,kl_3m_3} = \sum_{l_1m_1,l_2m_2}^{l_3m_3} \sum_{\tilde{k}} W^{(t)}_{l_1,k} W^{(t)}_{l_2,k\tilde{k}} C^{l_3m_3}_{l_1m_1,l_2m_2} Y^{m_1}_{l_1}(\hat{r}_{ji}) \cdot \cdot \left(g^{(t)}_i \middle\| \, g^{(t)}_j \middle\| \, S^{(t)}_{ij}\right)_{\tilde{k}l_2m_2} \,. \tag{12}$$

where the $e^{(0)}_{ij,klm}$ as input of the first layer is initialized by the edge distance embedding. A gate activation similar to Eq. 9 is also employed on the edge information $U^{(t)}_{ij}$ to increase model nonlinearity. Then, an element-wise multiplication is performed between each irrep and length-dependent weight to incorporate the distance information:

$$P^{(t)}_{ij,klm} = \text{Gate}\left(U^{(t)}_{ij}\right)_{klm} W^{(t)}_{kl}(r_{ji}) \,. \tag{13}$$

For the final output of the block, linear transformation, residual connection and layer normalization are employed:

$$e^{(t)}_{ij,klm} = \text{E3LayerNorm}\left(\sum_{\tilde{k}} W^{(t)}_{lk\tilde{k}} P^{(t)}_{ij,\tilde{k}lm} + \sum_{\tilde{k}} W^{(t)}_{Z_{ij}lk\tilde{k}} e^{(t-1)}_{ij,\tilde{k}lm}\right) + \sum_{\tilde{k}} \bar{W}^{(t)}_{lk\tilde{k}} e^{(t-1)}_{ij,\tilde{k}lm} \,, \tag{14}$$

where the $W_{Z_{ij}}$ are element-pair-wise learnable weights.

4. Hamiltonian output construction

Upon the model output of the corresponding irreps o_{ij,kl_3m_3} in the direct sum form, each orbital-pair-resolved subblock in tensorial form can be converted according to the Wigner-Eckart theorem using the tensor product expansion. For occasions without SOC, the conversion is:

$$H_{ij,l_1m_1l_2m_2} = \sum_{l_3=|l_1-l_2|}^{l_1+l_2} \sum_{m_3=-l_3}^{l_3} C^{l_1m_1,l_2m_2}_{l_3m_3} o_{ij,kl_3m_3} \tag{15}$$

where C is also the CG coefficient, this step can be regarded as the inverse operation of the tensor product contraction. The matrix form output represents the component with different group symmetries of the corresponding subblock. Notably, the number of channels in each segment of the edge-wise output irreps is predetermined according to the Wigner-Eckart theorem, such that all of them are used to construct the Hamiltonian matrix. Subblock matrices are then obtained as the summation of the tensor product form of the relevant irreps. For the SOC case, the relevant equations are given in Ref. [34]. The sparse overall Hamiltonian matrices are assembled from the subblocks for downstream calculations.

B. Model Validation and Benchmark

The performance of the MACE-H model was assessed by training the model separately on DFT data for a two-dimensional bilayer of Bi_2Te_3 and face-centered-cubic bulk Au. Details on the training data are provided in the Methods section. Both systems represent good benchmarks as Bi_2Te_3 features strong SOC with high geometrical complexity, and Au KS Hamiltonians calculated with all-electron DFT calculations require the model to handle many Hamiltonian blocks with high accuracy and numerical stability. To compare pseudopotential-based and

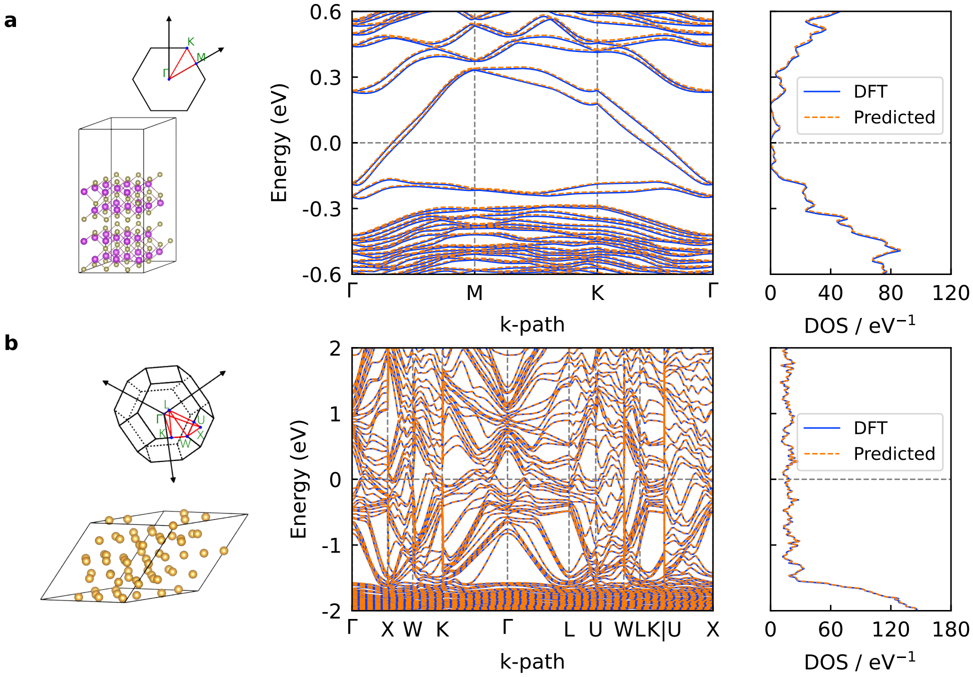


FIG. 3. **Demonstration of MACE-H for Materials Electronic Structure Prediction.** The predicted band structure along the high-symmetry k-path and the density of states in comparison with DFT results for a two-dimensional bilayer of Bi_2Te_3 (**a**) and face-centered-cubic bulk Au (**b**).

all-electron predictions on equal footing and to minimise the overhead of predicting matrix blocks associated with core states, we employ a core projection on the bulk gold data (Methods sections IV A and IV B).

MACE-H achieves an overall mean absolute error (MAE) of 0.278 meV with shift-and-scale operation for the Hamiltonian matrix elements of a hold-out test set for the shifted Bi_2Te_3 dataset with SOC. Detailed orbital pair-resolved error matrices are shown in Supplementary Fig. 1. While direct MAEs on Hamiltonian matrix elements have previously been commonly used to assess model performance in literature, we note they are not instructive in assessing the ability of the model to faithfully reproduce DFT-level band structures or densities-of-states (DOS) of materials. To quantify the performance in electronic property prediction, the eigenvalue error (EE) and electronic entropy error (EEE) metrics have been defined and employed (Methods section IV C). The corresponding values of EE and EEE for the hold-out test set Bi_2Te_3 configuration shown in Fig. 3a at a temperature of 1000 K are 14.6 meV and 7.53×10^{-8} meV·K^{-1}·Å^{-3}, respectively.

For the Au dataset, MACE-H with shift-and-scale operation reached an MAE of 0.269 meV for the Hamiltonian matrix elements in the test set. The corresponding EE and EEE errors for the test Au configuration in Fig. 3b at 1000 K are 2.5 meV and 2.19×10^{-7} meV·K^{-1}·Å^{-3}. For both systems, these are errors that produce bandstructures and DOS that are visually indistinguishable from the DFT reference calculations (see Fig. 3).

To directly compare MACE-H against the message passing model DeepH-E3, which has a similar architecture but only employs two-body messages, MACE-H and DeepH-E3 were trained on a previously published 2D materials dataset for materials with SOC and without SOC (noSOC) [30, 34]. This dataset includes DFT Kohn-Sham Hamiltonians calculated with OpenMX [39, 40] for noSOC monolayer and bilayer graphene, noSOC monolayer MoS_2, SOC bilayer bismuthene, SOC bilayer Bi_2Se_3, and SOC/noSOC bilayer Bi_2Te_3. The training data was generated by random displacements and interlayer shifting, with details explained in Methods section IV A. Moreover, we employed the same hyperparameter settings of the edge-update blocks for both the MACE-H and DeepH-E3 to illustrate the effectiveness of the higher-body messages of MACE-H, and also adopted similar settings for the message-passing process of MACE-H (with correlation order $\nu = 3$) and DeepH-E3, as hyperparameters available in Supplementary Table 1 and Section V.

On randomly split held-out test sets, MACE-H shows consistently lower MAE of matrix element prediction compared to DeepH-E3 (Fig. 4a, Supplementary Tables 2 and 3). This is the case for both noSOC and SOC data. As shown for the Bi_2Te_3 and bulk gold dataset in the Supplementary Material (Supplementary Table 4, Supplementary Figs. 2, and 3), this increase in accuracy holds independently of the correlation order ν, the choice of basis function, and the number of layers. We

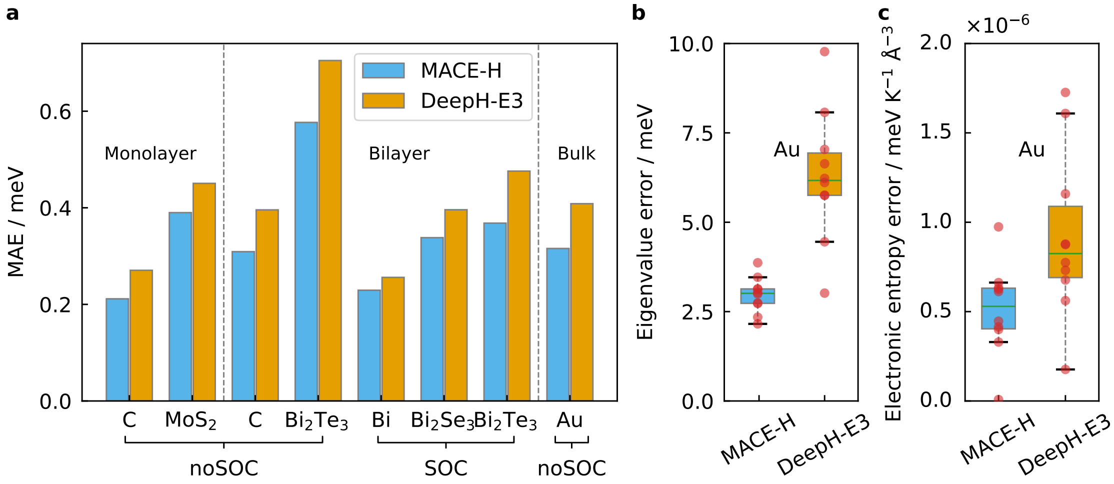


FIG. 4. **Model Performance Assessment of MACE-H Compared to DeepH-E3. a.** The test set mean absolute error of matrix elements for different datasets of 2D monolayers, shifted 2D bilayers, and bulk systems. The 2D material datasets are generated by the OpenMX package, the Au all-electron dataset is generated with FHI-aims. **b**) and **c**). The eigenvalue error and electronic entropy error for Au datasets. The eigenvalue error and electronic entropy error are defined in Section IV C. The hyperparameters are listed in Supplementary Table 8. The correlation order is $\nu = 3$.

.

attribute the improvement to the NDE block. Apart from higher nonlinearity, the NDE block relies on the tensor product between two node-level features to enrich the irreps with higher degrees or different parities, which intrinsically incorporates the 3-body messages regardless of the correlation order ν in MACE blocks. Consequently, MACE-H can, in fact, still achieve higher body messages even if $\nu = 1$, and outperforms DeepH-E3 for the shifted bilayer (see Supplementary Tables 4 and 5 for Bi_2Te_3 and Bi_2Se_3).

Improved accuracy for MACE-H compared to DeepH-E3 holds for the mono- and bilayer data and the bulk gold data. For the bulk gold system, the pristine MACE-H also achieved a lower MAE on the matrix elements of 0.316 meV, compared to 0.409 meV of DeepH-E3 (Supplementary Table 6). Further improvement of MACE-H by shift-and-scale the output irreps will be discussed in Section II E. To make a fair comparison with DeepH-E3, all figures except Fig. 3 use MACE-H models trained without the shift-and-scale operation. As assessments for downstream electronic property prediction for the bulk gold system, the average EE and EEE by MACE-H are 2.96 meV and 5.11×10^{-7} meV·K^{-1}·Å^{-3} compared to 6.28 meV and 9.17×10^{-7} meV·K^{-1}·Å^{-3} of those by DeepH-E3 for the configurations in Fig. 4b and 4c. Since we adopted the same edge update block for MACE-H as was used in DeepH-E3, we attribute the improved expressiveness to the many-body expansion in the message-passing phase of MACE-H, while DeepH-E3 only features 2-body messages.

C. Data Efficiency and Computational Cost

To evaluate the influence of the many-body expansion on data efficiency, we compared the MACE-H and DeepH-E3 models with the same setting of azimuthal numbers in irreps. Note that a MACE-H setting with $\nu = 1$ doesn't necessarily correspond to a pure 2-body message due to the NDE as explained in the last section. Fig. 5a shows that with the model depth equal to 3 layers ($T = 3$), MACE-H yields lower MAEs than DeepH-E3, and further improvement can be achieved by increasing the azimuthal number of irreps in the hidden states. Here, Deep-H uses spherical harmonics with l up to 4, MACE-H uses the same settings in the edge-update block and node-wise blocks with spherical harmonics with l up to 4, and hidden states with azimuthal number up to 2. The higher setting of MACE-H ("MACE-H high" in Fig. 5) uses spherical harmonics l up to 5 and hidden state azimuthal numbers up to 5. The correlation order ν equals 3 for both MACE-H models. With decreasing training data size, MACE-H retains a lower MAE than DeepH-E3 due to the higher expressiveness of the many-body expansion. For example, the MACE-H model with higher azimuthal numbers of irreps and trained on 20% of the data has an equivalent accuracy as DeepH-E3 trained with 40% of the data. With the higher-body-order message involved, MACE-H can accurately describe the chemical environment using fewer layers. Supplementary Fig. 4 considers different model depths. The 2-layer MACE-H provides higher accuracy and data efficiency than the 3-layer DeepH-E3. The difference between the same model with different depths decreases with the training data size. Among these, the MACE-H model with higher azimuthal numbers showed a

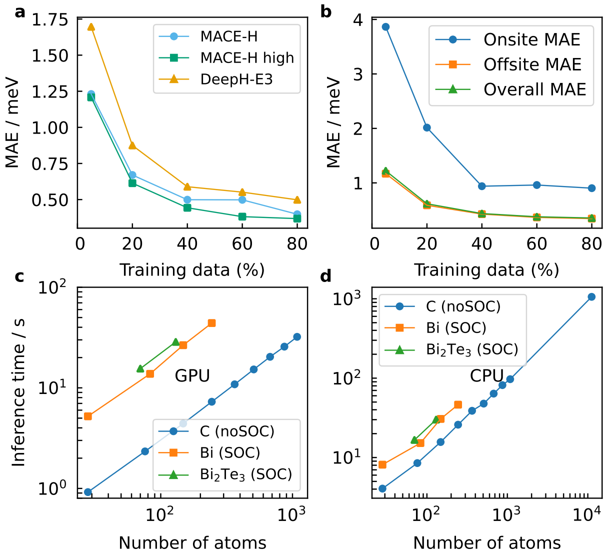


FIG. 5. **Data- and Computational Efficiency of MACE-H. a.** The matrix element MAE as a function of training data size of shifted Bi_2Te_3 for MACE-H and DeepH-E3. Here, Deep-H uses spherical harmonics with l up to 4, MACE-H uses the same settings in the edge-update block and node-wise blocks with spherical harmonics with l up to 4, and hidden states with azimuthal number up to 2. The MACE-H with notation "high" uses spherical harmonics l up to 5 and hidden state azimuthal numbers up to 5. The correlation order $\nu = 3$ for both MACE-H models. **b.** The onsite and offsite-resolved matrix element MAE as a function of Bi_2Te_3 training data size, which consists of a total of 256 configurations with 90 atoms in each. Furthermore, shown are the single batch inference time of MACE-H on a single GPU (see **c**) and 64 CPU cores (see **d**) for graphene bilayers without spin-orbital coupling and Bismuthene and Bi_2Te_3 bilayers with spin-orbital coupling.

marginal difference between the model depths of 2 and 3 across various data sizes. Furthermore, the 2-layer MACE-H with larger azimuthal numbers outperforms the 3-layer MACE-H with lower ones in terms of data efficiency and accuracy, indicating that increasing the azimuthal number of the irreps is more effective than increasing the model depth. The on-site blocks are the blocks between the same atoms in the central unit cell, while the rest are off-site blocks. On-site blocks converge faster with the number of training data provided than the off-site blocks (Fig. 5b), but they converge to larger MAE values. This is likely due to the large magnitude of Hamiltonian matrix elements on these blocks.

Fig. 5c and 5d show the inference time for different 2D material datasets using a single GPU and 64 CPU cores, respectively. The inference time of MACE-H is linear with the number of atoms in the system, manifesting the scalability to larger systems. For example, the inference time for a graphene bilayer with 1084 atoms takes 32.1 seconds on an Nvidia A100 GPU. Further acceleration on the GPU can be achieved by batched samples. MACE-H GPU inference only negligibly differs from DeepH-E3, as they share the same edge-update block, which accounts for most of the computational overhead in the model (Supplementary Fig. 5). Meanwhile, the additional many-body expansion provides no significant overhead. The inference time on a CPU is slower than on a GPU, but it is still efficient compared to outright DFT calculations. Due to current memory limitations on available GPU hardware, we perform the Hamiltonian matrix prediction for magic-angle graphene on a CPU. This takes ca. 18 minutes for a system of 11,164 atoms.

D. The Effect of Many-Body Expansion

So far, MACE-H has only been assessed for its performance on randomly held-out test set data of rattled and shifted monolayers and bilayers of two-dimensional materials and bulk gold. Following the approach of Gong *et al.* [34], we test the out-of-distribution performance of the model by performing predictions on unseen twisted bilayer configurations with varying twist angles. For the five bilayer datasets (C/graphene: noSOC, Bi_2Te_3: noSOC and SOC, Bi: soc, Bi_2Se_3) for which twisted configurations exist, MACE-H ($\nu = 3$, $L_{\max} = 5$) performs either as well or slightly worse than DeepH-E3 (Supplementary Fig. 6a). This effect is most significant for the Bi_2Te_3 dataset (with and without SOC), featuring more complex atomic environments. As a comparison, the MAE for SOC Bi_2Te_3 increases from 0.37/0.48 meV for the shifted bilayers to 1.25/0.70 meV for the twisted bilayers using MACE-H/DeepH-E3. For Bi_2Se_3, MACE-H and DeepH-E3 appear to perform similarly; however, the Bi_2Se_3 dataset is three times larger than the other datasets, indicating that larger dataset sizes are able to counteract this effect. For the twisted configurations, the many-body expansion seems to either not provide an advantage or reduce the accuracy of the model slightly, as shown in Supplementary Figs. 6a, 7, and 8. However, we note that even a matrix element MAE of 3 meV still yields band structures and DOS in the vicinity of the Fermi level that are effectively indistinguishable from the DFT reference. This effect, therefore, is acceptable in terms of model performance, but it is worth investigating further.

To investigate the difference between MACE-H and DeepH-E3 for the prediction of twisted bilayers, we analyse the models using the response to a slightly displaced atom in Bi_2Te_3 bilayer (see Fig. 6a) with the response defined as the prediction difference between the perturbed and the pristine conformations. Since the onsite matrix blocks are more straightforward to study the message passing of MACE-H with many-body expansion, we calculate the response error of the models for the onsite matrix blocks, while the response error is defined as the blockwise MAE of the difference between the model-predicted and DFT-predicted responses. In Fig. 6b, MACE-H outperforms DeepH-E3 for the shifted bilayers regardless

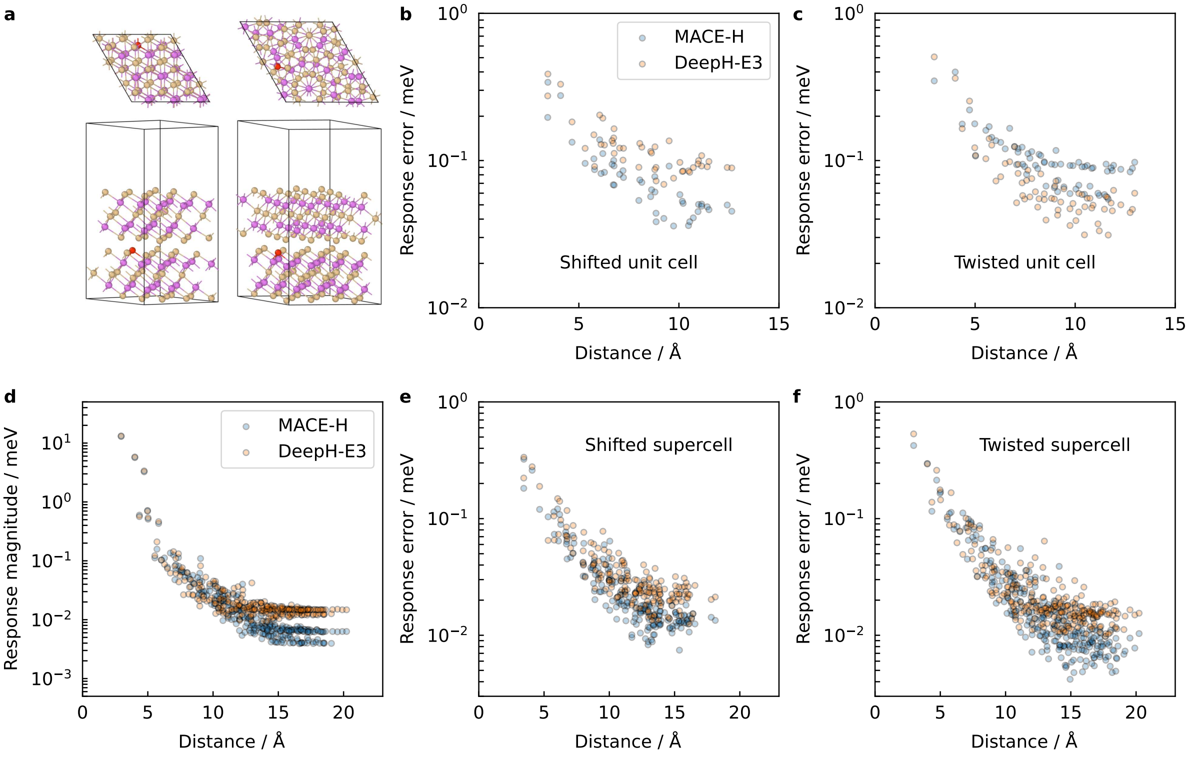


FIG. 6. **Analysis of Atom Perturbation Response of MACE-H and DeepH-E3. a.** The configuration of shifted and twisted Bi_2Te_3 bilayer unit cells with the red atom being the perturbed Te atom. **b.** The response error of onsite matrix blocks w.r.t. the distance from the perturbed atom for a shifted bilayer unit cell. **c.** The onsite response error for a twisted bilayer unit cell. **d.** The decay rate comparison for the 2×2×1 twisted supercell bilayer with varying distance using MACE-H and DeepH-E3. The onsite response error for the 2×2×1 shifted (**e**) and twisted (**f**) supercell counterparts. The onsite block components plotted in the figure are of different layers from the perturbed atoms.

of distance (the unperturbed configuration index is 72-0 in the dataset). The overall onsite response MAEs for the shifted Bi_2Te_3 bilayer are 9.6×10^{-2} and 0.13 meV for MACE-H and DeepH-E3, respectively. In contrast, MACE-H shows higher error compared to DeepH-E3 (Fig. 6b), particularly for matrix blocks further away from the perturbed atom (beyond 5 Å). The overall onsite response MAEs for the twisted Bi_2Te_3 bilayer are 0.13 and 9.8×10^{-2} meV for MACE-H and DeepH-E3, respectively. To exclude periodic image interaction, we constructed a $2 \times 2 \times 1$ supercell of the shifted and twisted Bi_2Te_3 bilayer. For the supercells (shifted and twisted), the differences between MACE-H and DeepH-E3 are in the noise of the model as the interactions decay towards large distance from the perturbed atom (Fig. 6e and f: 4.5×10^{-2} versus 4.8×10^{-2} meV for the shifted supercell and 4.14×10^{-2} versus 4.47×10^{-2} meV for the twisted supercell. Further analysis of the absolute response magnitude (Fig. 6d) shows that the decreased response error of MACE-H for the twisted supercell beyond 10 Å is a result of MACE-H attenuating faster with distance. The response magnitudes, defined as the mean absolute value of the matrix block, are 6.8×10^{-3} and 1.5×10^{-2} for the interlayer onsite blocks with distance beyond 10 Å using MACE-H and DeepH-E3, respectively. In summary, MACE-H shows a tendency for locality. In the case of the twisted structures in the original unit cells, this leads to larger errors. For the supercells, this effect is not significant and increased locality may even be an advantage. The analysis for the second twisted bilayer configuration in the dataset and its $2 \times 2 \times 1$ supercell also showed the same tendency (Supplementary Figs. 9 and 10). Our analysis also indicates that the interlayer onsite blocks showed higher sensitivity to different models than the intralayer onsite blocks (Supplementary Note 5.2).

To understand how the many-body expansion makes MACE-H more local and increases its sensitivity to shorter-range interactions, we refer back to the many-body expression (equation 3). Using the iterative node-wise tensor product of irreps, the effect of edge components with larger contributions (which are usually edges with shorter distances) in the aggregation process will be magnified, while the importance of those with lower contributions will be reduced. Supplementary Fig. 6b shows that by increasing the correlation order ν and maximum degree of hidden states $L_{\max}$, the MACE-H prediction accuracy is increased for the shifted bilayers and reduced for the twisted bilayers. On the contrary, lower correlation order (ν=1) and degree of hidden states ($L_{\max}$=1) leads to reduced accuracy for the shifted bilayers and higher accuracy for the twisted bilayers. The difference is statistically significant, and the influence is even more salient using the maximum MAE among the subblock elements as the measure (Supplementary Fig. 7). Supplementary Fig. 8 reports model performance on shifted and twisted Bi_2Te_3 as a function of correlation order ν and maximum hidden state degree $L_{\max}$, which further supports this observation. Moreover, the locality of the model is partially related to the choice of radial basis function and envelop function (Supplementary Tables 4, 5, and 7).

E. Modified Shift-and-Scale Operation

In MLIPs, the graph neural network outputs are usually shifted and scaled according to the mean value and standard deviation of the species-aware energy labels in the dataset to fit the output energy. In Hamiltonians, as values of matrix elements vary dramatically, the standard deviations within specific element-pair-resolved interaction blocks will also vary significantly. For instance, the standard deviation of different output irreps ranges from 5.42×10^{-16} to 28.33 for the Bi_2Te_3 with SOC training data. The large magnitude difference between subblocks poses a challenge for conserving the norm of the model output while maintaining numerical stability. The direct implementation of shifting and scaling the model output incurred noticeable numerical instability, as shown in Supplementary Fig. 11. This is because irreps with larger standard deviations are prone to having larger mean square errors, as they are used in the loss function, and in the back-propagation process, the gradient will be further magnified by multiplying by the standard deviation, which tends to cause gradient explosion. Conversely, the irreps with lower standard deviations will experience vanishing gradients.

We rectify the gradient flow in MACE-H by removing scaling in the back propagation step and applying scaling only in the feed-forward process (see Supplementary Note 6.1). The general formula for the scaling and shifting operation is expressed as follows:

$$o_{ij,klm} = e^{(t)}_{ij,klm}\sigma_{Z_{ij},k} + \mu_{Z_{ij},k} \tag{16}$$

where $\mu_{Z_{ij},k}$ and $\sigma_{Z_{ij},k}$ are the precalculated standard deviations and mean values of the corresponding target irreps, while the mean values will be set to zero for non-scalar vectors. More detailed discussions can be found in Supplementary Note 6.

The modified shift-and-scale operation in MACE-H improves the average and maximum test set MAE for both the shifted Bi_2Te_3 bilayer with SOC and bulk Au datasets (Figure 7) In particular, the minimum MAE between the sub-block elements decreases by more than two magnitudes for the SOC Bi_2Te_3 since values of coupling matrix elements between spin channels are significantly lower than values of spin-diagonal blocks. Above all, the resulting MAE for Bi_2Te_3 is 0.255 meV compared to 0.368 meV for MACE-H without the modified shift-and-scale implementation.

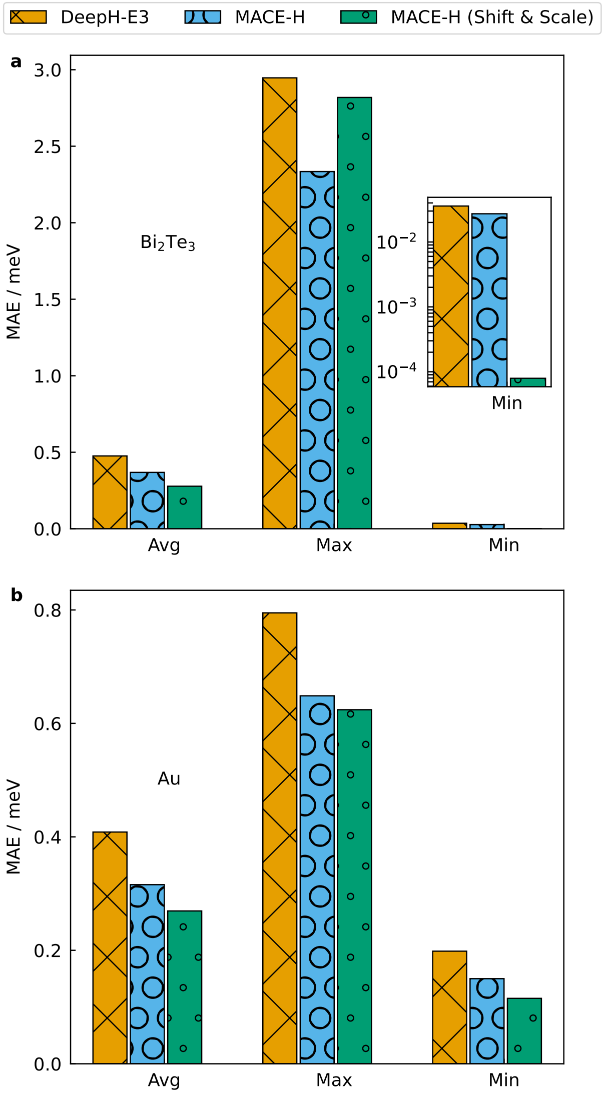


FIG. 7. **Effect of Shift-and-Scale Operation on Output Features.** The average, maximum, and minimum MAE of matrix elements for MACE-H with/without scale-and-shift operation compared to DeepH-E3 for shifted Bi_2Te_3 bilayer data (see **a**) and bulk Au (see **b**). The inset in panel **a** shows the magnified view of the minimum value of the error matrix.

F. Label-Free Accuracy Estimation of the Predicted Matrices

The MACE-H model does not strictly impose hermiticity of the real-space KS Hamiltonian matrix, but it is rather imposed post-prediction through symmetrisation. The degree to which the model violates hermiticity can

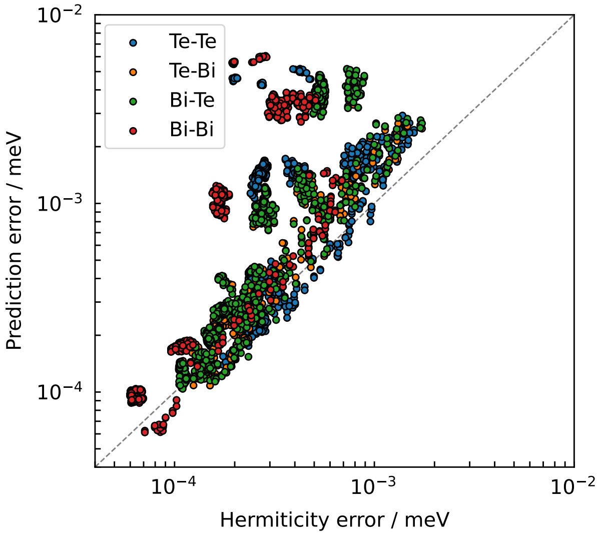


FIG. 8. **Correlation of Predicted Matrix Block Accuracy with Hermiticity of the Predicted Matrix.** The correlation between the absolute value of matrix element prediction error and the hermiticity in the twisted Bi_2Te_3 configuration resolved for different element-pairs.

be evaluated by the self-adjointness of each atom pair:

$$\mathrm{Error_{herm}} = \left\| H_{ij,\boldsymbol{R}} - H^{\dagger}_{ji,-\boldsymbol{R}} \right\| \quad (17)$$

where $H_{ij,\boldsymbol{R}}$ is the real-space Hamiltonian block between i and j with lattice translation vector $\boldsymbol{R}$, and $H^{\dagger}_{ji,-\boldsymbol{R}}$ is the conjugate transpose of the real-space Hamiltonian block.

Fig. 8 compares the block-wise prediction error to the hermiticity error, which shows a positive correlation. Subblocks with high prediction error are also likely to violate hermiticity. We can consider the prediction of the two hermitian conjugate blocks of the matrix as an in-built ensemble of two, as the two subblocks will be exposed to different stochastically optimised parameter blocks in the model. As such, the degree by which hermiticity is violated could serve as a computationally efficient metric in active learning workflows to identify subblocks with large errors without recourse to DFT ground truth labels.

III. DISCUSSION

In this work, we propose the MACE-H model, which combines many-body message passing with an equivariant graph neural network for predicting the DFT Kohn-Sham Hamiltonian matrix and, subsequently, electronic properties such as the band structure and electronic DOS. This model is generally suitable for extended materials with diverse chemical composition, as is shown here for a range of two-dimensional and bulk materials. The model achieves ab-initio-level accuracy across a range of materials datasets. The use of many-body messages and a node degree expansion block improves the prediction accuracy compared to GNNs that employ strictly two-body messages, such as DeepH-E3. Model validation using eigenvalue error and electron entropy error metrics also confirmed that the band structure derived from the predicted real-space KS Hamiltonian matrices is on par with DFT calculations. The use of many-body message passing leads to increased data efficiency of the models at no significant computational overhead for inference or training. We furthermore implement a modified shift-and-scale operation on individual output irreducible representations of the model that improves numerical stability and reduces noise in matrix subblocks with small values. Finally, the degree with which the model satisfies hermiticity can be used as an approximate gauge of the accuracy of different subblocks, as it is found to correlate with the prediction errors.

For the case of bulk gold predictions, we show that the MACE-H model is compatible with data generated by all-electron DFT packages such as FHI-aims. Firstly, KS DFT Hamiltonians can be pre-processed with a core projection to generate valence-only Hamiltonians that serve as training data. Secondly, accurate atomic orbital basis evaluations require a large cutoff radius that will lead to low sparsity in the Hamiltonian and matrix subblocks that span many orders of magnitude. Subblocks with very small values can lead to stability issues and background noise. To this end, sparsity can be enforced by truncation of matrix values below a threshold, and a modified shift-and-scale operation ensures smooth and stable training.

The model is already fully integrated into the DeepH data pipeline and is able to train directly on FHI-aims output data [41]. In the future, we plan to further improve this integration to harness the highly optimised parallel matrix algebra in FHI-aims and other codes for the diagonalisation of large KS Hamiltonian matrices, which forms the natural bottleneck in electronic structure predictions. The model can also be applied to the prediction of the density matrix in local orbital product basis representation, which offers the possibility to accelerate self-consistent-field convergence [22, 42].

Even when a deep integration with electronic structure software is achieved for Hamiltonian prediction message passing models, further implementation improvements will be required to improve the scalability and parallelisation as well as the memory utilisation of the model during training and prediction. This will unlock new electronic structure prediction capabilities for dynamics and high-throughput materials discovery as well as improve the transferability of materials surrogate models by retaining electronic structure information.

IV. METHODS

A. Training data

2D materials data: The Hamiltonian data are calculated using OpenMX [39, 40] and taken from Refs. [34], [30]. The monolayer graphene and MoS_2 configurations are generated from the ab initio molecular dynamics trajectory using the Vienna ab initio simulation package (VASP) [43] with PBE functional [44] and the projector-augmented wave (PAW) potential [45, 46], containing 450 and 500 supercells, respectively. The shifted bilayer graphene, bismuthene, Bi_2Se_3, and Bi_2Te_3 are constructed by shifting two relaxed neighboring van der Waals layers horizontally and imposing random perturbations below 0.1 Å along each direction, containing 300, 576, 576, and 256 geometries, respectively. Noteably, the Bi_2Te_3 dataset contains two different versions, one with SOC and one without SOC. As a hold out assessment, twisted bilayers are constructed by twisting between different layers for graphene, bismuthene, Bi_2Se_3, and Bi_2Te_3, containing 9, 4, 2, and 2 geometries, respectively. More detailed information on the 2D material datasets is summarised in Supplementary Table 8 and can also be found in ref. [30], [34]. The overlap matrices used for the band structure calculation of Bi_2Te_3 are calculated by OpenMX (version 3.9) with Bi8.0-s3p2d2 and Te7.0-s3p2d2 localized pseudo-atomic orbital (PAO) basis. The hyperparameter configuration files for the datasets can be found on Zenodo [47].

For the locality analysis of Bi_2Te_3, configuration "72_0" was used as a representative structure from the test set containing shifted structures and configurations "1-2" and "1-3" were taken to represent twisted bilayer structures. The Hamiltonians of unperturbed and perturbed unit cells and $2 \times 2 \times 1$ supercells of these configurations were calculated with OpenMX (version 3.9) using the same settings as described above. The perturbed atom is subjected to displacement by 0.1 Å in three directions simultaneously.

Gold data: An unlabelled dataset of 1000 Au(FCC) 4x4x4 supercell configurations was generated by performing Langevin molecular dynamics (MD) simulations with an Effective Medium Theory (EMT) potential provided in ASE 3.22.1 [48] at 1000 K for 100 ps and saving frames every 100 fs. From these configurations, the 200 most diverse datapoints were selected by Farthest Point Sampling (FPS) with configuration-averaged ACE descriptors [3, 49, 50]. Ten configurations were randomly selected for testing, and the remaining 190 configurations we randomly split into training and validation sets with a ratio of 4:1.

Hamiltonian and overlap matrices were obtained by performing DFT calculations with the FHI-aims electronic structure code (version 240926, commit df50e6022) [51] using the PBE functional [44] and the frozen-core approximation [52]. FHI-aims provides a range of numerical atomic basis sets and real-space integration grid settings for each chemical element. For more information, see [51]. A customised "tight" basis set was used for gold, which was modified by removing a subset of extra valence orbitals, resulting in a *tier1-sdfgh* basis set with a 6 Å radial cutoff. Additionally, radial and angular densities of the real-space integration grid were increased beyond "tight" settings. The calculations were performed with a $7 \times 7 \times 7$ k-point grid. Raw calculation data with further details on computational settings can be found on the NOMAD materials repository [53].

The core states were projected from DFT data as outlined in Section IV B. Furthermore, all the Hamiltonian and overlap blocks with an edge distance above 10 Å were set to zero, even though the maximum edge distance in unprojected data was equal to 12 Å arising from the basis function radial cutoff of 6 Å in DFT calculations. We found that keeping the edges up to 12 Å significantly slowed down both training and inference, even though edges beyond 10 Å were found to only marginally affect Hamiltonian eigenvalues as discussed in the Supplementary Note 8. Furthermore, the performance of the model deteriorated for edges beyond 10 Å because such edges correspond to very small absolute Hamiltonian values, which pose a challenge for the current model architecture. We plan to modify the architecture of the model in the future to better tackle systems where truncation of long edges may not be justified.

Finally, the valence-only Au data was converted to a DeepH-compatible format, which can be found on Zenodo [47].

B. Core-Orbital Projection

To speed up training and inference without sacrificing accuracy in the valence-energy region, the low-energy core states of gold were projected from the DFT data. This was achieved by first mapping the real-space Hamiltonian to reciprocal space on a dense k-point grid:

$$\boldsymbol{H}_{\boldsymbol{k}} = \sum_{\boldsymbol{R}} \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \boldsymbol{H}_{\boldsymbol{R}} \,, \tag{18}$$

where $\boldsymbol{H}_{\boldsymbol{R}}$ is the real-space Hamiltonian between the orbitals in the central unit cell and the unit cell at $\boldsymbol{R}$. The reciprocal-space Hamiltonians were then partitioned into $\boldsymbol{H}_{\boldsymbol{k},\mathrm{cc}}$, $\boldsymbol{H}_{\boldsymbol{k},\mathrm{vv}}$, and $\boldsymbol{H}_{\boldsymbol{k},\mathrm{cv}}$, corresponding to core-core, valence-valence, and core-valence blocks, respectively. An equivariance-preserving linear basis transformation $\bar{\boldsymbol{B}}_{\boldsymbol{k}}$ was performed to minimise core-valence coupling [54]:

$$\boldsymbol{H}_{\boldsymbol{k}} = \begin{pmatrix} \boldsymbol{H}_{\boldsymbol{k},\mathrm{cc}} & \boldsymbol{H}_{\boldsymbol{k},\mathrm{cv}} \\ \boldsymbol{H}^{\dagger}_{\boldsymbol{k},\mathrm{cv}} & \boldsymbol{H}_{\boldsymbol{k},\mathrm{vv}} \end{pmatrix} \xrightarrow[\bar{\boldsymbol{H}}_{\boldsymbol{k}} = \bar{\boldsymbol{B}}^{\dagger}_{\boldsymbol{k}}\, \boldsymbol{H}_{\boldsymbol{k}}\, \bar{\boldsymbol{B}}_{\boldsymbol{k}}]{\bar{\boldsymbol{B}} = \begin{pmatrix} \boldsymbol{S}^{-\frac{1}{2}}_{\boldsymbol{k},\mathrm{cc}} & -\boldsymbol{S}^{-1}_{\boldsymbol{k},\mathrm{cc}}\boldsymbol{S}_{\boldsymbol{k},\mathrm{cv}} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}} \bar{\boldsymbol{H}}_{\boldsymbol{k}} \approx \begin{pmatrix} \bar{\boldsymbol{H}}_{\boldsymbol{k},\mathrm{cc}} & \boldsymbol{0} \\ \boldsymbol{0} & \bar{\boldsymbol{H}}_{\boldsymbol{k},\mathrm{vv}} \end{pmatrix}, \tag{19}$$

where $\boldsymbol{S}_{\boldsymbol{k}}$ is the reciprocal-space overlap matrix with the same partitioning, and $\boldsymbol{I}$ is the identity matrix. The valence-only reciprocal-space Hamiltonians $\bar{\boldsymbol{H}}_{\boldsymbol{k},\mathrm{vv}}$ were mapped back to real space:

$$\bar{\boldsymbol{H}}_{\boldsymbol{R},\mathrm{vv}} = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{R}}\, \bar{\boldsymbol{H}}_{\boldsymbol{k},\mathrm{vv}} \,. \tag{20}$$

For the gold dataset, it was chosen to only include 5d, 6s, and 6p orbitals in the valence partition, reducing the dimension of matrix blocks from 43×43 to 9×9. The resulting valence-only Hamiltonians $\bar{\boldsymbol{H}}_{\boldsymbol{R},\mathrm{vv}}$ were used as ground-truth labels for gold in this work.

Although not strictly required for training the model, the same projection procedure was performed for overlap matrices to obtain the single-particle energies by solving the generalised eigenvalue problem:

$$\bar{\boldsymbol{H}}_{\boldsymbol{k},\mathrm{vv}}\bar{\boldsymbol{C}}_{\boldsymbol{k},\mathrm{vv}} = \bar{\boldsymbol{S}}_{\boldsymbol{k},\mathrm{vv}}\bar{\boldsymbol{C}}_{\boldsymbol{k},\mathrm{vv}}\boldsymbol{\epsilon}_{\boldsymbol{k},\mathrm{vv}} \,, \tag{21}$$

where $\bar{\boldsymbol{H}}_{\boldsymbol{k},\mathrm{vv}}$ and $\bar{\boldsymbol{S}}_{\boldsymbol{k},\mathrm{vv}}$ correspond to valence-only reciprocal-space Hamiltonian and overlap matrices in the transformed basis, respectively, and $\bar{\boldsymbol{C}}_{\boldsymbol{k},\mathrm{vv}}$ is a matrix containing eigenvectors in each column and $\boldsymbol{\epsilon}_{\boldsymbol{k},\mathrm{vv}}$ corresponds to a matrix which contains the resulting valence-only eigenvalues on the diagonal.

Further details, such as the effect of core projection on valence eigenvalues and the projection convergence, can be found in Supplementary Note 8.

C. Accuracy Metrics

In order to investigate model performance beyond Hamiltonian MAE, two additional physically inspired error metrics based on Hamiltonian eigenvalues were used.

To quantify the accuracy in the occupied eigenvalue region by ignoring less physically relevant high-energy states, a smeared eigenvalue error was defined as:

$$\Delta\epsilon(T) = \frac{\sum_{\boldsymbol{k}i} |\epsilon_{\boldsymbol{k}i} - \tilde{\epsilon}_{\boldsymbol{k}i}|\, w_{\boldsymbol{k}} f(\epsilon_{\boldsymbol{k}i}, \mu, T)}{\sum_{\boldsymbol{k}i} w_{\boldsymbol{k}} f(\epsilon_{\boldsymbol{k}i}, \mu, T)} \,, \tag{22}$$

where $w_{\boldsymbol{k}}$ is the weight of the k-point $\boldsymbol{k}$ on a dense k-point grid, and $\epsilon_{\boldsymbol{k}i}$ and $\tilde{\epsilon}_{\boldsymbol{k}i}$ denote DFT and predicted eigenvalues of the electronic band i, respectively. The denominator ensures the error is normalised with respect to the number of electrons in the unit cell. The chemical potential $\mu(T)$ required for the Fermi-Dirac distribution $f(E, \mu, T) = (1 + \exp\left[(E - \mu) / k_{\mathrm{B}}T\right])^{-1}$ at a given temperature T was obtained by enforcing charge neutrality via:

$$N_{\mathrm{el}} = \int_{-\infty}^{+\infty} \mathrm{DOS}(E)\, f(E, \mu, T)\, \mathrm{d}E \,, \tag{23}$$

where N_{el} is the number of electrons in the unit cell, k_{B} is the Boltzmann constant, and the same $\mathrm{DOS}(E) = \frac{1}{V}\sum_{\boldsymbol{k}i} \delta(E - \epsilon_{\boldsymbol{k}i})$ was used at any chosen temperature, with V being the volume of the unit cell.

To quantify how well the models predict eigenvalues close to the Fermi level, it was chosen to compare DFT and predicted electronic entropy densities:

$$s(T) = -k_{\mathrm{B}} \int_{-\infty}^{+\infty} \mathrm{DOS}(E)\, \left[f \ln f + (1 - f) \ln(1 - f)\right] \mathrm{d}E \,, \tag{24}$$

where $f \equiv f(E, \mu, T)$ is the Fermi-Dirac distribution, and the $\mathrm{DOS}(E)$ is assumed to be temperature-independent as above. As before, chemical potentials for both DFT and predicted electronic entropies were estimated from charge neutrality. The electronic entropy error is reported as the absolute error between DFT and predicted values.

V. DATA AVAILABILITY

The raw Au dataset is available in NOMAD (`https://doi.org/10.17172/NOMAD/2025.04.14-1`). The valence-only Au dataset, training configuration files to reproduce the models and the corresponding Python environment containers are available on Zenodo (`https://doi.org/10.5281/zenodo.15223696`). The 2D material datasets are available on Zenodo (`https://zenodo.org/records/7553640,https://doi.org/10.5281/zenodo.7553827,https://doi.org/10.5281/zenodo.7553843`) as released by ref. [30, 34].

VI. CODE AVAILABILITY

The MACE-H code in the current paper is available on GitHub: github.com/maurergroup/MACE-H.

[1] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, Combining machine learning and computational chemistry for predictive insights into chemical systems, Chem. Rev. **121**, 9816 (2021).

[2] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Machine learning force fields, Chem. Rev. **121**, 10142 (2021).

[3] R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, Phys. Rev. B **99**, 014104 (2019).

[4] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, Graph networks as a universal machine learning framework for molecules and crystals, Chem. Mater. **31**, 3564 (2019).

[5] D. Zhang, H. Bi, F.-Z. Dai, W. Jiang, L. Zhang, and H. Wang, Dpa-1: Pretraining of attention-based deep potential model for molecular simulation, arXiv preprint arXiv:2208.08236 (2022).

[6] J. Zeng, D. Zhang, D. Lu, P. Mo, Z. Li, Y. Chen, M. Rynik, L. Huang, Z. Li, S. Shi, *et al.*, Deepmd-kit v2: A software package for deep potential models, J. Chem. Phys. **159** (2023).

[7] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, Nat. Commun. **13**, 2453 (2022).

[8] A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth, and B. Kozinsky, Learning local equivariant representations for large-scale atomistic dynamics, Nat. Commun. **14**, 579 (2023).

[9] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csanyi, Mace: Higher order equivariant message passing neural networks for fast and accurate force fields, in *Advances in Neural Information Processing Systems*, Vol. 35, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, Inc., 2022) pp. 11423–11436.

[10] J. T. Frank, O. T. Unke, K.-R. Müller, and S. Chmiela, A euclidean transformer for fast and stable machine learned force fields, Nat. Commun. **15**, 6539 (2024).

[11] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, *et al.*, A foundation model for atomistic materials chemistry, arXiv preprint arXiv:2401.00096 (2023).

[12] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, Scaling deep learning for materials discovery, Nature **624**, 80 (2023).

[13] G. Seifert, Tight-binding density functional theory: an approximate Kohn-Sham DFT scheme., J. Phys. Chem. A **111**, 5609 (2007).

[14] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irle, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W.-z. Yu, and T. Frauenheim, DFTB+, a software package for efficient approximate density functional theory based atomistic simulations, J. Chem. Phys. **152**, 124101 (2020).

[15] P. O. Dral, B. Hourahine, and S. Grimme, Modern semiempirical electronic structure methods, J. Chem. Phys. **160**, 040401 (2024).

[16] S. Goedecker, Linear scaling electronic structure methods, Rev. Mod. Phys. **71**, 1085 (1999).

[17] W. Dawson, A. Degomme, M. Stella, T. Nakajima, L. E. Ratcliff, and L. Genovese, Density functional theory calculations of large systems: Interplay between fragments, observables, and computational complexity, WIREs Comp. Mol. Sci. **12**, e1574 (2022).

[18] C. Schattauer, M. Todorović, K. Ghosh, P. Rinke, and F. Libisch, Machine learning sparse tight-binding parameters for defects, npj Comput. Mater. **8**, 116 (2022).

[19] H. Li, C. Collins, M. Tanha, G. J. Gordon, and D. J. Yaron, A Density Functional Tight Binding Layer for Deep Learning of Chemical Hamiltonians, J. Chem. Theory Comput. **14**, 5764 (2018).

[20] A. McSloy, G. Fan, W. Sun, C. Hölzer, M. Friede, S. Ehlert, N.-E. Schütte, S. Grimme, T. Frauenheim, and B. Aradi, TBMaLT, a flexible toolkit for combining tight-binding and machine learning, J. Chem. Phys. **158**, 034801 (2023).

[21] Q. Gu, Z. Zhouyin, S. K. Pandey, P. Zhang, L. Zhang, and W. E, Deep learning tight-binding approach for large-scale electronic simulations at finite temperatures with ab initio accuracy, Nat. Commun. **15**, 6772 (2024).

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

[23] K. Schütt, P.-J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller, Schnet: A continuous-filter convolutional neural network for modeling quantum interactions, Advances in neural information processing systems **30** (2017).

[24] S. Sanyal, J. Balachandran, N. Yadati, A. Kumar, P. Rajagopalan, S. Sanyal, and P. Talukdar, Mt-cgcnn: Integrating crystal graph convolutional neural network with multitask learning for material property prediction, arXiv preprint arXiv:1811.05660 (2018).

[25] Y. Lin, K. Yan, Y. Luo, Y. Liu, X. Qian, and S. Ji, Efficient approximations of complete interatomic potentials for crystal property prediction, in *International Conference on Machine Learning* (PMLR, 2023) pp. 21260–21287.

[26] K. Yan, Y. Liu, Y. Lin, and S. Ji, Periodic graph transformers for crystal material property prediction, Advances in Neural Information Processing Systems **35**, 15066 (2022).

[27] Y. Liu, L. Wang, M. Liu, X. Zhang, B. Oztekin, and S. Ji, Spherical message passing for 3d graph networks, arXiv preprint arXiv:2102.05013 (2021).

[28] M. Geiger and T. Smidt, e3nn: Euclidean neural networks, arXiv preprint arXiv:2207.09453 (2022).

[29] O. T. Unke and H. Maennel, E3x: E(3)-equivariant deep learning made easy, arXiv preprint arXiv:2401.07595 (2024).

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

[31] O. Unke, M. Bogojeski, M. Gastegger, M. Geiger, T. Smidt, and K.-R. Müller, Se (3)-equivariant prediction of molecular wavefunctions and electronic densities, Advances in Neural Information Processing Systems **34**, 14434 (2021).

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

[33] J. Nigam, M. J. Willatt, and M. Ceriotti, Equivariant representations for molecular Hamiltonians and N-center atomic-scale properties, J. Chem. Phys. **156**, 014115 (2022).

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

[35] L. Zhang, B. Onat, G. Dusson, A. McSloy, G. Anand, R. J. Maurer, C. Ortner, and J. R. Kermode, Equivariant analytical mapping of first principles hamiltonians to accurate and transferable materials models, npj Comput. Mater. **8**, 158 (2022).

[36] E. Cignoni, D. Suman, J. Nigam, L. Cupellini, B. Mennucci, and M. Ceriotti, Electronic excited states from physically constrained machine learning, ACS Cent. Sci. **10**, 637 (2024).

[37] Y.-L. Liao and T. Smidt, Equiformer: Equivariant graph attention transformer for 3d atomistic graphs, arXiv preprint arXiv:2206.11990 (2022).

[38] Y.-L. Liao, B. Wood, A. Das, and T. Smidt, Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations, arXiv preprint arXiv:2306.12059 (2023).

[39] T. Ozaki, Variationally optimized atomic orbitals for large-scale electronic structures, Phys. Rev. B **67**, 155108 (2003).

[40] T. Ozaki and H. Kino, Numerical atomic basis orbitals from h to kr, Phys. Rev. B **69**, 195113 (2004).

[41] J. W. Abbott, C. M. Acosta, A. Akkoush, A. Ambrosetti, V. Atalla, A. Bagrets, J. Behler, D. Berger, B. Bieniek, J. Björk, *et al.*, Roadmap on advancements of the fhi-aims software package, arXiv preprint arXiv:2505.00125 (2025).

[42] L. Zhang, P. Mazzeo, M. Nottoli, E. Cignoni, L. Cupellini, and B. Stamm, A symmetry-preserving and transferable representation for learning the kohn-sham density matrix, arXiv preprint arXiv:2503.08400 (2025).

[43] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

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

[45] P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B **50**, 17953 (1994).

[46] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).

[47] V. Vitartas, C. Qian, J. R. Kermode, and R. J. Maurer, Dataset for MACE-H: Equivariant Hamiltonian Prediction with Many-Body Message Passing, `10.5281/zenodo.15223696` (2025).

[48] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, The atomic simulation environment—a Python library for working with atoms, J. Phys.: Condens. Matter **29**, 273002 (2017).

[49] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space, in *Advances in Neural Information Processing Systems*, Vol. 30 (Curran Associates, Inc., 2017).

[50] G. Dusson, M. Bachmayr, G. Csányi, R. Drautz, S. Etter, C. van der Oord, and C. Ortner, Atomic cluster expansion: Completeness, efficiency and stability, J. Comput. Phys. **454**, 110946 (2022).

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

[52] V. W.-z. Yu, J. Moussa, and V. Blum, Accurate frozen core approximation for all-electron density-functional theory, J. Chem. Phys. **154**, 224107 (2021).

[53] V. Vitartas, C. Qian, J. R. Kermode, and R. J. Maurer, FHI-aims Data for MACE-H: Equivariant Hamiltonian Prediction with Many-Body Message Passing, `10.17172/NOMAD/2025.04.14-1` (2025).

[54] D. N. Laikov, Intrinsic minimal atomic basis representation of molecular electronic wavefunctions, Int. J. Quantum Chem. **111**, 2851 (2011).

VII. ACKNOWLEDGEMENTS

The authors acknowledge funding through the UKRI Future Leaders Fellowship programme (MR/X023109/1), a UKRI frontier research grant (EP/X014088/1), and the EPSRC Centre for Doctoral Training in Modelling of Heterogeneous Systems (EP/S022848/1) at the University of Warwick.

High-performance computing resources were provided via the Scientific Computing Research Technology Platform of the University of Warwick, the EPSRC-funded HPC Midlands+ computing centre for access to Sulis (EP/P020232/1), and the Northern Ireland High Performance Computing (NI-HPC) service for access to Kelvin2 (EP/T022175/1).

Supporting Information

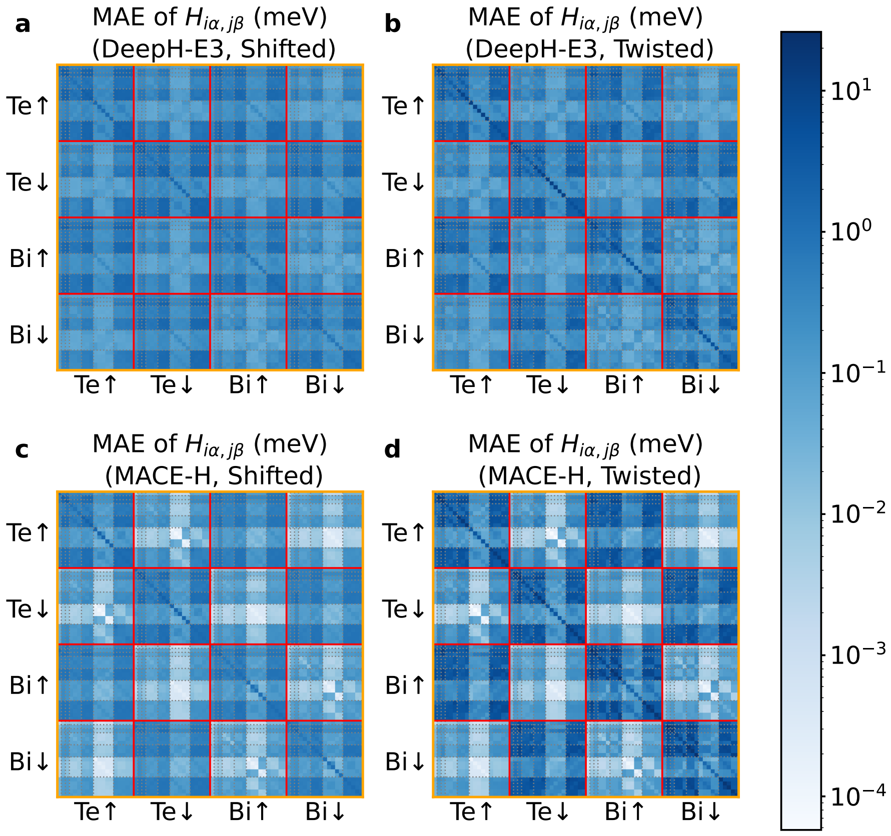
Equivariant Electronic Hamiltonian Prediction with Many-Body Message Passing

Chen Qian, Valdas Vitartas, James R. Kermode, and Reinhard J. Maurer

Contents

Supplementary Note 1 Hamiltonian Matrix MAE Analysis

To compare the performance of MACE-H and DeepH-E3, we first trained a series of models on shifted Bi_2Te_3 bilayers and then evaluated them using shifted test data and twisted bilayers. Supplementary Fig. 1 shows the corresponding analytic MAE matrices. For performance on shifted bilayers, MACE-H outperformed DeepH-E3, with the average, maximum, and minimum of the error matrix being 0.28/0.48 meV, 2.82/2.95 meV and $7.98 \times 10^{-5}/3.59 \times 10^{-2}$ meV. In Supplementary Fig. 1a, the error matrix blocks (segmented by the red lines) between different spin channels are usually of similar magnitudes to the ones between the same spin channels for DeepH-E3. Supplementary Fig. 1c shows that the MACE-H (employing shift and scale operation) manifested much lower errors between different spin channels. Moreover, the improvement of MACE-H is also manifested in the error matrix between the same spin channels. When applied to the twisted bilayers, the average, maximum, and minimum of the error matrices of MACE-H and DeepH-E3 are 1.31/0.70 meV, 26.11/11.11 meV and $5.78 \times 10^{-5}/2.59 \times 10^{-2}$ meV with noticeably larger errors of MACE-H, as shown in Supplementary Fig. 1b, d.

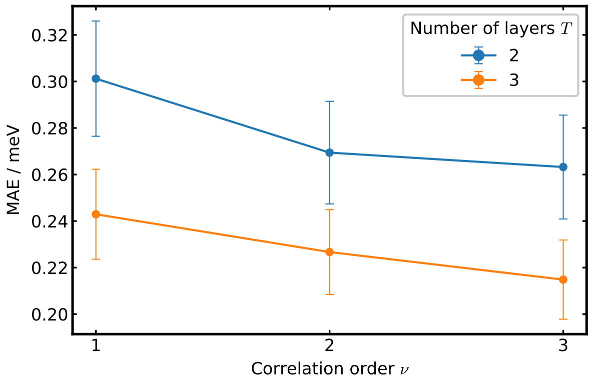


Supplementary Figure 1: **Hamiltonian matrix element mean absolute error (MAE) comparison between MACE-H and DeepH-E3 for shifted and twisted Bi_2Te_3.** The MAE matrix of **a**. shifted Bi_2Te_3 bilayer predicted with DeepH-E3, **b**. twisted Bi_2Te_3 bilayer predicted with DeepH-E3, **c**. shifted Bi_2Te_3 bilayer predicted with MACE-H, **d**. shifted Bi_2Te_3 bilayer predicted with MACE-H. The MACE-H model in **c** and **d** includes shift-and-scale operations (see Supplementary Note S6).

Supplementary Note 2 Additional Gold Results

Correlation Order vs Number of Layers

In order to investigate the effect of increasing correlation order ν for the gold dataset, six different models were trained by varying the correlation order ν and the number of message-passing layers T as shown in Supplementary Fig. 2. The model used in the main text corresponds to $\nu = 2$ and $T = 2$ unless stated otherwise.



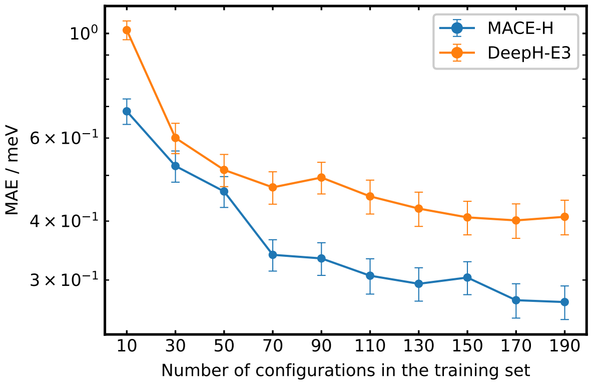
Supplementary Figure 2: **Hamiltonian mean absolute error (MAE) for a range of models trained on the gold dataset with different correlation orders ν and number of layers T.** The error bars were obtained by computing the standard deviation $\sigma_{\nu,T}$ of the MAE across configurations in the test set, and plotting $\pm\sigma_{\nu,T}$ intervals.

Learning Curves

To explore model convergence with respect to gold dataset size, MACE-H and DeepH-E3 models were trained with a range of training dataset sizes, and the MAE was computed for each model using the same test set as shown in Supplementary Fig. 3.

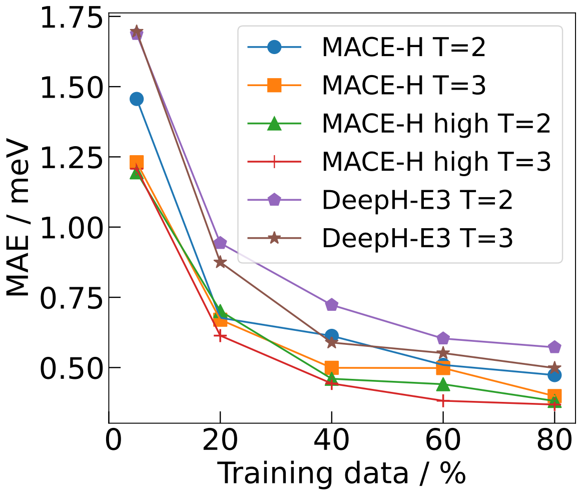
Supplementary Note 3 Data Efficiency of the Model

To assess the effect of different hyperparameter settings on MACE-H data efficiency, we adopted different settings. In Supplementary Fig. 4, Deep-H uses spherical harmonics with l up to 4, MACE-H uses the same settings in the edge-update block and specifies the node-wise MPNN block with spherical harmonics with l up to 4, and hidden states with azimuthal number up to 2. The higher setting of MACE-H uses spherical harmonics l up to 5 and hidden state azimuthal numbers up to 5. Models with two ($T = 2$) and three layers are compared. The correlation order $\nu = 3$ for all the MACE models. Benefiting from the higher expressive power of many-body expansion, MACE-H shows higher accuracy compared to DeepH-E3. While increasing the model depth helps to increase the data efficiency, the tighter setting of the many-body-expansion with larger correlation order ν and minimum azimuthal number L_{max} of hidden states seems to be more effective for higher



Supplementary Figure 3: **Learning curves for MACE-H and DeepH-E3 for the gold dataset.** The reported error bars correspond to $\pm\sigma_{N_{\text{train}}}$, where $\sigma_{N_{\text{train}}}$ is the standard deviation of the MAE across configurations in the test set for each model trained on N_{train} configurations.

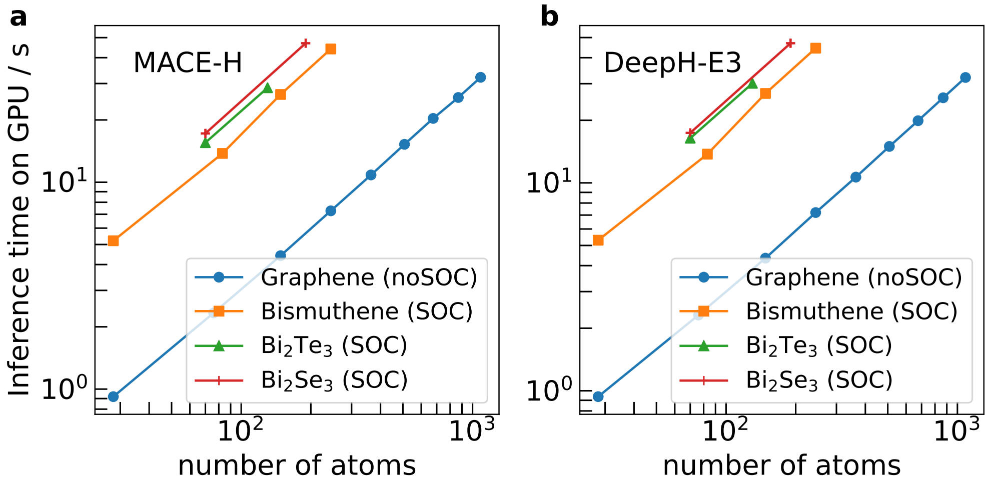
accuracy across different training data sizes. The overall Bi_2Te_3 dataset contains 256 configurations.



Supplementary Figure 4: **Test set MAE for DeepH-E3 and MACE-H with different settings for shifted SOC Bi_2Te_3 bilayers.** The label ”high” means higher settings of spherical harmonics and the hidden states of many-body expansion, T, stand for model depth. The correlation order is $\nu = 3$ for all the MACE-H models. The overall shifted SOC Bi_2Te_3 bilayers dataset contains 256 configurations.

Supplementary Note 4 Time-to-Solution Comparison: DeepH-E3 and MACE-H

To compare the inference time between DeepH-E3 and MACE-H, we perform tests using systems of various sizes in Supplementary Fig. 5. Since the most computationally costly operation is the edge-wise update, the overall inference time of MACE-H is comparable to DeepH-E3 as they share the same edge-update block. No significant overhead due to the many-body expansion and the node degree expansion block can be identified.



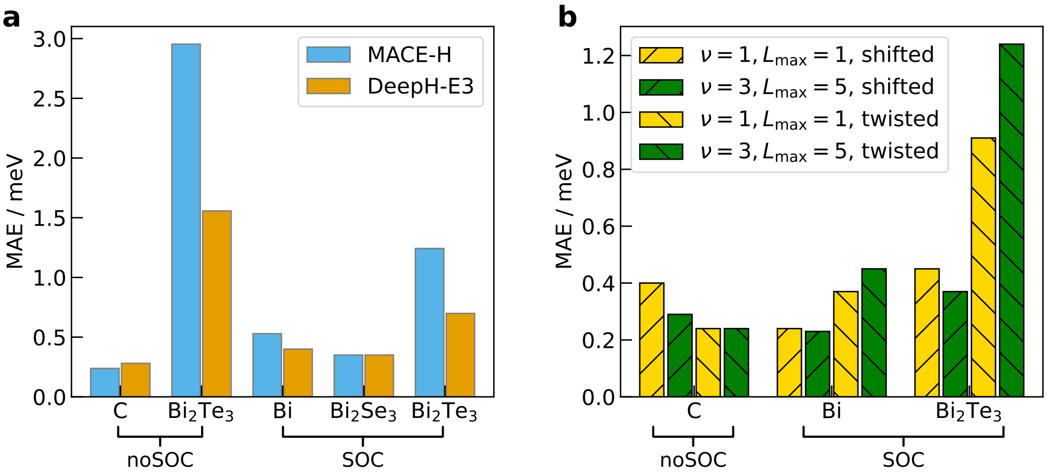
Supplementary Figure 5: **Inference time comparison between MACE-H and DeepH-E3 on a single A100 GPU.**

Supplementary Note 5 Many-Body-Expansion for Different Interaction Range

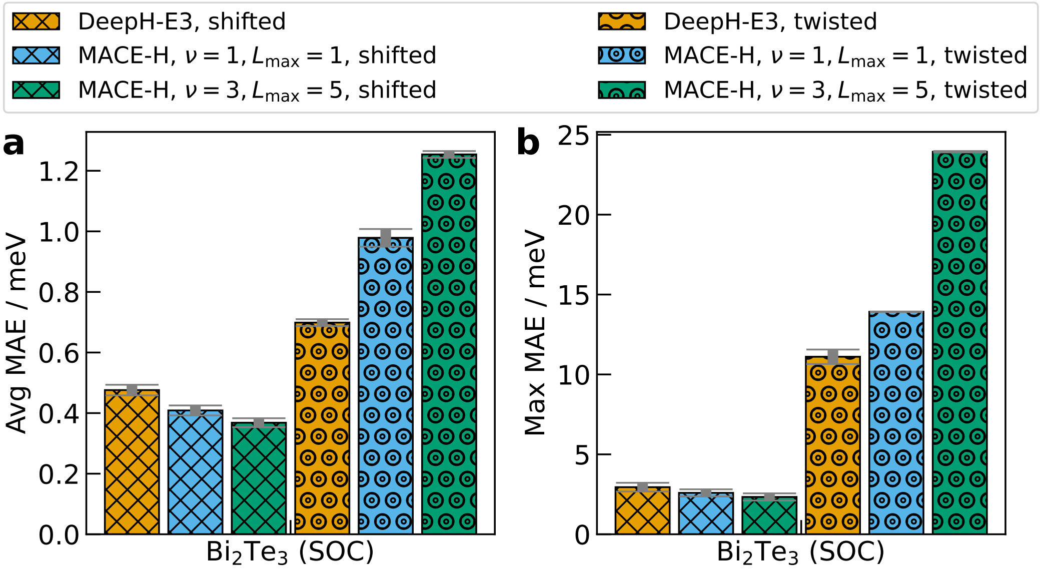
The Influence of the Many-Body Expansion for Bilayer Predictions

Supplementary Fig. 6a shows the performance of the model that was trained on shifted bilayers applied to twisted bilayer structures. While we found MACE-H predictions to be more accurate than DeepH-E3 for shifted bilayers, MACE-H either provides similarly accurate or slightly less accurate predictions for the twisted bilayers. We note that the Bi_2Se_3 dataset is three times larger than Bi_2Te_3. Since shifted and twisted bilayers have similar atomic environments, the difference may arise from differences in the range of interactions. To study the influence of the many-body expansion on the shifted and twisted bilayers, we evaluate models with different settings of the many-body expansion block in Supplementary Fig. 6b. It should be noted that to make a fair comparison with DeepH-E3, the MACE-H model used here does not incorporate the shift-and-scale operation. It shows that a tighter setting (using larger correlation order ν and the maximal azimuthal number of hidden states $L_{\max}$) of the many-body expansion will favour the performance on shifted bilayers in the test set but will bring further error for twisted bilayers, indicating the locality preference. This effect is even more distinguishable for the maximum values of the Hamiltonian MAE matrices in Supplementary Fig. 7, with the confidence interval showing the statistical significance. For predicting twisted bilayers with shifted bilayers as training data, it is recommended to use lower settings of ν and $L_{\max}$, and a larger dataset.

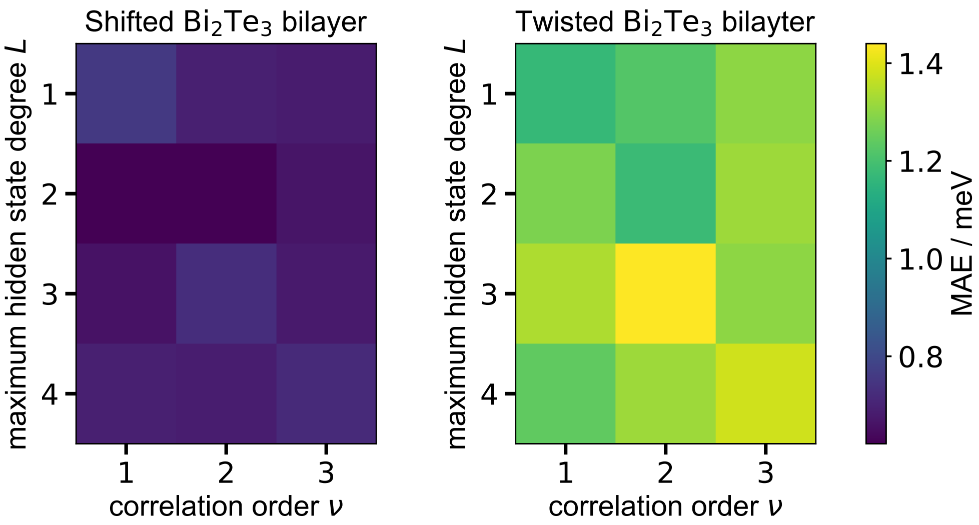
To study the effect of correlation order, ν, and maximal azimuthal number of many-body expansion hidden states, $L_{\max}$, Fig. 8 shows the MAE using different hyperparameter settings. The other settings here are the same as those used in Supplementary Fig. 3, i.e., we used spherical harmonics with l up to 4, hidden states with azimuthal number ($L_{\max}$) up to 2, and correlation order $\nu = 3$. To avoid excessive computational overhead, we only used 20% of the dataset as training samples, and no hyperparameter tuning was used with only fixed hyperparameters. The right panel shows the tendency for preferred locality with larger ν and L_{max}.



Supplementary Figure 6: **Effect of correlation order for MACE-H predictions. a.** Out-of-distribution MAE of matrix elements when applying the model trained on shifted bilayers to twisted bilayers for MACE-H and DeepH-E3. The correlation order $\nu = 3$ is used for MACE-H. **b.** Test set (shifted systems) and out-of-distribution (twisted systems) MAE of matrix elements using different settings of correlation order ν and L_{max} in the many-body expansion.



Supplementary Figure 7: **Average and maximum values of the Hamiltonian error matrices.** Data is shown for shifted and twisted Bi_2Te_3 bilayer using DeepH-E3 and MACE-H with different settings of correlation order ν and maximum azimuthal number of hidden state L_{max}. The confidence interval of a standard deviation $\sigma_{\nu,L_{\mathrm{max}}}$ is taken across the error matrix metric of individual samples.



Supplementary Figure 8: **The MAE for different settings of many-body expansion for shifted and twisted Bi_2Te_3 bilayers.**

Atom Perturbation Analysis of Hamiltonian Model Response

To further study the interaction range captured by MACE-H, we analyse how MACE-H predicted on-site matrix blocks at varying distances from a single atom change upon the displacement of this atom. We studied the response magnitude by the perturbation for the shifted (indexed by 72-0 in the dataset) and twisted (indexed 1-2 and 1-3) Bi_2Te_3 bilayer using DFT, MACE-H, and DeepH-E3. Supplementary Fig. 9 shows the comparison between the shifted 72-0 and the twisted 1-2 unit cells using different methods. To further exclude the image interactions, we also construct the $2 \times 2 \times 1$ supercell counterparts of the unit cells, and the result also shows a similar observation. The data for the second twisted geometry shown in Supplementary Fig. 10 shows slightly different interaction range dependence of the models than the geometry shown in Figure 6 of the main manuscript, although the overall trends are the same

For the shifted 71-0 bilayer unit cell, the overall onsite response error using MACE-H and DeepH-E3 are 9.6×10^{-2} and 13.3×10^{-2} meV, respectively. For the shifted 1-2 unit cell, the overall onsite response error using MACE-H and DeepH-E3 are 20.2×10^{-2} and 13.8×10^{-2} meV, respectively. However, for the supercell counterparts, the difference between MACE-H and DeepH-E3 regarding response error becomes negligible, with MACE-H showing a slightly lower error. We find that the change for supercells is mostly because MACE-H manifests lower response errors compared to DeepH-E3 for distances beyond 15 Å in the twisted supercells, which is a result of faster attenuation of the predicted response magnitude for longer distance regions, also showing the locality. Nevertheless, the larger matrix MAEs of twisted bilayers for MACE-H also exist in the supercells.

Here, the response to the perturbation ΔH_{ij} is defined as:

$$\Delta H_{ij} = H_{ij}^{\text{perturbed}} - H_{ij}^{\text{pristine}} \tag{5.0.1}$$

where $H_{ij}^{\text{perturbed}}$ and H_{ij}^{pristine} is the matrix block between atom i and j of the perturbed and pristine conformation. The response magnitude is defined as the mean absolute value of the response matrix elements:

$$\|\Delta H_{ij}\| = \frac{1}{N_{orb}^2} \sum_{k_1=1}^{N_{\text{orb}}} \sum_{k_2=1}^{N_{\text{orb}}} |\Delta H_{ij}|_{k_1 k_2} \tag{5.0.2}$$

where k_1, k_2 indicate the indices of the matrix element, N_{orb} is the number of orbitals in a Hamiltonian matrix block. The response error is defined as:

$$Error_{\Delta H_{ij}} = \frac{1}{N_{orb}^2} \sum_{k_1=1}^{N_{\text{orb}}} \sum_{k_2=1}^{N_{\text{orb}}} \left| \Delta H_{ij}^{\text{pred}} - \Delta H_{ij}^{DFT} \right|_{k_1 k_2} \tag{5.0.3}$$

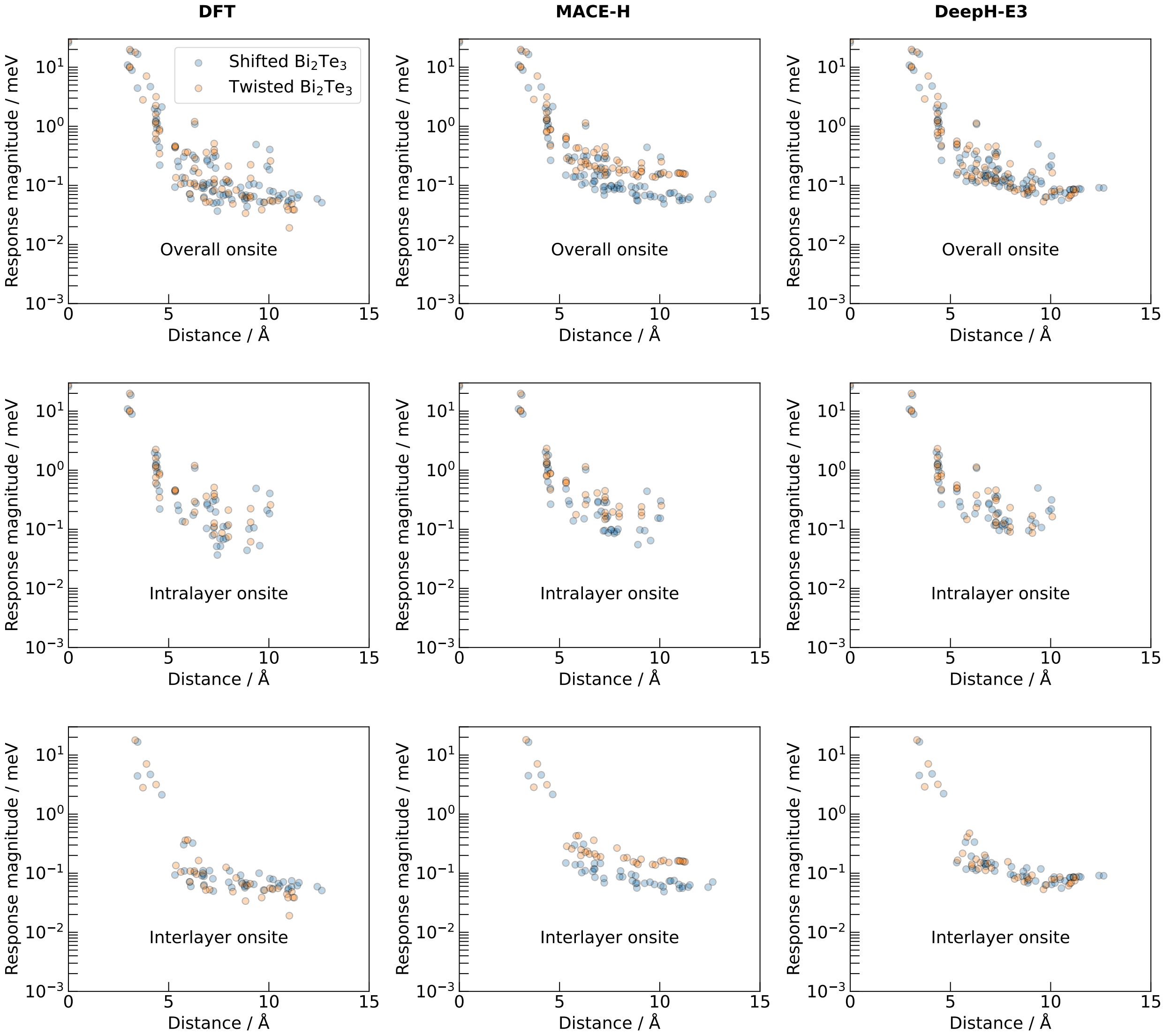
where $\Delta H_{ij}^{\text{pred}}$ and ΔH_{ij}^{DFT} are the response matrices using machine learning model prediction and DFT calculation, respectively.

Supplementary Note 6 Shift-and-Scale Operation on the Last-Layer Irreps

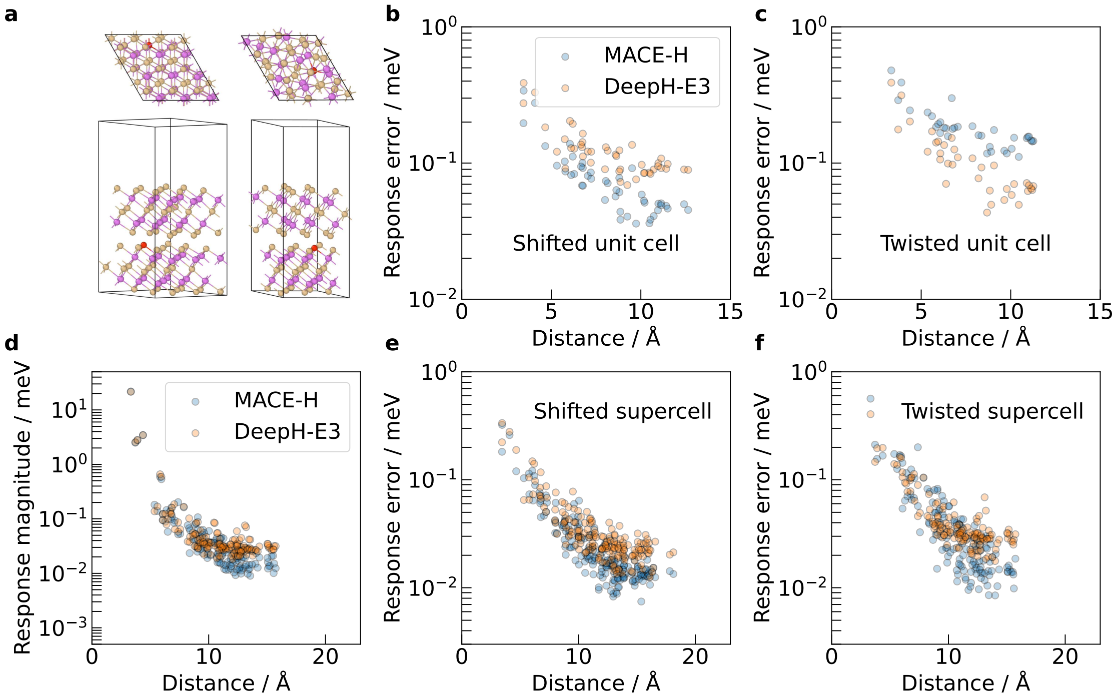
Numerical Instability for Large Matrix Value Ranges

Due to the large norm difference between the irreps of different element-orbital-pair resolved sub-blocks, adjusting the output accordingly may have the potential to accelerate the model convergence and increase the accuracy. However, irreps for different sub-blocks exhibit enormous differences over many orders of magnitude, which is a much more challenging scenario than than energy predictions in machine learning interatomic potentials. For example, Supplementary Fig. 11b shows that when we directly apply the shift-and-scale operation, the learning curve oscillates, the model fails to converge well, and the overall accuracy deteriorates. This is because as we scale the norms of the last layer irreps in the feed-forward process, it not only pushes the result closer to the magnitude of the target but also will scale the according gradients again in the backpropagation process, which will induce gradient explosion (ascribed to irreps with larger standard deviation) and disappearance (ascribed to the components with smaller standard deviation). Furthermore, the components with larger standard deviations will also have their errors magnified at early epochs after the scaling operation, which further contributes to the numerical instability. To tackle the issue, we first attempted to set the standard deviation components with magnitude lower than 1 to 1 (Supplementary Fig. 11c). This method alleviates the instability in the earlier stage, but the large oscillations persist in the later stage. Moreover, manually trimming the standard deviation with a threshold will not be suitable for a streamlined workflow since it is inconvenient to determine the optimal threshold. Thus, we adopted the method to allow the scaling operation in the feed-forward process but to sidestep the scaling factors in the gradient backpropagation process by redirecting the gradient flow:

$$o_{ij,klm} = o_{ij,klm}^{(t)} + (o_{ij,klm}^{(t)}(\sigma_{z_{i,j},k} - 1)).detach() + \mu_{Z_{i,j},k} \tag{6.0.1}$$

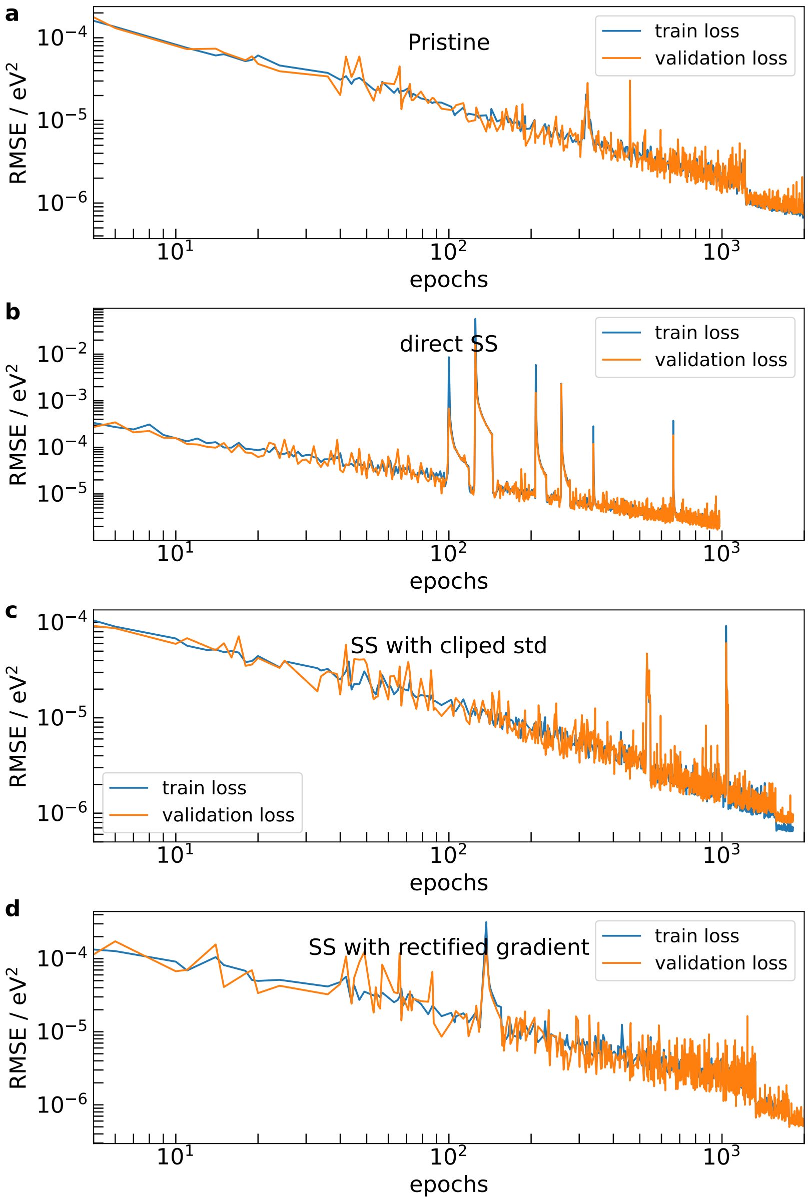


Supplementary Figure 9: **Comparison of the predicted perturbation response magnitude between shifted (indexed by 72-0 in the dataset) and twisted (indexed by 1-2 in the dataset) unit cell bilayers using DFT, MACE-H, and DeepH-E3.**



Supplementary Figure 10: **Locality analysis of MACE-H compared to DeepH-E3 for the twisted bilayer based on single-atom displacement perturbation. a.** The configuration of shifted (indexed by 72-0 in the dataset) and twisted (indexed by 1-2 in the dataset) Bi_2Te_3 bilayer unit cells with the red atom being the perturbed Te atom. **b.** The response error of onsite matrix blocks w.r.t. the distance from the perturbed atom for a shifted bilayer unit cell. **c.** The onsite response error for a twisted bilayer unit cell. **d.** The decay rate comparison for the 2×2×1 twisted bilayer with varying distance using MACE-H and DeepH-E3. The onsite response error for the 2×2×1 shifted (**e**) and twisted (**f**) supercell counterparts. The onsite block components plotted in the figure are of different layers from the perturbed atoms.

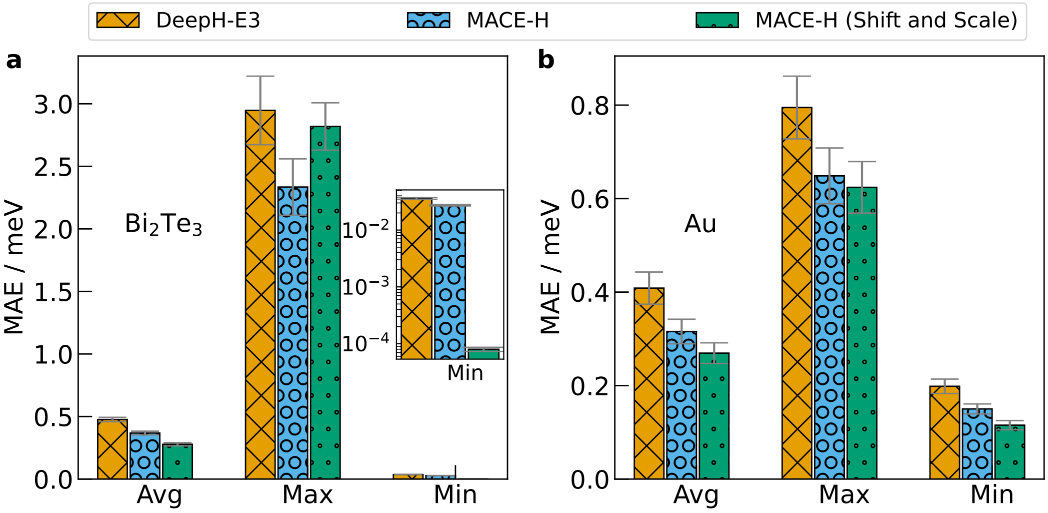
where $\mu_{Z_{i,j},k}$ and $\sigma_{z_{i,j},k}$ stand for the precalculated standard deviations and mean values of the corresponding target irreps, and the mean values will be set to zero for non-scalar vectors, the detach means that the gradient will be truncated for that term. Although the learning curve still has one observable oscillation at earlier epochs, the later stage converges faster than other methods with improved accuracy.



Supplementary Figure 11: **The learning curve convergence of the models for shifted $\mathbf{Bi_2Te_3}$. a** The original model training without shift-and-scale operation. **b** Direct application of the shift-and-scale operation to the last layer irreps. **c** Application of the shift-and-scale to the irreps with standard deviations of the norms lower than 1 manually set to 1. **d** Application of the shift-and-scale operation with a gradient bypass of the standard deviation in the backpropagation process.

Ablation Test for Shift-and-Scale Operation

To evaluate the effect by shifting and scaling the last layer according to the sub-block type resolved value distribution (i.e., the mean values for scalars and standard deviation for both scalars and vectors), we compared the average values of the MAE matrix using DeepH-E3, the pristine MACE-H, and the MACE-H with shift-and-scale operation. In Supplementary Fig. 12a, the shift-and-scale operation of MACE-H reduced the average MAE error from 0.32 meV of the pristine one to 0.27 meV. Due to the small magnitude of the SOC-involved matrix sub-blocks between different spin channels, the corresponding minimum value of the error matrix decreases by 3 magnitudes. A similar trend can also be observed for bulk Au in Supplementary Fig. 12b.

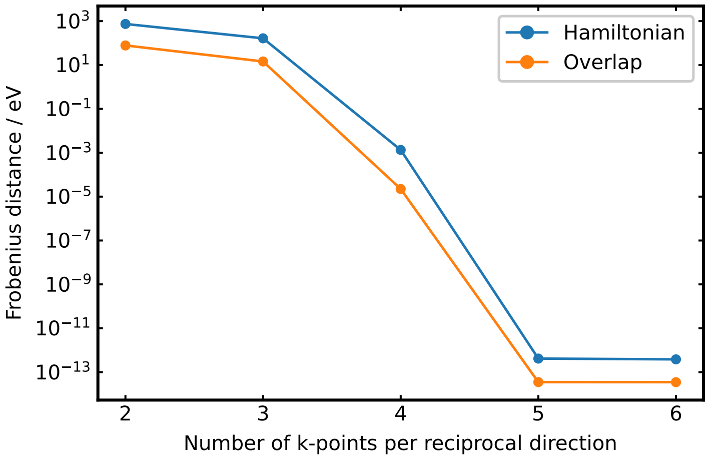


Supplementary Figure 12: **Average values of the Hamiltonian error matrices**. Data shown for **a**. Bi_2Te_3 bilayer and **b**. bulk Au using DeepH-E3, the pristine MACE-H, and the MACE-H with the shift-and-scale operation. The confidence interval of a standard deviation σ is taken across the error matrix metric of individual samples. The inset in **a** is the magnified view of the Min MAE for Bi_2Te_3

Supplementary Note 7 Core Projection and Sparsification

Reciprocal-to-Real Transformation Convergence

It was investigated how many k-points are required to accurately perform the inverse Fourier transform from reciprocal to real space for core-projected Hamiltonian and overlap matrices, as shown in Supplementary Fig. 13. The convergence study was performed for a single configuration in the training set called 'run_Aims_duy8hd_p', which can be found on NOMAD. Based on these results, it was decided to perform core projection for the whole gold dataset using the 6x6x6 k-point grid. Note that the reciprocal cell for each configuration in the gold dataset is the same and contains reciprocal-space vectors of equal length.



Supplementary Figure 13: **Convergence of reciprocal-to-real transformation (inverse Fourier transform) as a function of k-point grid for core-projected Hamiltonian and overlap matrices for one of the configurations in the training set.** The Frobenius distance corresponds to $\|\boldsymbol{H}_{n_k} - \boldsymbol{H}_{n_{k,\mathrm{ref}}}\|_{\mathcal{F}}$, where $n_{k,\mathrm{ref}}$ is the reference number of k-points per reciprocal dimension, which was set to 7, corresponding to a 7 × 7 × 7 k-point grid.

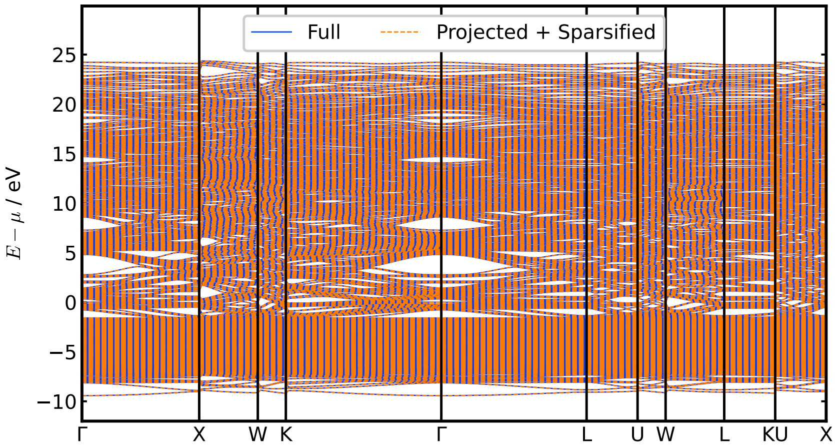
Eigenvalue Accuracy

Core states from Hamiltonian and overlap matrices in the gold dataset were projected out as discussed in the main text and Sec. 7. Furthermore, to decrease the computational cost during training and inference, the off-site matrix blocks corresponding to interactions spanning over 10 Å were set to zero. This was motivated by the fact that absolute Hamiltonian and overlap matrix elements for edges beyond 10 Å become too small to significantly affect Hamiltonian eigenvalues as quantitatively discussed below.

It was examined how accurate the resulting eigenvalues from processed (core-projected and sparsified) matrices are compared to eigenvalues obtained from original full-basis matrices. This was investigated for a single gold configuration in the training set ('run_Aims_duy8hd_p'). It was found that the eigenvalue and electronic entropy errors at 1000 K between the full and processed matrices are equal to 1.41×10^{-4} eV and 6.96×10^{-12} eV K^{-1} Å^{-3}, respectively. These values are much smaller than the model errors shown in Supplementary Fig. 4 in the main text. The valence electronic band structures using full and processed matrices are shown in Supplementary Fig. 14

Supplementary Note 8 Additional data

All data below (except where stated otherwise) employs MACE-H models without shift-and-scale operation.



Supplementary Figure 14: **Valence electronic band structure using full and processed (core-projected + sparsified) matrices for one of the configurations in the training set.** The bands from full and processed matrices are shifted with respect to their chemical potentials μ obtained with Fermi-Dirac smearing at 1000 K.

Supplementary Table 1: Hyperparameters of the models for different datasets in Fig. 4. The l_r, N_B, N_{layers}, ν, L_{MACE}, L_{node}, L_{hidden}, and L_{edge} stand for the learning rate, batch size, number of layers, correlation order, the largest azimuthal number of the intermediate MACE block irreps, the intermediate node-wise block irreps, the many-body expansion hidden states irreps, and the intermediate edge-wise block irreps

Dataset	MACE-H							DeepH-E3				
	l_r	N_B	N_{layers}	ν	L_{MACE}	L_{hidden}	L_{edge}	l_r	N_B	N_{layers}	L_{node}	L_{edge}
monolayer graphene (noSOC)	0.006	1	3	3	5	5	5	0.003	1	3	5	5
monolayer MoS_2 (noSOC)	0.006	1	3	3	5	5	5	0.005	1	3	5	5
bilayer graphene (noSOC)	0.003	1	3	3	5	5	5	0.003	1	3	5	5
bilayer Bismuthene (SOC)	0.01	1	3	3	5	5	5	0.005	1	3	5	5
bilayer Bi_2Se_3 (SOC)	0.01	1	3	3	5	5	5	0.005	1	3	5	5
bilayer Bi_2Te_3 (SOC)	0.008	2	3	3	5	5	5	0.004	2	3	5	5
bilayer Bi_2Te_3 (noSOC)	0.008	2	3	3	5	2	4	0.004	2	3	5	5
bulk Au (noSOC)	0.008	2	2	2	5	2	4	0.008	2	2	4	4

Supplementary Table 2: MAE (in meV) for Hamiltonian matrix prediction on 2D monolayers without SOC using DeepH, DeepH-E3 and our MACE-H. The basis sets used for both graphene and MoS_2 are up to d-orbital.

model	Graphene (nonSOC)	MoS_2 (nonSOC)			
	C-C	Mo-Mo	Mo-S	S-S	overall
DeepH	2.1	1.3	0.9	0.7	0.95
DeepH-E3	0.27	0.51	0.44	0.36	0.45
MACE-H (ours)	**0.21**	**0.42**	**0.39**	**0.32**	**0.39**

Supplementary Table 3: MAE (in meV) for Hamiltonian matrix prediction on 2D bilayers using DeepH, DeepH-E3 and our MACE-H. The basis sets are up to d-orbital, while the graphene and Bi_2Te_3 dataset doesn't involve SOC, the Bismuthene, Bi_2Te_3, Bi_2Se_3 involve SOC.

model	Graphene (nonSOC)		Bi_2Te_3 (nonSOC)		Bismuthene (SOC)		Bi_2Te_3 (SOC)		Bi_2Se_3 (SOC)	
	shifted	twisted	shifted	twisted	shifted	twisted	shifted	twisted		
DeepH	1.9	0.62	/	/	/	/	/	/	/	/
DeepH-E3	0.40	0.28	0.86	**1.56**	0.26	**0.40**	0.48	**0.70**	0.40	0.35
MACE-H (ours)	**0.31**	**0.24**	**0.58**	2.50	**0.23**	0.45	**0.37**	1.25	0.34	0.35

Supplementary Table 4: MAE (in meV) on bilayer Bi_2Te_3 with SOC using different settings of radial basis and many-body expansion and its relation to the locality. The values in the parentheses show the maximum error of the Hamiltonian matrix elements. The B and G in parentheses stand for Bessel and Gaussian radial basis set, respectively.

model	shifted Bi_2Te_3 (SOC)				twisted Bi_2Te_3 (SOC)			
	Te-Te	Te-Bi	Bi-Bi	overall	Te-Te	Te-Bi	Bi-Bi	overall
DeepH-E3	0.63 (3.23)	0.48 (2.34)	0.41 (2.08)	0.50 (3.23)	0.90 (11.84)	0.59 (4.74)	0.66 (7.11)	0.68 (11.84)
MACE-H (B, v=3, L=2)	0.51 (3.08)	0.37 (1.84)	0.34 (1.75)	0.40 (3.80)	1.26 (17.91)	1.10 (12.10)	1.51 (23.83)	1.24 (23.83)
MACE-H (G, v=3, L=5)	0.46 (2.33)	0.35 (1.66)	0.30 (1.62)	0.37 (2.33)	1.26 (17.91)	1.10 (12.10)	1.51 (23.83)	1.24 (23.83)
MACE-H (G, v=3, L=2)	0.50 (2.49)	0.38 (1.80)	0.34 (1.66)	0.40 (2.49)	1.16 (18.11)	1.06 (9.19)	1.48 (20.93)	1.19 (20.93)
MACE-H (G, v=1, L=1)	0.57 (2.87)	0.43 (2.04)	0.38 (1.89)	0.45 (2.87)	1.03 (12.70)	0.82 (6.42)	0.98 (12.46)	0.91 (12.70)

Supplementary Table 5: MAE (in meV) on bilayer Bi_2Se_3 with SOC using different settings of radial basis and many body expansion and its relation to the locality. The values in the parentheses show the maximum error of the Hamiltonian matrix elements. The B and G in parentheses stand for Bessel and Gaussian radial basis set, respectively.

model	shifted Bi_2Se_3 (SOC)				twisted Bi_2Se_3 (SOC)			
	Se-Se	Se-Bi	Bi-Bi	overall	Se-Se	Se-Bi	Bi-Bi	overall
DeepH-E3	0.39 (1.45)	0.45 (1.72)	0.32 (1.97)	0.40 (1.97)	0.37 (1.58)	0.38 (1.71)	0.30 (2.29)	0.35 (2.29)
MACE-H (B, v=3, L=2)	0.35 (1.33)	0.39 (1.57)	0.28 (1.69)	0.35 (1.69)	0.38 (1.95)	0.39 (1.84)	0.28 (2.41)	0.35 (2.41)
MACE-H (G, v=3, L=2)	0.33 (1.20)	0.37 (1.49)	0.29 (1.64)	0.34 (1.64)	0.42 (2.29)	0.52 (2.89)	0.47 (3.02)	0.48 (3.02)
MACE-H (G, v=3, L=5)	0.32 (1.25)	0.37 (1.40)	0.28 (1.69)	0.33 (1.69)	0.38 (2.37)	0.43 (2.41)	0.46 (3.38)	0.43 (3.38)
MACE-H (G, v=1, L=1)	0.35 (1.39)	0.39 (1.57)	0.29 (1.54)	0.35 (1.59)	0.40 (2.07)	0.46 (2.33)	0.46 (4.04)	0.45 (4.04)

Supplementary Table 6: The MAE (in meV) on Au data using DeepH-E3, MACE-H with/without shift-and-scale operation. The values in the parentheses show the maximum error of the Hamiltonian matrix elements.

	DeepH-E3	MACE-H	MACE-H (Shift & Scale)
Au	0.41 (0.80)	0.32 (0.65)	0.27 (0.62)

Supplementary Table 7: MAE (in meV) on bilayer graphene without SOC and Bismuthene with SOC using different settings of radial basis and many body expansion and its relation to the locality. Values in parentheses show the maximum error of the Hamiltonian matrix elements. The B and G in parentheses stand for Bessel and Gaussian radial basis set, respectively.

model	graphene (nonSOC)		Bismuthene (SOC)	
	shifted	twisted	shifted	twisted
DeepH-E3	0.40 (0.73)	0.28 (1.04)	0.26 (1.55)	0.40 (3.94)
MACE-H (B, v=3, L=5)	0.29 (0.52)	0.24 (0.78)	0.25 (1.4)	0.52 (6.29)
MACE-H (G, v=3, L=5)	0.31 (0.51)	0.24 (0.80)	0.23 (1.27)	0.45 (4.79)
MACE-H (G, v=1, L=1)	0.40 (0.75)	0.24 (0.59)	0.24 (1.50)	0.37 (3.39)

Supplementary Table 8: Summary of employed datasets

Dataset	system dimension	DFT package	number of layers	SOC	data size (train/val/test)	system size (atoms)
graphene	2D	OpenMx	monolayer	noSOC	450 (0.6/0.2/0.2)	72
MoS_2	2D	OpenMx	monolayer	noSOC	500 (0.6/0.2/0.2)	75
graphene	2D	OpenMx	bilayer	noSOC	300 (0.6/0.2/0.2)	64
Bismuthene	2D	OpenMx	bilayer	SOC	576 (0.4/0.2/0.2)	36
Bi_2Se_3	2D	OpenMx	bilayer	SOC	576 (0.4/0.2/0.2)	90
Bi_2Te_3	2D	OpenMx	bilayer	noSOC	256 (0.8/0.15/0.05)	90
Bi_2Te_3	2D	OpenMx	bilayer	SOC	256 (0.8/0.15/0.05)	90
Au	Bulk	FHI-aims	/	noSOC	200 (0.76/0.19/0.05)	64