Efficient and Equivariant Graph Networks for Predicting Quantum Hamiltonian

Haiyang Yu [1] **Zhao Xu** [1] **Xiaofeng Qian** [2 3 4 *] **Xiaoning Qian** [1 3 *] **Shuiwang Ji** [1 *]

Abstract

We consider the prediction of the Hamiltonian matrix, which finds use in quantum chemistry and condensed matter physics. Efficiency and equivariance are two important, but conflicting factors. In this work, we propose a SE(3)-equivariant network, named QHNet, that achieves efficiency and equivariance. Our key advance lies at the innovative design of QHNet architecture, which not only obeys the underlying symmetries, but also enables the reduction of number of tensor products by 92%. In addition, QHNet prevents the exponential growth of channel dimension when more atom types are involved. We perform experiments on MD17 datasets, including four molecular systems. Experimental results show that our QHNet can achieve comparable performance to the state of the art methods at a significantly faster speed. Besides, our QHNet consumes 50% less memory due to its streamlined architecture. Our code is publicly available as part of the AIRS library (https://github.com/divelab/AIRS).

1. Introduction

Deep learning has achieved significant progress in computational quantum chemistry in recent years. Existing deep learning methods have demonstrated their efficiency and expressiveness in tackling various challenging quantum mechanical simulation tasks. For example, deep graph learning methods can now accurately predict quantum properties of a molecule, such as molecular energy and the HOMO-LUMO gap (Schütt et al., 2017; Liu et al., 2022b; Gasteiger et al., 2020; Klicpera et al., 2020; Wang et al., 2022b; Liu et al., 2021; Wang et al., 2022a; Brandstetter et al., 2022; Yan et al., 2022). Recent deep generative models have also shown to be capable of generating new materials and molecules by faithfully learning the distribution of their structures (Simm & Hernandez-Lobato, 2020; Mansimov et al., 2019; Xu et al., 2021b; Shi et al., 2021; Xu et al., 2021a; Ganea et al., 2021; Luo & Ji, 2022). Recent efforts on modeling interaction of two and more molecules also shed light on protein-ligand docking for drug development (Corso et al., 2022; Ganea et al., 2022; Stärk et al., 2022; Zhang et al., 2022; Lu et al., 2022; Jiang et al., 2022; Liu et al., 2022a). Inspired by all these advancements, we aim to predict a more fundamental target in computational physics, quantum tensors, by developing a new deep graph learning model framework in this work.

Quantum tensors, such as Hamiltonian matrix, eigen wavefunctions, and eigen energies, can be used to describe molecular systems and their quantum states. Since quantum tensors contain the most critical information about molecular systems, many molecular properties can be directly derived from quantum tensors and wavefunctions. Unfortunately, obtaining precise quantum tensors is at considerably high cost. Density functional theory (DFT) (Hohenberg & Kohn, 1964; Kohn & Sham, 1965) and *ab initio* quantum chemistry methods (Szabo & Ostlund, 2012) are routinely used to calculate electronic wavefunctions, charge density, and total energy of molecules and solids. However, first-principles methods are computationally very expensive, limiting their use in small systems. Therefore, deep learning is believed to have the potential to accelerate quantum mechanical simulations if it can accurately and reliably predict quantum tensors (Schütt et al., 2019; Unke et al., 2021) .

Unlike invariant molecular properties such as energy and equivariant properties such as atomic forces, quantum tensors possess a higher rotation order to reflect the exact rotation of a molecule. Therefore, developing deep learning methods to predict quantum tensors is challenging and requires elaborate designs of model architectures. SE(3)-equivariant graph neural networks (Satorras et al., 2021; Schütt et al., 2021; Thomas et al., 2018; Anderson et al., 2019; Gasteiger et al., 2021) have the promising potential in predicting quantum tensors since they ensure the equivariance of permutation, translation, and rotation. Output quantum tensors are guaranteed to be permuted, translated,

[*]Equal Senior contribution [1]Department of Computer Science & Engineering, Texas A&M University, TX, USA [2]Department of Materials Science & Engineering, Texas A&M University, TX, USA [3]Department of Electrical & Computer Engineering, Texas A&M University, TX, USA [4]Department of Physics & Astronomy, Texas A&M University, TX, USA. Correspondence to: Shuiwang Ji <sji@tamu.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

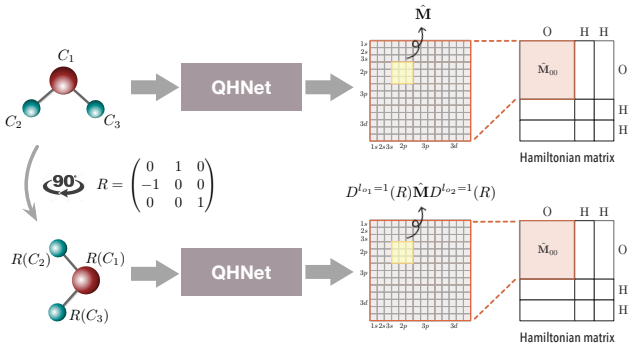

Figure 1. The rotation equivariance relationship between the input molecule and output Hamiltonian matrix.

and rotated in the same way as the input molecule shown in Figure 1. However, the efficiency of SE(3)-equivariant models are generally low as a trade-off for equivariance. In this work, we propose an efficient and equivariant graph network, named QHNet, for predicting quantum tensors including Hamiltonian matrices.

The high efficiency of the proposed QHNet is mainly because QHNet adopts much fewer tensor product (TP) operations than existing SE(3)-equivariant graph networks. Note that TP operation causes most of the time overheads though they are critical for learning quantum tensors. As a result, QHNet boosts efficiency by more than three times and reduces 50% of GPU memory, which is a significant advance compared to existing SE(3)-equivariant networks. With such an efficient architecture, our QHNet surprisingly achieves comparable performance to the state of the art methods.

In addition to better efficiency and prediction performance, QHNet has another advantage of more flexible applicability. The expansion module of QHNet outputs intermediate blocks with a fixed shape for all node pairs despite atom types. The fixed shape is predefined by full orbitals needed for a particular dataset. Then, the pairwise blocks of every two atoms are extracted to build the quantum tensor according to specific atomic orbitals. This design enables an easy extension to cases where more atom types are involved, *i.e.*, the model architecture is not affected by different types of atoms. Therefore, given the advantages of efficacy, efficiency, and applicability, our QHNet is an excellent choice for predicting quantum tensors such as quantum Hamiltonian, in more general molecular systems.

2. Background and Related Work

2.1. Quantum Tensors in Density Functional Theory

Quantum mechanical methods such as first-principles density functional theory (DFT) (Hohenberg & Kohn, 1964; Kohn & Sham, 1965) and *ab initio* quantum chemistry methods (Szabo & Ostlund, 2012), are widely used to calculate the electronic structure of solids and molecules, such as electronic wavefunctions, charge density, total energy etc. These quantities can in turn be used to derive many other physical and chemical properties, for examples, electronic, mechanical, optical, magnetic, and catalytic properties of molecules and solids. DFT maps a many-body interacting system onto a many one-body non-interacting system, allowing to efficiently solve the Schrödinger equation for complex materials and molecules. In principle, DFT with exact exchange-correlation energy functional yields exact ground-state density and total energy. The Schrödinger equation for an n-electron interacting system is given by,

$$\hat{\boldsymbol{H}}\Psi\left(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n\right)=E\Psi\left(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n\right), \tag{1}$$

where Ψ is the total wavefunction of all electrons, $\hat{\boldsymbol{H}}$ is the Hamiltonian operator, and E is the total energy. Here, the ionic degree of freedom is not considered. DFT essentially decouples the wavefunction from an interactive high dimensional function into a combination of non-interacting univariate wavefunctions as

$$\Psi\left(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n\right)=\psi_1\left(\boldsymbol{r}_1\right)\psi_2\left(\boldsymbol{r}_2\right)\ldots\psi_n\left(\boldsymbol{r}_n\right), \tag{2}$$

where n is the number of electrons, $\boldsymbol{r}_i$ is the coordinates of the i-th electron, and ψ_i is the wavefunction for the i-th electron. In molecular systems, ψ_i usually represents the i-th molecular orbital. Then ψ_i is approximated by a linear combination of predefined basis set ϕ_j as

$$\psi_i(\boldsymbol{r})=\sum_j \mathbf{C}_{ij}\phi_j(\boldsymbol{r}), \tag{3}$$

where $\mathbf{C}\in\mathbb{R}^{n\times n}$ denotes the molecular orbital coefficients of interest, although it could have complex values for solids etc. Derived from the Schrödinger equation in Eq. (1), the coefficients $\mathbf{C}$ for the molecular ground state satisfy the eigenvalue decomposition equation (Szabo & Ostlund, 2012),

$$\mathbf{HC}=\boldsymbol{\epsilon}\mathbf{SC}, \tag{4}$$

where $\mathbf{H}$ is the Hamiltonian matrix with $\mathbf{H}_{ij}=\langle\phi_i|\hat{\mathbf{H}}|\phi_j\rangle=\int\phi_i^*(\boldsymbol{r})\hat{\mathbf{H}}\phi_j(\boldsymbol{r})d\boldsymbol{r}$, $\mathbf{S}$ is the overlap matrix with $\boldsymbol{S}_{ij}=\langle\phi_i|\phi_j\rangle=\int\phi_i^*(\boldsymbol{r})\phi_j(\boldsymbol{r})d\boldsymbol{r}$ which represents the integral of a pair of the predefined basis functions, and each element in the diagonal matrix $\boldsymbol{\epsilon}$ denotes the energy of the corresponding molecular orbital. Note that $\mathbf{H},\mathbf{S},\epsilon\in\mathbb{R}^{n\times n}$ in the present work.

The ultimate goal of DFT is to calculate these quantum tensors $\mathbf{H}$, $\mathbf{C}$, and ϵ, from which other properties such as total energy and atomic forces etc. can be readily computed. By utilizing the self-consistent field (SCF) methods (Payne et al., 1992; Cances & Le Bris, 2000; Kudin et al., 2002),

DFT can approximate the solution iteratively with a time complexity of $O(n^3)$ for each step, resulting in significant cost for running DFT simulations for large systems. This is particularly true when multiple iterations have to be conducted to obtain the final quantum tensors in settings in which a high level of accuracy is required. With the recent advances and successes of machine learning approaches for scientific computing, we here propose to develop deep learning models to predict quantum tensors within a reasonable level of accuracy, thereby accelerating the optimization process in the electronic structure calculations with better accuracy.

2.2. Related Work

In recent years, graph neural networks have yielded promising results in quantum chemistry by solving problems such as precise prediction of molecular properties (Gilmer et al., 2017; Liu et al., 2022b; Fuchs et al., 2020; Batzner et al., 2022; Schütt et al., 2021; Godwin et al., 2021; Satorras et al., 2021; Unke & Meuwly, 2019; Gasteiger et al., 2020; Qiao et al., 2020). A molecule usually has various quantum properties that describe the molecule from different perspectives. We divide these properties into three categories based on their rotation orders. A typical example of the first category is molecular energy, including the total energy and the HOMO-LUMO gap, as molecular energy is invariant to molecular rotation. The force falls into the second category because force vectors rotate as a molecule rotates. Thus, force predictions have to be equivariant to rotations of the same order. The third category includes a higher rotation order of the matrix that indicates the exact rotation. The Hamiltonian matrix is an example in this category and is more challenging to predict. To predict molecular properties with different rotation orders, different variants of invariant and equivariant GNNs have been proposed accordingly.

Invariant GNNs use relative geometric information of molecules as input and predict invariant molecular properties. For instance, SchNet (Schütt et al., 2018) considers pairwise distances when performing message passing, while DimeNet (Gasteiger et al., 2020) also uses angles in addition to distance. SphereNet (Liu et al., 2022b) and ComENet (Wang et al., 2022a) incorporate distances, angles, and torsion angles to build more informative representations.

Equivariant GNNs use equivariant features and specific model architectures to predict the roto-equivariant molecular properties. For example, PaiNN (Schütt et al., 2021) maintains the equivariant features of rotation order $\ell = 1$ in the model architecture. Tenor field networks, SE3-transformers, and NequIP (Thomas et al., 2018; Fuchs et al., 2020; Batzner et al., 2022) use tensor products to incorporate lifted higher-order of equivariant features while prediction targets could be either invariant or equivariant with order $\ell = 1$.

To predict high-order quantum tensors, SchNorb (Schütt et al., 2019) follows SchNet (Schütt et al., 2018) to consider pairwise distances and incorporates direction information by applying pairwise direction vector on pairwise interaction features. It then constructs blocks of the Hamiltonian matrix. SchNorb considers system coordinates through the pairwise direction vector. However, it does not provide guarantee on yielding an equivariant matrix. In PhiSNet (Unke et al., 2021), tensor products (Thomas et al., 2018) are used to ensure equivariance. In this method, the equivariant atomic representation network is applied to extract equivariant features for each atom. The mixing layers are then applied to construct equivariant representations for each node pair. The Hamiltonian matrix is finally constructed with irreducible representations collected from the mapping from orbital interactions to channel indices. The DeepH (Li et al., 2022) uses the 3D GNN to predict the invariant Hamiltonian matrix block. It then applies the Wigner-D matrix to the predicted invariant matrix, thereby ensuring rotation equivariance.

3. Methodology

In this section, we describe our proposed SE(3) equivariant graph neural network, QHNet, for quantum tensor prediction tasks.

3.1. Tensor Field Networks

The tensor field network (TFN) (Thomas et al., 2018) is one of the commonly-used equivariant neural network architectures that achieve 3D rotation, translation, and permutation equivariance. TFN uses tensor product to combine two irreducible representations u and v of the rotation orders ℓ_1 and ℓ_2 with the Clebsch-Gordan (CG) coefficients (Griffiths & Schroeter, 2018) to produce a new irreducible representation of order ℓ_3 as

$$(u^{\ell_1} \otimes v^{\ell_2})^{\ell_3}_{m_3} = \sum_{m_1=-\ell_1}^{\ell_1} \sum_{m_2=-\ell_2}^{\ell_2} C^{(\ell_3,m_3)}_{(\ell_1,m_1),(\ell_2,m_2)} u^{\ell_1}_{m_1} v^{\ell_2}_{m_2},$$

where C denotes the CG matrix, ℓ_3 satisfies $|\ell_1 - \ell_2| \leq \ell_3 \leq \ell_1 + \ell_2$, and $\ell_1, \ell_2, \ell_3 \in \mathbb{N}$. Note that m denotes the m-th element in the irreducible representation with $-\ell \leq m \leq \ell$ and $m \in \mathbb{N}$. In the TFN (Thomas et al., 2018), each layer is composed of filter, convolution, self-interaction and nonlinear activation modules. First, spherical harmonic filters Y are applied to the node pair direction $\hat{r}_{ij}$, then combined with the pairwise distance r_{ij} to obtain the filter outputs F as

$$F^{(\ell_{\text{in}},\ell_f)}_{cm}(r_{ij}, \hat{r}_{ij}) = R^{(\ell_{\text{in}},\ell_f)}_c(r_{ij}) Y^{\ell_f}_m(\hat{r}_{ij}), \qquad (5)$$

where R is a multiple layer perceptron (MLP) that takes the embedding of pairwise distance as input, and c is the channel index. Then the convolution collects information

from the irreducible representations of other nodes and the pairwise spherical harmonics through the tensor product as

$$\tilde{V}_i^{\ell_{\text{out}}} = \sum_j (F^{(\ell_{\text{in}},\ell_f)}(r_{ij}, \hat{r}_{ij}) \otimes V_j^{\ell_{\text{in}}})^{\ell_{\text{out}}} \tag{6}$$

with $|\ell_{\text{in}} - \ell_f| \leq \ell_{\text{out}} \leq \ell_{\text{in}} + \ell_f$, and $\ell_{\text{in}}, \ell_{\text{out}}, \ell_f \in \mathbb{N}$. Then the self-interaction is a linear layer that combines features across the channel dimension to obtain irreducible representations with the same order ℓ as

$$V_{icm}^{\ell} = f_{\text{linear}}(V_i^{\ell})_{cm} = \sum_c w_{cc'} V_{ic'm}^{\ell}. \tag{7}$$

In the final nonlinear layer, a nonlinear function is applied to the invariant irreducible representations of $\ell = 0$. Meanwhile, the average norm $\sum_c \|V_{ic}^{\ell}\|$ is calculated over the channel for irreducible representations with $\ell > 1$, and V_i^{ℓ} is normalized with this average norm.

3.2. Norm Gate

The irreducible representations can be represented as the multiplication of their norm and direction vectors. When the norm of $\mathbf{x}_i$ is larger than other nodes, the message from node i has greater impact on node messaging updating. Instead of the layer norm operation in the original TFN layer, we perform norm gate learning to rescale the norm of irreducible representations before applying message passing and tensor product. The norm of order ℓ representations for node i, $\mathbf{x}_i^{\ell}$, is defined as

$$\mathbf{n}_i^{\ell} = \|\mathbf{x}_i^{\ell}\|. \tag{8}$$

We apply an MLP to learn the norm of new irreducible representations along with the new irreducible representations with order 0 as

$$\hat{\mathbf{x}}_i^0, \boldsymbol{\sigma}_i^1, \ldots \boldsymbol{\sigma}_i^{\ell_{\max}} = f_{\text{MLP}}(\mathbf{x}_i^0, \mathbf{n}_i^1, \ldots, \mathbf{n}_i^{\ell_{\max}}). \tag{9}$$

The irreducible representations with order higher than 0 are rescaled with the factor

$$\hat{\mathbf{x}}_{im}^{'\ell} = \boldsymbol{\sigma}_i^{\ell} \mathbf{x}_{im}^{\ell}. \tag{10}$$

Since $\mathbf{x}^0$ and the norm $\mathbf{n}_i^{\ell}$ are SE(3) invariant, the scale $\boldsymbol{\sigma}_i$ is also invariant. Therefore, the rescaled irreducible representations $\hat{\mathbf{x}}_i^{\ell}$ have the same SE(3) equivariance as $\mathbf{x}_i^{\ell}$. The illustration of norm gate is shown in Figure 2.A.

Usually, a self-interaction layer is applied after the norm gate to produce the representation $\hat{\mathbf{x}}_i^{\ell}$ as

$$\hat{\mathbf{x}}_i^{\ell} = f_{\text{linear}}(\hat{\mathbf{x}}_i^{'\ell}). \tag{11}$$

3.3. Node-Wise Interaction Layer

The filter module in the TFN layer controls influence of messages from other nodes. It takes pairwise distances as input and induces most of parameters in the TFN layer. Intuitively, when nodes are far from each other, their influence on each other should be weak.

In our node-wise interaction layers as illustrated in Figure 2.B, we further consider the similarity of pairwise nodes. We calculate the inner product of the pairwise irreducible representations of order ℓ as

$$\mathbf{I}_{ij}^{\ell} = \langle f_{\text{linear}}(\mathbf{x}_i^{\ell}), f_{\text{linear}}(\mathbf{x}_j^{\ell}) \rangle . \tag{12}$$

To increase expressive power, a self-interaction linear layer is applied before the inner product. The pairwise cosine similarity of irreducible representations are then fed into a MLP along with the irreducible representations of order 0 to calculate attentive scores as

$$a_{ij} = f_{\text{MLP}}(\mathbf{x}_i^0, \mathbf{x}_j^0, \mathbf{I}_{ij}^1, \ldots, \mathbf{I}_{ij}^{\ell_{\max}}). \tag{13}$$

Since $\mathbf{x}^0$ and the inner product I_{ij} are invariant, the attentive scores are also SE(3) invariant.

Even though the filter in the TFN layer assigns a weight for each $(\ell_{\text{in}}, \ell_f)$ according to Eq. (5), the weights are the same when the order of output features ℓ_{out} are different. To explicitly express the difference, we extend the filter to assign weights for each path $(\ell_{\text{in}}, \ell_f, \ell_{\text{out}})$. Furthermore, the filter assigns an attentive score for each path to consider pairwise similarities. Formally,

$$F_{cm}^{(\ell_{\text{in}},\ell_f,\ell_{\text{out}})}(r_{ij}, \hat{r}_{ij}) = a_{ijc}^{(\ell_{\text{in}},\ell_f,\ell_{\text{out}})} R_c^{(\ell_{\text{in}},\ell_f,\ell_{\text{out}})}(r_{ij}) Y_m^{\ell_f}(\hat{r}_{ij}). \tag{14}$$

Note that c denotes the channel index and m indexes the irreducible representations.

In the next step, we calculate the message $\mathbf{m}_{ij}$ from node j to node i. The irreducible representations $\mathbf{x}$ are first rescaled with a norm gate and self-interaction layer to obtain $\hat{\mathbf{x}}$. Then the rescaled representations cooperates with the filter to produce $\mathbf{m}_{ij}$ as

$$\mathbf{m}_{ijc}^{\ell_{\text{out}}} = \sum_{\ell_f, \ell_{\text{in}}} (F_c^{(\ell_{\text{in}},\ell_f,\ell_{\text{out}})}(r_{ij}, \hat{r}_{ij}) \otimes (\hat{\mathbf{x}}_{jc}^{\ell_{\text{in}}}))^{\ell_{\text{out}}}. \tag{15}$$

The output irreducible representations $\tilde{\mathbf{x}}$ are then obtained by aggregating the message $\mathbf{m}_{ij}$ and the self-connection and further updated with a self-interaction layer as

$$\tilde{\mathbf{x}}_i^{\ell} = f_{\text{linear}}(\hat{\mathbf{x}}_i^{\ell} + \sum_j \mathbf{m}_{ij}^{\ell}). \tag{16}$$

3.4. Building Quantum Tensor

Quantum tensors like Hamiltonian matrix characterize the pairwise relationship between atoms in molecular systems, and they can be divided into diagonal and non-diagonal

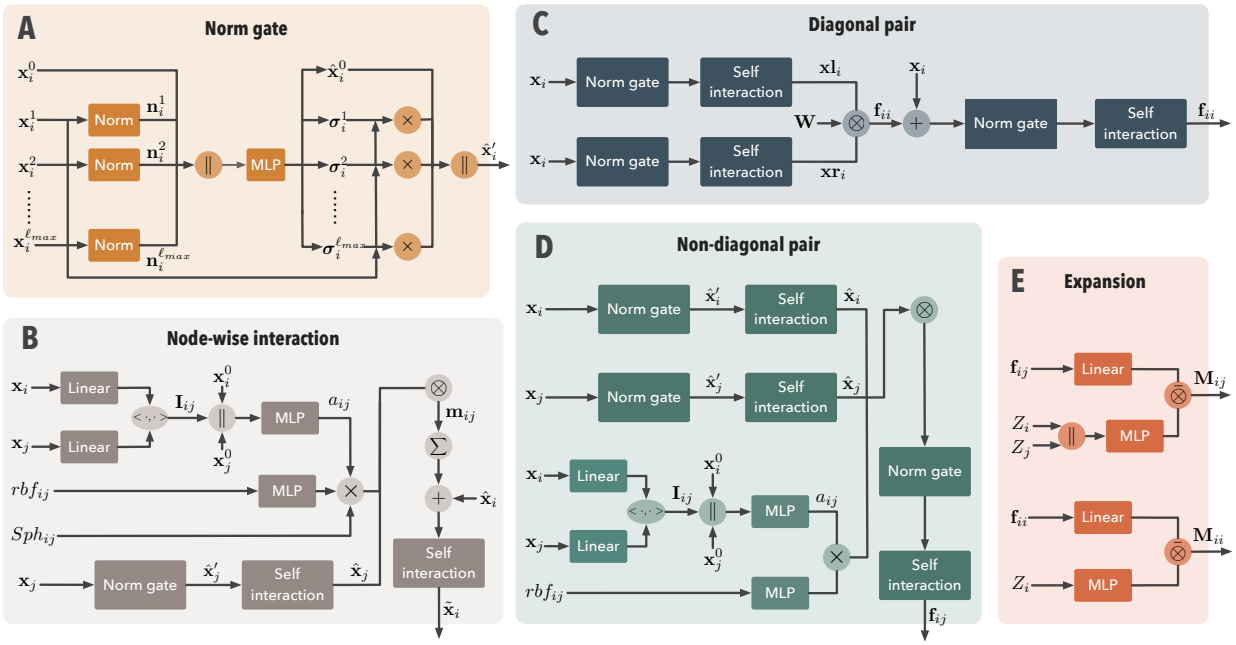

Figure 2. The message passing schema of our model. **A:** Norm gate. The norm of node representations $\mathbf{x}$ at different orders are concatenated and fed to an MLP to calculate scale factor σ at each order. Irreducible representations at different orders are rescaled with these factors and then be concatenated to output $\hat{\mathbf{x}}'$. Note that rescaling only applies to orders higher than 0. **B:** Node-wise interaction layer. The inner product of linearly transformed representations of nodes i and j is concatenated with their irreducible representation at order $\ell = 0$ and then fed to an MLP to calculate the attentive score a_{ij}. Next, the spherical harmonics of pairwise direction $\hat{r}_{ij}$, radius basis function (RBF) transformed pairwise distance r_{ij}, and attentive score a_{ij} are multiplied to produce the filter. A norm gate followed by a self-interaction layer is applied to $\mathbf{x}_j$ to produce $\hat{\mathbf{x}}_j$ which is used in a tensor product with the filter to obtain message $\mathbf{m}_{ij}$. Finally, messages are aggregated with $\hat{\mathbf{x}}_i$ and updated to output $\tilde{\mathbf{x}}_i$. **C:** Diagonal pair. Two norm gates with self-interaction are applied to node irreducible representation $\mathbf{x}_i$ in parallel to obtain $\mathbf{xl}_i$ and $\mathbf{xr}_i$. Then, a tensor product with parameters $\mathbf{W}$ is used to produce the diagonal pair representation $\mathbf{f}_{ii}$. The $\mathbf{f}_{ii}$ is skip-connected with $\mathbf{x}_i$ and processed by another norm gate with self-interaction before being outputted. **D:** Non-diagonal pair. The attentive score a_{ij} of nodes i and j is multiplied with the RBF and MLP transformed pairwise distance r_{ij} to produce the filter. Two norm gates with self-interaction are applied to $\mathbf{x}_i$ and $\mathbf{x}_j$ separately to obtain $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{x}}_j$. Then, the non-diagonal pair representation $\mathbf{f}_{ij}$ is produced by a tensor product applying to $\hat{\mathbf{x}}_i$, $\hat{\mathbf{x}}_j$, and a_{ij}. The $\mathbf{f}_{ij}$ is outputted after going through another norm gate with self-interaction. **E:** Expansion module. For a non-diagonal pair, node embeddings of atoms i and j are concatenated and transformed by an MLP, and then fed into an inverse of tensor product with linearly transformed pair representation $\mathbf{f}_{ij}$ to output the intermediate block matrix $\mathbf{M}_{ij}$ in a fixed shape. For a diagonal pair, inputs of expansion are pair representation $\mathbf{f}_{ii}$ and embedding of node i. Output is the intermediate block $\mathbf{M}_{ii}$.

blocks. The diagonal blocks consider Hamiltonian operators on a single atom, while the non-diagonal blocks are for pairs of two atoms. In this subsection, we explain how to produce pairwise irreducible representations and then build target quantum matrix block-by-block using tensor expansion with tensor product filters.

Non-Diagonal Pair. For each pair of non-diagonal nodes, a tensor product filter controls the influence of their irreducible representations onto node pair irreducible representation $\mathbf{f}_{ij}$. As shown in Figure 2.D, the filter first calculates attentive scores a_{ij} in the same way as Eq. (13) to consider pairwise similarity $\mathbf{I}^l_{ij}$ along with the irreducible representations of order 0. Next, it combines the attentive scores with $R(r_{ij})$ to produce a weight for each path $(l_{in_i}, l_{in_j}, l_{out})$ as

$$F_{cm}^{(\ell_{\text{in}_i},\ell_{\text{in}_j},\ell_{\text{out}})}(r_{ij}) = a_{ijc}^{(\ell_{\text{in}_i},\ell_{\text{in}_j},\ell_{\text{out}})} R_{cm}^{(\ell_{\text{in}_i},\ell_{\text{in}_j},\ell_{\text{out}})}(r_{ij}).$$

After obtaining the above weight as well as $\hat{\mathbf{x}}$ by rescaling the irreducible representations $\mathbf{x}$ with a norm gate followed by self-interaction, a tensor product filter is applied to produce the node pair irreducible representations $\mathbf{f}_{ij}$ as

$$\mathbf{f}_{ij}^{\ell} = (F^{(\ell_{\text{in}_i},\ell_{\text{in}_j},\ell_{\text{out}})}(r_{ij}) \quad \hat{\mathbf{x}}_i^{\ell_{\text{in}_i}} \otimes \hat{\mathbf{x}}_j^{\ell_{\text{in}_j}})^{\ell_{\text{out}}}. \qquad (17)$$

Diagonal Pair. Unlike non-diagonal pair, the diagonal pair fuses irreducible representations of the same node as illustrated in Figure 2.C. First, two norm gates with self-interaction are applied to the irreducible representation $\mathbf{x}_i$

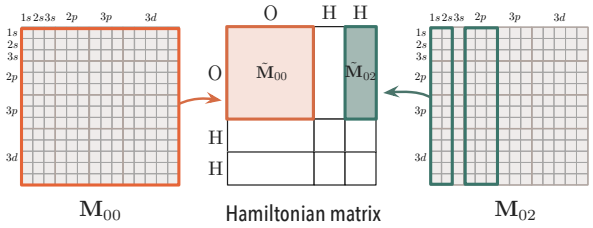

Figure 3. Construction of tensor matrix. The expansion module of our model outputs the intermediate block $\mathbf{M}$ with full orbitals for each node pair. Then, pair block $\tilde{\mathbf{M}}$ is selected from $\mathbf{M}$ according to orbitals of corresponding atoms. Finally, the Hamiltonian matrix of a molecule is built by combining $\tilde{\mathbf{M}}$ of all node pairs.

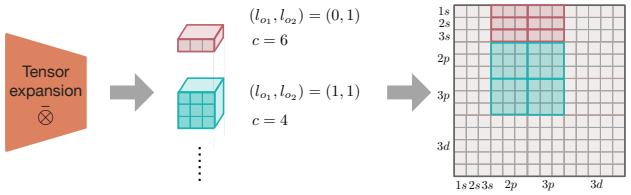

Figure 4. The relationship between the channel of equivariant matrix and the intermediate full orbital matrix. Here, c is the number of channel for equivariant matrices, and (l_{o1}, l_{o2}) is the rotation order of the equivariant matrix. Then, these equivariant matrices composed of the intermediate full orbital matrix for each node pair.

in parallel to produce two irreducible representations $\mathbf{xl}_i$ and $\mathbf{xr}_i$. Then, the pair representation $\mathbf{f}_{ii}^{\ell_{\text{out}}}$ is calculated by a tensor product with learnable parameters $\mathbf{W}$ for each path $(\ell_{\text{in}l}, \ell_{\text{in}r}, \ell_{\text{out}})$, which is defined as

$$\mathbf{f}_{ii}^{\ell_{\text{out}}} = (\mathbf{W}^{(\ell_{\text{in}_l}, \ell_{\text{in}_r}, \ell_{\text{out}})} \quad \mathbf{xl}_i^{\ell_{\text{in}_l}} \otimes \mathbf{xr}_i^{\ell_{\text{in}_r}})^{\ell_{\text{out}}}. \tag{18}$$

After that, a residual connection is added between the obtained diagonal pair representations and the input irreducible representations:

$$\mathbf{f}_{ii}^{\ell} = \mathbf{f}_{ii}^{\ell} + \mathbf{x}_i^{\ell}. \tag{19}$$

Finally, a norm gate followed by a self-interaction layer updates the pair features $\mathbf{f}_{ii}^{\ell}$ and outputs the final irreducible representations for diagonal pairs.

Construction of Tensor Matrix. After collecting irreducible representations for diagonal and non-diagonal pairs, the next step is to build the final Hamiltonian matrix. As illustrated in Figure 3, each entry in the matrix denotes the interaction between two orbitals. Specifically, pair block $\tilde{\mathbf{M}}_{ij}$ contains all the interactions between atoms i and j. Since atoms have different numbers of orbitals, pair blocks $\tilde{\mathbf{M}}_{ij}$'s are in different shapes that should be determined when constructing the Hamiltonian matrix. In our framework, we introduce an intermediate block $\mathbf{M}_{ij}$ with full orbitals and then extract $\tilde{\mathbf{M}}_{ij}$ from that based on corresponding atom orbitals. For example, there are four atoms H, C, N, and O, in the MD17 dataset. In this case, full orbitals contain 1s, 2s, 3s, 2p, 3p, and 3d, and $\mathbf{M} \in \mathbb{R}^{14\times14}$. When dealing with the hydrogen atom H, only 1s, 2s, 2p orbitals are selected to construct the Hamiltonian matrix. In this way, each node pair irreducible representation $\mathbf{f}_{ij}$ can be converted to an intermediate pair block $\mathbf{M}_{ij}$ with a predefined shape despite the atom type. This strategy is advantageous as it can be easily extended to different molecules.

To construct the intermediate pair blocks with full orbitals using pair irreducible representations, we apply a tensor expansion operation with the filter operation as shown in Figure 2.E. The tensor expansion is defined as

$$(\bar{\otimes} w^{\ell_3})_{(m_1,m_2)}^{(\ell_1,\ell_2)} = \sum_{m_3=-\ell_3}^{\ell_3} C_{(\ell_3,m_3)}^{(\ell_1,m_1),(\ell_2,m_2)} w_{m_3}^{\ell_3}, \tag{20}$$

where C is the CG matrix and $\bar{\otimes}$ denotes tensor expansion which is the inverse operation of tensor product. Since CG matrix can be transformed to a unitary matrix, when $w^{\ell_3} = (u^{\ell_1} \otimes v^{\ell_2})^{\ell_3}$, we deduce that $u^{\ell_1} \otimes v^{\ell_2} = \sum_{\ell_3} \bar{\otimes} w^{\ell_3}$ with $|\ell_1 - \ell_2| \le l_3 \le \ell_1 + \ell_2$. Then, the filter takes atom types as input to produce a weight for each path $(\ell_{\text{o1}}, \ell_{\text{o2}}, \ell_{\text{in}})$:

$$\begin{aligned} F_{ij}^{(\ell_{\text{o1}},\ell_{\text{o2}},\ell_{\text{in}})} &= f(Z_i, Z_j) \\ F_{ii}^{(\ell_{\text{o1}},\ell_{\text{o2}},\ell_{\text{in}})} &= f(Z_i), \end{aligned} \tag{21}$$

where Z denotes the embedding of atom types. Finally, the intermediate blocks $\mathbf{M}$ are built by the filter and node pair irreducible representations as

$$\begin{aligned} M_{iic}^{(\ell_{\text{o1}},\ell_{\text{o2}})} &= \sum_{\ell_{\text{in}},c'} F_{iicc'}^{(\ell_{\text{o1}},\ell_{\text{o2}},\ell_{\text{in}})} \bar{\otimes} \mathbf{f}_{iic'}^{\ell_{\text{in}}} \\ M_{ijc}^{(\ell_{\text{o1}},\ell_{\text{o2}})} &= \sum_{\ell_{\text{in}},c'} F_{ijcc'}^{(\ell_{\text{o1}},\ell_{\text{o2}},\ell_{\text{in}})} \bar{\otimes} \mathbf{f}_{ijc'}^{\ell_{\text{in}}}, \end{aligned} \tag{22}$$

where c' denotes the channel index in the input irreducible representations and c denotes the channel index in $\mathbf{M}$. For example, there are nine channels with $(\ell_{\text{o1}}, \ell_{\text{o2}}) = (0, 0)$, and four channel with $(\ell_{\text{o1}}, \ell_{\text{o2}}) = (1, 1)$. Note that a bias term is learned for node pair representation with $\ell_{\text{in}} = 0$. We provide a demonstration showing the mapping from the output of expansion module to the full orbital matrices in Figure 4.

In the tensor expansion module of PhiSNet (Schütt et al., 2019), a record has to be maintained to map the relationship of atom-orbital pair interactions so that a unique channel index in irreducible representations can be selected for each interaction. In contrast to determining the channel dimension by the number of atom-orbital pair interactions in PhiSNet, our model does not need to keep the record because our

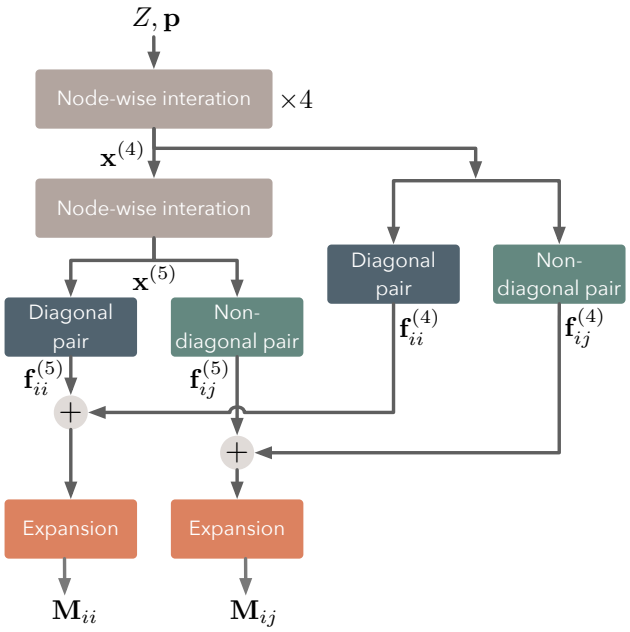

Figure 5. The whole architecture of the proposed QHNet. Inputs of the QHNet include atom types Z and atom positions $\mathbf{p}$. Five layers of node-wise interaction are used to learn SE(3)-equivariant irreducible representations for atoms. Then, diagonal and non-diagonal atom pairs are passed to distinctive modules to create pairwise representations $\mathbf{f}_{ii}$ and $\mathbf{f}_{ij}$. Finally, the expansion module converts pairwise representations to the intermediate blocks $\mathbf{M}$ that are post-processed to build the quantum tensor.

Table 1. Comparison of the total number and maximum sequential number of tensor products in PhiSNet and QHNet.

Methods	# of TP	# of sequential TP
PhiSNet	121	76
QHNet	9	6

channel dimension is fixed as a hyper-parameter. As a result, our model can effectively prevent the exponential growth of channel dimension when more atom types are involved.

3.5. Model Architecture

We now describe our whole framework. The model takes atom types $Z \in \mathbb{N}^{N\times 1}$ and their positions $\mathbf{p} \in \mathbb{R}^{N\times 3}$ as inputs. Here N denotes the number of atoms in a molecule. As shown in Figure 3.4, our network uses five node-wise interaction layers under the default setting to learn SE(3)-equivariant irreducible representations for atoms. Next, we take atom representations outputted by the last two layers to construct the diagonal and non-diagonal irreducible representations. Then, we combine the pairwise representations obtained after both layers with a sum operation and feed them into the expansion module.

The computation and time overheads of quantum tensor networks mainly come from message passing and tensor product (TP) operations. The computational cost of TP is much larger than a linear layer because it involves multiplication with CG matrix for each path sequentially. The time complexity analysis for the forward procedure of our module is shown as following: linear $O(C^2)$, self-interaction $O(C^2L)$, linear layer $O(C^2L^2)$ in the norm gate, and tensor product $O(CL^6)$. Here, C represents the number of channels, and L denotes the maximum rotation orders. Notably, in our experiments with QHNet and PhiSNet, we set $C = 128$ and $L = 4$. Therefore, the tensor product operation is as the most time-consuming operation in the model. In Table 1, we compare the total number of TP and the maximum number of sequential TP in our model and PhiSNet. The number of sequential TP is the number of passed TP for one variable from input to output. Since we use fewer TP operations, our model achieves higher efficiency and consumes less GPU memory.

4. Experiments

Dataset. We conduct experiments to evaluate the performance of QHNet on MD17 datasets (Schütt et al., 2019). MD17 consists of four datasets of small molecules, including water, ethanol, malondialdehyde, and uracil. Each dataset contains thousands of 3D molecular structures and their corresponding Hamiltonian matrices. The Hamiltonian matrices in MD17 were calculated by using the def2-SVP basis set (Weigend & Ahlrichs, 2005) with PBE (Perdew et al., 1996) density functionals. The statistics of MD17 dataset is shown in Table 5.

Software and Hardware. Our experiments are implemented based on PyTorch 1.11.0 (Paszke et al., 2019), PyTorch Geometric 2.1.0 (Fey & Lenssen, 2019), and e3nn (Geiger & Smidt, 2022). In our experiments, models are trained on a single 11GB Nvidia GeForce RTX 2080Ti GPU and Intel Xeon Gold 6248 CPU.

4.1. Overall Performance Evaluation

Setup. To assess the efficiency and efficacy of the proposed QHNet, we conduct experiments over four molecular systems collected in MD17 datasets, including water, ethanol, malondialdehyde, and uracil. Since the number of optimization steps plays an important role in model training on MD17, we set the total training steps to 200,000 for all experiments. We adopt the learning rate scheduler to speed up the convergence of model training. Specifically, the scheduler increases the learning rate gradually during the first 1,000 warm-up steps. The initial learning rate is 0, and the maximum learning rate is $5e^{-4}$. Then, the sched-

Table 2. Overall performance comparison in validation and test sets. Both PhiSNet and our QHNet are trained using a learning rate scheduler with a linearly decreasing learning rate to ensure the convergence with 1,000 warm-up steps and 200,000 total steps with double-precision floating-point. The unit for Hamiltonian $\mathbf{H}$ and eigen energies $\boldsymbol{\epsilon}$ is Hartree, denoted by E_h. Training 'Time' is in days. Training loss is described in Appendix A.1.

Dataset	Method	Time	Validation			Test		
			$\mathbf{H}$ $[10^{-6}E_h]\downarrow$	$\boldsymbol{\epsilon}$ $[10^{-6}E_h]\downarrow$	ψ $[10^{-2}]\uparrow$	$\mathbf{H}$ $[10^{-6}E_h]\downarrow$	$\boldsymbol{\epsilon}$ $[10^{-6}E_h]\downarrow$	ψ $[10^{-2}]\uparrow$
Water	PhiSNet	4.67d	7.87	29.81	99.99	15.67	–	99.94
	QHNet	1.27d	10.49	31.62	99.99	10.79	33.76	99.99
Ethanol	PhiSNet	7.45d	19.60	101.10	99.91	20.09	102.04	99.81
	QHNet	3.73d	20.45	81.29	99.99	20.91	81.03	99.99
Malondial-dehyde	PhiSNet	8.75d	21.89	104.76	99.96	21.31	100.60	99.89
	QHNet	3.33d	22.07	86.04	99.89	21.52	82.12	99.92
Uracil	PhiSNet	14.12d	18.89	146.14	99.87	18.65	143.36	99.86
	QHNet	3.31d	20.33	113.48	99.88	20.12	113.44	99.89

Table 3. Comparing PhiSNet and QHNet in terms of **training time per iteration** and **GPU memory consumption** on the MD17 datasets. Note that these are in double-precision floating-point.

Dataset	Batch Size	PhiSNet	QHNet	Speedup	Memory Efficiency
Water	10	2.02s / 3,839M	0.56s / 2,419M	3.61×	1.59×
Ethanol	5	3.22s / 8,673M	0.84s / 3,917M	3.83×	2.21×
Malondialdehyde	5	3.78s / 8,869M	0.88s / 3,925M	4.30×	2.25×
Uracil	3	6.10s / 9,135M	0.94s / 4,049M	6.49×	2.25×

uler reduces the learning rate linearly so that the learning rate reaches $1e^{-7}$ at the last step. Our competing baseline PhiSNet has five Pairmix layers that incorporate the representations from neighbor nodes to update node representations in the equivariant atomic representation network. For a fair comparison, QHNet uses five node-wise interaction layers to aggregate the messages from neighbor nodes to update node irreducible representations.

Note that the batch size differs for different molecules. Specifically, for QHNet, the batch size is set to 10 for water, ethanol, and malondialdehyde, while it is set to 5 for uracil. On the other hand, for PhiSNet, the batch size is as follows: 10 for water, 5 for ethanol and malondialdehyde, and 3 for uracil.

Model parameters are set in double-precision floating-point format when we run experiments for both the baseline and our QHNet. It is because the resolution of single-precision, $1e^{-6}$, is at a similar order of magnitude to the predicted error. Using double-precision can approximate at a higher resolution, and it provides more consistent and reliable evaluation with DFT algorithms where double-precision is required to ensure high accuracy. Note that we omit the task of predicting overlap matrix, since overlap matrix can be easily calculated without any errors in $10^{-3}s$ for a these molecules shown in Table 6.

Results. We provide the overall performance results in Table 2. Three metrics were applied to evaluate the accuracy of the predicted Hamiltonian matrices, including mean absolute error (MAE) of the Hamiltonian matrix elements $\mathbf{H}_{ij}$. Besides directly checking the prediction error of Hamiltonian matrices, it is also important to evaluate the error of the predicted orbital energy $\boldsymbol{\epsilon}$ and wavefunction ψ, which can be deduced from the Hamiltonian matrix $\mathbf{H}$ through Eqs. (3) and (4). Therefore, we report the MAE of the predicted energies and cosine similarity of coefficients for occupied molecular orbitals.

As reported in Table 2, QHNet can achieve competitive MAE on $\mathbf{H}$ in terms of both validation and test accuracy, and obtain better results on orbital energies. Specifically, on the water dataset, QHNet outperforms PhiSNet by a significant margin of $4.88 \times 10^{-5} E_h$. Although its obtained MAEs of predicted Hamiltonian matrices $\mathbf{H}$ are similar to the baseline, QHNet can better predict molecular properties from the predicted $\mathbf{H}$. For orbital energies, QHNet consistently outperforms other models on ethanol, malondialdehyde, and uracil datasets. We did not include the test MAE of $\boldsymbol{\epsilon}$ by PhiSNet on the water dataset in Table 2 as it is ten times worse than that by QHNet (33.76). All these

Table 4. Performance of QHNet trained on the mixed dataset. Note that this dataset includes only 500 water examples in the training set while there are 25,000 examples each for ethanol, malondialdehyde and uracil shown in Table 5 in Appendix A.2.

Dataset	$\mathbf{H}$ $[10^{-6}E_h]\downarrow$	$\boldsymbol{\epsilon}$ $[10^{-6}E_h]\downarrow$	ψ $[10^{-2}]\uparrow$
Water	190.03	304.48	99.99
Ethanol	51.68	130.68	99.97
Malondialdehyde	36.79	101.14	99.89
Uracil	34.77	124.86	99.80
Mixed Dataset	83.12	173.86	99.92

experimental results demonstrate the efficacy and better generalizability of QHNet.

4.2. Time and Memory Efficiency

We further compare the running time and GPU memory consumption of PhiSNet and QHNet. For each dataset, we set the same batch size for both models, and show the comparison in Table 3. The result indicates that the proposed QHNet is much more efficient than PhiSNet. Specifically, QHNet runs more than three times faster during training, requiring less than half of the GPU memory. As explained in Sec 3.5 and shown in Table 1, the high efficiency of our QHNet is because the total number and sequential number of tensor products are reduced to less than 14% of that in PhiSNet.

4.3. Performance on Mixed Dataset

Experiments in Sec 4.1 and Sec 4.2 focus on training and testing models on the same molecular system. Thus, the predicted Hamiltonian matrices have the same shape as training Hamiltonian matrices. In this subsection, we mix previously mentioned four datasets together while keeping the original split of training, validation, and testing sets. In this case, the mixed dataset contains four kinds of molecules, and QHNet is trained to predict the Hamiltonian matrices for multiple molecules rather than one. As described in Sec 3.4, we adopt the design of full orbital matrices with a fixed shape in the expansion module. Hence, QHNet can be easily extended to this mixed dataset and still achieve comparable performances, as demonstrated in Table 4. Note that such a mixed task is complex for PhiSNet with potential troubles for reasons we explain at the end of Sec 3.4. The flexible implementation of QHNet facilitates training a universal quantum tensor prediction network to further advance deep learning for quantum chemistry and condensed matter physics.

5. Conclusion

In this work, we presented a SE(3)-equivariant network—**QHNet**—to predict the Hamiltonian matrix with high accuracy and efficiency. QHNet is built carefully to maintain the underlying symmetry while eliminating 86% of tensor product operations to make it super streamlined compared to existing methods. We evaluated QHNet using the MD17 datasets to show that it can accelerate the training by more than three times while using less than 50% of the GPU memory. Additionally, our experimental results indicate that QHNet achieves competitive MAE on the Hamiltonian matrix and can derive more accurate molecular properties. Furthermore, by outputting full orbital matrices with a fixed shape and applying post-processing, QHNet is highly versatile and can be extended to datasets mixed with a variety of molecules. Therefore, we believe QHNet has great potential to efficiently predict accurate quantum tensors such as Hamiltonian matrix in a wide range of molecular systems.

Acknowledgments

This work was supported in part by National Science Foundation grants CCF-1553281, IIS-1812641, OAC-1835690, DMR-2119103, DMR-2103842, IIS-1908220, IIS-2006861, and IIS-2212419, and National Institutes of Health grant U01AG070112. Acknowledgment is made to the donors of the American Chemical Society Petroleum Research Fund for partial support of this research.

References

Anderson, B., Hy, T. S., and Kondor, R. Cormorant: Covariant molecular neural networks. *Advances in neural information processing systems*, 32, 2019.

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

Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E. J., and Welling, M. Geometric and physical quantities improve E(3) equivariant message passing. In *International Conference on Learning Representations*, 2022.

Cances, E. and Le Bris, C. On the convergence of SCF algorithms for the Hartree-Fock equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 34(4):749–774, 2000.

Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*, 2022.

Fey, M. and Lenssen, J. E. Fast Graph Representation Learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

Fuchs, F., Worrall, D., Fischer, V., and Welling, M. SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. *Advances in Neural Information Processing Systems*, 33:1970–1981, 2020.

Ganea, O.-E., Pattanaik, L., Coley, C. W., Barzilay, R., Jensen, K. F., Green, W. H., and Jaakkola, T. S. GeoMol: Torsional Geometric Generation of Molecular 3D Conformer Ensembles. *Advances in neural information processing systems*, 2021.

Ganea, O.-E., Huang, X., Bunne, C., Bian, Y., Barzilay, R., Jaakkola, T. S., and Krause, A. Independent SE(3)-equivariant models for end-to-end rigid protein docking. In *International Conference on Learning Representations*, 2022.

Gasteiger, J., Groß, J., and Günnemann, S. Directional Message Passing for Molecular Graphs. In *International Conference on Learning Representations (ICLR)*, 2020.

Gasteiger, J., Becker, F., and Günnemann, S. GemNet: Universal Directional Graph Neural Networks for Molecules. *Advances in Neural Information Processing Systems*, 34: 6790–6802, 2021.

Geiger, M. and Smidt, T. e3nn: Euclidean Neural Networks, 2022. URL https://arxiv.org/abs/2207.09453.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pp. 1263–1272. PMLR, 2017.

Godwin, J., Schaarschmidt, M., Gaunt, A. L., Sanchez-Gonzalez, A., Rubanova, Y., Veličković, P., Kirkpatrick, J., and Battaglia, P. Simple GNN Regularisation for 3D Molecular Property Prediction and Beyond. In *International conference on learning representations*, 2021.

Griffiths, D. J. and Schroeter, D. F. *Introduction to Quantum Mechanics*. Cambridge University Press, 2018.

Hohenberg, P. and Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.*, 136:B864–B871, 1964.

Jiang, H., Wang, J., Cong, W., Huang, Y., Ramezani, M., Sarma, A., Dokholyan, N. V., Mahdavi, M., and Kandemir, M. T. Predicting protein–ligand docking structure with graph neural network. *Journal of Chemical Information and Modeling*, 62(12):2923–2932, 2022.

Klicpera, J., Giri, S., Margraf, J. T., and Günnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. *arXiv preprint arXiv:2011.14115*, 2020.

Kohn, W. and Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140: A1133–A1138, 1965.

Kudin, K. N., Scuseria, G. E., and Cances, E. A black-box self-consistent field convergence algorithm: One step closer. *The Journal of Chemical Physics*, 116(19):8255–8261, 2002.

Li, H., Wang, Z., Zou, N., Ye, M., Xu, R., Gong, X., Duan, W., and Xu, Y. Deep-learning density functional theory Hamiltonian for efficient *ab initio* electronic-structure calculation. *Nature Computational Science*, 2(6):367–377, 2022. ISSN 2662-8457.

Liu, M., Fu, C., Zhang, X., Wang, L., Xie, Y., Yuan, H., Luo, Y., Xu, Z., Xu, S., and Ji, S. Fast Quantum Property Prediction via Deeper 2D and 3D Graph Networks. *arXiv preprint arXiv:2106.08551*, 2021.

Liu, M., Luo, Y., Uchino, K., Maruhashi, K., and Ji, S. Generating 3D Molecules for Target Protein Binding. In *Proceedings of The 39th International Conference on Machine Learning*, pp. 13912–13924, 2022a.

Liu, Y., Wang, L., Liu, M., Lin, Y., Zhang, X., Oztekin, B., and Ji, S. Spherical message passing for 3D molecular graphs. In *International Conference on Learning Representations*, 2022b.

Lu, W., Wu, Q., Zhang, J., Rao, J., Li, C., and Zheng, S. TANKBind: Trigonometry-Aware Neural NetworKs for Drug-Protein Binding Structure Prediction. *bioRxiv*, pp. 2022–06, 2022.

Luo, Y. and Ji, S. An Autoregressive Flow Model for 3D Molecular Geometry Generation from Scratch. In *International Conference on Learning Representations*, 2022.

Mansimov, E., Mahmood, O., Kang, S., and Cho, K. Molecular geometry prediction using a deep generative graph neural network. *Scientific reports*, 9(1):1–13, 2019.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32, 2019.

Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A., and Joannopoulos, J. D. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, 64:1045–1097, 1992.

Perdew, J. P., Burke, K., and Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865, 1996.

Qiao, Z., Welborn, M., Anandkumar, A., Manby, F. R., and Miller III, T. F. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *The Journal of Chemical Physics*, 153(12):124111, 2020.

Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) Equivariant Graph Neural Networks. In *International Conference on Machine Learning*, pp. 9323–9332. PMLR, 2021.

Schütt, K., Kindermans, P.-J., Felix, H. E. S., Chmiela, S., Tkatchenko, A., and Müller, K.-R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, pp. 991–1001, 2017.

Schütt, K., Unke, O., and Gastegger, M. Equivariant Message Passing for the Prediction of Tensorial Properties andMolecular Spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A., and Müller, K.-R. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018.

Schütt, K. T., Gastegger, M., Tkatchenko, A., Müller, K.-R., and Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature Communications*, 10(1):5024, 2019.

Shi, C., Luo, S., Xu, M., and Tang, J. Learning gradient fields for molecular conformation generation. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9558–9568. PMLR, 18–24 Jul 2021.

Simm, G. and Hernandez-Lobato, J. M. A generative model for molecular distance geometry. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8949–8958. PMLR, 13–18 Jul 2020.

Stärk, H., Ganea, O., Pattanaik, L., Barzilay, R., and Jaakkola, T. Equibind: Geometric deep learning for drug binding structure prediction. In *International Conference on Machine Learning*, pp. 20503–20521. PMLR, 2022.

Szabo, A. and Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation, 2012.

Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., and Riley, P. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Unke, O., Bogojeski, M., Gastegger, M., Geiger, M., Smidt, T., and Müller, K.-R. SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. *Advances in Neural Information Processing Systems*, 34: 14434–14447, 2021.

Unke, O. T. and Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *Journal of Chemical Theory and Computation*, 15(6):3678–3693, 2019.

Wang, L., Liu, Y., Lin, Y., Liu, H., and Ji, S. ComENet: Towards complete and efficient message passing for 3D molecular graphs. In *The 36th Annual Conference on Neural Information Processing Systems*, 2022a.

Wang, Z., Liu, M., Luo, Y., Xu, Z., Xie, Y., Wang, L., Cai, L., Qi, Q., Yuan, Z., Yang, T., and Ji, S. Advanced Graph and Sequence Neural Networks for Molecular Property Prediction and Drug Discovery . *Bioinformatics*, 38(9): 2579–2586, 2022b.

Weigend, F. and Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Physical Chemistry Chemical Physics*, 7(18):3297–3305, 2005.

Xu, M., Luo, S., Bengio, Y., Peng, J., and Tang, J. Learning Neural Generative Dynamics for Molecular Conformation Generation. In *International Conference on Learning Representations*, 2021a.

Xu, M., Wang, W., Luo, S., Shi, C., Bengio, Y., Gomez-Bombarelli, R., and Tang, J. An End-to-End Framework for Molecular Conformation Generation via Bilevel Programming. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11537–11547. PMLR, 18–24 Jul 2021b.

Yan, K., Liu, Y., Lin, Y., and Ji, S. Periodic Graph Transformers for Crystal Material Property Prediction. In *The 36th Annual Conference on Neural Information Processing Systems*, 2022.

Zhang, Y., Cai, H., Shi, C., Zhong, B., and Tang, J. E3Bind: An End-to-End Equivariant Network for Protein-Ligand Docking. *arXiv preprint arXiv:2210.06069*, 2022.

Table 5. The statistics of MD17 dataset (Schütt et al., 2019).

Dataset	# of structures	Train	Val	Test	# of atoms	# of orbitals	# of occupied molecular orbitals
Water	4,900	500	500	3,900	3	24	5
Ethanol	30,000	25,000	500	4,500	9	72	10
Malondialdehyde	26,978	25,000	500	1,478	9	90	19
Uracil	30,000	25,000	500	4,500	12	132	26

Table 6. Time comparison for single example among DFT algorithm, calculating overlap matrix, PhiSNet inference and QHNet inference. Note that we use batch size 64 to conduct inference for both PhiSNet and QHNet.

dataset	DFT[s]	overlap [10^{-4}s]	PhiSNet [10^{-2}s]	QHNet [10^{-2}s]
water	11.38	4.87	1.26	1.09
ethanol	25.11	8.71	8.83	7.11
malondialdehyde	40.63	7.27	9.15	7.92

A. Appendix.

A.1. Training Loss for QHNet

For training, we follow the implementation of PhiSNet (Unke et al., 2021), and use Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) as loss function to train the QHNet for predicting the Hamiltonian matrices.

$$\mathcal{L}(\mathbf{H}, \mathbf{H}^{GT}) = \left(\sqrt{\frac{1}{N^2} \sum_{i_1, i_2} (\mathbf{H}_{i_1, i_2} - \mathbf{H}^{GT}_{i_1, i_2})^2} + \frac{1}{N^2} \sum_{i_1, i_2} |\mathbf{H}_{i_1, i_2} - \mathbf{H}^{GT}_{i_1, i_2}| \right), \tag{23}$$

where $\mathbf{H}$ is the predicted Hamiltonian matrix, $\mathbf{H}^{GT}$ is the ground-truth Hamiltonian matrix, and the size of $\mathbf{H}$ and $\mathbf{H}^{GT}$ is $N \times N$.

A.2. Statistics of the dataset

We here provide the statistics of the MD17 datasets in Table 5. Note that the mixed dataset follows the same dataset split.

A.3. Accelerating the DFT algorithm.

To study the acceleration by using QHNet, we compare the inference time of QHNet with the time consumption of DFT algorithm. Here we use PySCF to conduct DFT algorithm with PBE correlation functional and def2-SVP basis set. Moreover, we select DIIS as the SCF algorithm for the DFT calculation. In Table 6, QHNet exhibited an acceleration of approximately 1000x on water, and approximately 300x on both ethanol and malondialdehyde. Note that we conduct our experiments on 11GB Nvidia GeForce RTX 2080Ti GPU and Intel Xeon Gold 6248 CPU.

Next, we focus onoptimization steps ratio when continuing to optimize the predicted Hamiltonian matrices using the DFT algorithm. It shows the time ratio to get the converged Hamiltonian matrix with machine learning model acceleration and reflects the quality of the predicted Hamiltonian matrices, with higher-quality matrices requiring fewer optimization steps. Note that we use PySCF on 50 random selected geometries in each molecular dataset to conduct DFT algorithm with PBE correlation functional and def2-SVP basis set, and select DIIS as the SCF algorithm for the DFT calculation.

Table 7. The ratio of optimization steps when taking predicted Hamiltonian matrices as the initial status of the DFT algorithm to accelerate the optimization.

dataset	QHNet	PhiSNet
water	0.494	0.497
ethanol	0.554	0.562
malondialdehyde	0.562	0.567

Table 8. Performance of QHNet with various training steps compared to the reported results in PhiSNet. In these experiments, QHNet are in single-precision floating-point following the setting in original PhiSNet experiments. * denotes the training is converged and EMA is not used in this experiment.

Dataset	Method	Training statgies	$\mathbf{H}$ $[10^{-6}E_h]\downarrow$	$\boldsymbol{\epsilon}$ $[10^{-6}E_h]\downarrow$	ψ $[10^{-2}]\uparrow$
Water	QHNet*	RLROP	10.36	36.21	99.99
	PhiSNet	RLROP	17.59	85.53	-
Ethanol	QHNet	LSW (10,000, 1,000,000)	12.78	62.97	99.99
	QHNet	RLROP	13.12	51.80	99.99
	PhiSNet	RLROP	12.15	62.75	-
Malondialdehyde	QHNet	LSW (10,000, 1,000,000)	11.97	55.57	99.94
	QHNet	RLROP	13.18	51.54	99.95
	PhiSNet	RLROP	12.32	73.50	-
Uracil	QHNet	LSW (10,000, 1,000,000)	9.96	66.75	99.95
	PhiSNet	RLROP	10.73	84.03	-

Table 9. Ablation study of QHNet.

Dataset	full model	wo attentive	wo NormGate
water	10.79	10.65	10.91
ethanol	20.91	-	39.26
malondialdehyde	21.52	30.89	40.79

A.4. Performance comparison

To compare with PhiSNet, we trained our QHNet using the same settings as PhiSNet, employing the Reduce Learning Rate On Plateau (RLROP) scheduler. We initialized the scheduler with a learning rate of $5e-4$ and continued training until it reached $1e^{-6}$. In order to limit the training steps, we set the maximum number of steps to $1,000,000$. The results are presented in Table 8. QHNet successfully converged on the water dataset. However, for the other three datasets, QHNet did not converge, and we therefore report the final results. Furthermore, we implemented a linear schedule with warmup, specifying the warmup steps and total training steps as LSW, and (10,000, 1,000,000) denotes a warmup period of 10,000 steps and a total training duration of 1,000,000 steps. Additionally, the batch size is set to 10 for all the experiments. Note that the results of PhiSNet comes from its original paper. For the ψ, since it is reported as 1.00 in the original PhiSNet paper, it can not compare without enough accurate digits. Therefore, we omit the number here.

A.5. Ablation studies on the model architecture

To study the various modules in QHNet, we conduct experiments to compare the QHNet without attentive scores or NormGate in node-wise interaction layers and report the MAE of the Hamiltonian matrix on the test set in Table 9.