MACHINE LEARNING
Science and Technology

PAPER • OPEN ACCESS

Towards harmonization of SO(3)-equivariance and expressiveness: a hybrid deep learning framework for electronic-structure Hamiltonian prediction

To cite this article: Shi Yin *et al* 2024 *Mach. Learn.: Sci. Technol.* **5** 045038

MACHINE
LEARNING
Science and Technology

CrossMark

OPEN ACCESS

RECEIVED
21 June 2024

REVISED
25 September 2024

ACCEPTED FOR PUBLICATION
30 October 2024

PUBLISHED
12 November 2024

PAPER

Towards harmonization of SO(3)-equivariance and expressiveness: a hybrid deep learning framework for electronic-structure Hamiltonian prediction

Shi Yin[1,*], Xinyang Pan[2], Xudong Zhu[1,2], Tianyu Gao[2], Haochong Zhang[1], Feng Wu[1,2] and Lixin He[1,2,*]

[1] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, People's Republic of China
[2] University of Science and Technology of China, Hefei, People's Republic of China
[*] Authors to whom any correspondence should be addressed.

E-mail: shiyin@iai.ustc.edu.cn and helx@ustc.edu.cn

Keywords: deep learning, electronic-structure Hamiltonian, SO(3)-equivariance

Abstract

Deep learning for predicting the electronic-structure Hamiltonian of quantum systems necessitates satisfying the covariance laws, among which achieving SO(3)-equivariance without sacrificing the non-linear expressive capability of networks remains unsolved. To navigate the harmonization between SO(3)-equivariance and expressiveness, we propose HarmoSE, a deep learning method synergizing two distinct categories of neural mechanisms as a two-stage encoding and regression framework. The first stage corresponds to group theory-based neural mechanisms with inherent SO(3)-equivariant properties prior to the parameter learning process, while the second stage is characterized by a non-linear 3D graph Transformer network we propose, featuring high capability on non-linear expressiveness. Their combination lies in the point that, the first stage predicts baseline Hamiltonians with abundant SO(3)-equivariant features extracted, assisting the second stage in empirical learning of equivariance; and in turn, the second stage refines the first stage's output as a fine-grained prediction of Hamiltonians using powerful non-linear neural mappings, compensating for the intrinsic weakness on non-linear expressiveness capability of mechanisms in the first stage. Our method enables precise, generalizable predictions while capturing SO(3)-equivariance under rotational transformations, and achieves state-of-the-art performance in Hamiltonian prediction tasks under multiple mean absolute error (MAE) metrics, such as the average MAE across all samples and matrix elements, the MAE for challenging samples, the MAE for different Hamiltonian blocks, and the MAE for the challenging blocks. It also demonstrates significant improvements in accuracy for downstream quantities, such as occupied orbital energy and the electronic wavefunction, as measured by MAE and cosine similarity, respectively.

1. Introduction

Recently, deep learning methods [1–6] have emerged as a promising trend for predicting the electronic-structure Hamiltonian, an essential physical quantity in understanding a wide range of properties, including electronic-structures, magnetic properties, optics, transport, and numerous other properties. These methods have offered a way to bypass the computationally exhaustive self-consistent steps of the traditional density functional theory (DFT) method [7, 8], thereby providing a viable pathway for the efficient simulation and design of large-scale atomic systems, laying the foundation for many down-stream applications [9] in the information and energy areas.

Despite these progresses, the Hamiltonian prediction task continues to present substantial challenges for deep learning techniques. High numerical accuracy is required to derive reasonable physical quantities, and furthermore, the fidelity of Hamiltonian predictions should not be confined to a specific coordinate system; rather, the results must demonstrate robust covariance and generalizability across various choices of reference frames. However, achieving 3D rotational equivariance, i.e. equivariance to the SO(3) group, is a tough target for a deep learning Hamiltonian prediction method. This difficulty arises because the Hamiltonian of each pair of atoms is usually high-dimensional, and its variation space under rotational disturbance is large. Consequently, it is difficult to cover the vast variability space they inhabit with rotations merely depending on parameter learning from discrete training samples. To address this, several works [6, 10] applied group theory-guided feature descriptors and tensor operators assuring inherent SO(3)-equivariance prior to the data-driven parameter learning process. Yet, to guarantee such SO(3)-equivariance independent to specific network parameters, these methods highly restricted the use of non-linear activation layers for SO(3)-equivariant features, leading to bottlenecks in expressiveness for complex non-linear mappings, limiting the accuracy achievable in predicting Hamiltonians. This dilemma, is also broadly prevalent in other 3D machine learning tasks where SO(3)-equivariance is highlighted, as analyzed by Zitnick *et al* [11].

To harmonize SO(3)-equivariance and expressiveness for the prediction of electronic-structure Hamiltonians, this paper proposes a two-stage encoding and regression framework, namely HarmoSE, which combines mechanisms boasting parameter-independent prior SO(3)-equivariance with mechanisms featuring flexibility in non-linear expressiveness, overcoming the respective challenges of each categories of mechanisms, i.e. the limited non-linear expressive capability for the former as well as the difficulty of learning SO(3)-equivariance from data for the latter, through effective complementary strategies. Specifically, the first stage corresponds to the neural mechanisms constructed based on group theory with prior equivariant properties of 3D atomic systems, predicting an approximate value of the Hamiltonian, with abundant SO(3)-equivariant features provided. In the second stage, a highly expressive graph Transformer network we design, with no restrictions on non-linear activations, takes over. This network dynamically learns the 3D structural patterns of the atomic systems, compensates for the expressiveness shortcomings of the first-stage network arising from limited non-linear mappings, and refines the Hamiltonian values predicted in the first stage to enhance accuracy. Although this stage might not possess a parameter-independent prior SO(3)-equivariance due to its non-linearity, it is capable of capturing SO(3)-equivariance through learning effective network parameters with the help of three pivotal mechanisms. First, instead of directly regressing the entire Hamiltonians, the second stage aims to refine the Hamiltonian predictions from the first stage with corrective adjustments in a cascaded manner. The scopes of adjustments are smaller, lowing down the difficulties on non-linear learning of SO(3)-equivariance. Second, the second-stage network incorporates covariant features, including SO(3)-equivariant features extracted by the first-stage network and SO(3)-invariant features engineered by geometric knowledge, with its inputs to assist in the implicit learning of SO(3)-equivariance. Third, as the core of Transformer, the attention mechanism has potentials to adapt to geometric condition variations such as coordinate transformations, through its dynamic weighting strategy. Collectively, the combination of the two categories of neural mechanisms in the two stages allows the framework to overcome the challenges of each individual mechanism and make precise, generalizable, and SO(3)-equivariant predictions, being much more effective than simply increasing the parameter count for the network from one stage and fine-tuning it alone.

Our method achieves state-of-the-art (SOTA) performance on Hamiltonian regression on both crystalline and molecular benchmark databases measured by multiple mean absolute error (MAE) metrics, comprehensively demonstrating the superiority of our method. Particularly, our SOTA performance in the twisted samples, which exhibit both SO(3)-equivariance effects and variations in van der Waals (vdW) interactions due to the inter-layer rotations, comprehensively confirms the robust capability of our model in capturing the intrinsic SO(3)-equivariance of Hamiltonians as well as its strong non-linear expressive power to generalize to complex and dynamic 3D geometric structures of atomic systems. Moreover, the Hamiltonians predicted by our method demonstrate significant accuracy advantages over the baseline method [12] when used to derive key physical quantities of electronic structures, such as occupied orbital energy and the electronic wavefunction.

2. Related work

In this part, we firstly overview deep learning studies on capturing rotational equivariance. After that, we segue into related works on deep Hamiltonian prediction, in which 3D rotational equivariance is pursued.

As representative researches on equivariance to discrete rotational group, Dieleman *et al* [13] introduced cyclic symmetry operations into CNNs to achieve rotational equivariance; Ravanbakhsh *et al* [14] explored parameter-sharing techniques for equivariance to discrete rotations; Kondor *et al* [15] developed equivariant representations via compositional methods and tensor theory; Zitnick *et al* [11], Passaro *et al* [16], Liao *et al* [17], and Wang *et al* [18] applied spherical harmonic bases for atomic modeling, focusing on rotational equivariance but limited to discrete sub-groups of SO(3) due to their discrete sampling strategy. These approaches were very effective on discrete symmetries but sub-optimal on handling continuous 3D rotations. Focusing on equivariance to continuous rotational group, Jaderberg *et al* [19] and Cohen *et al* [20] achieved considerable success in 2D image recognition tasks by modeling equivariance to in-plane rotations. However, their applications were limited within the scope of 2D tasks and did not fit for the more complex demands of equivariance to 3D continuous rotational group, i.e. SO(3), required in the Hamiltonian prediction task.

In the field of researches on equivariance to SO(3), approaches like DeepH [4] explored equivariance via a local coordinate strategy, which made inference within the fixed local coordinate systems built with neighboring atoms, then transferred the output according to equivariance rules to the corresponding global coordinates. However, due to a lack of in-depth exploration of SO(3)-equivariance at the neural mechanism level, this method faced challenges when the local coordinate system underwent rotational disturbances from non-rigid deformation, e.g. the inter-layer twist of bilayer structures. In contrast, methods like TFN [21], SE(3)-Transformer [22], E3NN [10], Equiformer [23], DeepHE3 [6], QHNet [12, 24], DEQH [25] and SELF-CON [26] incorporated SO(3)-equivariance into the neural network mechanisms through operations developed from the group theory to effectively model atomic systems. However, a common challenge across these methods lies in the fact that, to achieve inherent equivariance prior to the parameter learning process, they forbade the use of complex non-linear mappings such as *Sigmoid*, *Tanh*, *SiLU*, and *Softmax* for SO(3)-equivariant features whose degree $l > 1$. This significantly limits the network's expressive potential, creating a bottleneck in generalization performance. To promote expressiveness, these methods applied a gated activation function, where SO(3)-invariant features undergone through non-linear activation layers were used as gating coefficients that were multiplied with SO(3)-equivariant features. However, viewed from the perspective of equivariant features, this mechanism approximated to a linear operation and did not fundamentally improve their expressive capability. For these methods, this conflict between equivariance and expressiveness remains an unsolved problem.

3. Preliminary

In the study of symmetry on mathematical structures, an operation A is equivariant with respect to B if applying B before or after A has the same effect, expressed as: $A(B(x)) = B(A(x))$. The key equivariance properties of Hamiltonians are the 3D rotational equivariance with respect to reference frame. Specifically, when the reference frame rotates by a rotation matrix denoted as $\mathbf{R}$, the edge of an atom pairs (i,j) transforms from $\mathbf{r}_{ij}$ to $\mathbf{R}\cdot\mathbf{r}_{ij}$, and the Hamiltonians in the direct sum state transforms equivariantly from $\mathbf{h}_{ij}$ to $D(\mathbf{R})\cdot\mathbf{h}_{ij}$, where $D(\mathbf{R})$ is the Wigner-D matrix[3]. The requirements for the fitting and generalization capability of a neural network $f_{\mathrm{nn}}(\cdot)$ for Hamiltonian prediction can be formally expressed as: $f_{\mathrm{nn}}(\{\mathbf{r}_{ij}|i\in Nodes, ij\in Edges\})\cong\{\mathbf{h}_{ij}\}$; moreover, the requirement on SO(3)-equivariance can be represented as: $f_{nn}(\{\mathbf{R}\cdot\mathbf{r}_{ij}|i\in Nodes, ij\in Edges\})\cong\{D(\mathbf{R})\cdot\mathbf{h}_{ij}\}$. It is crucial that $f_{\mathrm{nn}}(\cdot)$ intrinsically captures SO(3)-equivariance to effectively generalize under rotational reference frames, and meanwhile, $f_{\mathrm{nn}}(\cdot)$ must also possess sufficient expressive power to generalize across different types and structures of atomic systems and make accurate predictions.

4. Method

As shown in figure 1, to harmonize SO(3)-equivariance and expressiveness for Hamiltonian prediction, we propose a hybrid framework with two encoding and regression stages, from which the first-stage network, a group theory-informed network possessing SO(3)-equivariance prior to the learning process, provides essential foundations to the second stage in mastering SO(3)-equivariance, whereas the second-stage network, with highly expressive non-linear mappings, enriches the expressiveness capabilities of the whole framework. The combination of these two stages not only enhances expressiveness but also ensures robust equivariance to rotations of reference frames, bringing accurate predictions for electronic-structure Hamiltonians despite rotational transformations.

4.1. Initial features

In our framework, the initial feature for the $i(1\leqslant i\leqslant N)$ th node is its node embedding, denoted as $\mathbf{z}_i$, a coordinate-independent SO(3)-invariant semantic embedding that marks its element type. Given the locality of the Hamiltonian [4], each atom j in the local set $\Omega(i)$ within the cutoff radius of an atom i form an edge with i. From each edge, a Hamiltonian is defined. The initial features for edge (i,j) include both SO(3)-invariant encodings and SO(3)-equivariant encodings. The former includes edge embeddings $\mathbf{z}_{ij}$ marking the types of interacting atom pairs, as well as the distance features $\mathbf{d}_{ij}$ in the form of Gaussian

[3] Here we present SO(3)-equivariance under the direct sum state due to its simple vector form. For the equivalent formulation under the matrix-formed direct product state, please refer to Gong *et al* [6].

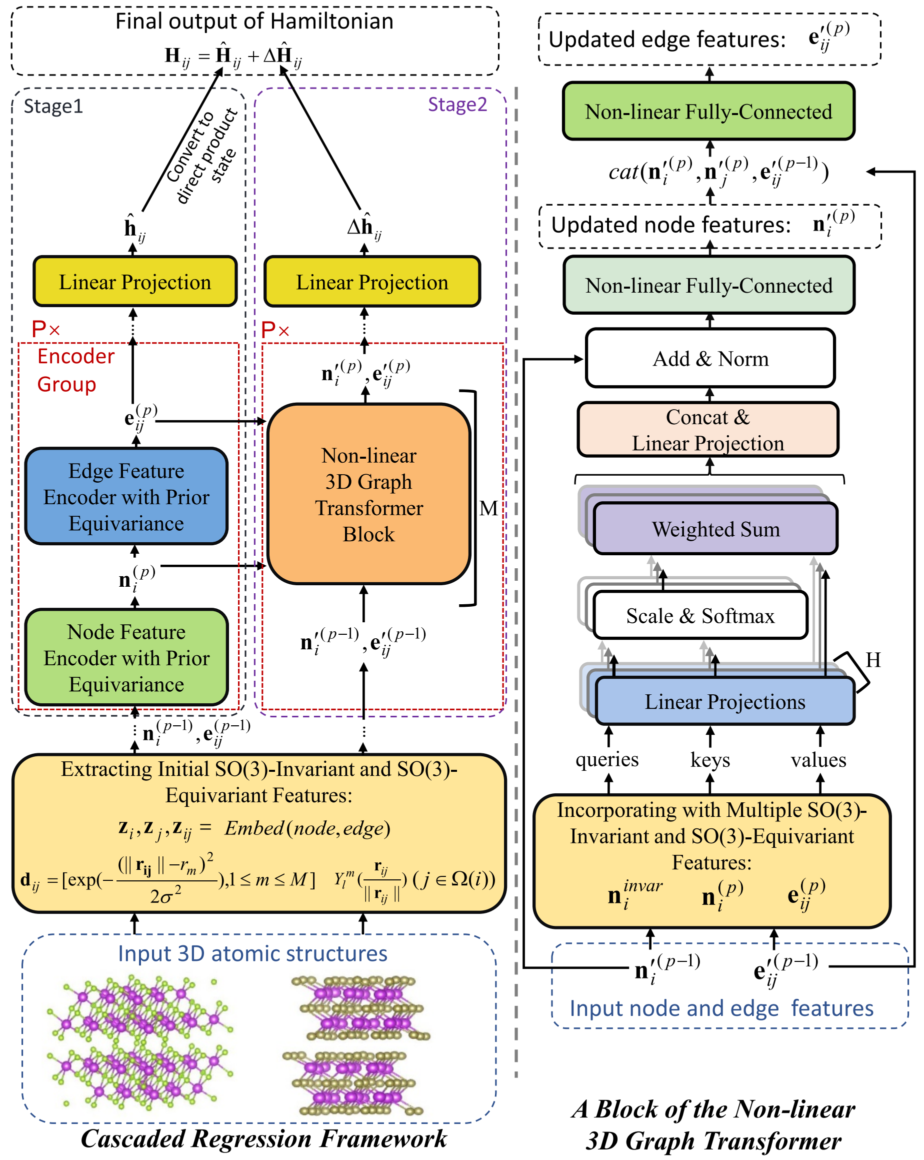

Final output of Hamiltonian
$\mathbf{H}_{ij} = \hat{\mathbf{H}}_{ij} + \Delta\hat{\mathbf{H}}_{ij}$
Stage1
Stage2
Convert to direct product state
$\hat{\mathbf{h}}_{ij}$
$\Delta\hat{\mathbf{h}}_{ij}$
Linear Projection
Linear Projection
P×
P×
Encoder Group
$\mathbf{e}_{ij}^{(p)}$
Edge Feature Encoder with Prior Equivariance
$\mathbf{n}_i^{(p)}$
Node Feature Encoder with Prior Equivariance
$\mathbf{n}_i'^{(p)}, \mathbf{e}_{ij}'^{(p)}$
Non-linear 3D Graph Transformer Block
M
$\mathbf{n}_i'^{(p-1)}, \mathbf{e}_{ij}'^{(p-1)}$
$\mathbf{n}_i^{(p-1)}, \mathbf{e}_{ij}^{(p-1)}$
Extracting Initial SO(3)-Invariant and SO(3)-Equivariant Features:
$\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_{ij} = Embed(node, edge)$
$\mathbf{d}_{ij} = [\exp(-\frac{(\|\mathbf{r_{ij}}\| - r_m)^2}{2\sigma^2}), 1 \le m \le M] \quad Y_l^m(\frac{\mathbf{r}_{ij}}{\|\mathbf{r}_{ij}\|}) \ (j \in \Omega(i))$
Input 3D atomic structures
Cascaded Regression Framework
Updated edge features: $\mathbf{e}_{ij}'^{(p)}$
Non-linear Fully-Connected
$cat(\mathbf{n}_i'^{(p)}, \mathbf{n}_j'^{(p)}, \mathbf{e}_{ij}'^{(p-1)})$
Updated node features: $\mathbf{n}_i'^{(p)}$
Non-linear Fully-Connected
Add & Norm
Concat & Linear Projection
Weighted Sum
Scale & Softmax
H
Linear Projections
queries
keys
values
Incorporating with Multiple SO(3)-Invariant and SO(3)-Equivariant Features:
$\mathbf{n}_i^{invar}$ $\mathbf{n}_i^{(p)}$ $\mathbf{e}_{ij}^{(p)}$
$\mathbf{n}_i'^{(p-1)}$ $\mathbf{e}_{ij}'^{(p-1)}$
Input node and edge features
A Block of the Non-linear 3D Graph Transformer

Figure 1. Left part: overview of the two-stage encoding and regression framework for Hamiltonian prediction. Right part: The internal architecture of the non-linear 3D graph Transformer network.

functions [4]; the latter is spherical harmonics, denoted as $Y_l^m(\frac{\mathbf{r}_{ij}}{||\mathbf{r}_{ij}||})$ [27], where $\frac{\mathbf{r}_{ij}}{||\mathbf{r}_{ij}||}$ describes the relative orientation between two atoms.

4.2. The first regression stage

In our framework, the primary role of the first encoding and regression stage is to extract node and edge representations with intrinsic SO(3)-equivariance independent to network parameters, and regress baseline Hamiltonian predictions, denoted as $\hat{\mathbf{h}}_{ij}$ in the direct sum state and $\hat{\mathbf{H}}_{ij}$ in the direct product state for each pair (i, j), establishing a firm foundation on SO(3)-equivariance in both feature level and the regression target level. The encoding process of the p-th encoder group can be described by the following equation:

$$\mathbf{n}_i^{(p)} = Gate\left(\mathbf{n}_i^{(p-1)} + \sum_{j \in \Omega(i)} EquiLin\left(\mathbf{n}_i^{(p-1)}, \mathbf{e}_{ij}^{(p-1)}, \mathbf{n}_j^{(p-1)}\right)\right)$$
$$\mathbf{e}_{ij}^{(p)} = Gate\left(EquiLin(\mathbf{n}_i^{(p)}, \mathbf{e}_{ij}^{(p-1)}, \mathbf{n}_j^{(p)})\right) \tag{1}$$

where $\mathbf{n}_i^{(p)}$ and $\mathbf{e}_{ij}^{(p)}$ respectively denote the node and edge features from the pth $(1 \leqslant p \leqslant P)$ encoder group, $Gate(\cdot)$ is the gated activation function introduced in the related work section, $EquiLin(\cdot)$ denotes a combination of tensor operators consisting of linear scaling, element-wise sum, direct sum, direct product, as well as the Clebsch–Gordan decomposition, possessing parameter-independent SO(3)-equivariance guaranteed by group theory. These operators, serving as our first-stage backbone, have been comprehensively developed in previous works [6, 10, 21], achieving maturity in their equivariance capabilities, yet facing inherent limitations in non-linear expressiveness cannot be easily resolved within their mechanisms. Therefore, we focus more on the complementary mechanisms of the hybrid two-stage encoding and regression framework as well as the design of the second-stage network, aiming at rectifying the intrinsic weaknesses in non-linear expressiveness of the first-stage network while embodying robust SO(3)-equivariant capability.

4.3. The second regression stage

The second encoding and regression stage of our framework is designed to fully exploit non-linear mappings to enhance the expressive capability of the whole framework while capturing SO(3)-equivariance. For that purpose, as shown in the right part of figure 1, we propose a 3D graph Transformer which effectively models 3D atomic structures and predicts non-linear correction terms that complement the predictions from the first stage, achieving high-precision Hamiltonian prediction. Yet, one key problem to solve is, non-linear projections might not have parameter-independent guarantee on SO(3)-equivariance, forcing the second stage to capture equivariance through learning effective network parameters from the data. Thus, the difficulty on empirical learning SO(3)-equivariance of Hamiltonians must be addressed. Our second-stage network adeptly resolves this issue, simultaneously capturing SO(3)-equivariance and enhancing the expressive capabilities. This is primarily attributed to three pivotal mechanisms we design.

First, the second stage works in a cascaded manner for regression, which means that its prediction target is not the entire Hamiltonian but a correction term $\Delta\hat{\mathbf{H}}_{ij}$, relative to the first stage's output, i.e. the initial Hamiltonian estimate $\hat{\mathbf{H}}_{ij}$. The sum of these two stages' outputs forms the final prediction of the Hamiltonian: $\mathbf{H}_{ij} = \hat{\mathbf{H}}_{ij} + \Delta\hat{\mathbf{H}}_{ij}$. Given that the predicted results of $\hat{\mathbf{H}}_{ij}$ are theoretically SO(3)-equivariant and numerically approximate to reasonable, the range of variations for the correction term $\Delta\hat{\mathbf{H}}_{ij}$, becomes smaller compared to $\hat{\mathbf{H}}_{ij}$. This reduces the complexity of the output space for the second stage and enhances the feasibility on implicitly mastering SO(3)-equivariance through data-driven learning for non-linear modules.

Second, several theoretical-guaranteed covariant features, including both SO(3)-equivariant and SO(3)-invariant features, are integrated into the input features of each Transformer block in the second stage to aid it in capturing equivariance, as shown in equation (2):

$$\widetilde{\mathbf{n}}_i'^{(p)} = \mathbf{n}_i'^{(p-1)} + \alpha\mathbf{n}_i^{(p)} + \beta\mathbf{n}_i^{\mathrm{invar}}, \quad \widetilde{\mathbf{e}}_{ij}'^{(p)} = \mathbf{e}_{ij}'^{(p-1)} + \lambda\mathbf{e}_{ij}^{(p)} \tag{2}$$

where α, β and λ are hyper-parameters, $\mathbf{n}_i'^{(p-1)}$ and $\mathbf{e}_{ij}'^{(p-1)}$ denote the outputs of the Transformer at the $p-1th$ encoder group, $\widetilde{\mathbf{n}}_i'^{(p)}$ and $\widetilde{\mathbf{e}}_{ij}'^{(p)}$ respectively serve as the input node and edge features for the subsequent modules of the Transformer at the $p\ th$ encoder group, $\mathbf{n}_i^{(p)}$ and $\mathbf{e}_{ij}^{(p)}$ are the SO(3)-equivariant node and edge features, respectively, from the corresponding encoder group of the first-stage network. Besides SO(3)-equivariant features, since as demonstrated by literature [28–30], SO(3)-invariant features also facilitate the learning of SO(3)-equivariance, we also develop a SO(3)-invariant node feature, i.e. $\mathbf{n}_i^{\mathrm{invar}}$, aggregated from multiple SO(3)-invariant features, such as node embeddings $\mathbf{z}_i$, edge embeddings $\mathbf{z}_{ij}$, distance features $\mathbf{d}_{ij}$, and triplet angle feature θ_{ijk} formed by node i as well as two of its local atoms j and k, in the way like:

$$\mathbf{n}_i^{\mathrm{invar}} = \sum_{(j,k)\in\Omega(i)} FC(cat(\mathbf{z}_i, \mathbf{z}_j, \mathbf{z}_k, \mathbf{d}_{ij}, \mathbf{d}_{ik}, \mathbf{c}_{ijk}))) \tag{3}$$

where $cat(\cdot)$ is the concatenation operator, $FC(\cdot)$ denotes fully-connected layers with non-linear activations, $\mathbf{c}_{ijk} = [\cos(\theta_{ijk}), \cos(\theta_{ijk}), \ldots]$ is a vector extended by duplication, which serves to amplify the angle features for $FC(\cdot)$. To reduce the quadratic complexity, i.e. $O(|\Omega(i)|^2)$ when sampling $(j,k)\in\Omega(i)$, we arrange the set $\Omega(i)$ as an array and only extract adjacent element pairs as tuples (j,k), lowing down the sampling complexity to $O(|\Omega(i)|)$ to efficiently compute $\mathbf{n}_i^{\mathrm{invar}}$. In equation (2), $\mathbf{n}_i^{\mathrm{invar}}$ is directly merged into node features, and since node features are then merged into edge features in the subsequent modules, it also enhances the learning of edge features. With the help of these covariant features, the second-stage network, even a non-linear one, can also capture the SO(3)-equivariant properties inherent in the Hamiltonian.

Third, we design a multi-head attention mechanism to learn node and edge representations of the 3D atomic systems. The capability to dynamically focus on related geometric features enables robust adaptability to diverse geometric conditions, from structural variants to coordinate transformations. Specifically, the attention mechanism firstly learns dynamic weights, i.e. $\alpha_{ij}^{(p)}$ for the edge (i,j) at the p th encoder group, based on the interactive relationship between the current atom i and its local atoms $j \in \Omega(i)$, as shown in equation (4):

$$\begin{aligned}
\mathbf{q}_{ij}^{h(p)} &= \mathbf{W}_q^h \cdot cat\left(\widetilde{\mathbf{n}}_i^{\prime(p)}, \widetilde{\mathbf{e}}_{ij}^{\prime(p)}\right), \\
\mathbf{k}_{ij}^{h(p)} &= \mathbf{W}_k^h \cdot cat\left(\widetilde{\mathbf{n}}_j^{\prime(p)}, \widetilde{\mathbf{e}}_{ij}^{\prime(p)}\right), \\
\alpha_{ij}^{(p)} &= softmax\left(\frac{\left(\mathbf{q}_{ij}^{h(p)}\right)^T \cdot \mathbf{k}_{ij}^{h(p)}}{\sqrt{d_h}}\right)
\end{aligned} \tag{4}$$

where $h(1 \leqslant h \leqslant H)$ is the head index, d_h is the dimension of features, $\mathbf{W}_q^h$ and $\mathbf{W}_k^h$ are parameter matrices to calculate queries and keys, i.e. $\mathbf{q}_{ij}^{h(p)}$ and $\mathbf{k}_{ij}^{h(p)}$, respectively. Here the scale factor $\sqrt{d_h}$ in the denominator is used to prevent $softmax(\cdot)$ from gradient saturation, and the multiple heads aim at enhancing the model capacity. Based on $\alpha_{ij}^{(p)}$, the node features are updated flexibly through the structural information embedded in its local sets, as shown in equation (5):

$$\begin{aligned}
\mathbf{v}_i^{h(p)} &= \sum_{j \in \Omega(i)} \alpha_{ij}^{(p)} \cdot \left(\mathbf{W}_v^h \cdot cat\left(\widetilde{\mathbf{n}}_j^{\prime(p)}, \widetilde{\mathbf{e}}_{ij}^{\prime(p)}\right)\right), \\
\mathbf{n}_i^{\prime(p)} &= FC\left(LN\left(\mathbf{W}_o \cdot cat\left(\mathbf{v}_i^{1(p)}, \ldots, \mathbf{v}_i^{H(p)}\right) + \mathbf{n}_i^{\prime(p-1)}\right)\right)
\end{aligned} \tag{5}$$

where $LN(\cdot)$ is the layer normalization operator, $FC(\cdot)$ denotes fully-connected layers with non-linear activations. Based on $\mathbf{n}_i^{\prime(p)}$, the edge representations are updated as:

$$\mathbf{e}_{ij}^{\prime(p)} = FC\left(cat\left(\mathbf{n}_i^{\prime(p)}, \mathbf{n}_j^{\prime(p)}, \mathbf{e}_{ij}^{\prime(p-1)}\right)\right) \tag{6}$$

when repeatedly stacking operations in equations (5) and (6) in an alternate manner, local patterns can incrementally spread to a larger scale through the neighbors of neighboring atoms. Nevertheless, given the Hamiltonian's locality, there's typically no need for information transfer over very long distances. Finally, the correction term outputted by the second stage is regressed from the edge features $\mathbf{e}_{ij}^{\prime(P)}$ encoded by the last encoder group, in the way like:

$$\Delta\hat{\mathbf{H}}_{ij} = DStoDP\left(\Delta\hat{\mathbf{h}}_{ij}\right) = DStoDP\left(FC\left(\mathbf{e}_{ij}^{\prime(P)}\right)\right) \tag{7}$$

where $DStoDP(\cdot)$ is the conversion operation from the vector-formed direct sum state to the matrix-formed direct product state, which is more commonly used in the down-stream computational tasks based on Hamiltonians.

4.4. Training

Denote the ground truth Hamiltonian label for the atom pair (i, j) as $\mathbf{H}_{ij}^*$, in the first stage, parameters of the first-stage network are optimized by minimizing $MSE(\hat{\mathbf{H}}_{ij}, \mathbf{H}_{ij}^*)$, while in the second stage, parameters of the second-stage network are optimized by minimizing $MSE(\Delta\hat{\mathbf{H}}_{ij}, (\mathbf{H}_{ij}^* - \hat{\mathbf{H}}_{ij}))$.

4.5. Asymptotic time complexity

The time complexity for both the first-stage and second-stage networks is $\mathcal{O}(BN\overline{E})$, where B refers to the number of basic blocks, N is the total number of atoms in the system, and $\overline{E}$ represents the average number of neighboring atoms within a cutoff distance from a given atom. When N is small, connections are typically formed between each of the atom pairs, resulting in $\overline{E}$ being approximately equal to N. However, as N grows larger, due to the locality of Hamiltonian interactions, the number of neighbors per atom tends to stabilize, meaning $\overline{E}$ remains roughly constant and does not scale significantly with N. Thus, for large atomic systems with very high N, $\overline{E} \ll N$, and likewise, $B \ll N$. In such scenarios, the time complexity of our framework approaches $\mathcal{O}(N)$. In contrast, the traditional DFT method [8] requires T iterative steps to decompose $N \times N$ matrices, resulting in a time complexity of $\mathcal{O}(TN^3)$. As N increases, the cubic dependence on N leads to a considerable computational cost, especially since T, the number of iterations needed to achieve self-consistency, is often quite large. This makes DFT simulations for large systems computationally prohibitive within reasonable time constraints. On the other hand, our approach offers a much more efficient solution for large-scale atomic simulations by providing linear time complexity with respect to N and avoiding the need for multiple iterative steps.

5. Experiments

5.1. Experimental conditions

We evaluate our method on both crystalline material structures with spatial periodicity and single-cluster molecular systems.

The experiments for crystalline materials involve six benchmark databases, including monolayer graphene (abbreviated as *MG*), monolayer MoS2 (*MM*), bilayer graphene (*BG*), bilayer bismuthene (*BB*), bilayer Bi2Te3 (*BT*), and bilayer Bi2Se3 (*BS*), which were released by the DeepH projects [4, 6]. These databases are diverse and representative for periodic systems, as they cover atomic structures with strong chemical bonds within individual layers and weak vdW interactions between two layers; and include varied degrees of spin–orbit coupling (SOC), featuring both strong SOC samples like *BT*, *BB* and *BS*, and others with weak SOC. These atomic structures hold significant potential and value in the information science and technology sectors. A concise overview of these databases is presented in figure 2. Predicting the Hamiltonian accurately for these atomic structures poses a significant challenge due to the presence of structural deformations caused by thermal motions and inter-layer twists, as shown in figure 4. It is worth noting, the twisted structures have become a research hotspot due to their potentials for new electrical and quantum topological properties [31–33]. During the twist transformations, the relative rotation between atoms and the coordinate system brings the corresponding SO(3)-equivariant effects; meanwhile, the change in orientations between two layers of atoms causes variations in vdW interactions. These combined effects present a challenge to both of the equivariance and expressiveness capabilities of the regressor. In our experiments, the twisted subsets are even challenging as there are no such samples in the training set. For experiments on these databases, the training, validation, and testing sets as well as the data pre-processing protocols and optimizer (Adam) we use are the same as [6].

The experiments for molecular systems involve two benchmark databases, which consist of structures containing elements such as C, H, O, N, and F. The *QS* database consists of numerous stable molecular structures, while the *QD* database is composed of trajectories capturing multiple temporal frames from molecular structures undergoing thermal motion. The statistical details of these databases are shown in figure 3. For the *QS* database, the 'ood' strategy from the official settings [24] is employed to divide the training, validation, and test sets, ensuring that the number of atoms in the samples does not overlap between these subsets. This strategy aims at evaluating the model's ability to generalize across the number of atoms. For the *QD* database, and the 'mol' strategy is applied to partition the data, ensuring that no thermal motion samples from the same temporal trajectory appear in multiple subsets, i.e. the training, validation and testing sets. This strategy is designed to evaluate the model's generalization performance to new temporal trajectories. For experiments on *QS* and *QD*, the training, validation, and testing sets as well as the data pre-processing protocols and optimizer (AdamW) we use are the same as [12].

We use a fixed random seed for all procedures involving randomness, such as parameter initialization, data loader, as well as the rigid rotational augmentation introduced during training. Except for the network modules we need to compare with, i.e. those from DeepHE3 [6] or QHNet [12], the hyper-parameters of our framework are determined based on model selection on the validation sets; for DeepHE3 and QHNet, we adopt the optimal hyper-parameters officially provided.

5.2. Experimental results on the crystalline material databases

On the six crystalline material databases from DeepH, we design a set of experimental settings that include not only our complete method but also ablation and comparison settings as follows:

- DeepHE3 [6]. This experimental setting evaluates the performance of using only the DeepHE3 architecture to predict Hamiltonians. DeepHE3 will serve as the first-stage network of our framework in the following experiments on crystalline material databases. Although our first stage is designed with the flexibility to employ any combination of operators with prior SO(3)-equivariance, for a clear and fair comparison, we here opt to use the architecture of DeepHE3 which consists of node and edge encoders with abundant priorly-equivariant operators [10], ensuring our experimental results can be reliably compared to the established SOTA method on the databases from DeepH. We re-reproduce DeepHE3 with its latest resources[4], and the results are slightly better than those reported in their original paper.
- $S1_{\mathrm{DeepHE3}} + S2_{\mathrm{DeepHE3}}$. This experimental setting arranges two DeepHE3 networks as a two-stage (stage1 and stage2 abbreviated as $S1$ and $S2$, respectively, hereafter) cascaded regression framework for Hamiltonian prediction.

[4] https://github.com/Xiaoxun-Gong/DeepH-E3.

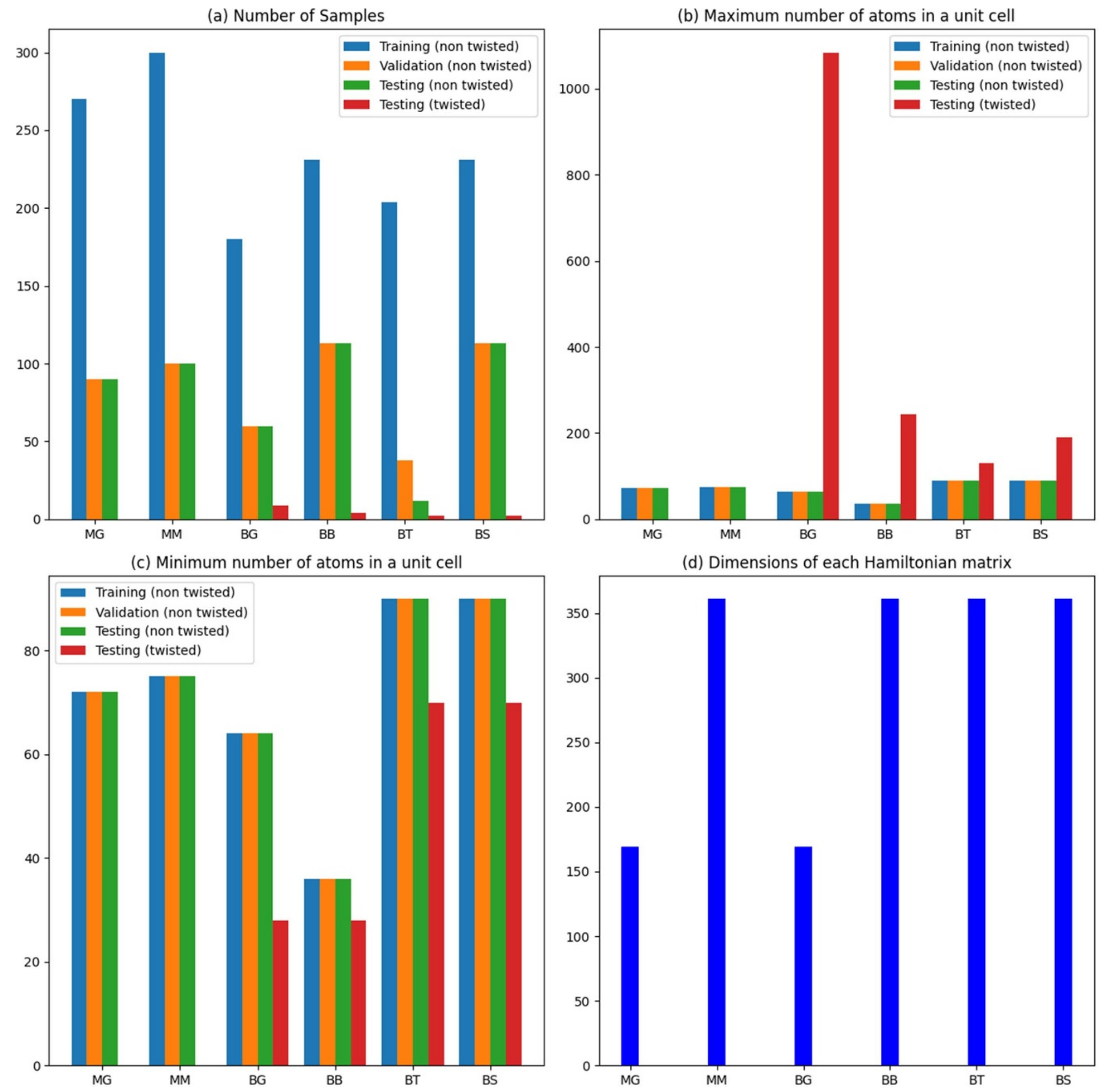


Figure 2. Visualization of the statistical information for the six crystal databases: monolayer graphene (MG), monolayer MoS2 (MM), bilayer graphene (BG), bilayer bismuthene (BB), bilayer Bi2Te3 (BT), and bilayer Bi2Se3 (BS). The four subfigures show: (a) the number of training, validation, and testing samples (including both non-twisted and twisted samples); (b) the maximum number of atoms in a unit cell; (c) the minimum number of atoms in a unit cell; and (d) the dimensions of the Hamiltonian matrix for each database.

- GFormer. This experimental setting evaluates the performance of using only the proposed non-linear graph Transformer (abbreviated as GFormer) architecture to predict Hamiltonians. To facilitate the empirical learning of SO(3)-equivariance for non-linear modules, rigid rotational data augmentation on the training samples is introduced.
- $S1_{\mathrm{GFormer}} + S2_{\mathrm{GFormer}}$. This experimental setting arranges two of the non-linear graph Transformer networks as a two-stage cascaded regression framework for Hamiltonian prediction.
- $S1_{\mathrm{DeepHE3}} + S2^{-\mathrm{cas}}_{\mathrm{GFormer}}$. This experimental setting retains the two-stage encoding framework, only removing the cascaded regression strategy at the output level by directly taking the second stage to predict the entire Hamiltonian targets. This setup is used to exactly examine the necessity of the cascaded regression strategy, thus only removing the prediction results of the first stage network at the output level, while retaining the first stage's support for the second stage at the feature level.
- $S1_{\mathrm{DeepHE3}} + S2^{-\mathrm{cov}}_{\mathrm{GFormer}}$. This experimental setting retains the two-stage regression framework, only removing the mechanism of flowing features with prior covariance into the input layers of Transformer blocks in the second stage. This setup is used to examine the necessity of these covariant features in assisting the non-linear graph Transformer network at learning SO(3)-equivariance.
- $S1_{\mathrm{DeepHE3}} + S2^{-\mathrm{att}}_{\mathrm{GFormer}}$. This experimental setting retains the two-stage encoding and regression framework and their corporation at both feature level as well as output level, only removing the attention mechanism

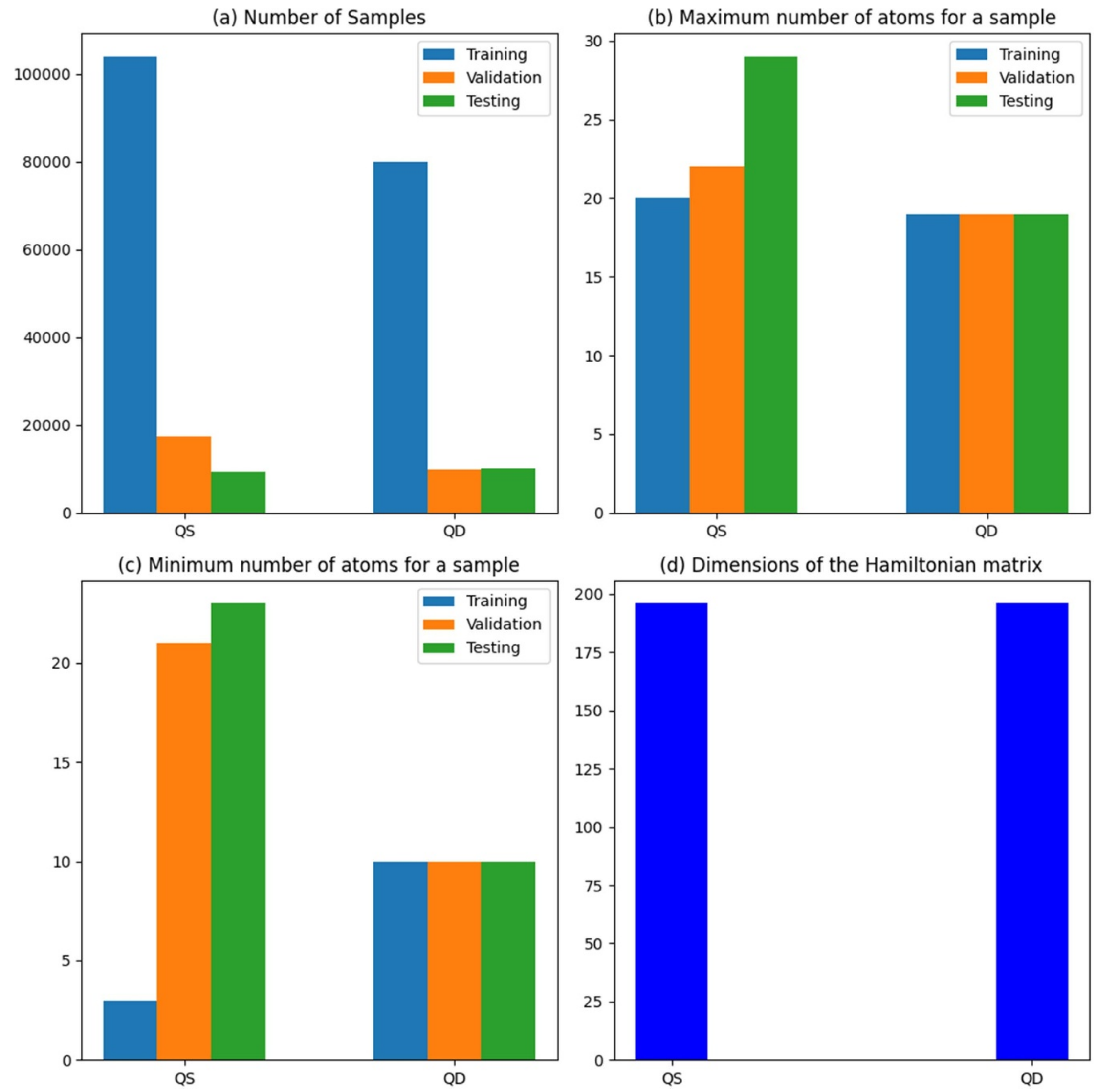


Figure 3. Visualization of the statistical information for the two molecular databases from QH9: QH9-Stable (QS) and QH9-Dynamic (QD). The subplots illustrate: (a) the number of training, validation, and testing samples; (b) the maximum number of atoms in a sample; (c) the minimum number of atoms in a sample ; and (d) the dimensions of the Hamiltonian matrix. For the QS and QD databases, the 'ood' strategy and the 'mol' strategy [24] are respectively applied to split the training, validation, and testing sets.

from the Transformer and replacing it by mixing neighboring features by stationary averages similar to DeepHE3. This setup is used to examine the necessity of the multi-head attention mechanism.

- *Ours*@$(S1_{\text{DeepHE3}} + S2_{\text{GFormer}})$. An implementation of our whole framework with mechanisms of DeepHE3 as well as the proposed graph Transfomer respectively serve as the two encoding and regression stages.

Experimental results of our complete method as well as the compared experimental settings on the six benchmark databases are presented in tables 1–3, respectively detailing the results for monolayer structures, as well as the results for non-twisted and twisted samples of bilayer structures. In these tables, the MAE metric is used as the accuracy metric. Besides the classical MAE metric, denoted as MAE^{H}_{all}, which measures the average error among all testing samples, we also record $MAE^{H}_{\text{cha\_s}}$, the MAE for the most challenging sample where the baseline (DeepHE3) performs the worst. In addition to taking the Hamiltonian of each edge as a whole for accuracy statistics, since the Hamiltonian matrix in the direct product state is constituted by several basic blocks based on the angular momentum of interacting orbitals, we also conduct fine-grained accuracy statistics on these basic blocks. The MAE metrics (denoted as MAE^{H}_{block}) of our method and DeepHE3 on different blocks of the Hamiltonian matrix of the six structures are presented in figures 5 and 6, where figure 5 illustrates the results on monolayer structures, while figure 6 is dedicated to the bilayer structures. In these figures, we specifically highlight the MAE values (denoted as $MAE^{H}_{\text{cha\_b}}$) on the Hamiltonian block where the baseline model DeepHE3 performs the worst for comparison on challenging

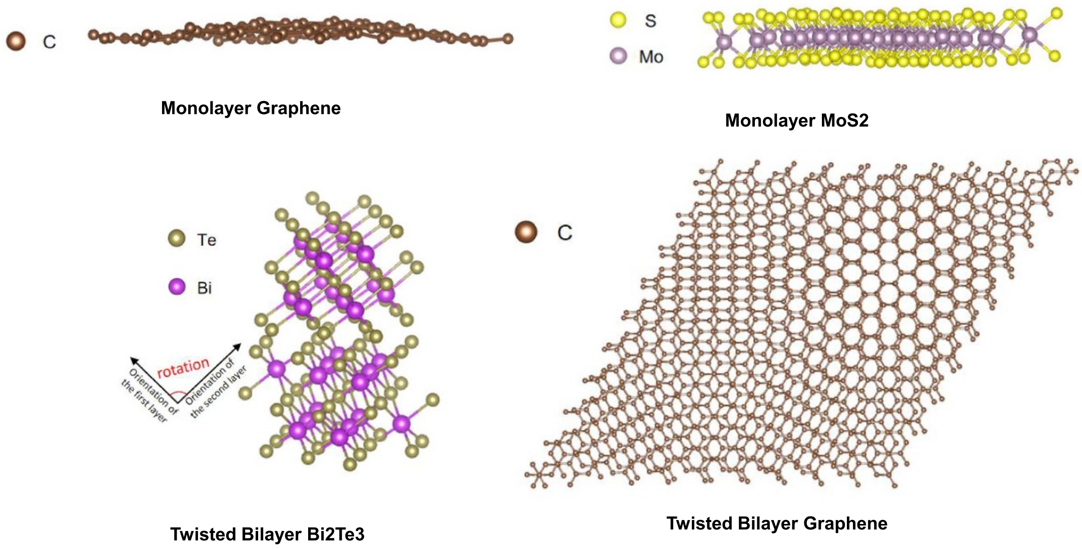


Figure 4. Visualization of challenging testing samples, which exhibit structural deformations caused by thermal motions and inter-layer twists, calling for strong capabilities on expressiveness and SO(3)-equivariance of a regression model.

Table 1. Comparison of experimental results measured by MAE^{H}_{all} as well as $MAE^{H}_{\mathrm{cha\_s}}$ (meV) on the monolayer structures, including monolayer graphene (MG) and monolayer MoS2 (MM). Bold is used to indicate the best performance.

	MG		MM	
	MAE (↓)			
Method	MAE^{H}_{all}	$MAE^{H}_{\mathrm{cha\_s}}$	MAE^{H}_{all}	$MAE^{H}_{\mathrm{cha\_s}}$
DeepHE3	0.251	0.357	0.406	0.574
$S1_{\mathrm{DeepHE3}} + S2_{\mathrm{DeepHE3}}$	0.239	0.338	0.392	0.499
GFormer	0.816	0.897	1.025	1.250
$S1_{\mathrm{GFormer}} + S2_{\mathrm{GFormer}}$	0.653	0.720	0.911	0.923
$S1_{\mathrm{DeepHE3}} + S2^{-\mathrm{cas}}_{\mathrm{GFormer}}$	0.774	0.880	0.927	0.958
$S1_{\mathrm{DeepHE3}} + S2^{-\mathrm{cov}}_{\mathrm{GFormer}}$	0.243	0.328	0.384	0.414
$S1_{\mathrm{DeepHE3}} + S2^{-\mathrm{att}}_{\mathrm{GFormer}}$	0.221	0.297	0.319	0.366
$Ours@(S1_{\mathrm{DeepHE3}} + S2_{\mathrm{GFormer}})$	**0.176**	**0.267**	**0.233**	**0.293**

blocks. All presented results are the mean values from ten independent repeat experiments. Since a fixed random seed is used, the standard deviation is less than 0.008 meV, a small value that can be ignored.

Observed from tables 1–3, and figures 5 and 6, we find that the proposed hybrid approach significantly improves the prediction accuracy beyond what is achievable solely with the priorly-equivariant mechanisms in DeepHE3 or the non-linear mechanisms in GFormer, demonstrating the effectiveness on leveraging the complementarity of the two categories of neural mechanisms to overcome their respective challenges comprehensively analyzed in previous sections. On one hand, the highly expressive non-linear mechanisms in the Transformer effectively compensate for the limitations in non-linear expressiveness of DeepHE3's mechanisms, thus significantly lowering down the MAE^{H}_{all} and $MAE^{H}_{\mathrm{cha\_s}}$ of DeepHE3 by up to 42% and 48%, respectively, and decreasing MAE^{H}_{block} for the vast majority of basic blocks, particularly for those blocks where DeepHE3 performs the worst, with the maximum reduction in $MAE^{H}_{\mathrm{cha\_b}}$ exceeding 66%. On the other hand, the mechanisms of DeepHE3 help the non-linear mechanisms in the Transformer to better learn equivariance from the data and reduces the difficulty for the Transformer on regressing SO(3)-equivariant targets, thus also significantly promoting the results from merely using the Transformer network. From the experimental results, it is observed that without any prior information on covariance, solely relying on the non-linear Transformer architecture to learn SO(3)-equivariance from data is extremely challenging despite rotational augmentation, due to the complexity and high-dimensionality nature of Hamiltonians as shown in figure 2(d). And since SO(3)-equivariance is strongly linked to the intrinsic mathematical structure of Hamiltonians, the weakness on capturing SO(3)-equivariance results in inadequate modeling of Hamiltonians, leading to inaccurate predictions, especially for samples that exhibit obvious SO(3)-equivariant effects, such as twisted samples. Our framework mitigates this challenge by incorporating mechanisms with prior covariance, which helps the non-linear mechanisms in the Transformer to learn

Table 2. Experimental results on the non-twisted subsets (marked with superscripts *nt*) of the bilayer structures including bilayer graphene (BG), bilayer bismuthene (BB), bilayer Bi2Te3 (BT), and bilayer Bi2Se3 (BS). The symbol ↓ indicates that lower values of the corresponding metrics imply better accuracy. The units for the MAE metrics are given in meV. Bold is used to indicate the best performance.

	BG^{nt}		BB^{nt}	
	MAE (↓)			
Method	MAE^{H}_{all}	$MAE^{H}_{\text{cha\_s}}$	MAE^{H}_{all}	$MAE^{H}_{\text{cha\_s}}$
DeepHE3	0.389	0.453	0.274	0.304
$S1_{\text{DeepHE3}} + S2_{\text{DeepHE3}}$	0.372	0.434	0.268	0.292
GFormer	1.295	1.483	0.886	0.949
$S1_{\text{GFormer}} + S2_{\text{GFormer}}$	0.786	0.828	0.785	0.816
$S1_{\text{DeepHE3}} + S2^{-\text{cas}}_{\text{GFormer}}$	0.854	0.920	0.802	0.873
$S1_{\text{DeepHE3}} + S2^{-\text{cov}}_{\text{GFormer}}$	0.365	0.427	0.249	0.281
$S1_{\text{DeepHE3}} + S2^{-\text{att}}_{\text{GFormer}}$	0.348	0.419	0.243	0.286
$Ours@(S1_{\text{DeepHE3}} + S2_{\text{GFormer}})$	**0.287**	**0.362**	**0.172**	**0.198**

	BT^{nt}		BS^{nt}	
	MAE (↓)			
Method	MAE^{H}_{all}	$MAE^{H}_{\text{cha\_s}}$	MAE^{H}_{all}	$MAE^{H}_{\text{cha\_s}}$
DeepHE3	0.447	0.480	0.397	0.424
$S1_{\text{DeepHE3}} + S2_{\text{DeepHE3}}$	0.435	0.471	0.389	0.410
GFormer	1.018	1.230	1.352	1.483
$S1_{\text{GFormer}} + S2_{\text{GFormer}}$	0.903	0.982	0.862	0.922
$S1_{\text{DeepHE3}} + S2^{-\text{cas}}_{\text{GFormer}}$	0.930	1.026	0.898	0.960
$S1_{\text{DeepHE3}} + S2^{-\text{cov}}_{\text{GFormer}}$	0.439	0.466	0.392	0.401
$S1_{\text{DeepHE3}} + S2^{-\text{att}}_{\text{GFormer}}$	0.385	0.414	0.348	0.375
$Ours@(S1_{\text{DeepHE3}} + S2_{\text{GFormer}})$	**0.294**	**0.321**	**0.282**	**0.308**

Table 3. Experimental results on the twisted subsets (marked with superscripts *t*) of the bilayer structures including bilayer graphene (BG), bilayer bismuthene (BB), bilayer Bi2Te3 (BT), and bilayer Bi2Se3 (BS). The symbol ↓ indicates that lower values of the corresponding metrics imply better accuracy. The units for the MAE metrics are given in meV. Bold is used to indicate the best performance.

	BG^{t}		BB^{t}	
	MAE (↓)			
Method	MAE^{H}_{all}	$MAE^{H}_{\text{cha\_s}}$	MAE^{H}_{all}	$MAE^{H}_{\text{cha\_s}}$
DeepHE3	0.264	0.429	0.468	0.602
$S1_{\text{DeepHE3}} + S2_{\text{DeepHE3}}$	0.257	0.423	0.460	0.595
GFormer	0.982	1.153	1.784	1.921
$S1_{\text{GFormer}} + S2_{\text{GFormer}}$	0.841	0.873	1.426	1.680
$S1_{\text{DeepHE3}} + S2^{-\text{cas}}_{\text{GFormer}}$	0.801	0.863	1.213	1.569
$S1_{\text{DeepHE3}} + S2^{-\text{cov}}_{\text{GFormer}}$	0.312	0.441	0.530	0.697
$S1_{\text{DeepHE3}} + S2^{-\text{att}}_{\text{GFormer}}$	0.278	0.428	0.504	0.669
$Ours@(S1_{\text{DeepHE3}} + S2_{\text{GFormer}})$	**0.227**	**0.403**	**0.438**	**0.578**

	BT^{t}		BS^{t}	
	MAE (↓)			
Method	MAE^{H}_{all}	$MAE^{H}_{\text{cha\_s}}$	MAE^{H}_{all}	$MAE^{H}_{\text{cha\_s}}$
DeepHE3	0.831	0.850	0.370	0.390
$S1_{\text{DeepHE3}} + S2_{\text{DeepHE3}}$	0.826	0.843	0.358	0.381
GFormer	2.682	2.827	1.785	2.037
$S1_{\text{GFormer}} + S2_{\text{GFormer}}$	2.190	2.379	1.624	1.953
$S1_{\text{DeepHE3}} + S2^{-\text{cas}}_{\text{GFormer}}$	1.892	1.937	1.569	1.892
$S1_{\text{DeepHE3}} + S2^{-\text{cov}}_{\text{GFormer}}$	0.928	0.943	0.415	0.456
$S1_{\text{DeepHE3}} + S2^{-\text{att}}_{\text{GFormer}}$	0.837	0.846	0.392	0.421
$Ours@(S1_{\text{DeepHE3}} + S2_{\text{GFormer}})$	**0.774**	**0.794**	**0.336**	**0.365**

equivariance from the data and reduces the difficulty for non-linear regression of SO(3)-equivariant targets, bringing satisfactory performance of the non-linear modules. The results on twisted samples demonstrate that, our method, despite the extensive use of non-linear operators, can still capture the symmetry properties

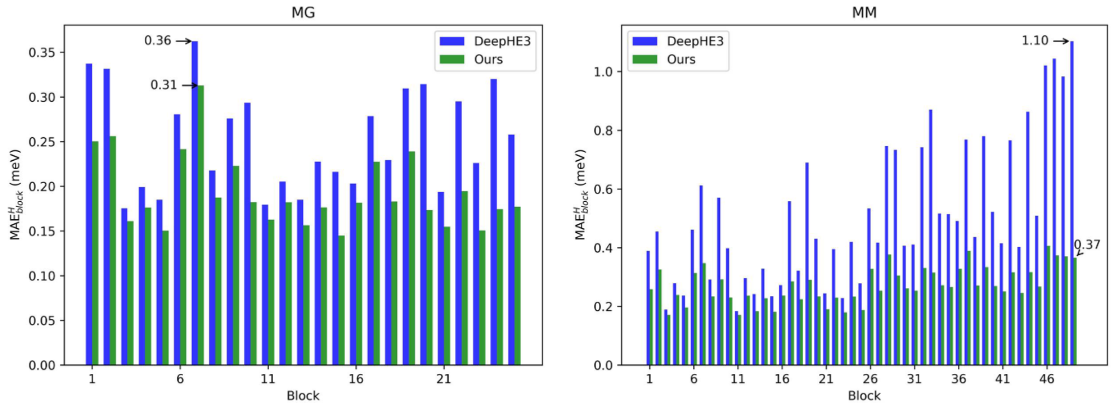


Figure 5. The MAE values (denoted as MAE^{H}_{block}) on various basic blocks of the Hamiltonian matrix in direct product state on the experimental monolayer structures, including monolayer graphene (MG) and Monolayer MoS2 (MM). The MAE values ($MAE^{H}_{\text{cha\_b}}$) on the Hamiltonian block where the baseline model DeepHE3 performs the worst are highlighted for comparison.

of Hamiltonians and make equivariant predictions under rotational operations. In contrast to this, the experimental results from $S1_{\text{DeepHE3}} + S2_{\text{DeepHE3}}$ and $S1_{\text{GFormer}} + S2_{\text{GFormer}}$ shows that simply scaling up the parameters for *DeepHE*3 or *GFormer* and fine-tuning it alone only yields limited improvements. This indicates that the bottlenecks encountered by these two categories of neural mechanisms might not be fully overcome through scaling up their sizes, further highlighting the superiority and necessity of our hybrid framework. Furthermore, as shown in table 4, our hybrid approach is also more efficient in inference compared to $S1_{\text{DeepHE3}} + S2_{\text{DeepHE3}}$. This is because the tensor product and Clebsch–Gordan decomposition operators in *DeepHE*3 incur significant computational overhead, whereas our second-stage graph Transformer network is more lightweight. This demonstrates that our hybrid encoding and decoding framework not only offers advantages in accuracy over simply expanding the *DeepHE*3 architecture, but also in efficiency.

As fine-grained ablation studies, by comparing the results of the three experimental settings, i.e. $S1_{\text{DeepHE3}} + S2^{-\text{cas}}_{\text{GFormer}}$, $S1_{\text{DeepHE3}} + S2^{-\text{cov}}_{\text{GFormer}}$, and $S1_{\text{DeepHE3}} + S2^{-\text{att}}_{\text{GFormer}}$, with our complete method, we could observe that the cascaded regression mechanism, the covariant feature integration mechanism, as well the multi-head attention mechanism, all contribute significantly to the performance of our method. The cascaded regression mechanism, by reducing the output space of the non-linear network, eases the difficulties on non-linear regression of Hamiltonians with SO(3)-equivariance; the covariant feature integration mechanism, through leveraging theoretical-guaranteed covariant features from DeepHE3 and geometric knowledge, successfully assists the non-linear network in learning SO(3)-equivariance; the multi-head attention mechanism, by assigning dynamic weights when fusing features, adapts to the wide variation range of geometric conditions, including both thermal deformations and twists. Under the combination of these mechanisms, networks from the two stage complement each other effectively, making our framework possess both excellent expressive capability and SO(3)-equivariant performance to achieve good results.

In figure 7, we illustrate the ground truths and prediction errors for two Hamiltonian matrices sampled from the monolayer graphene (MG) testing set. One matrix is formed by a single C atom with itself, and the other is formed by two C atoms approximately 5 Å apart. In each of the two sub-figures, the left heatmap shows the ground truth matrices, while the right two heatmaps compare the errors of the DeepHE3 method and our method. From these sub-figures, it is clear that our method brings smaller errors overall for the matrix elements in both cases, with the error reduction being particularly pronounced in the latter case.

5.3. Experimental results on the molecular databases

On the molecular databases, i.e. *QS* and *QD*, we compare the results of the following two experimental settings:

- QHNet [12]. A SOTA neural architecture composed of abundant prior SO(3)-equivariant modules, showing superior performance on the databases from QH9. The results of QHNet used for comparison are copied from the original paper [24]. For consistency in units, the units of MAE are converted from 10^{-6} Hartree (E_h) in the original literature to meV.
- *Ours*@($S1_{\text{QHNet}} + S2_{\text{GFormer}}$). An implementation of our whole framework with mechanisms of QHNet as well as the proposed graph Transfomer respectively serve as the two encoding and regression stages.

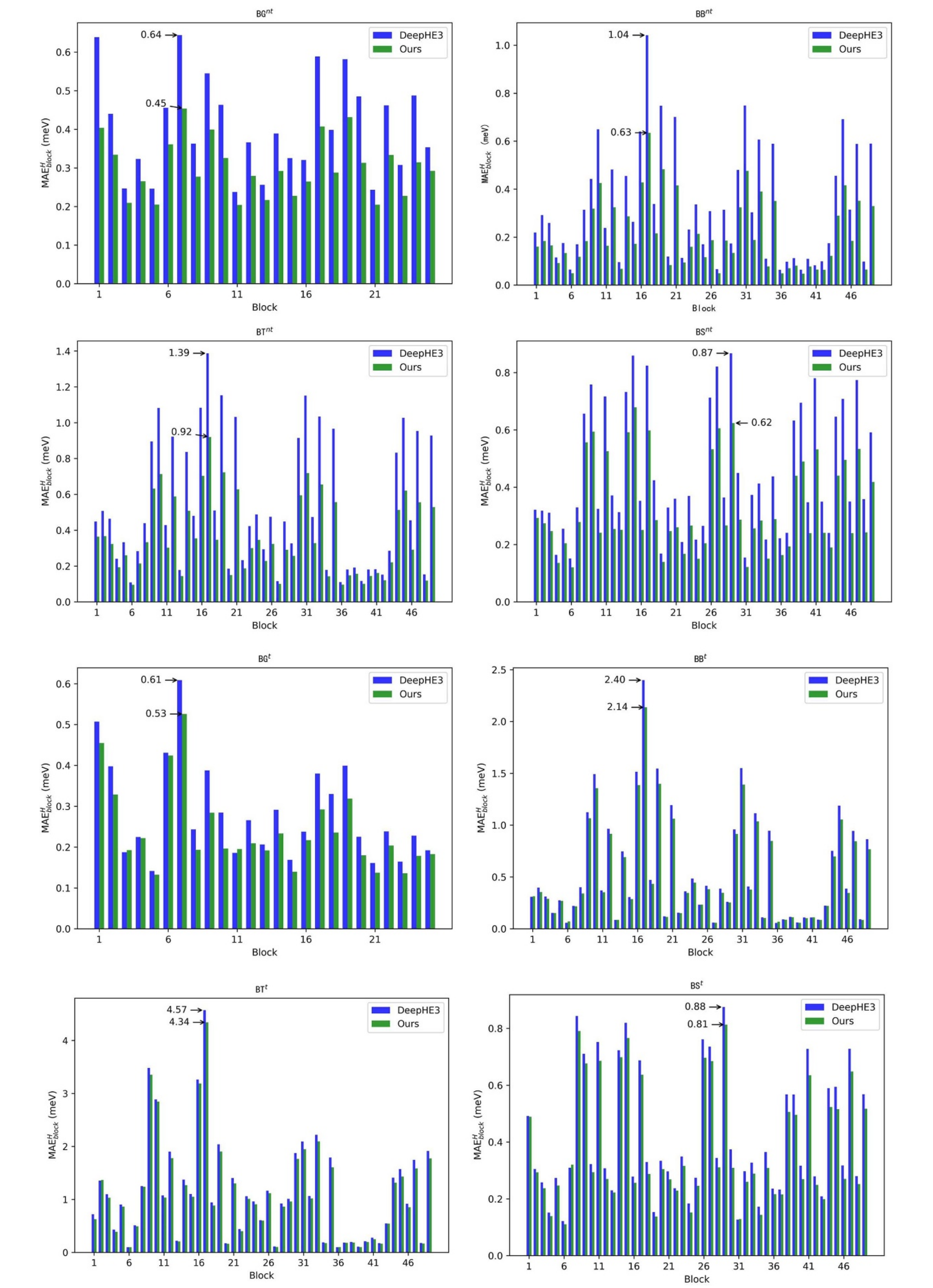
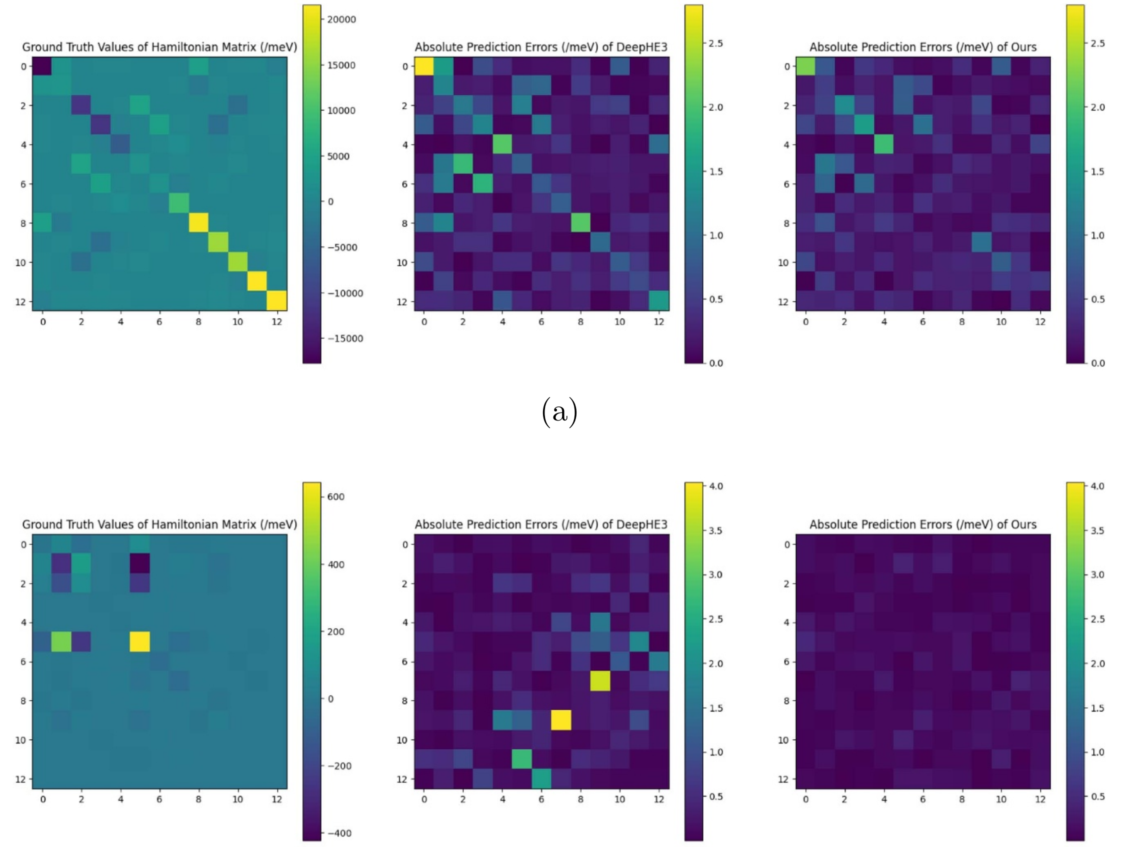


Figure 6. The MAE values (denoted as MAE^{H}_{block}) on various basic blocks of the Hamiltonian matrix in direct product state on the non-twisted (marked with superscripts nt) and twisted (marked with superscripts t) bilayer structures, including bilayer graphene (BG), bilayer bismuthene (BB), bilayer Bi2Te3 (BT), and bilayer Bi2Se3 (BS). The MAE values ($MAE^{H}_{\text{cha\_b}}$) on the Hamiltonian block where the baseline model DeepHE3 performs the worst are highlighted for comparison.

The experimental results of QHNet and our method, implemented as $Ours@(S1_{\text{QHNet}} + S2_{\text{GFormer}})$, are presented in tables 5 and 6, corresponding to the QS and QD databases, respectively. For these two databases, the evaluation MAE metrics outlined in the original work of QHNet [24] are adopted. These metrics include MAE^{H}_{all}, which represents the average MAE across all matrix elements, MAE^{H}_{diag}, which measures the MAE for Hamiltonian elements formed by an atom with itself, and $MAE^{H}_{\text{non\_diag}}$, which focuses on the MAE between

Table 4. Average inference time (ms/sample) on monolayer graphene (MG) and Monolayer MoS2 (MM) testing sets using a single RTX 4090 GPU with Pytorch 2.0.1.

	MG	MM
Method	Time	Time
DeepHE3	0.116	0.108
$S1_{\mathrm{DeepHE3}} + S2_{\mathrm{DeepHE3}}$	0.239	0.217
GFormer	0.018	0.019
$S1_{\mathrm{GFormer}} + S2_{\mathrm{GFormer}}$	0.064	0.070
$Ours@(S1_{\mathrm{DeepHE3}} + S2_{\mathrm{GFormer}})$	0.144	0.153



(a)

(b)

Figure 7. Visualization of the ground truth values of Hamiltonian matrices and the prediction errors. Sub-figure (a) and sub-figure (b) correspond to two sampled Hamiltonian matrices from the monolayer graphene (MG) test samples. (a) corresponds the Hamiltonian matrix formed by a single C atom with itself, and (b) corresponds the Hamiltonian matrix formed by two C atoms approximately 5 Å apart. In each subfigure, the left heatmap shows the ground truth matrix, the middle shows the absolute prediction error of the DeepH method, and the right shows the absolute prediction error of Our method.

Hamiltonians of different atoms. Additionally, we report the MAE (MAE^{ϵ}) of the occupied orbital energies ϵ, derived from the predicted Hamiltonians and compared against the ground truth values; and the cosine similarity ($Sim(\psi)$) between the predicted and ground truth electronic wavefunctions ψ calculated from Hamiltonians. Our method's results are the average of 10 independent runs. A fixed random seed for all randomized processes across experiments on both datasets is adopted, and as a result, the standard deviations of the Hamiltonian MAE across repeated experiments are very small, namely below 0.013 meV for *QS* and *QD*.

From tables 5 and 6, it is evident that our method significantly improves the Hamiltonian prediction accuracy over the baseline QHNet, as reflected in the metrics MAE^{H}_{all}, MAE^{H}_{diag}, and $MAE^{H}_{\mathrm{non\_diag}}$. Additionally, for the key physical quantities derived from the Hamiltonian, such as the occupied orbital energies ϵ and the electronic wavefunctions ψ, our method also achieves substantial improvements compared to QHNet. These two quantities are critical for understanding various electronic properties, including optical characteristics, conductivity, and chemical reactivity of atomic systems, as they directly

Table 5. Experimental results on the QH9-Stable (QS) database, which is divided according to the 'ood' strategy [24]. The symbol ↓ indicates that lower values of the corresponding metrics imply better accuracy, while ↑ indicates that higher values represent better performance. The units for the MAE metrics are given in meV, whereas $Sim(\psi)$ represents the cosine similarity, which is dimensionless. Bold is used to indicate the best performance.

	MAE (↓)				
Method	MAE^H_{all}	MAE^H_{diag}	$MAE^H_{\mathrm{non\_diag}}$	MAE^{ϵ}	$Sim(\psi)$ (↑)
QHNet	1.962	3.040	1.902	17.528	0.937
$Ours@(S1_{\mathrm{QHNet}} + S2_{\mathrm{GFormer}})$	**1.451**	**2.323**	**1.401**	**13.029**	**0.951**

Table 6. Experimental results on the QH9-Dynamic (QD) database, which is divided according to the 'ood' strategy [24]. The symbol ↓ indicates that lower values of the corresponding metrics imply better accuracy, while ↑ indicates that higher values represent better performance. The units for the MAE metrics are given in meV, whereas $Sim(\psi)$ represents the cosine similarity, which is dimensionless. Bold is used to indicate the best performance.

	MAE (↓)				
Method	MAE^H_{all}	MAE^H_{diag}	$MAE^H_{\mathrm{non\_diag}}$	MAE^{ϵ}	$Sim(\psi)$ (↑)
QHNet	4.733	11.347	4.182	264.483	0.792
$Ours@(S1_{\mathrm{QHNet}} + S2_{\mathrm{GFormer}})$	**3.465**	**8.510**	**3.062**	**175.024**	**0.846**

determine how electrons are distributed and interact within a system. The substantial improvements in both ϵ and ψ demonstrate that our method not only excels in Hamiltonian prediction but also holds great potential in capturing the essential physical properties of complex atomic systems.

5.4. Discussion on the scalability of our methods

The results presented in sections 5.2 and 5.3 underscore the strong scalability of our method. As shown in figure 2, the testing sets BG^t, BB^t, BT^t, and BS^t all contain unit cells that are larger than those in the training set, with BG^t being particularly notable, as its largest unit cell holds approximately 17 times more atoms than the largest in the training data. Similarly, as seen in figure 3, in the *QS* dataset, the 'ood' split strategy leads to a testing set with more atoms than the training set. The high accuracy achieved by our method on these testing sets highlights its impressive scalability and generalization to varying atomic sizes. Additionally, the improvements observed on the *QD* dataset, where the training, validation, and test sets are partitioned by distinct thermal motion trajectories using the 'mol' strategy [24], further demonstrate the robustness of our approach in handling previously unseen thermal motion sequences.

6. Limitations and future works

Despite the significant overall improvements achieved by our method across various databases and scenarios, there are still certain cases where the performance is not entirely satisfactory. For example, in the BT^t dataset, although there is some improvement compared to the baseline approach, the MAE of our method for the most challenging block remains higher than 4 meV, as shown in figure 6. This indicates that while our hybrid framework has made considerable progress, there is still room for further improvement in its generalization capabilities.

We summarize the limitations of our method as follows: First, while the non-linear operators, with the support of the prior SO(3)-equivariant operators, can capture SO(3)-equivariance and accurately regress SO(3)-equivariant Hamiltonian corrections in most cases, the SO(3)-equivariance learned from data-driven learning might still face challenges in rotational generalization when dealing with extremely sparse and exaggerated rotations in the data. Second, each type of operators, i.e. those with prior equivariance and those with non-linearity, has their distinct roles in our framework, and they could not replace each other, which introduces some additional computational overhead. In future research, we plan to explore the mathematical theory underlying SO(3)-equivariant non-linear representation learning to merge the advantages of both types of operators into a single, unified mechanism that theoretically harmonizes strict SO(3)-equivariance with non-linear expressiveness, further enhancing the computational efficiency and generalization capabilities for Hamiltonian prediction. Additionally, our method can be applied to scenarios such as material simulation, material design, atomic system dynamics modeling, and molecular medicine discovery, offering an efficient and high-precision tool for electronic structure calculations in these downstream tasks.

7. Conclusion

Deep learning for regressing electronic-structure Hamiltonian faces a pivotal challenge to capture SO(3)-equivariance without compromising neural expressiveness. To solve this, we propose a hybrid two-stage encoding and regression framework, where the first stage employs neural mechanisms inherent with SO(3)-equivariance properties prior to the learning process based on group theory, yielding baseline Hamiltonians with series of equivariant features assisting the subsequent stage on capturing SO(3)-equivariance. The second stage, leveraging the proposed non-linear 3D graph Transformer network for fine-grained structural analysis of 3D atomic systems, learns SO(3)-equivariant patterns from training data with the help of the first stage, while in turn, refines the initial Hamiltonian predictions via enhanced network expressiveness. Such a combination allows for accurate, generalizable Hamiltonian predictions while upholding good equivariant performance against rotational transformations. Our methodology demonstrates SOTA performance in Hamiltonian prediction, validated through **eight** benchmark databases, showing good potentials in high-performance deep modeling of atomic systems.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://drive.google.com/file/d/1jk5VP34I3z2_WkdJl6NTu0RICbpmTItM/view?usp = sharing.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant Number 12134012) and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant Number XDB0500201).

ORCID iDs

Shi Yin https://orcid.org/0009-0006-5459-5259
Xudong Zhu https://orcid.org/0000-0003-2006-7109
Lixin He https://orcid.org/0000-0003-2050-134X

References

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

[2] Unke O, Bogojeski M, Gastegger M, Geiger M, Smidt T and Müller K-R 2021 Se(3)-equivariant prediction of molecular wavefunctions and electronic densities *Advances in Neural Information Processing Systems* vol 34 pp 14434–47

[3] Gu Q, Zhang L and Feng J 2022 Neural network representation of electronic structure from *ab initio* molecular dynamics *Sci. Bull.* **67** 29–37

[4] Li H, Wang Z, Zou N, Ye M, Xu R, Gong X, Duan W and Xu Y 2022 Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation *Nat. Comput. Sci.* **2** 367–77

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

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

[7] Hohenberg P and Kohn W 1964 Inhomogeneous electron gas *Phys. Rev.* **136** B864

[8] Kohn W and Jeu Sham L 1965 Self-consistent equations including exchange and correlation effects *Phys. Rev.* **140** A1133

[9] Zhang X 2023 *et al* Artificial intelligence for science in quantum, atomistic, and continuum systems (arXiv:2307.08423)

[10] Geiger M and Smidt T E 2022 e3nn: Euclidean neural networks *CoRR* (arXiv:2207.09453)

[11] Zitnick L, Das A, Kolluru A, Lan J, Shuaibi M, Sriram A, Ulissi Z W and Wood B M 2022 Spherical channels for modeling atomic interactions *Conf. on Neural Information Processing Systems*

[12] Yu H, Xu Z, Qian X, Qian X and Ji S 2023 Efficient and equivariant graph networks for predicting quantum hamiltonian *Int. Conf. on Machine Learning* pp 40412–24

[13] Dieleman S, De Fauw J and Kavukcuoglu K 2016 Exploiting cyclic symmetry in convolutional neural networks *Int. Conf. on Machine Learning Workshop* vol 48 pp 1889–98

[14] Ravanbakhsh S, Schneider J G and Póczos B 2017 Equivariance through parameter-sharing *Int. Conf. on Machine Learning* pp 2892–901

[15] Kondor R, Truong Son H, Pan H, Anderson B M and Trivedi S 2018 Covariant compositional networks for learning graphs *Int. Conf. on Learning Representations Workshop*

[16] Passaro S and Lawrence Zitnick C 2023 Reducing SO(3) convolutions to SO(2) for efficient equivariant gnns *Int. Conf. on Machine Learning* pp 27420–38

[17] Liao Y-L, Wood B M, Das A and Smidt T E 2024 Equiformerv2: improved equivariant transformer for scaling to higher-degree representations *Int. Conf. on Learning Representations*

[18] Wang Y *et al* 2024 Universal materials model of deep-learning density functional theory hamiltonian *Sci. Bull.* **69** 2514–21

[19] Jaderberg M, Simonyan K, Zisserman A and Kavukcuoglu K 2015 Spatial transformer networks *Conf. on Neural Information Processing Systems* pp 2017–25

[20] Cohen T S and Welling M 2017 Steerable CNNS *Int. Conf. on Learning Representations*

[21] Thomas N, Smidt T, Kearnes S, Yang L, Li L, Kohlhoff K and Riley P 2018 Tensor field networks: rotation-and translation-equivariant neural networks for 3D point clouds (arXiv:1802.08219)

[22] Fuchs F, Worrall D E, Fischer V and Welling M 2020 Se(3)-transformers: 3D roto-translation equivariant attention networks *Conf. on Neural Information Processing Systems*

[23] Liao Y-L and Smidt T E 2023 Equiformer: Equivariant graph attention transformer for 3d atomistic graphs *Int. Conf. on Learning Representations*

[24] Yu H, Liu M, Luo Y, Strasser A, Qian X, Qian X and Ji S 2023 QH9: a quantum Hamiltonian prediction benchmark for QM9 molecules *Conf. on Neural Information Processing Systems*

[25] Wang Z, Liu C, Zou N, Zhang H, Wei X, Huang L, Lijun W and Shao B 2024 Infusing self-consistency into density functional theory hamiltonian prediction via deep equilibrium models *CoRR*, (arXiv:2406.03794)

[26] Zhang H, Liu C, Wang Z, Wei X, Liu S, Zheng N, Shao B and Liu T-Y 2024 Self-consistency training for density-functional-theory hamiltonian prediction *Int. Conf. on Machine Learning*

[27] Schrödinger E 1926 Quantisierung als eigenwertproblem *Ann. Phys., Lpz.* **384** 361–76

[28] Wang H, Zhang L, Han J and Weinan E 2018 Deepmd-kit: a deep learning package for many-body potential energy representation and molecular dynamics *Comput. Phys. Commun.* **228** 178–84

[29] Zhang Y, Hu C and Jiang B 2019 Embedded atom neural network potentials: efficient and accurate machine learning with a physically inspired representation *J. Phys. Chem. Lett.* **10** 4962–7

[30] Zhang Y and Jiang B 2023 universal machine learning for the response of atomistic systems to external fields *Nat. Commun.* **14** 6424

[31] Cao Y, Fatemi V, Fang S, Watanabe K, Taniguchi T, Kaxiras E and Jarillo-Herrero P 2018 Unconventional superconductivity in magic-angle graphene superlattices *Nature* **556** 43–50

[32] Wang C, Zhang X-W, Liu X, He Y, Xu X, Ran Y, Cao T and Di X 2024 Fractional chern insulator in twisted bilayer MoTe$_2$ *Phys. Rev. Lett.* **132** 036501

[33] He M *et al* 2024 Dynamically tunable moiré exciton rydberg states in a monolayer semiconductor on twisted bilayer graphene *Nat. Mater.* **23** 224–9