

## A Additional Preliminary

### A.1 Kohn-Sham density functional theory (DFT).

DFT is a cornerstone of quantum chemistry and materials science, offering a computationally tractable framework for approximating the electronic structure of many-body systems. As directly solving the many-electron Schrödinger equation [37] is prohibitively expensive for systems with many particles, DFT reformulates the problem by expressing ground-state properties as a functional of the electron density  $\rho(\mathbf{r})$ . Since  $\rho(\mathbf{r})$  depends only on three spatial coordinates, this significantly reduces the computational complexity regardless of the number of electrons.

In practical applications, DFT is most commonly implemented via the Kohn–Sham formulation [1, 12], which introduces a system of non-interacting electrons that reproduces the true ground-state density of the interacting system. The KF-DFT for  $N$ -particle system uses the parameterization  $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_{i=1}^N \psi_i(\mathbf{r}_i)$  using single-particle orbitals  $\{\psi_i\}_{i=1}^N$  where  $\rho(\mathbf{r}) = \sum_{n=1}^N |\psi_i(\mathbf{r}_i)|^2$ . The KS-equations are:

$$\left[ -\frac{1}{2} \nabla^2 + V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}[\rho](\mathbf{r}) + V_{\text{xc}}[\rho](\mathbf{r}) \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \quad (15)$$

where  $\nabla^2$  is the Laplacian operator for kinetic energy,  $V_{\text{ext}}$  is the external potential from the nuclei,  $V_{\text{H}}$  is the Hartree potential,  $V_{\text{xc}}$  is the exchange-correlation potential, and  $\psi_i, \epsilon_i$  are the Kohn–Sham orbitals and their corresponding energies.

### A.2 Group theory

We briefly introduce key concepts from group theory, providing the necessary background to understand the motivation behind equivariant designs in Hamiltonian prediction.

**Groups.** A *group*  $G$  is a non-empty set equipped with a binary operation  $\circ : G \times G \rightarrow G$ , satisfying the following axioms:

- **Associativity:** For all  $a, b, c \in G$ ,  $(a \circ b) \circ c = a \circ (b \circ c)$ .
- **Identity:** There exists an identity element  $e \in G$  such that for all  $a \in G$ ,  $e \circ a = a \circ e = a$ .
- **Inverse:** For each  $a \in G$ , there exists an inverse  $a^{-1} \in G$  such that  $a \circ a^{-1} = a^{-1} \circ a = e$ .
- **Closure:** For all  $a, b \in G$ , the result  $a \circ b$  is also in  $G$ .

We typically simplify notation by omitting  $\circ$ , writing  $ab$  instead of  $a \circ b$ . Groups can be finite or infinite, discrete or continuous, compact or non-compact. Group theory [43] provides a rigorous mathematical foundation for describing symmetries in physical systems [55], an essential perspective for designing physically informed machine learning models.

**Group actions.** To make group theory applicable to structures in mathematics and physics, we define how groups act on sets. A group  $G$  is said to act on a set  $\mathcal{X}$  if there exists a function  $\cdot_{\mathcal{X}} : G \times \mathcal{X} \rightarrow \mathcal{X}$ <sup>1</sup> such that for all  $g, h \in G$  and  $x \in \mathcal{X}$ , the following conditions hold: (1)  $e \cdot_{\mathcal{X}} x = x$ , where  $e$  is the identity element of  $G$ , and (2)  $(gh) \cdot_{\mathcal{X}} x = g \cdot_{\mathcal{X}} (h \cdot_{\mathcal{X}} x)$ . These conditions ensure that the group operation is compatible with its action on the set. In practice, such group actions often correspond to geometric transformations, such as translations, rotations, or reflections, on points or functions defined over Euclidean space.

**Equivariance.** When working with structured data such as molecular geometries or fields, it is often desirable that the operations we apply preserve the symmetries of the underlying space. A function  $f : \mathcal{X} \rightarrow \mathcal{Y}$  is said to be *equivariant* with respect to group actions  $\cdot_{\mathcal{X}}$  on the input space  $\mathcal{X}$  and  $\cdot_{\mathcal{Y}}$  on the output space  $\mathcal{Y}$ , if it satisfies

$$f(g \cdot_{\mathcal{X}} x) = g \cdot_{\mathcal{Y}} f(x), \quad \forall g \in G, x \in \mathcal{X}. \quad (16)$$

This means that applying a group transformation before or after the function yields consistent results. In physical systems, such as those governed by rotational symmetry (e.g.,  $\text{SO}(3)$  in molecules),

<sup>1</sup>The subscript of group action operator for indicating the input and output space is often omitted when clear from context.

equivariance ensures that rotated inputs yield correspondingly rotated outputs. This property is critical in designing models that generalize well across symmetrically equivalent configurations.

**Group representations.** Group actions on vector spaces are formalized through the notion of representations. A *representation* of a group  $G$  on a vector space  $V$  is a homomorphism  $\rho : G \rightarrow \text{GL}(V)$ , where  $\text{GL}(V)$  denotes the group of invertible linear transformations on  $V$ . This mapping associates each group element with a linear operator that describes how the group acts on the space. The vector space  $V$  is referred to as the *representation space*, and its dimension defines the representation’s size.

In finite-dimensional real spaces  $V = \mathbb{R}^d$ , these representations are often expressed as invertible matrices  $\mathbf{D}(g) \in \mathbb{R}^{d \times d}$ . For instance, the group  $\text{SO}(3)$  of 3D rotations is represented by orthogonal matrices  $\mathbf{g} \in \text{SO}(3)$  and its 3D rotation matrix  $\mathbf{R} \in \mathbb{R}^{3 \times 3}$  acting on vectors  $\mathbf{x} \in \mathbb{R}^3$  via matrix multiplication  $\mathbf{D}(\mathbf{g})\mathbf{x} = \mathbf{R}\mathbf{x}$ .

Two representations  $\mathbf{D}(g)$  and  $\mathbf{D}'(g)$  are said to be *equivalent* if they are related by a change of basis:

$$\mathbf{D}'(g) = \mathbf{Q}^{-1}\mathbf{D}(g)\mathbf{Q}, \quad (17)$$

for some invertible matrix  $\mathbf{Q}$ . Such equivalence indicates that the two representations describe the same group action under different coordinate systems.

### A.3 Irreducible representations

**Irreducible representations (Irreps).** A representation  $\rho : G \rightarrow \text{GL}(V)$  of a group  $G$  on a vector space  $V$  is called *irreducible* if it does not contain any nontrivial invariant subspace. More formally, a subspace  $W \subseteq V$  is said to be invariant under the representation  $\rho$  if for all group elements  $g \in G$  and all vectors  $w \in W$ , the action satisfies:

$$\rho(g)w \in W. \quad (18)$$

Then, the representation  $\rho$  is irreducible if the only invariant subspaces it admits are the trivial subspace  $\{0\}$  and the entire representation space  $V$ .

Irreducible representations play a fundamental role because any finite-dimensional unitary representation of a compact group can be uniquely decomposed into a direct sum of irreducible representations, up to isomorphism [56]. This decomposition is foundational in analyzing the symmetry properties and constructing equivariant neural network architectures.

For the rotation group  $\text{SO}(3)$ , the irreps are completely characterized by non-negative integers  $\ell \in \mathbb{N}_0$ . Each irreducible representation  $D^{(\ell)} : \text{SO}(3) \rightarrow \mathbb{C}^{(2\ell+1) \times (2\ell+1)}$  has dimension  $2\ell+1$  and is explicitly realized by the action on spherical harmonics  $Y_m^\ell(\theta, \phi)$ . For a function  $f : S^2 \rightarrow \mathbb{C}$  can be expanded in the basis of spherical harmonics as:

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{lm} Y_m^l(\theta, \phi), \quad (19)$$

where  $f_{lm} \in \mathbb{C}$  is the complex coefficients of the spherical harmonics. The action of a rotation  $\mathbf{R} \in \text{SO}(3)$  on  $f$  is defined as:

$$(\mathbf{R} \cdot f)(\theta, \phi) = f(\mathbf{R}^{-1} \cdot (\theta, \phi)) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \sum_{m'=-l}^l D_{mm'}^{(\ell)}(\mathbf{R}) f_{\ell m'} Y_m^\ell(\theta, \phi), \quad (20)$$

where  $D^{(\ell)}(\mathbf{R})$  is the Wigner D-matrix corresponding to the irreducible representation labeled by  $\ell$ , acting linearly on the vector of expansion coefficients  $\mathbf{f}^{(\ell)} = (f_{-\ell}, f_{-\ell+1}, \dots, f_{\ell}) \in \mathbb{C}^{(2\ell+1)}$ . Explicitly, under rotation:

$$\mathbf{f}^{(\ell)} \xrightarrow{\mathbf{R}} D^{(\ell)}(\mathbf{R})\mathbf{f}^{(\ell)}. \quad (21)$$

Thus, irreps serve as the fundamental building blocks in constructing functions and features that transform equivariantly under the action of  $\text{SO}(3)$ .

**Clebsch–Gordan tensor product.** Given two irreps  $D^{(\ell_1)}, D^{(\ell_2)}$  of  $\text{SO}(3)$ , their tensor product  $D^{(\ell_1)} \otimes D^{(\ell_2)}$  acts on the tensor product space and is generally reducible. It decomposes into a direct

928 sum of irreps:

$$D^{(\ell_1)} \otimes_{\text{CG}} D^{(\ell_2)} \cong \bigoplus_{\ell=|\ell_1-\ell_2|}^{\ell_1+\ell_2} D^{(\ell)}, \quad (22)$$

929 where the decomposition is governed by Clebsch–Gordan (CG) coefficients. CG coefficients enable  
 930 a basis transformation from the product space to irreps, playing crucial roles in quantum angular  
 931 momentum coupling and equivariant neural networks.

932 Explicitly, for vectors  $\mathbf{u}^{(\ell_1)} \in \mathbb{C}^{(2\ell_1+1)}$  and  $\mathbf{v}^{(\ell_2)} \in \mathbb{C}^{(2\ell_2+1)}$ , their CG tensor product produces an  
 933 irreducible feature  $\mathbf{w}^{(\ell)} \in \mathbb{C}^{(2\ell+1)}$ :

$$w_m^{(\ell)} = \sum_{m_1, m_2} C_{(\ell_1, m_1), (\ell_2, m_2)}^{(\ell, m)} u_{m_1}^{(\ell_1)} v_{m_2}^{(\ell_2)}. \quad (23)$$

934 This construction ensures that  $\mathbf{w}^{(\ell)}$  transforms correctly under  $\mathbf{R} \in \text{SO}(3)$ :

$$\left( D^{(\ell_1)}(\mathbf{R}) \mathbf{u}^{(\ell_1)} \right) \otimes_{\text{CG}} \left( D^{(\ell_2)}(\mathbf{R}) \mathbf{v}^{(\ell_2)} \right) = D^{(\ell)}(\mathbf{R}) \mathbf{w}^{(\ell)}. \quad (24)$$

935 Thus, CG tensor products ensure equivariant transformations, fundamental to maintaining consistency  
 936 and symmetry in equivariant neural architectures.

937 **Tensor field networks (TFN).** The TFN [47] is a widely used architecture that achieves SE(3)-  
 938 equivariance by explicitly encoding features as irreps of SO(3) and using spherical harmonics to  
 939 process directional information. Each feature vector at a node is decomposed into spherical tensor  
 940 components  $\mathbf{V}^{(\ell)} \in \mathbb{C}^{(2\ell+1)}$ , where  $\ell$  denotes the order of the irreducible representation. The key  
 941 component of the TFN layer is equivariant convolution and self-interaction. For the equivariant  
 942 convolution, the filter function is generated by the spherical harmonic functions  $Y_m^\ell(\hat{\mathbf{r}}_{ij})$  are applied  
 943 to the direction between atoms  $i$  and  $j$ , *i.e.*,  $\hat{\mathbf{r}}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/\|\mathbf{r}_i - \mathbf{r}_j\|$ , and are modulated by a  
 944 learnable radial function  $R(\|\mathbf{r}_{ij}\|)$ , implemented as an MLP to produce the filter:

$$F^{(\ell_{\text{in}}, \ell_f)}(\mathbf{r}_{ij}) = R^{(\ell_{\text{in}}, \ell_f)}(\|\mathbf{r}_{ij}\|) Y_m^{\ell_f}(\hat{\mathbf{r}}_{ij}). \quad (25)$$

945 The equivariant convolution aggregates information from neighboring nodes set  $\mathcal{N}$  using the tensor  
 946 product between the filters and neighboring features:

$$\tilde{V}_i^{(\ell_{\text{out}})} = \sum_{j \in \mathcal{N}} \left( F_c^{(\ell_{\text{in}}, \ell_f)}(\mathbf{r}_{ij}) \otimes V_j^{(\ell_{\text{in}})} \right)^{(\ell_{\text{out}})}, \quad (26)$$

947 where  $\ell_{\text{out}} \in \{|\ell_{\text{in}} - \ell_f|, \dots, (\ell_{\text{in}} + \ell_f)\}$ . This convolution ensures that the result transforms  
 948 equivariantly under rotation.

949 The self-interaction step applies a linear transformation over the channel dimension to each irrep  
 950 independently:

$$\tilde{V}_{icm}^{(\ell)} = \sum_{c'} w_{cc'} V_{ic'm}^{(\ell)}, \quad (27)$$

951 preserving equivariance due to the linearity and diagonal structure in the irrep indices. By construc-  
 952 tion, all operations in TFNs are equivariant to SE(3), allowing them to model tensorial and directional  
 953 quantities while respecting physical symmetries. This makes them especially well-suited for applica-  
 954 tions in molecular modeling, quantum chemistry, and materials science, where equivariance is crucial  
 955 for generalization and physical accuracy.

#### 956 A.4 Symmetry and equivariance of RH-DFT

957 **Spherical harmonics and atomic orbital basis.** To numerically solve the Kohn–Sham equations,  
 958 electronic wavefunctions are typically expanded in a basis of atomic orbitals. These orbitals are  
 959 constructed using spherical harmonics  $Y_m^\ell(\theta, \phi)$ , which form a complete orthonormal basis on the  
 960 sphere  $\mathbb{S}^2$  and transform under the irreducible representations  $D^{(\ell)}$  of the rotation group SO(3). An  
 961 atomic orbital basis function takes the form [57]:

$$\phi_{n\ell m}(\mathbf{r}) = R_{n\ell}(\|\mathbf{r}\|) Y_m^\ell(\theta, \phi), \quad (28)$$

where  $R_{n\ell}(\|\mathbf{r}\|)$  is a radial function (e.g., Gaussian [41] or Slater-type [40, 15]), and  $Y_m^\ell$  captures the angular component. Here,  $n$  is the principal quantum number,  $\ell$  is the orbital angular momentum quantum number, and  $m \in \{-\ell, \dots, \ell\}$  is the magnetic quantum number [44].

These basis functions are central to Kohn–Sham DFT calculations, as molecular orbitals are expressed as linear combinations of atomic orbitals:

$$\psi_i(\mathbf{r}) = \sum_{\alpha=(n\ell m)} C_{\alpha i} \phi_{n\ell m}(\mathbf{r} - \mathbf{r}_\alpha), \quad (29)$$

where  $\mathbf{C} \in \mathbb{R}^{B \times O}$  is the orbital coefficient matrix,  $B$  is the number of basis functions, and  $O$  is the number of occupied orbitals.

**Block Hamiltonian matrix.** The Hamiltonian matrix in RH-DFT can be decomposed according to the angular momentum quantum numbers  $\ell$  grouping associated with the atomic orbitals. Specifically, Let  $\mathbf{H}^{(\ell, \ell')} \in \mathbb{C}^{(2\ell+1) \times (2\ell'+1)}$  denote the submatrix representing the coupling between two angular momentum numbers  $\ell$  and  $\ell'$ . Under a rotation  $\mathbf{R} \in \text{SO}(3)$ , each such sub-matrix  $\mathbf{H}^{(\ell, \ell')}$  transforms equivariantly according to the  $\text{SO}(3)$  irreps as follows:

$$\mathbf{H}^{(\ell, \ell')}[\mathcal{D}(\mathbf{R})\mathbf{C}] = D^{(\ell)}(\mathbf{R})(\mathbf{H}^{(\ell, \ell')}[\mathbf{C}])D^{(\ell')}(\mathbf{R})^{-1}, \quad (30)$$

where atomic orbitals are grouped by their  $\ell$  and  $\ell'$ ,  $\mathcal{D}(\mathbf{R})$  is block diagonal Wigner-D matrix, and  $D^\ell$  and  $D^{\ell'}$  are Wigner D-matrices for  $\ell$  and  $\ell'$ , respectively. These submatrices  $\mathbf{H}^{(\ell, \ell')}$ , called orbital-wise Hamiltonian blocks, are the minimal building blocks of the full Hamiltonian that respect  $\text{SO}(3)$ -equivariance and have  $(2\ell+1) \times (2\ell'+1)$  elements which corresponds to the combination of the magnetic quantum number  $m$  and  $m'$ .

**Full Hamiltonian matrix.** In practice, basis functions are grouped not only by their angular momentum, but also by the atom they are centered on. For atom  $i$ , we denote its angular momentum support as  $L_i = \{\ell_1, \ell_2, \dots\}$ , where each  $\ell_k$  corresponds to its atomic shell (e.g.,  $s, p, d$ ) represented in the basis set.

From this atomic partitioning, the Hamiltonian can be reorganized into block form where each submatrix  $\mathbf{H}_{ij}^{(\ell, \ell')}$  corresponds to the interaction between angular momentum  $\ell \in L_i$  of atom  $i$  and  $\ell' \in L_j$  of atom  $j$ . The full Hamiltonian  $\mathbf{H}$  is thus a structured matrix composed of  $K \times K$  blocks, where  $K = \sum_i |L_i|$  is the total number of irreps (angular momentum groups) across all atoms, where  $|L_i|$  denotes the cardinality of the set.

To construct the full Hamiltonian matrix, we first form the atom-pair matrix  $\mathbf{H}_{ij}$  for every atom-pair  $(i, j)$ . This atom-pair matrix gathers all interactions between the angular momentum orbital groupings on the two atoms:

$$\mathbf{H}_{ij} = \begin{bmatrix} \mathbf{H}_{ij}^{(\ell_0, \ell'_0)} & \mathbf{H}_{ij}^{(\ell_0, \ell'_1)} & \dots \\ \mathbf{H}_{ij}^{(\ell_1, \ell'_0)} & \mathbf{H}_{ij}^{(\ell_1, \ell'_1)} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad \ell_k \in L_i, \ell'_k \in L_j, \quad (31)$$

where  $\mathbf{H}_{ij}^{(\ell, \ell')} \in \mathbb{C}^{(2\ell+1) \times (2\ell'+1)}$  denotes the interaction between angular momentum  $\ell \in L_i$  of atom  $i$  and  $\ell' \in L_j$  of atom  $j$ .

Next, we place these atom-pair blocks into a global block matrix to obtain the full RH Hamiltonian matrix: The full Hamiltonian matrix  $\mathbf{H} \in \mathbb{C}^{n \times n}$ , where  $n = \sum_\ell (2\ell+1)$ , is assembled by concatenating the atom-pair blocks:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{ii} & \mathbf{H}_{ji} & \dots \\ \mathbf{H}_{ij} & \mathbf{H}_{jj} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad n = \sum_{i \in \mathcal{M}} \sum_{\ell \in L_i} (2\ell+1), \quad (32)$$

for every atom-pair  $(i, j)$  in the molecule  $\mathcal{M}$ .

For better intuition, we illustrate a schematic structure of the Hamiltonian matrix in Figure 4a. Here, blocks corresponding to  $s, p$ , and  $d$  orbitals are arranged according to their angular momentum, highlighting the blockwise symmetry pattern induced by the underlying spherical harmonics basis.

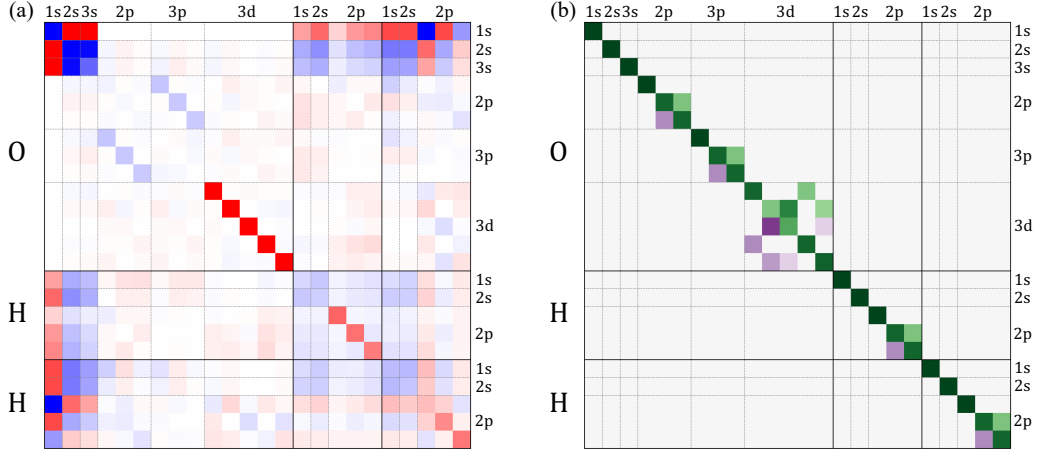


Figure 4: (a) Schematic illustration of the full Hamiltonian matrix  $\mathbf{H}$  for a water molecule ( $\text{H}_2\text{O}$ ). Color intensity indicates the magnitude of matrix elements, with red representing larger values and blue representing smaller values. (b) Schematic illustration of the full Wigner D-matrix  $\mathcal{D}$  corresponding to  $\mathbf{H}$ , where green denotes larger values and purple denotes smaller values. Gray solid and dashed lines separate molecular blocks and orbital blocks, respectively, corresponding to submatrices defined by atomic orbital pairs.

**Rotational equivariance of full Hamiltonian.** The global action of a rotation  $\mathbf{R}$  on the basis is encoded by the block-diagonal matrix as follows Equation (3):

$$\mathcal{D}(\mathbf{R}) = D^{(\ell_1)}(\mathbf{R}) \oplus \dots \oplus D^{(\ell_K)}(\mathbf{R}) = \begin{bmatrix} D^{(\ell_1)}(\mathbf{R}) & 0 & \dots \\ 0 & D^{(\ell_2)}(\mathbf{R}) & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad (33)$$

where each block  $D^\ell(\mathbf{R})$  corresponds to the Wigner D-matrix for angular momentum  $l$ . The Hamiltonian matrix as a whole then transforms under rotation as:

$$\mathbf{H}[\mathbf{C}] \xrightarrow{\mathbf{R}} \mathbf{H}[\mathbf{R} \cdot \mathbf{C}] = \mathbf{R} \cdot \mathbf{H}[\mathbf{C}] = \mathcal{D}(\mathbf{R}) \mathbf{H}[\mathbf{C}] \mathcal{D}(\mathbf{R})^{-1}. \quad (34)$$

This global equivariance condition guarantees that the Hamiltonian transforms consistently with molecular rotations, preserving the physical structure of the system. Leveraging this symmetry is critical for constructing SE(3)-equivariant machine learning models, which not only respect physical laws but also improve model generalization and interpretability in quantum chemistry tasks. For better intuition, we illustrate a schematic structure of the Wigner D-matrix in Figure 4b.

## 1009 B Training and sampling algorithms, and fine-tuning objective

### 1010 B.1 Training and sampling algorithms

1011 We investigate two types of prior distributions for the initial Hamiltonian  $\mathbf{H}_0$ : Gaussian orthogonal  
 1012 ensemble (GOE) and tensor expansion-based (TE) priors. For the GOE prior, we sample each matrix  
 1013 entry independently as  $M_{ij} \sim \mathcal{N}(0, \sigma^2)$ , where we set  $\sigma^2 = 1.0$  for MD17 and  $\sigma^2 = 0.1$  for QH9.

1014 For TE prior, we sample each irreducible component  $\mathbf{w}^l$  by drawing its radial norm from  
 1015  $\text{LogNormal}(1, \sigma^2 = 0.1)$  and applying a uniform  $\text{SO}(3)$  rotation to construct the full equivari-  
 1016 ant basis, which is then mapped to the Hamiltonian space via tensor expansion.

1017 These choices provide flexibility in modeling different types of initial distributions, depending on  
 1018 the target dataset and symmetry constraints. Formal definitions and proofs of equivariance for both  
 1019 priors are provided in [Appendix C.1](#) and [Appendix C.4](#).

1020 The training and sampling procedures based on the above rotationally invariant priors are summarized  
 1021 in [Algorithm 1](#) and [Algorithm 2](#), respectively.

---

#### Algorithm 1 QHFlow training procedure

---

**Require:** Dataset of molecular configurations  $\{\mathcal{M}_i\}$ , target Hamiltonians  $\mathbf{H}_{1,\mathcal{M}}$ , overlap matrix  $\mathbf{S}_{\mathcal{M}}$ , model  $f_\theta$

- 1: **for** each training step **do**
- 2:   Sample minibatch  $\mathcal{B} = \{\mathcal{M}_i\}_{i=1}^B$
- 3:   **for** each  $\mathcal{M}_i$  in  $\mathcal{B}$  **do**
- 4:     Sample initial Hamiltonian  $\mathbf{H}_0 \sim p_0$  and time  $t \sim \mathcal{U}(0, 1)$
- 5:     Compute interpolated Hamiltonian:  $\mathbf{H}_{t,\mathcal{M}} = (1 - t)\mathbf{H}_0 + t\mathbf{H}_{1,\mathcal{M}}$
- 6:     Predict  $\mathbf{H}_{1,\mathcal{M}}^{(\theta)} = f_\theta(\mathbf{H}_{t,\mathcal{M}}, \mathbf{S}_{\mathcal{M}}, \mathcal{M}, t)$
- 7:     Compute flow matching loss:  $\mathcal{L}_i = \|\mathbf{H}_{1,\mathcal{M}}^{(\theta)} - \mathbf{H}_{1,\mathcal{M}}\|^2$
- 8:   **end for**
- 9:   Compute  $\mathcal{L} = \frac{1}{B} \sum_{i=1}^B \mathcal{L}_i$ , update  $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$
- 10: **end for**

---



---

#### Algorithm 2 QHFlow sampling procedure

---

**Require:** Molecular configuration  $\mathcal{M}$ , initial Hamiltonian  $\mathbf{H}_0 \sim p_0$ , model  $f_\theta$ , time discretization  $\{t_0 = 0, t_1, \dots, t_K = 1\}$

- 1: Initialize  $\mathbf{H}_{t_0} = \mathbf{H}_0$
- 2: **for**  $k = 0$  to  $K - 1$  **do**
- 3:   Predict target Hamiltonian:  $\mathbf{H}_{1,\mathcal{M}}^{(\theta)} \leftarrow f_\theta(\mathbf{H}_{t_k}, \mathbf{S}, \mathcal{M}, t_k)$
- 4:   Compute conditional vector field:  $v_{t_k, \theta} \leftarrow (\mathbf{H}_{1,\mathcal{M}}^{(\theta)} - \mathbf{H}_{t_k}) / (1 - t_k)$
- 5:   Update Hamiltonian:  $\mathbf{H}_{t_{k+1}} \leftarrow \mathbf{H}_{t_k} + (t_{k+1} - t_k) \cdot v_{t_k, \theta}$
- 6: **end for**
- 7: Output  $\hat{\mathbf{H}}_1 = \mathbf{H}_{t_K}$

---

1022 These algorithms ensure that training aligns the predicted vector field with the true conditional flow,  
 1023 while inference produces a final Hamiltonian via integration through the learned ODE trajectory. This  
 1024 design ensures  $\text{SE}(3)$ -equivariance and physical consistency throughout the training and prediction  
 1025 processes.

### 1026 B.2 Fine-tuning objective

1027 After pre-training with the flow matching objective, we further fine-tune QHFlow to enhance its  
 1028 spectral accuracy by introducing energy alignment objectives. To implement the energy alignment  
 1029 objective, we adopt the weighted alignment loss (WALoss) from WANet [25].

1030 **Property of the RH-equation.** The relationship between the Hamiltonian matrix  $\mathbf{H}$ , the coefficient  
 1031 matrix  $\mathbf{C}$ , and the orbital energy matrix  $\epsilon$  is governed by the Roothaan–Hall equation:

$$\mathbf{H}\mathbf{C} = \mathbf{S}\mathbf{C}\epsilon, \quad (35)$$

1032 where  $\mathbf{S}$  is overlap matrix. Using orthonormality condition  $\mathbf{C}^\top \mathbf{S} \mathbf{C} = \mathbf{I}$ , this relation simplifies to:

$$\mathbf{C}^\top \mathbf{H} \mathbf{C} = \epsilon. \quad (36)$$

1033 This identity forms the foundation for designing spectral alignment objectives.

1034 **Weighted alignment loss (WALoss).** WALoss encourages the predicted Hamiltonian  $\hat{\mathbf{H}}$  to match the  
 1035 spectral structure of the converged SCF Hamiltonian  $\mathbf{H}^*$ . Let  $\mathbf{C}^*$  and  $\epsilon^*$  denote the eigenvectors and  
 1036 eigenvalues of  $\mathbf{H}^*$ , respectively. We define WALoss as:

$$\mathcal{L}_{\text{WALoss}} = \left\| \text{diag} \left( (\mathbf{C}^*)^\top \hat{\mathbf{H}} \mathbf{C}^* \right) - \epsilon^* \right\|_2, \quad (37)$$

1037 where  $\text{diag}(\cdot)$  extracts the diagonal elements and  $\|\cdot\|_2$  denotes the L2 norm.

1038 To emphasize physically important states, we place larger weights on eigenvalues up to the Lowest  
 1039 unoccupied molecular orbital (LUMO) *i.e.*, the  $k + 1$  lowest-energy orbitals. By projecting  $\hat{\mathbf{H}}$  onto  
 1040 the fixed eigenbasis  $\mathbf{C}^*$ , WALoss promotes alignment with the ground-truth spectrum. However, it is  
 1041 important to note that  $(\mathbf{C}^*)^\top \hat{\mathbf{H}} \mathbf{C}^*$  does not exactly diagonalize  $\hat{\mathbf{H}}$  unless  $\hat{\mathbf{H}} = \mathbf{H}^*$ . Nevertheless,  
 1042 minimizing this discrepancy helps guide  $\hat{\mathbf{H}}$  toward the correct eigenspectrum, improving orbital  
 1043 energy prediction and SCF behavior.

1044 **Overall fine-tuning objective.** Our final finetuning objective combines the flow matching loss with  
 1045 WALoss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CFM}} + \lambda_{\text{WA}} \mathcal{L}_{\text{WA}}, \quad (38)$$

1046 where  $\lambda_{\text{WA}}$  is the weighting coefficients controlling the contributions of the energy alignment terms,  
 1047 and we chose  $\lambda_{\text{WA}} = 2.0$  for the finetuning.

## 1048 C Proofs and additional details

### 1049 C.1 Proof of SE(3)-equivariance of GOE

1050 Let  $\mathbf{M} \in \mathbb{R}^{n \times n}$  be a symmetric matrix drawn from the Gaussian orthogonal ensemble (GOE),  
 1051 characterized by:

$$\mathbb{E}[M_{ij}] = 0, \quad \text{Var}(M_{ij}) = \sigma^2, \quad \text{with } M_{ij} = M_{ji}, \text{ and independent entries for } i \leq j. \quad (39)$$

1052 Let  $\mathbf{D} = \mathcal{D}(\mathbf{R})$  denote an orthogonal matrix representing a Wigner D-transformation corresponding  
 1053 to a rotation  $\mathbf{R} \in \text{SO}(3)$ , satisfying orthogonality *i.e.*  $\mathbf{D}\mathbf{D}^\top = \mathbf{I}$ . We define the rotated matrix as:

$$\mathbf{M}' = \mathbf{D}\mathbf{M}\mathbf{D}^\top, \quad \text{with } M'_{ij} = \sum_{k,l} D_{ik} M_{kl} D_{jl}. \quad (40)$$

1054 We aim to show that the distribution of  $\mathbf{M}'$  is the same as that of  $\mathbf{M}$ , *i.e.*, that GOE is invariant under  
 1055 conjugation by orthogonal transformations:

$$\rho(\mathbf{M}) = \rho(\mathbf{D}\mathbf{M}\mathbf{D}^\top). \quad (41)$$

1056 To establish this, it suffices to verify that the first and second moments of the entries in  $\mathbf{M}'$  match  
 1057 those of GOE.

1058 **Mean.** Since  $\mathbb{E}[M_{kl}] = 0$  for all  $k, l$ , we have:

$$\mathbb{E}[M'_{ij}] = \sum_{k,l} D_{ik} D_{jl} \mathbb{E}[M_{kl}] = 0. \quad (42)$$

1059 **Covariance.** We now compute the covariance of the entries  $M'_{ab}$  and  $M'_{cd}$ :

$$\mathbb{E}[M'_{ab} M'_{cd}] = \mathbb{E} \left[ \left( \sum_{i,j} D_{ai} D_{bj} M_{ij} \right) \left( \sum_{k,l} D_{ck} D_{dl} M_{kl} \right) \right] \quad (43)$$

$$= \sum_{i,j,k,l} D_{ai} D_{bj} D_{ck} D_{dl} \mathbb{E}[M_{ij} M_{kl}]. \quad (44)$$

1060 Using the independence and zero mean of GOE entries, we know:

$$\mathbb{E}[M_{ij} M_{kl}] = \sigma^2 \delta_{ik} \delta_{jl}. \quad (45)$$

1061 Thus,

$$\mathbb{E}[M'_{ab} M'_{cd}] = \sigma^2 \sum_{i,j} D_{ai} D_{bj} D_{ci} D_{dj} \quad (46)$$

$$\begin{aligned} &= \sigma^2 \left( \sum_i D_{ai} D_{ci} \right) \left( \sum_j D_{bj} D_{dj} \right) \\ &= \sigma^2 (\mathbf{D}\mathbf{D}^\top)_{ac} (\mathbf{D}\mathbf{D}^\top)_{bd} \\ &= \sigma^2 \delta_{ac} \delta_{bd}. \end{aligned} \quad (47)$$

1062 This matches the covariance structure of the original GOE matrix. Since both  $\mathbf{M}$  and  $\mathbf{M}'$  are jointly  
 1063 Gaussian with zero mean and identical covariances, we conclude:

$$\mathbf{D}\mathbf{M}\mathbf{D}^\top \sim \mathbf{M}, \quad \text{i.e., GOE is invariant under orthogonal conjugation.} \quad (48)$$

1064 This proves that the GOE prior is  $\text{SO}(3)$ -invariant, even for the high-rank Wigner D-matrices  $\ell \geq 2$ .



## 1065 C.2 Proof of the SE(3)-equivariance of tensor expansion operation

1066 Since the  $\mathbf{w}^{(\ell)}$  is the irrep vector, the  $m$ -th entry  $w_m^{(\ell)}$  transforms under rotation  $\mathbf{R} \in \text{SO}(3)$  as:

$$w_m^{(\ell)} \xrightarrow{\mathbf{R}} (\mathbf{R} \cdot \mathbf{w}^{(\ell)})_m = \left( \sum_{m'} D_{mm'}^{(\ell)}(\mathbf{R}) w_{m'}^{(\ell)} \right), \quad (49)$$

1067 where  $D^{(\ell)}(\mathbf{R}) \in \mathbb{C}^{(2\ell_1+1) \times (2\ell_2+1)}$  is the Wigner D-matrix.

1068 By substituting above Equation (49) in Equation (9), the entry of rotated tensor expansion is

$$\left( \bar{\otimes} \mathbf{w}^{(\ell)} \right)_{(m_1, m_2)}^{(\ell_1, \ell_2)} \xrightarrow{\mathbf{R}} \left( \bar{\otimes} [\mathbf{R} \cdot \mathbf{w}^{(\ell)}] \right) = \sum_m C_{(\ell_1, m_1), (\ell_2, m_2)}^{(\ell, m)} \left[ \sum_{m'} D_{mm'}^{(\ell)}(\mathbf{R}) w_{m'}^{(\ell)} \right], \quad (50)$$

1069 and we can exchange the order of summation:

$$\left( \bar{\otimes} \mathbf{w}^{(\ell)} \right)_{(m_1, m_2)}^{(\ell_1, \ell_2)} \xrightarrow{\mathbf{R}} \sum_{m'} \left( \sum_m C_{(\ell_1, m_1), (\ell_2, m_2)}^{(\ell, m)} D_{mm'}^{(\ell)}(\mathbf{R}) \right) w_{m'}^{(\ell)}. \quad (51)$$

1070 Now we can use a crucial identity from representation theory:

1071 **Theorem 2.** *The Clebsch–Gordan coefficients provide a basis change between the product representation  $D^{(\ell_1)}(\mathbf{R}) \otimes D^{(\ell_2)}(\mathbf{R})$  and the irreducible representation  $D^{(\ell)}(\mathbf{R})$ :*

$$\sum_m C_{(\ell_1, m_1), (\ell_2, m_2)}^{(\ell, m)} D_{mm'}^{(\ell)}(\mathbf{R}) = \sum_{m'_1, m'_2} D_{m_1 m'_1}^{(\ell_1)}(\mathbf{R}) D_{m_2 m'_2}^{(\ell_2)}(\mathbf{R}) C_{(\ell_1, m'_1), (\ell_2, m'_2)}^{(\ell, m')}. \quad (52)$$

1073 We can substitute the identity, then we have:

$$\left( \bar{\otimes} \mathbf{w}^{(\ell)} \right)_{(m_1, m_2)}^{(\ell_1, \ell_2)} \xrightarrow{\mathbf{R}} \sum_{m'_1, m'_2} D_{m_1 m'_1}^{(\ell_1)}(\mathbf{R}) D_{m_2 m'_2}^{(\ell_2)}(\mathbf{R}) \left( \sum_{m'} C_{(\ell_1, m'_1), (\ell_2, m'_2)}^{(\ell, m')} w_{m'}^{(\ell)} \right), \quad (53)$$

1074 and the last sum implies the

$$\sum_{m'} C_{(\ell_1, m'_1), (\ell_2, m'_2)}^{(\ell, m')} w_{m'}^{(\ell)} = \left( \bar{\otimes} \mathbf{w}^{(\ell)} \right)_{(m'_1, m'_2)}^{(\ell_1, \ell_2)} \quad (54)$$

1075 Therefore,

$$\left( \bar{\otimes} \mathbf{w}^{(\ell)} \right)_{(m_1, m_2)}^{(\ell_1, \ell_2)} \xrightarrow{\mathbf{R}} \sum_{m'_1, m'_2} D_{m_1 m'_1}^{(\ell_1)}(\mathbf{R}) D_{m_2 m'_2}^{(\ell_2)}(\mathbf{R}) \left( \bar{\otimes} \mathbf{w}^{(\ell_3)} \right)_{(m'_1, m'_2)}^{(\ell_1, \ell_2)}, \quad (55)$$

1076 in the matrix form,

$$\left( \bar{\otimes} \mathbf{w}^{(\ell)} \right) \xrightarrow{\mathbf{R}} D^{(\ell_1)}(\mathbf{R}) \left( \bar{\otimes} \mathbf{w}^{(\ell)} \right) D^{(\ell_2)}(\mathbf{R})^{-1}. \quad (56)$$

## 1077 C.3 Proof of the invariance of the tensor expansion based (TE) prior

1078 To construct an SO(3)-invariant distribution over matrices via tensor expansion, we begin with a  
1079 rotationally invariant distribution over irreducible features  $\mathbf{w}^{(\ell)} \in \mathbb{C}^{(2\ell+1)}$ . We define this distribution  
1080 by factorizing it into two independent components:

- 1081 • A magnitude  $r = \|\mathbf{w}^{(\ell)}\|$  drawn from a radial distribution, e.g.,  $r \sim \text{LogNormal}(1, \sigma^2)$ .
- 1082 • A direction sampled uniformly on the sphere  $\mathbb{S}^2$ , induced by a Haar-uniform rotation  
1083  $\mathbf{R} \in \text{SO}(3)$ , applied to a canonical unit-norm feature vector  $\mathbf{w}_0^{(\ell)}$ , such that  $\mathbf{w}^{(\ell)} =$   
1084  $r D^{(\ell)}(\mathbf{R}) \mathbf{w}_0^{(\ell)}$ .

1085 This construction ensures that  $\mathbf{w}^{(\ell)}$  is distributed isotropically, i.e.,  $p(\mathbf{w}^{(\ell)}) = p(D^{(\ell)}(\mathbf{R})\mathbf{w}^{(\ell)})$  for  
 1086 any  $\mathbf{R} \in \text{SO}(3)$ . Now, from the tensor expansion defined as:

$$\mathbf{H}^{(\ell_1, \ell_2)} = \left( \bar{\otimes} \mathbf{w}^{(\ell)} \right)^{(\ell_1, \ell_2)}, \quad (57)$$

1087 we know from the previous subsection that this construction transforms under  $\text{SO}(3)$  as:

$$\mathbf{H}^{(\ell_1, \ell_2)} \xrightarrow{\mathbf{R}} D^{(\ell_1)}(\mathbf{R}) \mathbf{H}^{(\ell_1, \ell_2)} D^{(\ell_2)}(\mathbf{R})^{-1}. \quad (58)$$

1088 To prove invariance, consider the pushforward distribution:

$$p(\mathbf{H}^{(\ell_1, \ell_2)}) = p(\mathbf{w}^{(\ell)}) = p(D^{(\ell)}(\mathbf{R})\mathbf{w}^{(\ell)}) \quad \forall \mathbf{R} \in \text{SO}(3). \quad (59)$$

1089 Because  $\mathbf{H}^{(\ell_1, \ell_2)}$  is a deterministic, equivariant function of  $\mathbf{w}^{(\ell)}$ , and  $p(\mathbf{w}^{(\ell)})$  is rotation-invariant, the  
 1090 induced distribution  $p(\mathbf{H}^{(\ell_1, \ell_2)})$  must also be invariant under conjugation by  $D^{(\ell_1)}(\mathbf{R})$  and  $D^{(\ell_2)}(\mathbf{R})$ .  
 1091 Thus:

$$p(\mathbf{H}^{(\ell_1, \ell_2)}) = p(D^{(\ell_1)}(\mathbf{R}) \mathbf{H}^{(\ell_1, \ell_2)} D^{(\ell_2)}(\mathbf{R})^{-1}). \quad (60)$$

1092 This proves that the prior distribution constructed via tensor expansion from rotationally invariant  
 1093  $\mathbf{w}^{(\ell)}$  yields an  $\text{SO}(3)$ -equivariant (conjugation-invariant) distribution over matrix-valued outputs.

#### 1094 C.4 Proof of the invariance of the multiple tensor expansion prior

1095 We now generalize the single-irrep vector expansion to construct a full Hamiltonian matrix  $\mathbf{H} \in \mathbb{R}^{n \times n}$   
 1096 composed of multiple irrep vectors components. Let the atomic orbital basis consist of a set of irrep  
 1097 vectors indexed by  $L = \{\ell_1, \ell_2, \dots, \ell_B\}$ , where each angular momentum  $\ell_i$  corresponds to a  
 1098 subspace of dimension  $(2\ell_i + 1)$ .

1099 For each irrep vector rank  $\ell_i$ , we define a random irrep vector  $\mathbf{w}^{(\ell_i)} \in \mathbb{C}^{(2\ell_i+1)}$ , sampled indepen-  
 1100 dently from a rotationally invariant distribution:

$$\mathbf{w}^{(\ell_i)} = r_i D^{(\ell_i)}(\mathbf{R}_i) \mathbf{w}_0^{(\ell_i)}, \quad r_i \sim \text{LogNormal}(1, \sigma^2), \quad \mathbf{R}_i \sim \text{Uniform}(\text{SO}(3)). \quad (61)$$

1101 We then define a factorized prior:

$$p(\mathbf{w}^{(\ell_1)}, \dots, \mathbf{w}^{(\ell_B)}) = \prod_{i=1}^B p_i(\mathbf{w}^{(\ell_i)}), \quad (62)$$

1102 with each  $p_i(\mathbf{w}^{(\ell_i)})$  being  $\text{SO}(3)$ -invariant.

1103 Using the tensor expansion, each pair  $(\ell_i, \ell_j)$  generates a block of the Hamiltonian matrix via:

$$\mathbf{H}^{(\ell_i, \ell_j)} = \bar{\otimes}_{\ell_i, \ell_j} \mathbf{w}^{(\ell_k)}, \quad \text{for some } \ell_k \in \{|\ell_i - \ell_j|, \dots, \ell_i + \ell_j\}. \quad (63)$$

1104 The full Hamiltonian is then assembled as:

$$\mathbf{H} = \bigoplus_{i,j} \mathbf{H}^{(\ell_i, \ell_j)}. \quad (64)$$

1105 From the single-component proof, we know that each block transforms as:

$$\mathbf{H}^{(\ell_i, \ell_j)} \xrightarrow{\mathbf{R}} D^{(\ell_i)}(\mathbf{R}) \mathbf{H}^{(\ell_i, \ell_j)} D^{(\ell_j)}(\mathbf{R})^{-1}. \quad (65)$$

1106 Therefore, under a global rotation  $\mathbf{R} \in \text{SO}(3)$ , the full matrix  $\mathbf{H}$  transforms as:

$$\mathbf{H} \xrightarrow{\mathbf{R}} \mathcal{D}(\mathbf{R}) \mathbf{H} \mathcal{D}(\mathbf{R})^{-1}, \quad (66)$$

1107 where  $\mathcal{D}(\mathbf{R})$  is the block-diagonal rotation operator acting on the full basis.

1108 Since each  $\mathbf{w}^{(\ell_i)}$  is sampled independently from an  $\text{SO}(3)$ -invariant distribution, the joint distribution  
 1109 of all irreducible components  $p(\mathbf{H})$  remains invariant under this global conjugation:

$$p(\mathbf{H}) = p\left(\mathcal{D}(\mathbf{R}) \mathbf{H} \mathcal{D}(\mathbf{R})^{-1}\right), \quad \forall \mathbf{R} \in \text{SO}(3). \quad (67)$$

1110 This completes the proof that the full Hamiltonian matrix constructed via multiple tensor expansions,  
 1111 using independently sampled spherical features from invariant priors, defines an  $\text{SO}(3)$ -invariant  
 1112 distribution over matrices.

## D Implementation details

### D.1 Model implementation

We build our model upon the official QHNet [23] and DEQHNet [58] codebases, which implement SE(3)-equivariant GNN for Hamiltonian matrix prediction. QHNet is selected as our backbone for its balance of architectural simplicity and strong predictive performance. Although more recent models such as WANet [25] and SPHNet [26] have demonstrated improvements, their codebases are not publicly available and thus are not considered here. Notably, our method is model-agnostic and could be paired with more expressive backbones if desired.

We extend QHNet by incorporating additional physically meaningful inputs: the current Hamiltonian matrix  $\mathbf{H}_t$ , the overlap matrix  $\mathbf{S}$ , the molecular configuration  $\mathcal{M}$ , and the time  $t$ . Unlike the original QHNet, which processes only  $\mathcal{M}$ , our model is designed to handle  $\mathbf{H}_t$ , and  $\mathbf{S}$  as well, following symmetry-preserving strategies inspired by DEQHNet and OrbNet-Equi [59].

**Project Block.** To map the equivariant matrix features  $\mathbf{H}$  and  $\mathbf{S}$  into the SO(3) irrep vectors, we used *diagonal reduction* in OrbNet [59] based on the Wigner-Eckart theorem [46]. For atom  $i$ , we first extract the atom-wise matrices,  $\mathbf{H}_i := \mathbf{H}_{ii}$  (see Equation (31)). Each element of  $\mathbf{H}_i$  carries two orbital index,  $(\ell, m)$  and  $(\ell', m')$ . We project these matrices onto rank- $\ell_o$  irrep vectors  $\mathbf{h}_i^{(\ell_o)} = [h_i^{(\ell_o, -\ell_o)}, \dots, h_i^{(\ell_o, \ell_o)}] \in \mathbb{C}^{(2\ell_o+1)}$  by

$$h_i^{(\ell_o, m_o)} = \sum_{(\ell, \ell')} \sum_{(m, m')} \mathbf{H}_i^{(\ell, m; \ell', m')} Q_{i, (\ell_o, m_o)}^{(\ell, m; \ell', m')}, \quad (68)$$

where  $\mathbf{H}_i^{(\ell, m; \ell', m')}$  is an element of submatrix  $\mathbf{H}_i^{(\ell, \ell')}$  that corresponds to the  $m$  and  $m'$  index, and a set of atom-wise projection weights  $Q_{i, (\ell_o, m_o)}^{(\ell, m; \ell', m')}$  which plays the role of Clebsch-Gordan-like projection weights. Because the map in Equation (68) is an SO(3) tensor contraction, the resulting feature vectors remain equivariant. The same procedure is applied to overlap matrix  $\mathbf{S}$ , giving two parallel streams of SE(3)-equivariant atomic features that are subsequently fused in the network.

**Construction of projection weights.** To construct the coefficients atom-wise projection coefficients, we compute three-center integrals involving orbital basis function  $\Phi_i^{(\ell, m)}$  and auxiliary spherical-type basis functions  $\tilde{\Phi}_i^{(\ell, m)}$ :

$$Q_{i, (\ell_o, m_o)}^{(\ell, m; \ell', m')} = \int \left( \Phi_i^{(\ell, m)}(\mathbf{r}) \right)^* \Phi_i^{(\ell', m')}(\mathbf{r}) \tilde{\Phi}_i^{(\ell_o, m_o)}(\mathbf{r}) d\mathbf{r}. \quad (69)$$

In practice, the angular parts of these integrals are related with the Clebsch-Gordan coefficients due to their relation to spherical harmonics:

$$Q_{i, (\ell_o, m_o)}^{(\ell, m; \ell', m')} \propto \int_{\mathbb{S}^2} Y_m^\ell(\hat{\mathbf{r}}) Y_{m'}^{\ell'}(\hat{\mathbf{r}}) (Y_{m_o}^{\ell_o}(\hat{\mathbf{r}}))^* d\mathbf{r}. \quad (70)$$

### D.2 Training objective

Our training objective adopts a residual learning strategy, following prior works such as WANet and SHNet. Rather than directly predicting the converged Hamiltonian, the model learns to approximate the residual between the initial and converged Hamiltonians using a time-dependent vector field within the conditional flow matching framework. We define the residual target as done in prior works:

$$\mathbf{H}_{1, \mathcal{M}} := \mathbf{H}_{\mathcal{M}}^* - \mathbf{H}_{\mathcal{M}}^{(0)}, \quad (71)$$

where  $\mathbf{H}_{\mathcal{M}}^{(0)}$  is the initial guess of Hamiltonian matrix<sup>2</sup>, and  $\mathbf{H}_{\mathcal{M}}^*$  is the SCF solution. The final prediction of Hamiltonian is obtained by summing the predicted residual and the initial Hamiltonian guess.

The conditional flow matching loss is defined as:

$$\mathcal{L} = \mathbb{E}_{(\mathbf{H}, \mathcal{M}) \sim \mathcal{A}, t \sim \mathcal{U}(0, 1), \mathbf{H}_t \sim p_t} \left[ \|v_{t, \theta}(\mathbf{H}_t, \mathcal{M}) - u_t(\mathbf{H}_t, \mathcal{M} \mid \mathbf{H}_{1, \mathcal{M}})\|_2^2 \right] \quad (72)$$

$$= \mathbb{E}_{(\mathbf{H}, \mathcal{M}) \sim \mathcal{A}, t \sim \mathcal{U}(0, 1), \mathbf{H}_t \sim p_t} \left[ \frac{1}{(1-t)^2} \left\| \mathbf{H}_{1, \mathcal{M}}^{(\theta)} - \mathbf{H}_{1, \mathcal{M}} \right\|_2^2 \right], \quad (73)$$

<sup>2</sup>We use the minao initialization from the PySCF package, and this guess does not contains SCF steps.

1149 where the second line follows from the analytical form of the conditional velocity field  $u_t$ , showing  
1150 that the model is penalized based on the squared residual error at each time step. This formulation  
1151 encourages smooth convergence from the initial guess to the target solution. To simplify the objective,  
1152 we omit the  $1/(1 - t)^2$  time-scaling for our implementation.

## E Experimental study settings

### E.1 Dataset preparation

To demonstrate the effectiveness of flow-matching-based training, we conduct experiments on two molecular datasets: MD17 and QH9. The MD17 represents a relatively simple task compared to the QH9, focusing solely on small systems and their conformational space. The PubChemQH9 is not considered since their dataset and codebase are not publicly released.

**MD17.** The MD17 [49, 19] dataset consists of quantum chemical simulations for four small organic molecules: water ( $\text{H}_2\text{O}$ ), ethanol ( $\text{C}_2\text{H}_5\text{OH}$ ), malondialdehyde ( $\text{CH}_2(\text{CHO})_2$ ), and uracil ( $\text{C}_4\text{H}_4\text{N}_2\text{O}_2$ ). It provides a comprehensive set of molecular properties, including geometries, total energies, forces, Kohn–Sham Hamiltonian matrices, and overlap matrices. All reference computations were implemented via the ORCA electronic structure package [60] using the PBE exchange–correlation functional [52, 53] and the def2-SVP Gaussian-type orbital (GTO) basis set. We follow the standard data split protocol used in prior work [19, 20, 23] to divide each molecule’s conformational data into training, validation, and test sets. The detailed dataset statistics are summarized in Table 5 and MOs in the table imply molecular orbitals (*i.e.* s,p,d,f)

Table 5: The statistics of MD17 dataset [19].

Dataset	# of structures	Train	Val	Test	# of atoms	# of orbitals	# of occupied MOs
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

**QH9.** The QH9 dataset [24] is a large-scale quantum chemistry benchmark designed to support the training and evaluation of machine learning models for Hamiltonian matrix prediction across diverse chemical structures. It is based on the QM9 [50, 51] molecular dataset and includes 130,831 Hamiltonian matrices from stable molecular geometries, as well as 2,698 molecular dynamics trajectories. The dataset covers small organic molecules composed of up to nine heavy atoms (C, N, O, and F). All Hamiltonians were computed using the PySCF [61] quantum chemistry package with the B3LYP [54] exchange–correlation functional and the def2-SVP Gaussian-type orbital (GTO) basis set. We provide the detailed dataset statistics in Table 6.

QH9 is organized into two main subsets, QH9-stable and QH9-dynamic-300k, and provides four standard evaluation splits: stable-id, stable-ood, dynamic-300k-geo, and dynamic-300k-mol. The id split randomly partitions the QH9-stable subset into training, validation, and test sets, while the ood split is constructed based on molecular size. Specifically, it groups molecules with 3–20 atoms in the training set, molecules with 21–22 atoms in the validation set, and molecules with 23–29 atoms in the test set. This ood setup allows for a rigorous evaluation of out-of-distribution (OOD) generalization in terms of molecular complexity.

The geo and mol splits are based on molecular dynamics (MD) trajectories, where each molecule is associated with 100 geometry snapshots. In the geo split, geometries are randomly assigned within each molecule: 80 for training, 10 for validation, and 10 for testing. Thus, while all molecules are present across the splits, the specific conformations are disjoint, enabling an assessment of geometric generalization. In contrast, the mol split divides the 2,698 molecules themselves into disjoint sets with an 80/10/10 train/validation/test ratio. All 100 geometries of a molecule are assigned to the same subset. This split presents a more challenging task than the geo split, as it requires generalization to entirely unseen molecular identities rather than just new conformations.

### E.2 Evaluation Metrics

To comprehensively evaluate the performance of Hamiltonian prediction models, we adopt several metrics that measure accuracy, physical fidelity, and downstream impact on quantum chemical properties. Below are detailed descriptions of each metric used in our experiments.

Table 6: The statistics of QH9 dataset [24].

Dataset	# of structures	# of Molecules	Train	Val	Test
Stable-id	130,831	130,831	104,664	13,083	13,084
Stable-ood	130,831	130,831	104,001	17,495	9,335
Dynamic-300k-geo	269,800	2,698	215,840	26,980	26,980
Dynamic-300k-mol	269,800	2,698	215,800	26,900	27,100

**Hamiltonian MAE.** This metric measures the mean absolute error (MAE) between the predicted Hamiltonian matrix  $\hat{\mathbf{H}}$  and the ground-truth matrix  $\mathbf{H}^*$ :

$$\text{MAE}(\mathbf{H}) = \frac{1}{n^2} \sum_{i,j=1}^n \left\| \hat{\mathbf{H}}_{ij} - \mathbf{H}_{ij}^* \right\|_2^2, \quad (74)$$

where  $n$  is the length of the Hamiltonian matrix ( $\mathbf{H} \in \mathbb{R}^{n \times n}$ ), and  $\|\cdot\|_2^2$  is the Frobenius norm. This metric directly reflects the quality of the predicted Hamiltonian in element-wise terms.

**Occupied orbital energy MAE ( $\epsilon_{\text{occ}}$ ).** This metric measures the MAE between the predicted and true orbital energies for only the occupied orbitals. Let  $\hat{\epsilon}$  and  $\epsilon^*$  be predicted and the ground-truth generalized eigenvalues of the Hamiltonian, respectively, and let  $\mathcal{I}_{\text{occ}}$  be the set of indices of the occupied orbitals:

$$\text{MAE}(\epsilon_{\text{occ}}) = \frac{1}{\text{card}(\mathcal{I}_{\text{occ}})} \sum_{i \in \mathcal{I}_{\text{occ}}} \|\hat{\epsilon}_i - \epsilon_i^*\|_2^2. \quad (75)$$

This is particularly important for capturing the energy spectrum relevant to ground-state electronic properties. We identify the occupied orbitals by selecting the  $\lfloor N/2 \rfloor$  lowest eigenvalues, where  $N$  is the number of electrons in the system.

**Orbital coefficient similarity score ( $S_C$ ).** To compare the predicted and ground-truth molecular orbital coefficient matrices  $\hat{\mathbf{C}}$  and  $\mathbf{C}^*$ , we use a cosine similarity-based score averaged over all columns:

$$S_C(\hat{\mathbf{C}}, \mathbf{C}^*) = \left\langle \hat{\mathbf{C}}, \mathbf{C}^* \right\rangle = \frac{\sum_i \mathbf{C}_i \mathbf{C}_i^*}{\|\hat{\mathbf{C}}\| \|\mathbf{C}^*\|}, \quad (76)$$

where  $\mathbf{C}_i$  is the  $i$ -th column vector of the coefficient matrix and  $\langle \cdot, \cdot \rangle$  denotes the inner product. This metric evaluates how well the predicted orbitals align with the ground truth.

**HOMO, LUMO, and energy gap MAE ( $\epsilon_{\text{HOMO}}$ ,  $\epsilon_{\text{LUMO}}$ ,  $\epsilon_{\Delta}$ ).** These metrics quantify the predictive accuracy of specific frontier orbital energies, which are critical for determining electronic properties such as reactivity and charge transport. The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies, along with their difference (the HOMO–LUMO gap), are particularly sensitive to model generalization, especially in challenging settings like QH9, where predictions are evaluated on unseen molecular structures or conformations. These metrics are computed as follows:

$$\epsilon_{\text{HOMO}} = |\hat{\epsilon}_{\text{HOMO}} - \epsilon_{\text{HOMO}}^*|, \quad (77)$$

$$\epsilon_{\text{LUMO}} = |\hat{\epsilon}_{\text{LUMO}} - \epsilon_{\text{LUMO}}^*|, \quad (78)$$

$$\epsilon_{\Delta} = |(\hat{\epsilon}_{\text{LUMO}} - \hat{\epsilon}_{\text{HOMO}}) - (\epsilon_{\text{LUMO}}^* - \epsilon_{\text{HOMO}}^*)|. \quad (79)$$

Here,  $\epsilon_{\text{HOMO}}$  corresponds to the  $\lfloor N/2 \rfloor$ -th lowest eigenvalue, and  $\epsilon_{\text{LUMO}}$  to the  $(\lfloor N/2 \rfloor + 1)$ -th, where  $N$  is the number of electrons in the system.

**SCF Iter Ratio:** This metric measures the reduction in the number of SCF iterations required when using a predicted Hamiltonian as the initial guess, relative to a standard reference initialization (e.g., from a minimal basis or atomic superposition guess). It is defined as

$$\text{SCF Iter Ratio} = \frac{\text{Iter}_{\text{pred}}}{\text{Iter}_{\text{ref}}}, \quad (80)$$

where  $\text{Iter}_{\text{pred}}$  is the number of SCF iterations with the predicted initialization, and  $\text{Iter}_{\text{ref}}$  is the number of iterations under the reference setup. In our setting, we used the reference value, which is measured from the conventional DFT metric to compare the acceleration of our method.

1226 **SCF T Ratio:** This represents the reduction in total wall-clock time spent during SCF convergence  
 1227 using the predicted Hamiltonian. It reflects actual runtime savings in quantum chemical simulations:

$$\text{SCF T Ratio} = \frac{T_{\text{SCF-pred}}}{T_{\text{SCF-ref}}}, \quad (81)$$

1228 where  $T_{\text{SCF-pred}}$  and  $T_{\text{SCF-ref}}$  denote the SCF runtimes under predicted and reference initializations,  
 1229 respectively.

1230 **Inf T Ratio:** This metric captures the inference overhead of the predictive model itself. It is the  
 1231 fraction of time required to generate the predicted Hamiltonian relative to the baseline SCF time:

$$\text{Inf T Ratio} = \frac{T_{\text{inference}}}{T_{\text{SCF-ref}}}, \quad (82)$$

1232 where  $T_{\text{inference}}$  is the runtime of the Hamiltonian prediction model.

1233 **Total T Ratio:** This quantifies the net cost of using the predictive model, combining both inference  
 1234 and SCF convergence times. It provides a holistic view of overall computational efficiency:

$$\text{Total T Ratio} = \frac{T_{\text{inference}} + T_{\text{SCF-pred}}}{T_{\text{SCF-ref}}}. \quad (83)$$

1235 **Note.** All energy values are reported in units of Hartree ( $1E_h = 27.211eV$ ), and similarity scores  
 1236 are unitless with values in  $[0, 1]$ , where higher is better. While the SCF Iter Ratio is a deterministic  
 1237 metric reflecting algorithmic convergence behavior, the wall-clock timing ratios (SCF T, Inf T,  
 1238 and Total T) are subject to variability due to system-level factors. In our experiments, all timing  
 1239 evaluations were conducted on A100-80GB GPUs. However, due to resource allocation via SLURM  
 1240 and shared CPU environments, precise time measurements were not always reproducible. As such,  
 1241 reported time ratios should be interpreted as indicative trends rather than absolute benchmarks.

### 1242 E.3 Experimental setup

1243 **Environment.** Experiments were conducted using a single GPU per model. For the MD17 dataset, we  
 1244 used NVIDIA RTX 3090 and A5000 GPUs, while for the QH9 dataset, experiments were performed  
 1245 on NVIDIA A100 80GB GPUs. Our implementation is based on PyTorch 2.1.2 and PyG 2.3.0, both  
 1246 compiled with CUDA 12.1. Detailed package versions [61–63] and environment specifications will  
 1247 be released upon publication for full reproducibility.

1248 Training was performed on a SLURM-managed cluster, where slight variability in runtime was  
 1249 observed due to system-wide GPU and CPU resource contention. Random seeds were fixed where  
 1250 possible to ensure training stability; however, minor non-determinism remains due to the inherent  
 1251 variability of distributed computing environments.

1252 Training QHFlow took approximately 2–4 days on MD17 and 7–10 days on QH9, which is slightly  
 1253 longer than QHNet due to the added complexity of incorporating the current Hamiltonian and orbital  
 1254 information as inputs. Finetuning on QH9 required an additional 3–4 days. While the total training  
 1255 time could be further reduced with optimized infrastructure, our experiments were conducted under a  
 1256 SLURM-managed environment with periodic reinitialization every 72 hours. Importantly, QHFlow is  
 1257 a model-agnostic framework; with more scalable architectures, training efficiency can be significantly  
 1258 improved.

1259 **Hyperparameters.** For fair comparison, we adopted the same hyperparameters used by the baseline  
 1260 model QHNet [23] whenever possible. Our QHFlow shares the majority of its architecture and training  
 1261 settings with QHNet to ensure that performance improvements are attributable to the proposed flow  
 1262 matching design rather than hyperparameter tuning. A summary of the key hyperparameters used  
 1263 across different datasets is provided in Table 7.

Table 7: Training hyperparameters of QHFlow used across datasets.

Hyperparameter	Description	QH9	MD17 (Water)	MD17 (Others)
Learning Rate	Initial learning rate	1e-3	5e-4	1e-3
Minimum Learning Rate	Minimum learning rate	1e-7	1e-9	1e-9
Batch Size	Number of molecules per batch	32	10	5
Scheduler	Learning rate scheduler	Polynomial	Polynomial	Polynomial
LR Warmup Steps	Warmup steps to linearly increase learning rate	1,000	1,000	1,000
Max Steps	Maximum number of training steps	260,000	200,000	200,000
Fine-tuning LR	Initial learning rate of fine-tuning stage	1e-5	-	-
Fine-tuning Minimum LR	Minimum learning rate of fine-tuning stage	1e-7	-	-
Finetuning Steps	Maximum number of Finetuning steps	60,000	-	-
Prior Distribution	Prior distribution for flow matching	GOE / TE	GOE	GOE
Using $H_t$ Block	Use time-dependent Hamiltonian $H_t$ as input	True	True	True
Using $S$ Block	Use overlap matrix $S$ as input	True	False	False
Model Order	Maximum degree of spherical harmonics	4	4	4
Embedding Dimension	Node feature embedding dimension	128	128	128
Bottle Hidden Size	Hidden size of bottleneck layer	32	32	32
Number of GNN Layers	Number of graph neural network layers	5	5	5
Max Radius	Cutoff radius for neighbor search	15	15	15
Sphere Channels	Number of channels in spherical basis	128	128	128
Edge Channels	Number of channels for edge features	32	32	32



## 1264 F Additional experimental results and limitations

### 1265 F.1 Ablation study of the finetuning objective.

1266 We conduct additional experiments to evaluate various fine-tuning strategies to improve the Hamiltonian prediction performance of QHFlow after standard flow matching pretraining. In this section, we  
1267 describe each strategy in detail and provide empirical observations.

1269 For the WALoss implementation, we follow the coefficient settings proposed in WANet [25] ( $\lambda_{WA} =$   
1270 2.0). During fine-tuning, we train for an additional 60,000 steps starting from the pretrained model  
1271 with TE prior, using an initial learning rate of  $1 \times 10^{-5}$  and a polynomial learning rate scheduler.

1272 **Full FM + WA Loss training from scratch (WA-Full).** In this strategy, we train the model from  
1273 scratch by jointly optimizing the flow matching objective and WALoss for 260,000 steps, matching  
1274 the default training steps as done in WANet. However, this approach significantly degraded the  
1275 Hamiltonian error. We hypothesize that imposing strong spectral constraints too early disrupts the  
1276 learning of the global Hamiltonian structure, leading to unstable convergence and poor generalization.

1277 **WA Loss finetuning (WA-FT).** After completing standard flow matching pre-training, we fine-tune  
1278 the model solely using WALoss, which encourages alignment between the predicted Hamiltonian’s  
1279 eigenstructure and the ground-truth SCF solutions. This approach improves orbital energy prediction  
1280 and SCF convergence compared to pure flow matching training, confirming the value of adding  
1281 spectral supervision during fine-tuning.

1282 **Summary.** Table 8 summarizes the quantitative results of all fine-tuning strategies. Here, w/o-FT  
1283 refers to QHFlow trained solely with the flow matching objective using the TE prior. Among the  
1284 approaches, WA-FT achieves the best balance between Hamiltonian prediction accuracy and practical  
1285 DFT acceleration, demonstrating the effectiveness of spectral alignment in the fine-tuning stage.

### 1286 F.2 Full results of ablation study

1287 For completeness, we provide the full ablation study results in Table 9, and Table 10, including  
1288 additional metrics beyond those presented in the main text. While the main manuscript focused  
1289 primarily on Hamiltonian prediction error ( $H$ ), occupied orbital energy error ( $\epsilon_{occ}$ ), and electronic  
1290 density accuracy ( $S_c$ ) for clarity, here we also report the LUMO ( $\epsilon_{LUMO}$ ), HOMO ( $\epsilon_{HOMO}$ ), and  
1291 HOMO-LUMO gap ( $\epsilon_{\Delta}$ ) energy errors.

1292 Overall, the trends observed in the full table are consistent with those discussed in the main text. In  
1293 particular, the improvements in Hamiltonian accuracy and density prediction are accompanied by  
1294 consistent reductions in orbital energy errors ( $\epsilon_{LUMO}$ ,  $\epsilon_{HOMO}$ ) and gap errors ( $\epsilon_{\Delta}$ ), further demonstrat-  
1295 ing the broad effectiveness of our proposed methods. These results reinforce the conclusion that flow  
1296 matching enhance both Hamiltonian structure prediction and downstream electronic properties.

1297 Also, to better visualize the role of the prior distributions, we provide interpolation trajectories from  
1298 the GOE and TE prior toward target Hamiltonians. Figure 5 shows example interpolations for both  
1299 settings, highlighting how the initial distribution affects the learned flow dynamics. As seen, the TE  
1300 prior results in smoother and more structured trajectories compared to the GOE prior, which starts  
1301 from unstructured noise.

### 1302 F.3 Analysis of error distribution of the predictions

1303 To better understand the properties of QHFLOW, we move beyond single-number summaries and  
1304 visualise the entire error distribution for six key quantities Hamiltonian MAE ( $H$ ), occupied-orbital  
1305 MAE ( $\epsilon_{occ}$ ),  $\epsilon_{HOMO}$ ,  $\epsilon_{LUMO}$ ,  $\epsilon_{\Delta}$ , and  $S_c$  on all four QH9 splits (id, ood, geo, mol). Figure 6 shows  
1306 log-scale violin plots for QHNet\* (grey), QHFLOW (orange), and its WA-finetuned variant (cyan).  
1307 (only the similarity score is kept on a linear axis.) Each violin displays the full empirical distribution;  
1308 the thick dashed line marks the median, and the two thinner lines divide the inter-quartile ranges.

1309 Across every split and metric, the QHFLOW violins are both *lower* and *narrower*. Median Hamiltonian  
1310 error is reduced by roughly one order of magnitude on id and geo, and by a factor of 3–5 on the  
1311 tougher ood and mol sets. In addition, the long error tails of QHNet\* largely vanish, demonstrating  
1312 that our flow objective mitigates the occasional catastrophic failure modes that plague point-wise  
1313 regression models.

Table 8: Ablation experimental results on QH9 dataset Finetuning. Results shown in **bold** denote the best result in each column, whereas those that are underlined indicate the second best.

Dataset	Type	$H$ [ $\mu E_h$ ] $\downarrow$	$\epsilon_{\text{occ}}$ [ $\mu E_h$ ] $\downarrow$	$\mathcal{S}_c$ [%] $\uparrow$	$\epsilon_{\text{LUMO}}$ [ $\mu E_h$ ] $\downarrow$	$\epsilon_{\text{HOMO}}$ [ $\mu E_h$ ] $\downarrow$	$\epsilon_{\Delta}$ [ $\mu E_h$ ] $\downarrow$
QH9-stable (id)	WA-Full	62.69	153.82	98.96	<u>240.07</u>	<u>117.26</u>	<u>251.81</u>
	w/o-FT	<b>22.95</b> $\pm 0.001$	<u>119.67</u> $\pm 0.211$	<u>99.51</u> $\pm 0.001$	437.96 $\pm 7.452$	179.48 $\pm 0.098$	553.87 $\pm 7.454$
	WA-FT	<u>23.85</u> $\pm 0.003$	<b>101.92</b> $\pm 0.279$	<b>99.56</b> $\pm 0.002$	<b>187.48</b> $\pm 6.434$	<b>92.22</b> $\pm 0.234$	<b>206.15</b> $\pm 6.197$
QH9-stable (ood)	WA-Full	57.11	110.34	98.12	416.08	<u>113.55</u>	<u>387.83</u>
	w/o-FT	<b>20.01</b> $\pm 0.001$	<u>84.54</u> $\pm 0.007$	<u>99.04</u> $\pm 0.003$	321.20 $\pm 1.497$	130.74 $\pm 0.043$	395.83 $\pm 1.510$
	WA-FT	<u>20.55</u> $\pm 0.002$	<b>72.64</b> $\pm 0.018$	<b>99.16</b> $\pm 0.006$	<b>171.24</b> $\pm 0.273$	<b>77.96</b> $\pm 0.095$	<b>179.57</b> $\pm 0.271$
QH9-dynamic (300k-geo)	WA-Full	75.38	108.96	98.77	<u>231.63</u>	<u>114.14</u>	<u>219.20</u>
	w/o-FT	<b>25.94</b> $\pm 0.001$	<u>103.11</u> $\pm 0.031$	<u>99.59</u> $\pm 0.001$	425.18 $\pm 1.119$	175.18 $\pm 0.255$	547.33 $\pm 1.168$
	WA-FT	<u>27.12</u> $\pm 0.002$	<b>89.03</b> $\pm 0.213$	<b>99.65</b> $\pm 0.001$	<b>136.63</b> $\pm 4.661$	<b>84.17</b> $\pm 0.211$	<b>154.68</b> $\pm 4.449$
QH9-dynamic (300k-mol)	WA-Full	103.47	692.17	97.02	<u>1046.21</u>	760.85	<u>1408.36</u>
	w/o-FT	<b>45.91</b> $\pm 0.001$	<u>442.56</u> $\pm 0.171$	<u>98.65</u> $\pm 0.001$	1344.68 $\pm 2.338$	479.71 $\pm 0.150$	1605.03 $\pm 2.286$
	WA-FT	<u>46.60</u> $\pm 0.003$	<b>424.75</b> $\pm 0.324$	<b>98.74</b> $\pm 0.001$	<b>912.10</b> $\pm 2.941$	<b>403.51</b> $\pm 1.861$	<b>1047.88</b> $\pm 2.683$

Table 9: Full experimental results on QH9 dataset along the initial distribution. Results shown in **bold** denote the best result in each column.

Dataset	Prior	$H$ $\downarrow$ [ $10^{-6} E_h$ ]	$\epsilon_{\text{occ}}$ $\downarrow$ [ $10^{-6} E_h$ ]	$\mathcal{S}_c$ $\uparrow$ [ $10^{-2}$ ]	$\epsilon_{\text{LUMO}}$ $\downarrow$ [ $10^{-6} E_h$ ]	$\epsilon_{\text{HOMO}}$ $\downarrow$ [ $10^{-6} E_h$ ]	$\epsilon_{\Delta}$ $\downarrow$ [ $10^{-6} E_h$ ]
QH9-stable (id)	GOE	25.93 $\pm 0.001$	154.65 $\pm 1.097$	99.39 $\pm 0.001$	638.03 $\pm 7.744$	220.49 $\pm 0.192$	764.46 $\pm 7.635$
	TE	<b>22.95</b> $\pm 0.001$	<b>119.61</b> $\pm 0.211$	<b>99.51</b> $\pm 0.001$	<b>437.96</b> $\pm 7.452$	<b>179.48</b> $\pm 0.098$	<b>553.87</b> $\pm 7.454$
QH9-stable (ood)	GOE	21.93 $\pm 0.001$	<b>87.32</b> $\pm 0.012$	98.95 $\pm 0.002$	382.09 $\pm 13.87$	<b>120.16</b> $\pm 0.070$	432.93 $\pm 13.84$
	TE	<b>20.01</b> $\pm 0.001$	<u>84.54</u> $\pm 0.007$	<b>99.04</b> $\pm 0.003$	<b>321.20</b> $\pm 1.497$	130.74 $\pm 0.043$	<b>395.83</b> $\pm 1.510$
QH9-dynamic (300k-geo)	GOE	29.39 $\pm 0.001$	122.14 $\pm 0.050$	99.49 $\pm 0.001$	<b>618.75</b> $\pm 3.089$	215.41 $\pm 0.569$	<b>756.45</b> $\pm 3.371$
	TE	<b>25.94</b> $\pm 0.001$	<b>103.11</b> $\pm 0.031$	<b>99.59</b> $\pm 0.001$	425.18 $\pm 1.119$	<b>175.18</b> $\pm 0.255$	547.33 $\pm 1.168$
QH9-dynamic (300k-mol)	GOE	46.78 $\pm 0.001$	<b>419.68</b> $\pm 0.001$	<b>98.65</b> $\pm 0.001$	1409.07 $\pm 3.759$	<b>478.12</b> $\pm 0.214$	1718.80 $\pm 3.907$
	TE	<b>45.91</b> $\pm 0.001$	<u>442.56</u> $\pm 0.171$	<b>98.65</b> $\pm 0.001$	<b>1344.68</b> $\pm 2.338$	479.71 $\pm 0.150$	<b>1605.03</b> $\pm 2.286$

Figure 7 plots  $\log H$  against  $\log \epsilon_{\text{occ}}$  for every test geometry. QHFLOW occupies the bottom-left corner of each panel, while QHNet\* points fan out towards larger joint errors, especially in ood and mol, where chemical diversity is highest. These results show the QHFLOW superior performance on both the geometric and unseen molecules.

Finally, Figure 8 compares Hamiltonian error versus atom count. QHFLOW remains flat up to the largest molecules in QH9 (29atoms). This suggests that modeling a *distribution* over Hamiltonians confers a form of size transferability.

#### F.4 Limitations

Our study demonstrates flow matching only on top of QHNet, mainly because its code is publicly available and its training cost is moderate. Although the approach should, in principle, be compatible with other SE(3) equivariant backbones (e.g., WANet or SPHNet), those implementations are not open-sourced, so we could not verify the transferability of our gains. Furthermore, the flow formulation increases the training time by approximately 10–20% compared to the vanilla QHNet, due to the need to process intermediate Hamiltonians. Finally, all experiments are limited to gas-phase molecules of up to 29 atoms and to two common XC functionals (PBE and B3LYP). Extending the method to much larger systems, periodic solids, or higher-accuracy levels of theory (e.g., hybrid or double-hybrid functionals) is left for future work.

Table 10: Full predictive variance results on QH9 dataset. WA-FT implies the finetune the model with WA loss. Metrics for Ours and Ours (WA-FT) are means $\pm$ std over five random seeds. Results shown in **bold** denote the best result in each column, whereas those that are underlined indicate the second best.

Dataset	Model	$H$ [ $\mu E_h$ ] $\downarrow$	$\epsilon_{\text{occ}}$ [ $\mu E_h$ ] $\downarrow$	$\mathcal{S}_c$ [%] $\uparrow$	$\epsilon_{\text{LUMO}}$ [ $\mu E_h$ ] $\downarrow$	$\epsilon_{\text{HOMO}}$ [ $\mu E_h$ ] $\downarrow$	$\epsilon_{\Delta}$ [ $\mu E_h$ ] $\downarrow$
QH9-stable (id)	QHNet*	77.72	963.45	94.80	18257.34	1546.27	17822.62
	WANet	80.00	833.62	96.86	-	-	-
	SPHNet	45.48	334.28	97.75	-	-	-
	<b>Ours</b>	<b>22.95</b> $\pm 0.001$	<b>119.67</b> $\pm 0.211$	<b>99.51</b> $\pm 0.001$	<b>437.96</b> $\pm 7.452$	<b>179.48</b> $\pm 0.098$	<b>553.87</b> $\pm 7.454$
	<b>Ours (WA-FT)</b>	<b>23.85</b> $\pm 0.003$	<b>101.92</b> $\pm 0.279$	<b>99.56</b> $\pm 0.002$	<b>187.48</b> $\pm 6.434$	<b>92.22</b> $\pm 0.234$	<b>206.15</b> $\pm 6.197$
QH9-stable (ood)	QHNet*	69.69	884.97	93.01	25848.83	1045.99	25370.10
	SPHNet	43.33	186.40	98.16	-	-	-
	<b>Ours</b>	<b>20.01</b> $\pm 0.001$	<b>84.54</b> $\pm 0.007$	<b>99.04</b> $\pm 0.003$	<b>321.20</b> $\pm 1.497$	<b>130.74</b> $\pm 0.043$	<b>395.83</b> $\pm 1.510$
	<b>Ours (WA-FT)</b>	<b>20.55</b> $\pm 0.002$	<b>72.64</b> $\pm 0.018$	<b>99.16</b> $\pm 0.006$	<b>171.24</b> $\pm 0.273$	<b>77.96</b> $\pm 0.095$	<b>179.57</b> $\pm 0.271$
QH9-dynamic (300k-geo)	QHNet*	88.36	1170.50	93.65	23269.41	2040.06	22407.96
	WANet	74.74	416.57	99.68	-	-	-
	SPHNet	52.18	100.88	99.12	-	-	-
	<b>Ours</b>	<b>25.94</b> $\pm 0.001$	<b>103.11</b> $\pm 0.031$	<b>99.59</b> $\pm 0.001$	<b>425.18</b> $\pm 1.119$	<b>175.18</b> $\pm 0.255$	<b>547.33</b> $\pm 1.168$
	<b>Ours (WA-FT)</b>	<b>27.12</b> $\pm 0.002$	<b>89.03</b> $\pm 0.213$	<b>99.65</b> $\pm 0.001$	<b>136.63</b> $\pm 4.661$	<b>84.17</b> $\pm 0.211$	<b>154.68</b> $\pm 4.449$
QH9-dynamic (300k-mol)	QHNet*	121.39	5554.36	86.02	53505.09	4352.76	50424.86
	SPHNet	108.19	1724.10	91.49	-	-	-
	<b>Ours</b>	<b>45.91</b> $\pm 0.001$	<b>442.56</b> $\pm 0.171$	<b>98.65</b> $\pm 0.001$	<b>1344.68</b> $\pm 2.338$	<b>479.71</b> $\pm 0.150$	<b>1605.03</b> $\pm 2.286$
	<b>Ours (WA-FT)</b>	<b>46.60</b> $\pm 0.003$	<b>424.75</b> $\pm 0.324$	<b>98.74</b> $\pm 0.001$	<b>912.10</b> $\pm 2.941$	<b>403.51</b> $\pm 1.861$	<b>1047.88</b> $\pm 2.683$

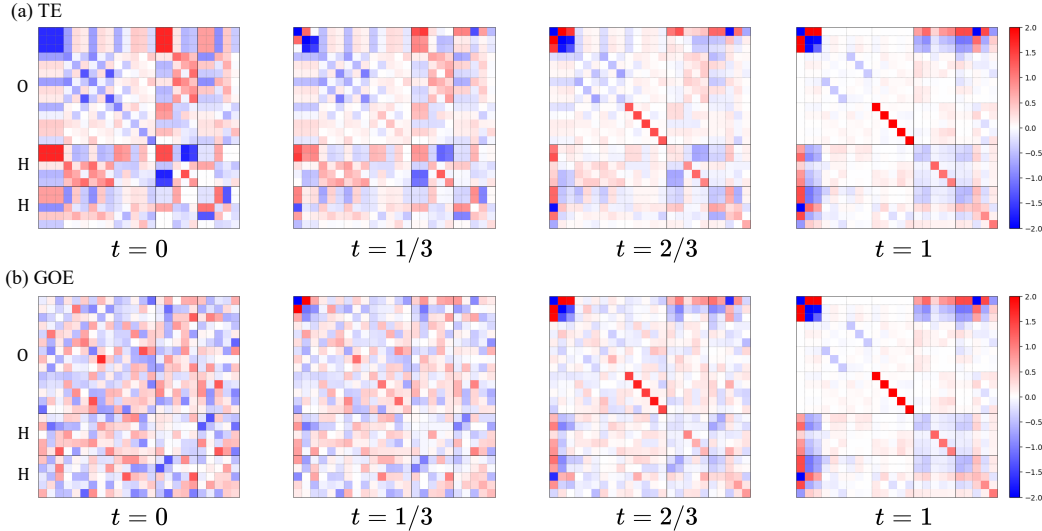


Figure 5: Visualization of Hamiltonian interpolation trajectories from different prior distributions of H<sub>2</sub>O: (a) tensor expansion-based (TE) prior and (b) Gaussian orthogonal ensemble (GOE) prior. Color intensity indicates the magnitude of Hamiltonian elements across flow time  $t$ .

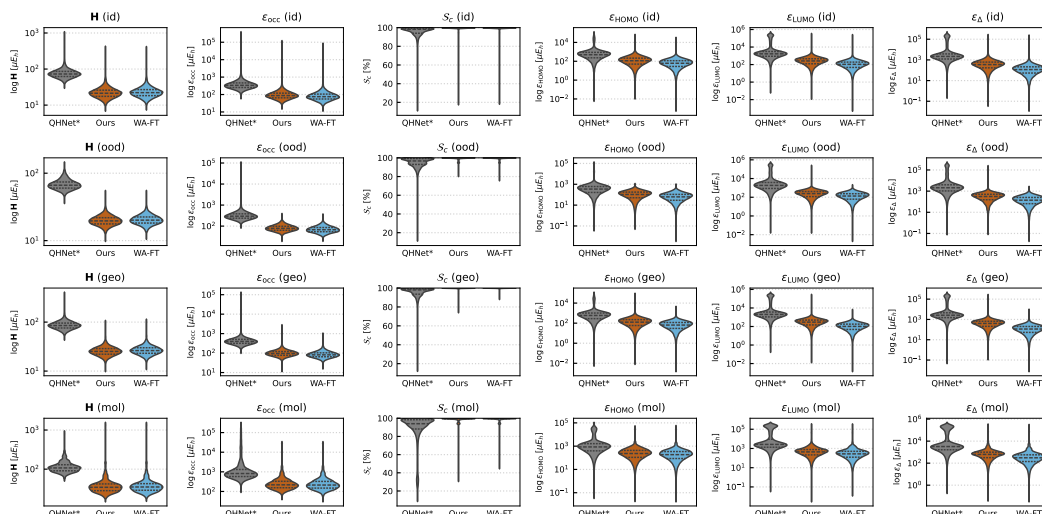


Figure 6: Violin plots compare QHNet\* (gray), QHFLOW(orange), and its weighted-alignment fine-tune (WA-FT, cyan) on six metrics: Hamiltonian MAE ( $\mathbf{H}$ ), occupied-orbital MAE ( $\epsilon_{\text{occ}}$ ),  $\epsilon_{\text{HOMO}}$ ,  $\epsilon_{\text{LUMO}}$ ,  $\epsilon_{\Delta}(\Delta)$ , and coefficient similarity ( $\mathcal{S}_c$ ). Results are shown for the four evaluation splits — id, ood, geo, mol. All axes are logarithmic except  $\mathcal{S}_c$ . Thick dashed lines mark the median; thin dashed lines indicate the quartiles. Across every split the QHFLOW violins are both lower and narrower than those of QHNet\*, and WA-FT delivers an additional (though smaller) improvement on the energy-related metrics.

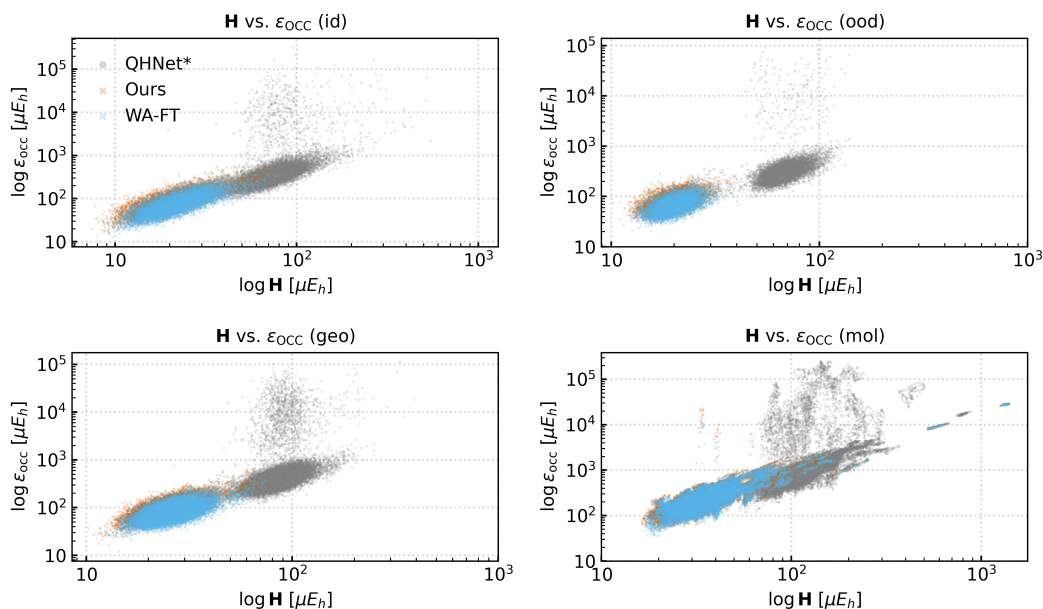


Figure 7: Scatter plots of  $\log \mathbf{H}$  versus  $\log \epsilon_{\text{occ}}$  for each QH9 split. Points for QHNet\* (gray) extend to the upper right, revealing frequent large joint errors. QHFLOW (orange) and WA-FT (cyan) cluster tightly in the lower-left region, indicating that the flow-matching model achieves simultaneously low Hamiltonian and occupied-orbital errors, even on the challenging ood and mol splits.

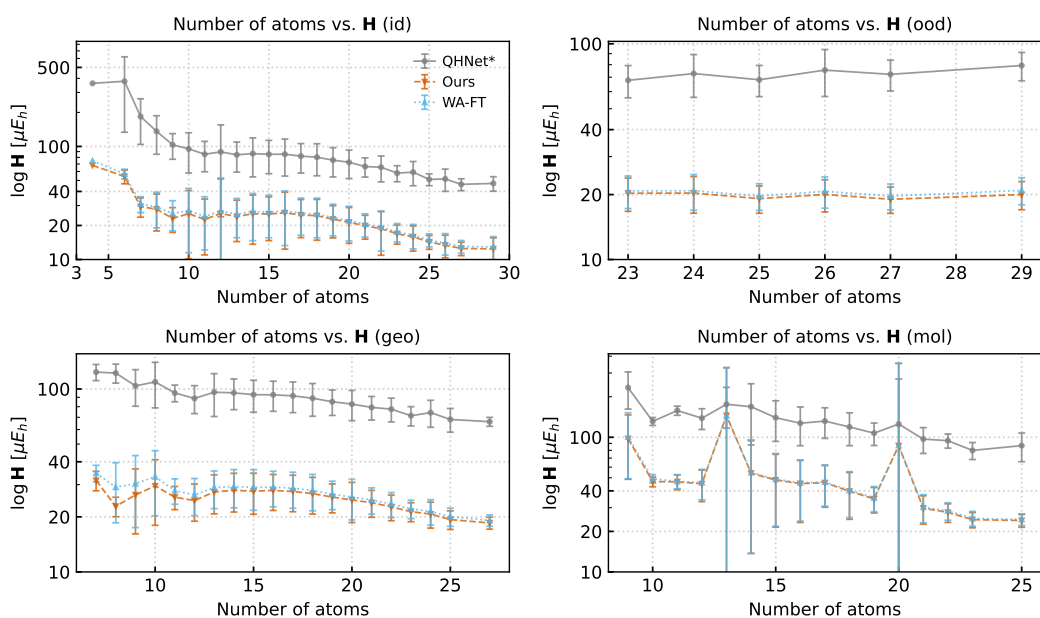


Figure 8: Mean log Hamiltonian error (markers) and one-standard-deviation bands (vertical bars) versus atom count for the four QH9 splits. QHNet\* (gray) errors rise with molecular size in ood, whereas QHFlow (orange) and WA-FT (cyan) remain nearly flat up to 29 atoms. This suggests that modelling the full Hamiltonian distribution with symmetry-aware priors confers robust transferability to larger, more complex molecules.