PAIRFORMER-H: AN E(3)-EQUIVARIANT PAIRFORMER FOR ELECTRONIC HAMILTONIAN PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning electronic Hamiltonians from atomic structures offers a route to electronic-property prediction without costly self-consistent density functional theory (DFT) calculations. However, accurately capturing how the surrounding atomic environment shapes orbital couplings between atom pairs remains a central challenge. We introduce Pairformer-H, a sparse E(3)-equivariant architecture that incorporates the Hamiltonian's orbital-block structure into pairwise information exchange through learned orbital-operator compositions. Along each path connecting a target pair through an intermediate atom, the model composes two pair operators with an environment-dependent operator acting on the intermediate orbital space. This composition contracts over intermediate orbital indices, producing a contribution with the orbital dimensions of the target pair. Adaptive aggregation combines contributions across periodicity-consistent paths to update the target-pair representation. Across seven datasets without spin–orbit coupling and two with spin–orbit coupling, Pairformer-H achieves Hamiltonian matrix-element mean absolute errors of 0.166–0.356 meV and the lowest reported errors among the evaluated baselines on five without spin–orbit coupling datasets and both with spin–orbit coupling datasets. Relative to the best baseline for each dataset, errors decrease by 51.7% on bilayer Bi_2Te_3 without spin–orbit coupling and 35.4% on bulk T_d-WTe_2. Representative band structures, densities of states, and optical responses closely agree with DFT references. These results show that aligning a network's internal interactions with the orbital-block structure of the Hamiltonian yields accurate predictions across materials and spin settings.

1 INTRODUCTION

Predicting how a material conducts electricity or responds to light requires understanding its electronic structure. Density functional theory (DFT) is widely used to determine electron energies and states from atomic species and positions (Kohn & Sham, 1965; Hohenberg & Kohn, 1964; Jones, 2015). In its Kohn–Sham formulation, an effective Hamiltonian represented in a chosen basis yields electronic states, band structures, and densities of states. Obtaining this Hamiltonian requires self-consistent-field (SCF) iterations that update the Hamiltonian and electron density until convergence. As illustrated in Fig. 1(b), repeated eigenvalue calculations make this procedure expensive for large structures and high-throughput materials screening (Kohn, 1996; Goedecker, 1999).

Hamiltonian learning replaces this iterative construction with a direct mapping from atomic structure to the converged Hamiltonian. Given atomic species, positions, and lattice vectors, neural networks predict the matrix for subsequent electronic-property calculations (Li et al., 2022; Gong et al., 2023; Zhong et al., 2023; Wang et al., 2024; Zhouyin et al., 2025; Yin et al., 2025; Qian et al., 2026). Recent advances use geometric representations and equivariant architectures to encode chemical environments while respecting physical symmetry. Equivariance is particularly important because the prediction target is expressed in an orbital basis: rotating or reflecting a structure transforms its Hamiltonian matrix according to the transformation of that basis, leaving physical predictions unchanged. Enforcing this relationship provides a principled foundation for learning Hamiltonians.

The Hamiltonian's internal organization offers a further architectural opportunity. In a localized atomic-orbital basis, each atom carries a set of localized basis functions, or orbitals. As shown in Fig. 1(a), the Hamiltonian is partitioned into blocks associated with atom pairs, with rows indexing

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

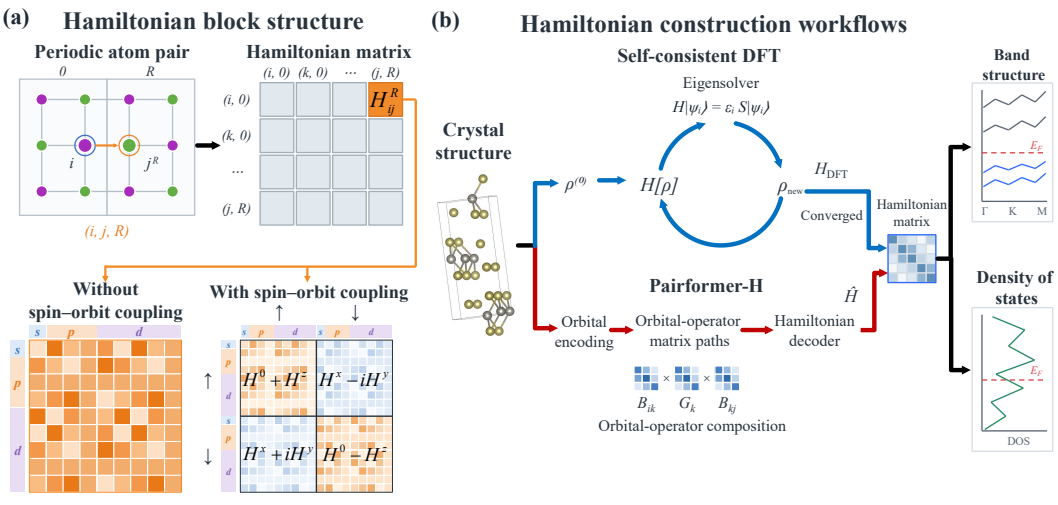


Figure 1: Structure and construction workflows of the Kohn–Sham Hamiltonian for periodic materials. (a) Hamiltonian blocks associated with periodic atomic pairs and their orbital and spin structure without and with spin–orbit coupling. (b) Hamiltonian construction using self-consistent DFT and Pairformer-H.

orbitals on one atom and columns indexing orbitals on the other. For periodic materials, a lattice translation specifies the relevant image of the second atom. Although each block is indexed by two atoms, its values depend on the surrounding chemical environment. Our central insight is that this orbital-block structure should guide the network's internal interactions. We implement this principle by composing learned operators through intermediate orbital spaces. For a target pair i, j and an intermediate atom k, learned pair operators B_{ik} and B_{kj} connect the corresponding orbital spaces, while an environment-dependent operator G_k acts within the orbital space of k. Their composition, $M_{i\to k\to j} = B_{ik} G_k B_{kj}$, contracts over intermediate orbital indices and produces a matrix with the orbital dimensions of the target pair. Each path thus supplies an orbital-resolved contribution to the pair update, bringing the algebraic structure of the prediction target into the message-passing operation itself.

We introduce Pairformer-H, a sparse $E(3)$-equivariant architecture that implements orbital-operator composition on a graph of directed periodic atom pairs. Inspired by the pair-representation interactions of Pairformer in AlphaFold 3 (Abramson et al., 2024), the model updates each target pair by adaptively aggregating operator compositions through intermediate atoms. Paths satisfy periodic closure constraints: their constituent lattice translations sum to the target-pair translation, ensuring that they connect the correct periodic images. An equivariant mapping converts the aggregated operator into a pair-feature update. Orbital-shell token encoding conditions the initial atomic and pair representations on element identities and orbital types. Successive layers combine operator-path interactions with atomic updates and endpoint-to-pair feedback, and a final decoder produces Hamiltonian blocks with Hermiticity enforced. The architecture supports settings both without and with spin–orbit coupling, the interaction between electronic spin and orbital motion. Fig. 1(b) summarizes the prediction workflow.

We evaluate Pairformer-H on seven without spin–orbit coupling datasets and two with spin–orbit coupling datasets, spanning two-dimensional monolayer and bilayer materials and a newly constructed bulk T_d-WTe_2 dataset. Pairformer-H achieves the lowest reported Hamiltonian matrix-element mean absolute error (MAE) among the evaluated baselines on five of the seven without spin–orbit coupling datasets and both with spin–orbit coupling datasets. Relative to the best reported baseline for each system, MAE decreases by 51.7% on bilayer Bi_2Te_3 without spin–orbit coupling (versus MACE-H) and by 35.4% on bulk T_d-WTe_2 (versus SLEM). An ablation on bilayer Bi_2Te_3 shows that removing the complete matrix-path module increases MAE from 0.2370 to 0.5874 meV, supporting its contribution to prediction accuracy. On representative test structures, band structures,

densities of states, and optical responses calculated using the predicted Hamiltonians closely agree with DFT references, with overlap matrices and downstream calculation settings held fixed.

Our main contributions are:

- **Orbital-structured Hamiltonian learning.** We introduce Pairformer-H, a sparse $E(3)$-equivariant architecture that incorporates the Hamiltonian's orbital-block structure into internal network interactions. Orbital-shell token encoding embeds element identities and orbital types into atomic and pair representations to construct learnable orbital operators.
- **Orbital-operator matrix paths.** We propose periodicity-consistent path updates that compose two pair operators with an environment-dependent intermediate-atom operator through orbital-index contraction. Adaptive aggregation and an equivariant mapping convert these path contributions into pair updates, explicitly capturing how multicenter environments shape orbital couplings.
- **Validation across materials and spin settings.** Across all datasets, Pairformer-H achieves the lowest reported MAE among evaluated baselines on five without spin–orbit coupling and both with spin–orbit coupling datasets, reducing errors by up to 51.7%. Component ablations and representative band-structure, density-of-states, and optical-response calculations further validate the architecture.

2 PAIRFORMER-H

Given the lattice matrix $\mathbf{L}$, atomic numbers Z_i, and Cartesian coordinates $\mathbf{r}_i$, we represent a periodic atomic structure as $\mathcal{X} = (\mathbf{L}, \{(Z_i, \mathbf{r}_i)\}_{i=1}^{N})$. For a lattice translation $\mathbf{R} = \mathbf{L}\mathbf{n}$, where $\mathbf{n} \in \mathbb{Z}^3$, we define a directed periodic atomic pair $e = (i, j, \mathbf{R})$ with relative displacement $\mathbf{d}_e = \mathbf{r}_j + \mathbf{R} - \mathbf{r}_i$. Pairformer-H constructs a periodic computational graph using a cutoff radius r_c. We denote the edge set used for feature updates by $\mathcal{E}_{\mathrm{msg}}$ and the target pair set for Hamiltonian prediction by $\mathcal{E}_H$. A pair $e = (i, i, \mathbf{0})$ corresponds to an on-site term.

As shown in Fig. 1(a), we consider settings without and with spin–orbit coupling, denoted by noSOC and SOC, respectively. In a local orbital basis, the noSOC Hamiltonian blocks are defined as

$$\left[H_e^{\mathrm{noSOC}}\right]_{\mu\nu} = \left\langle \phi_{i\mu}^{\mathbf{0}} \middle| \hat{H}_{\mathrm{KS}}^{\mathrm{noSOC}} \middle| \phi_{j\nu}^{\mathbf{R}} \right\rangle, \tag{1}$$

where μ and ν index the local orbitals of atoms i and j, respectively. Let n_i denote the number of orbitals associated with atom i. The noSOC Hamiltonian blocks considered in this work are real matrices, $H_e^{\mathrm{noSOC}} \in \mathbb{R}^{n_i \times n_j}$. The SOC setting explicitly includes spin degrees of freedom, with Hamiltonian blocks defined in an orbital–spin basis as

$$\left[H_e^{\mathrm{SOC}}\right]_{\mu s,\nu s'} = \left\langle \phi_{i\mu}^{\mathbf{0}} \otimes s \middle| \hat{H}_{\mathrm{KS}}^{\mathrm{SOC}} \middle| \phi_{j\nu}^{\mathbf{R}} \otimes s' \right\rangle, \tag{2}$$

where $s, s' \in \{\uparrow, \downarrow\}$ are spin indices. We use spin-major basis ordering, placing all spin-up orbitals before all spin-down orbitals, yielding complex matrices $H_e^{\mathrm{SOC}} \in \mathbb{C}^{2n_i \times 2n_j}$. Both noSOC and SOC Hamiltonians satisfy the Hermiticity constraint $H_{ij}^{\mathbf{R}} = (H_{ji}^{-\mathbf{R}})^{\dagger}$.

Pairformer-H predicts the corresponding matrix blocks from the atomic structure and target pair indices:

$$\mathcal{F}_\theta : (\mathcal{X}, \mathcal{E}_H) \longmapsto \left\{ \widehat{H}_e \right\}_{e \in \mathcal{E}_H}. \tag{3}$$

2.1 ORBITAL-LEVEL TOKEN ENCODING

We denote atomic and pair features by $\mathbf{s}_i$ and $\mathbf{z}_e$, respectively. To incorporate element identities and orbital types explicitly into the initial representations, we use atomic orbital shells as tokens, as shown in Fig. 2(a). Let $\mathcal{S}_i$ denote the set of valid shells of atom i, as determined by an element-dependent mask. Shell a, with angular momentum l_a, contains $2l_a + 1$ orbitals and is initialized as

$$\mathbf{t}_{ia}^{(0)} = \mathrm{Emb}_Z(Z_i) + \mathrm{Emb}_{\mathrm{shell}}(a) + \mathrm{Emb}_l(l_a). \tag{4}$$

The shell-index embedding distinguishes different shells with the same angular momentum. Intra-atomic multi-head self-attention and residual feed-forward networks first update the valid shell tokens to obtain $\mathbf{t}_{ia}$, which are then aggregated by attention pooling to obtain an atomic encoding $\mathbf{u}_i$.

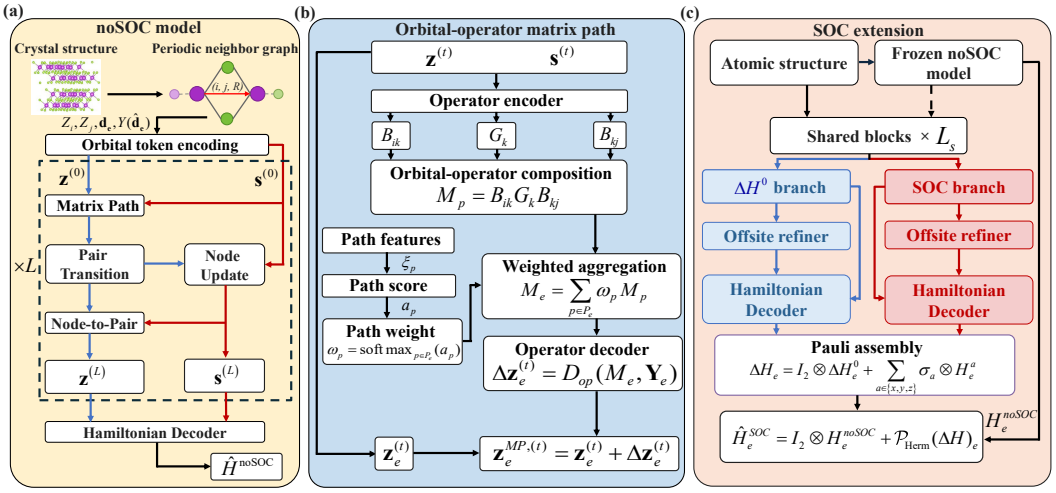


Figure 2: Architecture of Pairformer-H. (a) Orbital encoding, feature updates, and Hamiltonian decoding in the noSOC model. (b) Orbital-operator matrix paths that update pair representations through operator composition, path-weighted aggregation, and residual connections. (c) The SOC extension initialized from a noSOC checkpoint and fine-tuned.

For a directed periodic atomic pair e, the encoder combines the shell representations of the two endpoint atoms with scalar context to construct shell-pair tokens:

$$\mathbf{p}_{e,ab} = f_{\text{pair}}(\mathbf{W}_{\text{src}}\mathbf{t}_{ia} + \mathbf{W}_{\text{dst}}\mathbf{t}_{jb} + \text{MLP}_g(\mathbf{g}_e)), \qquad a \in \mathcal{S}_i, \quad b \in \mathcal{S}_j. \tag{5}$$

Here, f_{pair} comprises layer normalization and a residual feed-forward update, while $\mathbf{g}_e$ encodes radial features, the cutoff envelope, edge type, and relative fractional coordinates. Multi-head attention pooling then aggregates the valid shell-pair tokens into an invariant pair encoding $\mathbf{u}_e$.

The atomic and pair encodings are transformed by learnable mappings and added to their respective base representations as residual updates:

$$\begin{aligned} \mathbf{s}_i^{(0)} &= \mathbf{s}_i^{\text{base}} + \gamma_s\, \iota_{0e}(\mathbf{W}_s\mathbf{u}_i + \mathbf{b}_s), \\ \mathbf{z}_e^{(0)} &= \mathbf{z}_e^{\text{base}} + \gamma_z\, \text{TP}_\theta(\mathbf{u}_e, \mathbf{Y}_e). \end{aligned} \tag{6}$$

Here, $\mathbf{s}_i^{\text{base}}$ and $\mathbf{z}_e^{\text{base}}$ are the base atomic and pair representations, respectively. The map ι_{0e} injects the atomic encoding into even-parity scalar channels. The equivariant tensor product TP_θ combines the pair encoding with the spherical-harmonic features $\mathbf{Y}_e$ to generate equivariant features. The coefficients γ_s and γ_z are learnable. For on-site terms with zero displacement, only the $l = 0$ spherical-harmonic component is retained. Details of the base geometric representations and orbital-shell pooling are provided in Appendix B.1.

2.2 Sparse Equivariant Pairformer Backbone

Pairformer-H maintains pair representations only for directed periodic pairs in $\mathcal{E}_{\text{msg}}$ within the cutoff radius and constructs orbital-operator matrix paths from adjacent edges satisfying periodic closure.

Each layer first updates pair representations through orbital-operator matrix paths, then aggregates pair information to update atomic representations, and finally feeds the endpoint atomic information back into the pair representations:

$$\left(\mathbf{s}^{(t+1)}, \mathbf{z}^{(t+1)}\right) = \mathcal{B}_{\theta_t}\left(\mathbf{s}^{(t)}, \mathbf{z}^{(t)}; \mathcal{G}_{\text{msg}}\right), \tag{7}$$

where $\mathbf{s}^{(t)}$ and $\mathbf{z}^{(t)}$ denote the atomic and pair representations at layer t, and $\mathcal{G}_{\text{msg}}$ is the sparse periodic computational graph. Each stage uses equivariant mappings and residual connections. A complete proof of the model's $E(3)$ equivariance is provided in Appendix C.

As shown in Fig. 2(a), the noSOC model stacks L backbone layers. The final representations $\mathbf{s}^{(L)}$ and $\mathbf{z}^{(L)}$ are converted into orbital matrices by the Hamiltonian decoder, followed by Hermitian projection to obtain the predictions. The Hamiltonian decoder is defined in detail in Appendix B.3. As shown in Fig. 2(c), the SOC model initializes transferable parameters from the best noSOC model checkpoint. The model comprises L_s shared layers followed by independently parameterized spin-independent and spin-dependent correction branches. Through additional off-site refinement layers and decoders, the two branches predict the spin-independent correction ΔH_e^0 and three spin-dependent components H_e^a, respectively, where $a \in \{x, y, z\}$. The last shared backbone layer and both correction branches are fine-tuned jointly. For each target pair e, these components are assembled in the Pauli basis to form the full correction matrix in spin space:

$$\Delta H_e = I_2 \otimes \Delta H_e^0 + \sum_{a\in\{x,y,z\}} \sigma_a \otimes H_e^a, \tag{8}$$

where I_2 is the two-dimensional identity matrix in spin space, σ_a are the Pauli matrices, and $\otimes$ denotes the Kronecker product. Both ΔH_e^0 and H_e^a act in orbital space, yielding a $2n_i \times 2n_j$ matrix ΔH_e.

The assembled corrections undergo Hermitian projection and are added to the baseline generated by the frozen noSOC model to obtain the final prediction:

$$\widehat{H}_e^{\mathrm{SOC}} = I_2 \otimes H_e^{\mathrm{noSOC}} + \mathcal{P}_{\mathrm{Herm}}(\Delta H)_e. \tag{9}$$

Here, H_e^{noSOC} is the orbital matrix block predicted by the frozen model, and $I_2 \otimes H_e^{\mathrm{noSOC}}$ replicates it in the two spin-diagonal blocks. The operator $\mathcal{P}_{\mathrm{Herm}}$ enforces Hermiticity between correction blocks associated with forward and reverse periodic pairs, so that the projected corrections satisfy $\Delta H_{ij}^{\mathbf{R}} = (\Delta H_{ji}^{-\mathbf{R}})^\dagger$. Thus, the SOC model learns spin-independent corrections and spin-dependent components relative to the noSOC baseline, while the parameters of the noSOC model generating this baseline remain frozen.

2.3 Orbital-Operator Matrix Paths

To explicitly model how multicenter environments affect orbital couplings, we introduce orbital-operator matrix paths. As shown in Fig. 2(b), the matrix-path module at layer t takes the current atomic representations $\mathbf{s}^{(t)}$ and pair representations $\mathbf{z}^{(t)}$ as input. The module performs operator encoding, operator composition, path-weighted aggregation, and operator decoding in sequence, then updates the pair representations through a residual connection. The operator encoder normalizes the input features to obtain $\widetilde{\mathbf{s}}_k^{(t)}$ and $\widetilde{\mathbf{z}}_e^{(t)}$, and constructs the atom-pair operator and the propagation operator of the intermediate atom:

$$B_e = P_i \, \mathcal{D}_{\mathrm{orb}} \left[\mathrm{TP}_B \left(\widetilde{\mathbf{z}}_e^{(t)}, \mathbf{Y}_e \right) \right] P_j, \tag{10}$$

$$G_k = \mathcal{N}\Big(P_k \, \mathcal{D}_{\mathrm{orb}} \left[\mathrm{TP}_G \left(W_L \widetilde{\mathbf{s}}_k^{(t)}, W_R \widetilde{\mathbf{s}}_k^{(t)} \right) \right] P_k + \gamma_k I_k \Big). \tag{11}$$

Here, B_e is the orbital operator for pair e, and G_k is the propagation operator generated from the environmental features of intermediate atom k. For a directed periodic atom pair $e = (i, j, \mathbf{R})$, the rows and columns of B_e correspond to individual orbitals of atoms i and $j^{\mathbf{R}}$, respectively, with each matrix element encoding a learnable coupling feature between a pair of orbitals. For example, $[B_e]_{s,p_x}$ represents the coupling feature between the s orbital of atom i and the p_x orbital of atom $j^{\mathbf{R}}$, while the corresponding features for the p_y and p_z orbitals are represented by two other matrix elements. The rows and columns of G_k both correspond to individual orbitals of the intermediate atom k, and each matrix element $[G_k]_{\alpha\beta}$ represents a learnable mixing weight between orbitals α and β, determined by the local environment of atom k. The term $\mathbf{Y}_e$ denotes spherical-harmonic features, TP_B and TP_G are equivariant tensor products, and W_L and W_R are learnable equivariant mappings. The map $\mathcal{D}_{\mathrm{orb}}$ converts irreducible coefficients into orbital matrices, while P_i, P_j, and P_k mask invalid orbital channels of the corresponding atoms. The coefficient γ_k is predicted from scalar atomic-environment features, I_k is the identity on the valid orbital space, and $\mathcal{N}$ denotes magnitude normalization.

For target pair e, we consider two-step paths satisfying $(i, k, \mathbf{R}_1), (k, j, \mathbf{R}_2) \in \mathcal{E}_{\mathrm{msg}}$. The periodically closed path set is

$$\mathcal{P}_e = \{(k, \mathbf{R}_1, \mathbf{R}_2) \mid \mathbf{R}_1 + \mathbf{R}_2 = \mathbf{R}\}. \tag{12}$$

Each path $p \in \mathcal{P}_e$ connects atom i in the reference cell to atom j translated by $\mathbf{R}$, through a periodic image of atom k. The path operator is constructed by matrix multiplication:

$$M_p = B_{ik} G_k B_{kj}. \tag{13}$$

The product contracts over the intermediate orbital indices, allowing G_k to directly modulate orbital coupling between the two atom-pair operators.

The path features $\boldsymbol{\xi}_p$ combine environmental, elemental, geometric, and operator information to compute invariant path scores. Head h predicts the path score $a_p^{(h)} = f_h(\boldsymbol{\xi}_p)$ and applies softmax normalization over paths connecting the same target pair:

$$\omega_p^{(h)} = \mathrm{softmax}_{p \in \mathcal{P}_e}\left(a_p^{(h)}\right). \tag{14}$$

The noSOC backbone, shared SOC layers, and spin-independent correction branch use single-head aggregation, yielding $M_e = \sum_{p \in \mathcal{P}_e} \omega_p M_p$. The spin-dependent branch uses two-head aggregation in the joint spin–orbital space to separately weight the scalar and spin-vector components. The aggregation is detailed in Appendix B.2.5, and its equivariance is established in Appendix C.4.

The operator decoder re-encodes the aggregated operator into irreducible coefficients and combines them with spherical-harmonic features to generate a pair update. A residual connection then produces the output of the matrix-path module:

$$\begin{aligned} \Delta \mathbf{z}_e^{(t)} &= D_{\mathrm{op}}(M_e, \mathbf{Y}_e), \\ \mathbf{z}_e^{\mathrm{MP},(t)} &= \mathbf{z}_e^{(t)} + \Delta \mathbf{z}_e^{(t)}. \end{aligned} \tag{15}$$

Here, D_{op} combines orbital re-encoding $\mathcal{E}_{\mathrm{orb}}$ with the output mapping $\mathcal{F}_{\mathrm{out}}$:

$$D_{\mathrm{op}}(M_e, \mathbf{Y}_e) = \mathcal{F}_{\mathrm{out}}\left(\mathcal{E}_{\mathrm{orb}}(M_e), \mathbf{Y}_e; Z_i, Z_j, \widetilde{\mathbf{z}}_e^{(t),0e}\right). \tag{16}$$

The term $\widetilde{\mathbf{z}}_e^{(t),0e}$ denotes the even-parity scalar components of the normalized pair features, and $\mathcal{F}_{\mathrm{out}}$ combines element-pair conditioning, an equivariant tensor product, and scalar gating. The updated representation $\mathbf{z}_e^{\mathrm{MP},(t)}$ is then passed to the pair transition module. Implementation details of the orbital-operator matrix paths are provided in Appendix B.2.

3 Results

3.1 Benchmark Datasets

We evaluate Pairformer-H on seven noSOC datasets and two SOC datasets. The noSOC datasets comprise monolayer graphene, bilayer graphene, monolayer MoS_2, bilayer bismuthene, bilayer Bi_2Se_3, bilayer Bi_2Te_3, and bulk T_d-WTe_2. For SOC evaluation, we use bilayer Bi_2Te_3 and bilayer Bi_2Se_3 datasets. Except for bulk T_d-WTe_2, all datasets are obtained from the public benchmarks accompanying DeepH-E3 or constructed through additional calculations on the atomic configurations provided in these benchmarks (Gong et al., 2023).

We additionally construct a bulk T_d-WTe_2 noSOC dataset, extending the evaluation from two-dimensional monolayer and bilayer materials to three-dimensional periodic crystals. This dataset assesses the model's ability to describe intralayer and interlayer orbital couplings under three-dimensional periodic boundary conditions and to predict Hamiltonian changes induced by atomic displacements and lattice deformations. The initial crystal structure is obtained from the ICSD 653169 entry hosted by AFLOW (Curtarolo et al., 2012; Zagorac et al., 2019).

Each sample contains element identities, atomic coordinates, the lattice, and a reference Hamiltonian in a local orbital basis. Each dataset is divided into training, validation, and test sets for parameter optimization, model selection, and final evaluation, respectively. Dataset details are in Appendix D.

3.2 Hamiltonian Prediction Accuracy

We evaluate Hamiltonian prediction accuracy using the mean absolute error (MAE) over valid Hamiltonian matrix elements on the test set. We also conduct ablation studies of orbital encoding and matrix paths and evaluate model stability using three random seeds. Detailed results are

Table 1: Test-set MAE of Hamiltonian matrix elements for noSOC systems, reported in meV. Bold values indicate the lowest MAE for each dataset. An asterisk (*) indicates the mean over three runs with random seeds 11, 22, and 42.

Dataset	DeepH-E3	MACE-H	SLEM	Pairformer-H
Monolayer graphene	0.2516	0.2108	**0.1571**	0.1662
Bilayer graphene	0.3653	0.3069	**0.2233**	0.2336
Monolayer MoS_2	0.3892	0.3679	0.3180	**0.2520**
Bilayer bismuthene	0.3263	0.3092	0.3597	**0.2338**
Bilayer Bi_2Se_3	0.4483	0.3821	0.7380	**0.3099**
Bilayer Bi_2Te_3	0.5739	0.5159	0.6114	**0.2490***
Bulk T_d-WTe_2	1.0322	0.9545	0.5141	**0.3319**

Table 2: Test-set MAE of Hamiltonian matrix elements for the SOC systems, reported in meV. Bold values indicate the lowest MAE.

Dataset	DeepH-E3	MACE-H	Pairformer-H
Bilayer Bi_2Te_3	0.5036	0.2620	**0.2396**
Bilayer Bi_2Se_3	0.3973	0.3838	**0.3555**

provided in Appendix F. We select DeepH-E3, MACE-H, and SLEM as baseline models. Model configurations, optimization settings, and the evaluation protocol for Pairformer-H and the baselines are provided in Appendix E.

As shown in Table 1, Pairformer-H achieves sub-meV accuracy on all seven noSOC datasets and obtains the lowest MAE on five of them. For bilayer Bi_2Te_3, Pairformer-H achieves a mean MAE of 0.2490 meV across three random seeds, reducing the error by approximately 51.7% relative to the best baseline, MACE-H, whose MAE is 0.5159 meV. Notably, Pairformer-H performs particularly well on systems containing transition metals or heavy elements, highlighting its architectural strengths in modeling complex orbital couplings. For monolayer and bilayer graphene, SLEM remains slightly more accurate, but the absolute MAE differences are below 0.011 meV in both cases. Pairformer-H nevertheless outperforms DeepH-E3 and MACE-H on both graphene datasets.

We further evaluate the applicability of Pairformer-H to three-dimensional periodic crystals using bulk T_d-WTe_2. Pairformer-H achieves a test MAE of 0.3319 meV, which remains the lowest among all compared methods. This result shows that Pairformer-H retains its accuracy advantage on the evaluated three-dimensional periodic test set with structural perturbations.

Table 2 presents the SOC results for bilayer Bi_2Te_3 and bilayer Bi_2Se_3. SLEM does not provide a publicly reported SOC model for explicitly spin-resolved Hamiltonians, so we do not include SLEM in this comparison. Pairformer-H achieves the lowest test MAE on both SOC datasets, with an MAE of 0.2396 meV for bilayer Bi_2Te_3 and 0.3555 meV for bilayer Bi_2Se_3, demonstrating its accuracy advantage in SOC Hamiltonian prediction.

3.3 Electronic-Structure Reconstruction and Optical Response

To assess whether accurate Hamiltonian prediction translates into accurate electronic-property calculations, we compute band structures, densities of states (DOS), and optical conductivities using the Hamiltonians predicted by Pairformer-H and compare them with reference results obtained from DFT Hamiltonians.

Figure 3 shows the crystal structures, band structures, and DOS of bilayer Bi_2Te_3 in the noSOC and SOC settings. In both settings, the reconstructed bands closely match the reference dispersions, while the principal DOS peak positions and overall profiles also agree well. For the samples shown, the mean absolute errors in band eigenvalues are 1.58 and 4.45 meV for the noSOC and SOC settings, respectively, evaluated over reference states within ± 1.5 eV of the corresponding DFT Fermi level. Computational settings are detailed in Appendix H.1. Additional band structures and densities of states, together with quantitative band-energy errors, are provided in Appendix H.2.

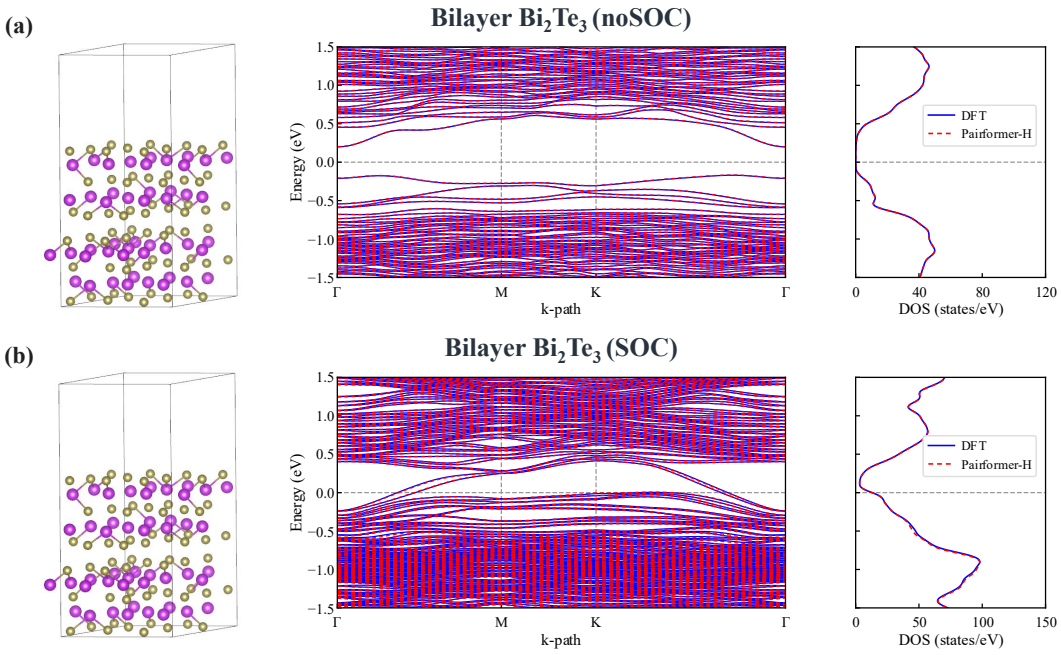


Figure 3: Band structures and densities of states reconstructed from Hamiltonians calculated by DFT and predicted by Pairformer-H: (a) bilayer Bi_2Te_3 in the noSOC setting; (b) bilayer Bi_2Te_3 in the SOC setting.

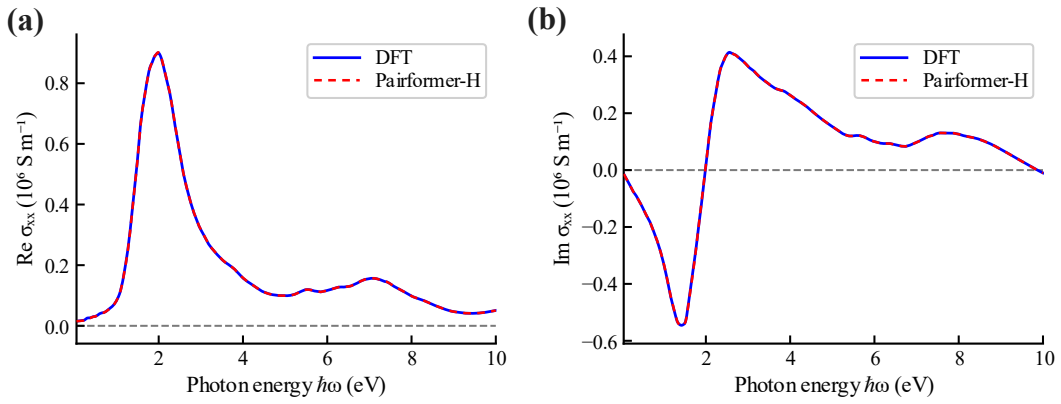


Figure 4: Interband optical conductivity of bilayer Bi_2Te_3 in the noSOC setting, calculated from Hamiltonians obtained using DFT and predicted by Pairformer-H: (a) real and (b) imaginary components of σ_{xx}.

To further evaluate the predicted Hamiltonians in downstream property calculations, we compute the optical conductivity σ_{xx} along the x direction, using the same overlap matrix and response-calculation settings for the predicted and reference Hamiltonians. As shown in Fig. 4, over the photon-energy range of 0.05–10 eV, the predicted results accurately reproduce the principal peak positions and intensities of the real component, as well as the extrema and sign changes of the imaginary component. For the structure shown, the relative L_1 errors of the real and imaginary components are 0.1524% and 0.1680%, respectively. These errors are defined as the integral of the absolute spectral deviation divided by the integral of the absolute reference spectrum. Optical-conductivity calculation settings, additional results, and quantitative error statistics are provided in Appendix H.3.

4 Related Work

Hamiltonian Learning. Direct prediction of Hamiltonian matrices in local orbital bases from atomic structures has been extensively studied for both periodic materials and molecular systems. Message-passing neural networks learn structural representations by exchanging and aggregating information on atomic graphs (Gilmer et al., 2017), providing a foundation for graph-based Hamiltonian prediction. For periodic materials, DeepH-E3 applies equivariant modeling to first-principles Hamiltonian prediction (Gong et al., 2023). Subsequent studies extend the scope of Hamiltonian learning: xDeepH targets magnetic systems (Li et al., 2023), DeepH-hybrid predicts hybrid-functional Hamiltonians (Tang et al., 2024), and the universal DeepH model explores Hamiltonian prediction across materials (Wang et al., 2024). Meanwhile, SLEM uses strictly local equivariant representations (Zhouyin et al., 2025), and MACE-H (Qian et al., 2026) extends the higher-body-order equivariant message-passing framework of MACE (Batatia et al., 2022) to electronic Hamiltonian prediction. TraceGrad enhances equivariant nonlinear expressiveness through invariants and their gradients (Yin et al., 2025), while NextHAM combines physical references with an equivariant Transformer to learn Hamiltonian corrections (Yin et al., 2026). For molecular systems, SchNOrb jointly learns electronic Hamiltonians and overlap matrices, connecting atomic structures to molecular orbitals and related electronic properties (Schütt et al., 2019). PhiSNet predicts electronic-structure matrices through an equivariant network that accounts for orbital transformations under molecular rotations (Unke et al., 2021). QHNet improves computational efficiency by reducing tensor-product operations (Yu et al., 2023b). SPHNet introduces adaptive sparse gating of atom-pair interactions and tensor-product paths to improve the efficiency and scalability of molecular Hamiltonian prediction (Luo et al., 2025). QH9 provides a benchmark for systematic evaluation across molecules and conformations (Yu et al., 2023a). Other work combines effective-Hamiltonian learning with physics-based calculations to predict molecular electronic excitation properties (Cignoni et al., 2024).

Orbital Representations and Pair Interactions. Orbital information and pair representations provide complementary perspectives for modeling interatomic interactions. OrbNet and its extension, OrbNet Denali, use orbital features obtained from low-cost quantum calculations to predict molecular energies through graph neural networks (Qiao et al., 2020; Christensen et al., 2021). Nigam et al. generalize atom-centered equivariant representations to multicenter properties, constructing symmetry-adapted descriptors for Hamiltonian prediction in local orbital bases (Nigam et al., 2022). For pair-representation learning, Pairformer in AlphaFold 3 updates pair representations through triangle interactions (Abramson et al., 2024). Uni-Mol and Uni-Mol2 combine atomic and pair representations to incorporate interatomic geometry into molecular representation learning, with Uni-Mol2 adopting a two-track Transformer to jointly update both representations (Zhou et al., 2023; Ji et al., 2024). Pairformer-H combines orbital representations with pair interactions, using matrix compositions of orbital operators along paths satisfying periodic closure constraints to model how multicenter environments influence orbital couplings.

5 Conclusion

We introduce Pairformer-H, a sparse $E(3)$-equivariant architecture for Hamiltonian prediction in periodic materials in both noSOC and SOC settings. The model incorporates the orbital-block structure of the Hamiltonian into its architecture. Through orbital-level token encoding and orbital-operator matrix paths subject to periodic closure constraints, it organizes information exchange between atom pairs to explicitly model multicenter orbital couplings. Across seven noSOC datasets and two SOC datasets, Pairformer-H achieves mean absolute errors below $1\,\mathrm{meV}$ for Hamiltonian matrix elements. Band structures and densities of states reconstructed from the predicted Hamiltonians for representative systems, together with the interband optical conductivity calculated for bilayer Bi_2Te_3, closely agree with DFT reference results. These results demonstrate that Pairformer-H can provide accurate Hamiltonians for electronic-structure and downstream property calculations in periodic materials. Future work will explore joint training across materials and extend this framework to molecular Hamiltonian prediction across diverse molecular structures and chemical environments.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation and method implementation. Their role in method implementation was limited to assisting with the writing of the code. We did not use generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, clean and reformat datasets, support qualitative and thematic data analysis, or interpret results. Additionally, we used generative AI tools to edit the manuscript to improve its readability. We have reviewed all AI-assisted work. The authors verified and tested the LLM-generated code for correctness and manually checked the AI-assisted translations and AI-edited portions of the manuscript. We take responsibility for the final content of this work, including text, claims, and other artifacts produced with the assistance of generative AI.

REPRODUCIBILITY STATEMENT

Upon acceptance, we will publicly release all source code and newly constructed datasets used in this work. The code will include scripts for data processing, model implementation, training, evaluation, and downstream property calculations to support reproducibility and further research. Details of the model implementation and training objectives are provided in Appendix B, and the equivariance proofs are presented in Appendix C. Dataset sources, construction procedures, and reference-calculation settings are described in Appendix D. Dataset splits, model and baseline configurations, optimization settings, and evaluation protocols are provided in Appendix E. Ablation studies and results across random seeds are reported in Appendix F, while computational settings for electronic structures and optical responses are detailed in Appendix H.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024.

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *NeurIPS*, 2022.

L. Calderín, V. V. Karasiev, and S. B. Trickey. Kubo–Greenwood electrical conductivity formulation and implementation for projector augmented wave datasets. *Computer Physics Communications*, 221:118–142, 2017.

Anders S. Christensen, Sai Krishna Sirumalla, Zhuoran Qiao, Michael B. O'Connor, Daniel G. A. Smith, Feizhi Ding, Peter J. Bygrave, Animashree Anandkumar, Matthew Welborn, Frederick R. Manby, and Thomas F. Miller, III. OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy. *The Journal of Chemical Physics*, 155(20):204103, 2021.

Edoardo Cignoni, Divya Suman, Jigyasa Nigam, Lorenzo Cupellini, Benedetta Mennucci, and Michele Ceriotti. Electronic excited states from physically constrained machine learning. *ACS Central Science*, 10(3):637–648, 2024.

Stefano Curtarolo, Wahyu Setyawan, Gus L. W. Hart, Michal Jahnatek, Roman V. Chepulskii, Richard H. Taylor, Shidong Wang, Junkai Xue, Kesong Yang, Ohad Levy, Michael J. Mehl, Harold T. Stokes, Denis O. Demchenko, and Dane Morgan. AFLOW: An automatic framework for high-throughput materials discovery. *Computational Materials Science*, 58:218–226, 2012.

Juan José Esteve-Paredes and Juan José Palacios. A comprehensive study of the velocity, momentum and position matrix elements for Bloch states: Application to a local orbital basis. *SciPost Physics Core*, 6(1):002, 2023.

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *ICML*, 2017.

Stefan Goedecker. Linear scaling electronic structure methods. *Reviews of Modern Physics*, 71(4): 1085–1123, 1999.

Xiaoxun Gong, He Li, Nianlong Zou, Runzhang Xu, Wenhui Duan, and Yong Xu. General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian. *Nature Communications*, 14(1):2848, 2023.

P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871, 1964.

Xiaohong Ji, Zhen Wang, Zhifeng Gao, Hang Zheng, Linfeng Zhang, Guolin Ke, and Weinan E. Exploring molecular pretraining model at scale. In *NeurIPS*, 2024.

R. O. Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of Modern Physics*, 87(3):897–923, 2015.

W. Kohn. Density functional and density matrix method scaling linearly with the number of atoms. *Physical Review Letters*, 76(17):3168–3171, 1996.

W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, 1965.

Risi Kondor, Zhen Lin, and Shubhendu Trivedi. Clebsch–Gordan Nets: a fully Fourier space spherical convolutional neural network. In *NeurIPS*, 2018.

Ryogo Kubo. Statistical-mechanical theory of irreversible processes. I. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12 (6):570–586, 1957.

Chi-Cheng Lee, Yung-Ting Lee, Masahiro Fukuda, and Taisuke Ozaki. Tight-binding calculations of optical matrix elements for conductivity using nonorthogonal atomic orbitals: Anomalous Hall conductivity in bcc Fe. *Physical Review B*, 98(11):115115, 2018.

Yung-Ting Lee, Chi-Cheng Lee, Masahiro Fukuda, and Taisuke Ozaki. Unfolding optical transition weights of impurity materials for first-principles LCAO electronic structure calculations. *Physical Review B*, 102(7):075143, 2020.

He Li, Zun Wang, Nianlong Zou, Meng Ye, Runzhang Xu, Xiaoxun Gong, Wenhui Duan, and Yong Xu. Deep-learning density functional theory Hamiltonian for efficient ab initio electronic-structure calculation. *Nature Computational Science*, 2(6):367–377, 2022.

He Li, Zechen Tang, Xiaoxun Gong, Nianlong Zou, Wenhui Duan, and Yong Xu. Deep-learning electronic-structure calculation of magnetic superstructures. *Nature Computational Science*, 3(4): 321–327, 2023.

Erpai Luo, Xinran Wei, Lin Huang, Yunyang Li, Han Yang, Zaishuo Xia, Zun Wang, Chang Liu, Bin Shao, and Jia Zhang. Efficient and scalable density functional theory Hamiltonian prediction through adaptive sparsity. In *ICML*, 2025.

Jigyasa Nigam, Michael J. Willatt, and Michele Ceriotti. Equivariant representations for molecular Hamiltonians and N-center atomic-scale properties. *The Journal of Chemical Physics*, 156(1): 014115, 2022.

T. Ozaki. Variationally optimized atomic orbitals for large-scale electronic structures. *Physical Review B*, 67(15):155108, 2003.

John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 1996.

Chen Qian, Valdas Vitartas, James R. Kermode, and Reinhard J. Maurer. Equivariant electronic Hamiltonian prediction with many-body message passing. *npj Computational Materials*, 12(1): 169, 2026.

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

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

Zechen Tang, He Li, Peize Lin, Xiaoxun Gong, Gan Jin, Lixin He, Hong Jiang, Xinguo Ren, Wenhui Duan, and Yong Xu. A deep equivariant neural network approach for efficient hybrid density functional calculations. *Nature Communications*, 15(1):8815, 2024.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Oliver Thorsten Unke, Mihail Bogojeski, Michael Gastegger, Mario Geiger, Tess Smidt, and Klaus-Robert Müller. SE(3)-equivariant prediction of molecular wavefunctions and electronic densities. In *NeurIPS*, 2021.

Yuxiang Wang, Yang Li, Zechen Tang, He Li, Zilong Yuan, Honggeng Tao, Nianlong Zou, Ting Bao, Xinghao Liang, Zezhou Chen, Shanghua Xu, Ce Bian, Zhiming Xu, Chong Wang, Chen Si, Wenhui Duan, and Yong Xu. Universal materials model of deep-learning density functional theory Hamiltonian. *Science Bulletin*, 69(16):2514–2521, 2024.

Shi Yin, Xinyang Pan, Fengyan Wang, and Lixin He. TraceGrad: a framework learning expressive SO(3)-equivariant non-linear representations for electronic-structure Hamiltonian prediction. In *ICML*, 2025.

Shi Yin, Zujian Dai, Xinyang Pan, and Lixin He. Advancing universal deep learning for electronic-structure Hamiltonian prediction of materials. In *ICLR*, 2026.

Haiyang Yu, Meng Liu, Youzhi Luo, Alex Strasser, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. QH9: A quantum Hamiltonian prediction benchmark for QM9 molecules. In *NeurIPS*, 2023a.

Haiyang Yu, Zhao Xu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Efficient and equivariant graph networks for predicting quantum Hamiltonian. In *ICML*, 2023b.

D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, and S. Rehme. Recent developments in the inorganic crystal structure database: theoretical crystal structure data and related features. *Journal of Applied Crystallography*, 52(5):918–925, 2019.

Yang Zhong, Hongyu Yu, Mao Su, Xingao Gong, and Hongjun Xiang. Transferable equivariant graph neural networks for the Hamiltonians of molecules and solids. *npj Computational Materials*, 9(1):182, 2023.

Gengmo Zhou, Zhifeng Gao, Qiankun Ding, Hang Zheng, Hongteng Xu, Zhewei Wei, Linfeng Zhang, and Guolin Ke. Uni-Mol: A universal 3D molecular representation learning framework. In *ICLR*, 2023.

Zhanghao Zhouyin, Zixi Gan, Shishir Kumar Pandey, Linfeng Zhang, and Qiangqiang Gu. Learning local equivariant representations for quantum operators. In *ICLR*, 2025.

A PRELIMINARIES

A.1 SINGLE AND PAIR REPRESENTATIONS

AlphaFold 3 organizes different types of biomolecular information using tokens. Standard amino acid residues and nucleotides are each represented by a single token, whereas other components, such as ligands, are primarily tokenized at the heavy-atom level. The model maintains single and pair representations to describe individual tokens and relationships between tokens, respectively (Abramson et al., 2024).

For a system containing n tokens, these representations are denoted by

$$\mathbf{S} = (\mathbf{s}_i)_{i=1}^{n} \in \mathbb{R}^{n \times c_s}, \qquad \mathbf{Z} = (\mathbf{z}_{ij})_{i,j=1}^{n} \in \mathbb{R}^{n \times n \times c_z}, \tag{17}$$

where c_s and c_z are the corresponding feature dimensions. The single representation $\mathbf{s}_i$ is indexed by an individual token, while the pair representation $\mathbf{z}_{ij}$ is indexed by an ordered token pair. Pair representations provide dedicated units for storing and updating relational information, allowing the network to organize information exchange directly between different token pairs.

A.2 TRIANGLE INTERACTIONS

Pairformer exchanges information among pair representations involving three tokens through triangular multiplicative updates and triangle attention. For a target representation $\mathbf{z}_{ij}$, a third token k connects it to two additional pairwise relationships, introducing context beyond the target token pair itself.

Triangular multiplicative updates have outgoing and incoming forms. Normalized pair features are mapped to $\mathbf{a}_{ij}$ and $\mathbf{b}_{ij}$ through gated linear projections. The core contractions of the two updates are

$$\mathbf{u}_{ij}^{\text{out}} = \sum_{k=1}^{n} \mathbf{a}_{ik} \odot \mathbf{b}_{jk}, \qquad \mathbf{u}_{ij}^{\text{in}} = \sum_{k=1}^{n} \mathbf{a}_{ki} \odot \mathbf{b}_{kj}, \tag{18}$$

where $\odot$ denotes elementwise multiplication, and the two update types use their own projection parameters. The contraction outputs undergo normalization, output projection, and gating before being added to the target pair representation through residual connections.

Triangle attention further adaptively aggregates related pair features. In attention around the starting node, the target pair (i, j) attends to pair features (i, k) sharing the same starting node, while the representation of the third edge (j, k) provides an attention bias. Attention around the ending node instead aggregates features (k, j) and constructs the bias from (k, i). Together, these operations allow the target pair representation to incorporate information from the other pairwise relationships within each triangle.

A.3 THE PAIRFORMER NETWORK

Pairformer consists of repeated layers. Each layer first applies outgoing and incoming triangular multiplicative updates, triangle attention around the starting and ending nodes, and a pair transition. The updated pair representation then supplies a bias for attention over the single representations, followed by a single transition. Each sublayer uses a residual connection, and the transition blocks use SwiGLU (Abramson et al., 2024).

Denoting the complete pair update within a layer by $\mathcal{F}_{\text{pair}}^{(\ell)}$, the computation can be summarized as

$$\begin{aligned} \mathbf{Z}^{(\ell+1)} &= \mathcal{F}_{\text{pair}}^{(\ell)}\big(\mathbf{Z}^{(\ell)}\big), \\ \widetilde{\mathbf{S}}^{(\ell)} &= \mathbf{S}^{(\ell)} + \text{AttentionPairBias}^{(\ell)}\big(\mathbf{S}^{(\ell)}, \mathbf{Z}^{(\ell+1)}\big), \\ \mathbf{S}^{(\ell+1)} &= \widetilde{\mathbf{S}}^{(\ell)} + \text{Transition}_{s}^{(\ell)}\big(\widetilde{\mathbf{S}}^{(\ell)}\big). \end{aligned} \tag{19}$$

For each head, single attention with pair bias computes

$$\alpha_{ij}^{(h)} = \mathrm{softmax}_j\left(\frac{(\mathbf{q}_i^{(h)})^{\mathsf{T}}\mathbf{k}_j^{(h)}}{\sqrt{d_h}} + b_{ij}^{(h)}\right), \qquad b_{ij}^{(h)} = \mathrm{Linear}_h\left(\mathrm{LN}(\mathbf{z}_{ij})\right). \tag{20}$$

Here, queries, keys, and values are generated from the single representations, and d_h is the dimension of each attention head. The pair bias $b_{ij}^{(h)}$ modulates information aggregation between tokens. The attention output is gated and projected to produce the update to the single representation.

Within an AlphaFold 3 Pairformer layer, pair representations influence single representations through attention biases, while single representations do not directly feed back into pair updates. Our work draws on its use of pair representations to organize triangle interactions and redesigns the representations and interactions for periodic materials through sparse $E(3)$-equivariant features, orbital-operator composition, and information exchange between atoms and atom pairs.

B Technical Details

We denote the features of atom i and directed periodic atomic pair e by $\mathbf{s}_i$ and $\mathbf{z}_e$, respectively, and their collective representations by $\mathbf{s}$ and $\mathbf{z}$. For $e = (i, j, \mathbf{R})$, the reverse pair is denoted by $\bar{e} = (j, i, -\mathbf{R})$. The operator TP denotes an equivariant tensor product, Norm denotes normalization grouped by irreducible representation, LN denotes layer normalization applied to scalar features, and ι_{0e} embeds scalar features into even-parity scalar channels. Equivariant operations in the model are implemented using e3nn (Geiger & Smidt, 2022).

B.1 Base Geometric Representations and Orbital-Shell Pooling

The base atomic representation $\mathbf{s}_i^{\mathrm{base}}$ comprises the scaled atomic number $Z_i/100$, a constant term of 1, the atomic fraction of the corresponding element in the structure, and a learnable element embedding. These features populate the even-parity scalar channels $0e$, while all remaining irreducible channels are initialized to zero.

For an atomic pair e in the computational graph, let $d_e = \|\mathbf{d}_e\|$ denote the interatomic distance. The radial basis functions are defined as

$$\phi_q(d_e) = f_c(d_e)\exp\left[-\beta(d_e - \mu_q)^2\right], \tag{21}$$

where μ_q is the center of the qth radial basis function, β controls its width, and the cutoff envelope is

$$f_c(d) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\frac{\pi d}{r_c}\right)\right], & d < r_c, \\ 0, & d \geq r_c. \end{cases} \tag{22}$$

The radial basis outputs form the vector $\boldsymbol{\phi}_e$. The geometric feature vector $\mathbf{g}_e$ contains the radial features, cutoff envelope, edge type, and Fourier encodings of the fractional-coordinate displacement. This displacement is defined as $\boldsymbol{\delta}_e = \mathbf{L}^{-1}\mathbf{d}_e$. Its cth component is encoded using $\sin(2\pi h\delta_{e,c})$ and $\cos(2\pi h\delta_{e,c})$, where h indexes the frequencies. Directional information is represented by the spherical-harmonic features $\mathbf{Y}_e$. For zero displacement, only the scalar component with $l = 0$ is retained.

The symmetrized geometric context is defined as

$$\bar{\mathbf{g}}_e = \frac{\mathbf{g}_e + \mathbf{g}_{\bar{e}}}{2}. \tag{23}$$

The base atomic representations of the two endpoints are combined symmetrically to produce an intermediate feature:

$$\mathbf{u}_e^{\mathrm{base}} = \frac{1}{2}\left[\mathrm{TP}_{nn}\left(\mathbf{s}_i^{\mathrm{base}}, \mathbf{s}_j^{\mathrm{base}}; \bar{\mathbf{g}}_e\right) + \mathrm{TP}_{nn}\left(\mathbf{s}_j^{\mathrm{base}}, \mathbf{s}_i^{\mathrm{base}}; \bar{\mathbf{g}}_e\right)\right]. \tag{24}$$

The base atomic-pair representation is then constructed as

$$\begin{aligned}\mathbf{z}_e^{\mathrm{base}} = \; & \mathrm{TP}_{ng}\left(\mathbf{u}_e^{\mathrm{base}}, \mathbf{Y}_e; \bar{\mathbf{g}}_e\right) \\ & + \mathrm{TP}_{rg}\left(f_{\mathrm{rad}}(\boldsymbol{\phi}_e), \mathbf{Y}_e; \bar{\mathbf{g}}_e\right) + \iota_{0e}\left[f_{\mathrm{sc}}(\bar{\mathbf{g}}_e)\right].\end{aligned} \tag{25}$$

Here, TP_{nn} combines the endpoint features, while TP_{ng} and TP_{rg} couple the intermediate and radial features, respectively, with directional information. All three are equivariant tensor-product mappings (Thomas et al., 2018). The invariant features following the semicolon parameterize the tensor-product weights. The mapping f_{rad} converts radial basis outputs into scalar features, while f_{sc} maps the symmetrized geometric context to scalar features injected directly into the $0e$ channels. The symmetrization applies to the geometric context and endpoint feature combination, while directional information is retained through $\mathbf{Y}_e$.

Orbital-shell token initialization and intra-atomic multi-head self-attention are described in Section 2.1. Only valid orbital shells participate in these operations. The updated shell tokens are pooled using scalar attention weights to obtain the atomic encoding:

$$\mathbf{u}_i = \sum_{a \in \mathcal{S}_i} \alpha_{ia} \mathbf{t}_{ia}, \qquad \alpha_{ia} = \mathrm{softmax}_{a \in \mathcal{S}_i}\left(\mathbf{w}^{\mathsf{T}} \mathbf{t}_{ia}\right), \tag{26}$$

where $\mathcal{S}_i$ is the set of valid orbital shells of atom i, and $\mathbf{w}$ is a learnable scoring vector.

For an atomic pair e, the shell-pair tokens are constructed and updated as

$$\begin{aligned}\mathbf{p}_{e,ab}^{(0)} &= \mathrm{LN}\left[\mathbf{W}_{\mathrm{src}} \mathbf{t}_{ia} + \mathbf{W}_{\mathrm{dst}} \mathbf{t}_{jb} + \mathrm{MLP}_g(\mathbf{g}_e)\right], \\ \mathbf{p}_{e,ab} &= \mathbf{p}_{e,ab}^{(0)} + \mathrm{FFN}_p\left[\mathrm{LN}\left(\mathbf{p}_{e,ab}^{(0)}\right)\right], \qquad a \in \mathcal{S}_i, \quad b \in \mathcal{S}_j.\end{aligned} \tag{27}$$

All valid shell pairs are pooled through multi-head attention to obtain an invariant pair encoding:

$$\begin{aligned}\mathbf{u}_e &= \mathbf{W}_o \bigoplus_h \left[\sum_{\substack{a \in \mathcal{S}_i \\ b \in \mathcal{S}_j}} \alpha_{e,ab}^{(h)} \mathbf{V}_h \mathbf{p}_{e,ab}\right], \\ \alpha_{e,ab}^{(h)} &= \mathrm{softmax}_{a \in \mathcal{S}_i,\, b \in \mathcal{S}_j}\left[f_{\mathrm{pool}}^{(h)}(\mathbf{p}_{e,ab})\right].\end{aligned} \tag{28}$$

Here, $\bigoplus_h$ denotes concatenation over attention heads, $f_{\mathrm{pool}}^{(h)}$ is the scoring map for the hth pooling head, $\mathbf{V}_h$ is the value projection, and $\mathbf{W}_o$ is the output projection.

B.2 Orbital-Operator Representation and Matrix-Path Implementation

B.2.1 Operator Encoder

The operator encoder converts atomic and pair representations at layer t into the propagation operator G_k and the atom-pair operator B_e, respectively. The input features are first normalized by irreducible representation:

$$\widetilde{\mathbf{s}}_k^{(t)} = \mathrm{Norm}_s\left(\mathbf{s}_k^{(t)}\right), \qquad \widetilde{\mathbf{z}}_e^{(t)} = \mathrm{Norm}_z\left(\mathbf{z}_e^{(t)}\right). \tag{29}$$

For a directed periodic atom pair $e = (i, j, \mathbf{R})$, an equivariant tensor product combines the normalized pair features with the spherical-harmonic features $\mathbf{Y}_e$. The resulting irreducible coefficients are converted into an orbital matrix:

$$\mathbf{c}_e^B = \mathrm{TP}_B\left(\widetilde{\mathbf{z}}_e^{(t)}, \mathbf{Y}_e\right), \qquad \widetilde{B}_e = \mathcal{D}_{\mathrm{orb}}\left(\mathbf{c}_e^B\right). \tag{30}$$

Here, TP_B is a learnable equivariant tensor-product mapping, and $\mathcal{D}_{\mathrm{orb}}$ converts irreducible coefficients into orbital matrices through the fixed basis transformations described in Section B.2.2.

In the noSOC operator space, define the valid-orbital mask as

$$P_i = \mathrm{diag}\left(m_{i1}, \ldots, m_{id_{\mathrm{op}}}\right), \qquad m_{i\mu} = \begin{cases} 1, & \mu \text{ is a valid orbital of atom } i, \\ 0, & \text{otherwise.} \end{cases} \tag{31}$$

Here, d_{op} is the padded operator dimension. Applying the masks of the two endpoint atoms gives

$$B_e = P_i \widetilde{B}_e P_j, \qquad [B_e]_{\mu\nu} = m_{i\mu} m_{j\nu} [\widetilde{B}_e]_{\mu\nu}. \tag{32}$$

The row index μ identifies an individual orbital of atom i, and the column index ν identifies an individual orbital of atom $j^{\mathbf{R}}$. Each valid matrix element $[B_e]_{\mu\nu}$ encodes a learnable coupling feature for that specific orbital pair, generated from the current pair representation and directional features. An orbital index distinguishes both the shell and the component within that shell; it does not denote an entire shell.

For example, an s shell on atom i and a p shell on atom $j^{\mathbf{R}}$ define a 1×3 block:

$$B_e^{s,p} = \begin{pmatrix} [B_e]_{s,p_x} & [B_e]_{s,p_y} & [B_e]_{s,p_z} \end{pmatrix}. \tag{33}$$

Its three entries encode the s–p_x, s–p_y, and s–p_z coupling features separately. Likewise, a p shell and a d shell define a 3×5 block whose 15 entries correspond to 15 individual orbital pairs. Different shells with the same angular momentum also have distinct orbital indices.

For the intermediate atom k, two independently parameterized equivariant linear mappings transform its normalized environmental representation:

$$\mathbf{a}_k^L = W_L \widetilde{\mathbf{s}}_k^{(t)}, \qquad \mathbf{a}_k^R = W_R \widetilde{\mathbf{s}}_k^{(t)}. \tag{34}$$

An equivariant tensor product then generates irreducible coefficients, which are converted into an orbital matrix:

$$\mathbf{c}_k^G = \mathrm{TP}_G\left(\mathbf{a}_k^L, \mathbf{a}_k^R\right), \qquad \widetilde{A}_k = \mathcal{D}_{\mathrm{orb}}\left(\mathbf{c}_k^G\right). \tag{35}$$

Because both inputs originate from the same atomic representation, this construction incorporates second-order interactions between its environmental features.

The unnormalized propagation matrix combines the masked orbital matrix with an environment-dependent identity term:

$$A_k = P_k \widetilde{A}_k P_k + \gamma_k I_k, \qquad \gamma_k = \tanh\left[f_\gamma\left(\widetilde{\mathbf{s}}_k^{(t),0e}\right)\right]. \tag{36}$$

Here, f_γ maps the even-parity scalar features to a scalar coefficient, and I_k is the identity on the valid orbital subspace. In the padded representation,

$$[I_k]_{\alpha\beta} = m_{k\alpha}\delta_{\alpha\beta}. \tag{37}$$

Magnitude normalization, defined in Section B.2.3, gives

$$G_k = \frac{A_k}{1+\rho(A_k)}. \tag{38}$$

Its matrix elements are therefore

$$[G_k]_{\alpha\beta} = \frac{m_{k\alpha} m_{k\beta} [\widetilde{A}_k]_{\alpha\beta} + \gamma_k m_{k\alpha}\delta_{\alpha\beta}}{1+\rho(A_k)}. \tag{39}$$

Both indices of G_k identify individual orbitals of the intermediate atom k. Each matrix element $[G_k]_{\alpha\beta}$ is a learnable orbital-mixing weight determined by the local environment of that atom. In the product $B_{ik}G_kB_{kj}$, a diagonal element $[G_k]_{\alpha\alpha}$ weights contributions that use the same intermediate orbital on both sides. An off-diagonal element $[G_k]_{\alpha\beta}$ instead weights contributions involving different intermediate orbitals α and β. The identity term contributes only to the diagonal entries associated with valid orbitals.

These orbital indices describe the noSOC setting. In the SOC operator space, each index specifies an orbital and a spin component. Entries of B_e then encode coupling features between endpoint orbital–spin channels, while entries of G_k weight combinations of orbital–spin channels on the intermediate atom. The mask is extended to $P_i^{\mathrm{SOC}} = I_2 \otimes P_i^{\mathrm{orb}}$, and the same encoding and masking construction applies.

B.2.2 Irreducible Representations of Orbital Operators

For two orbital shells with angular momenta l_a and l_b, the corresponding matrix block decomposes into irreducible representations with angular momenta $\ell = |l_a - l_b|, \ldots, l_a + l_b$ and parity $\pi_{ab} = (-1)^{l_a+l_b}$. Let Q_{ab} denote the fixed Clebsch–Gordan basis transformation (Kondor et al., 2018). The matrix block B_{ab} and its irreducible coefficients satisfy

$$\mathbf{c}_{ab} = Q_{ab}\,\mathrm{vec}(B_{ab}), \qquad \mathrm{vec}(B_{ab}) = Q_{ab}^{-1}\mathbf{c}_{ab}. \tag{40}$$

Here, vec denotes matrix vectorization. Combining the transformations for all ordered shell pairs defines the orbital-matrix encoding map $\mathcal{E}_{\mathrm{orb}}$ and decoding map $\mathcal{D}_{\mathrm{orb}}$. The former converts orbital matrices into irreducible coefficients, while the latter performs the reverse transformation.

The noSOC operators are real-valued orbital matrices. For operators in the joint spin–orbital space, the scalar spin component retains the orbital angular momentum ℓ. The vector spin components undergo the additional coupling

$$\ell \otimes 1 = \bigoplus_{J=|\ell-1|}^{\ell+1} J. \tag{41}$$

Spin transforms as an axial vector, so this coupling preserves the orbital parity π_{ab}.

For each ℓ, let S_ℓ denote the fixed Clebsch–Gordan transformation between the coupled representation and the orbital–spin product basis:

$$\mathbf{v}_{ab,\ell} = S_\ell \mathbf{c}^v_{ab,\ell}, \qquad \mathbf{c}^v_{ab,\ell} = S_\ell^{-1}\mathbf{v}_{ab,\ell}. \tag{42}$$

Here, $\mathbf{c}^v_{ab,\ell}$ concatenates the coefficients of all allowed J channels. The vector $\mathbf{v}_{ab,\ell}$ contains the product-basis coefficients for the $2\ell+1$ orbital components and three spin-vector components. The internal spin-vector component ordering is (y, z, x).

After spin decoupling and orbital-matrix decoding, let C^0_{ab}, C^x_{ab}, C^y_{ab}, and C^z_{ab} denote the resulting orbital matrices. With spin-major basis ordering, the internal SOC operator is assembled as

$$B^{\mathrm{SOC}}_{ab} = \frac{1}{\sqrt{2}} \begin{pmatrix} C^0_{ab} + C^z_{ab} & C^x_{ab} - \mathrm{i}C^y_{ab} \\ C^x_{ab} + \mathrm{i}C^y_{ab} & C^0_{ab} - C^z_{ab} \end{pmatrix}. \tag{43}$$

The factor $1/\sqrt{2}$ normalizes the internal operator basis. It applies only to internal operator assembly and is not used in the Pauli assembly of the final Hamiltonian corrections in Section 2.2.

Within each ordered shell pair, scalar coefficients precede the coupled spin-vector coefficients. Complex coefficients are stored as separate real and imaginary parts. Encoding applies the corresponding inverse transformations. Thus, $\mathcal{E}_{\mathrm{orb}}$ and $\mathcal{D}_{\mathrm{orb}}$ are inverse transformations within the adopted operator representation space.

Invalid orbitals are excluded by the mask matrix P_i. In the SOC operator space, the orbital mask is extended as

$$P_i^{\mathrm{SOC}} = I_2 \otimes P_i^{\mathrm{orb}}. \tag{44}$$

Below, P_i and the identity on the valid subspace are defined in the orbital or joint spin–orbital space, as appropriate.

B.2.3 Magnitude Normalization of the Propagation Operator

For the unnormalized propagation matrix A_k constructed in Section B.2.1, the normalization map is defined as

$$G_k = \mathcal{N}(A_k) = \frac{A_k}{1 + \rho(A_k)}, \qquad \rho(A) = \sqrt{\frac{1}{d_{\mathrm{op}}^2} \sum_{u,v=1}^{d_{\mathrm{op}}} |A_{uv}|^2 + \epsilon}. \tag{45}$$

Here, d_{op} is the complete padded operator dimension, rather than the number of valid basis functions of atom k, and ϵ is a numerical stabilizer. For complex-valued SOC operators, the magnitude is computed using squared complex moduli.

B.2.4 MATRIX-PATH COMPOSITION

For a target pair $e = (i, j, \mathbf{R})$, a path $p = (k, \mathbf{R}_1, \mathbf{R}_2) \in \mathcal{P}_e$ satisfies the message-edge conditions and periodic closure constraint defined in Section 2.3, with $\mathbf{R}_1 + \mathbf{R}_2 = \mathbf{R}$. The operators on its two edges are denoted by B_{ik} and B_{kj}, with lattice translations specified by p. The path operator and its matrix elements are

$$\begin{aligned} M_p &= B_{ik} G_k B_{kj}, \\ [M_p]_{\mu\nu} &= \sum_{\alpha,\beta} [B_{ik}]_{\mu\alpha} [G_k]_{\alpha\beta} [B_{kj}]_{\beta\nu}. \end{aligned} \tag{46}$$

In the noSOC setting, μ and ν index individual orbitals of the endpoint atoms, while α and β index individual orbitals of the intermediate atom. The entry $[B_{ik}]_{\mu\alpha}$ encodes a coupling feature between source orbital μ and intermediate orbital α. Similarly, $[B_{kj}]_{\beta\nu}$ encodes a coupling feature between intermediate orbital β and target orbital ν. The entry $[G_k]_{\alpha\beta}$ provides an environment-dependent weight for their product.

For fixed endpoint orbitals μ and ν, each term in the sum is the contribution associated with an intermediate orbital pair (α, β). For example,

$$[B_{ik}]_{s,p_x} [G_k]_{p_x,p_y} [B_{kj}]_{p_y,s} \tag{47}$$

contributes to the endpoint s–s entry of the path matrix through the p_x and p_y orbitals of atom k. The off-diagonal entry $[G_k]_{p_x,p_y}$ modulates the combination of these two intermediate orbital channels. Summing over all valid intermediate orbital pairs produces the path contribution $[M_p]_{\mu\nu}$ for the target orbital pair.

In the joint spin–orbital space, the indices μ, ν, α, β additionally include spin components, and the matrix product contracts over the valid intermediate orbital–spin channels.

Computations use uniformly padded matrices with invalid channels set to zero. The resulting valid subblock is therefore equivalent to matrix multiplication within the valid basis-function spaces.

B.2.5 PATH SCORING AND WEIGHTED AGGREGATION

The path-scoring context contains the $0e$ features of the two constituent atomic pairs and the intermediate atom. It also includes element embeddings for the source, intermediate, and target atoms, together with the following geometric and operator-magnitude features:

$$\mathbf{q}_p = \begin{bmatrix} \log(1 + d_{ik}) \\ \log(1 + d_{kj}) \\ \log(1 + d_{ij}) \\ -\widehat{\mathbf{d}}_{ik}^{\mathsf{T}} \widehat{\mathbf{d}}_{kj} \\ \log[1 + \rho(B_{ik})] \\ \log[1 + \rho(B_{kj})] \\ \log[1 + \rho(G_k)] \\ \log(1 + |\mathcal{P}_e|) \end{bmatrix}. \tag{48}$$

The path features include the even-parity scalar features of the two constituent atom pairs and the intermediate atom, the element embeddings of all three atoms, and geometric and operator-magnitude information. The three distances and the angle at the intermediate atom describe the path geometry. The three operator magnitudes characterize feature strength, while the path count indicates how many paths connect the same target atom pair. These features jointly determine the aggregation weight of each path. The displacements include the lattice translations of the corresponding periodic pairs. The quantities d_{ik}, d_{kj}, and d_{ij} are their magnitudes. The fourth component is the cosine of the angle at the intermediate atom, with the unit direction defined as zero for zero displacement. When orbital-state updates are enabled, the pooled orbital representations of the two constituent atomic pairs are also included. These feature groups are separately transformed and concatenated to form the invariant scoring context $\boldsymbol{\xi}_p$ in Section 2.3.

Head h predicts the path score $a_p^{(h)} = f_h(\boldsymbol{\xi}_p)$. The scores are normalized over the complete set of paths connecting the same target directed periodic atomic pair:

$$\omega_p^{(h)} = \mathrm{softmax}_{p \in \mathcal{P}_e} \left(a_p^{(h)} \right). \tag{49}$$

The implementation uses segment softmax grouped by target pair. Chunking changes only the execution of the path computation and does not alter the softmax normalization domain. No additional division by the number of paths is applied after weighted aggregation.

The noSOC backbone, shared SOC layers, and spin-independent correction branch use single-head aggregation, yielding $M_e = \sum_{p\in\mathcal{P}_e} \omega_p M_p$. The spin-dependent branch separately aggregates scalar and spin-vector components in the joint spin–orbital space. Writing each path operator as $M_p = I_2 \otimes M_p^0 + \sum_{a\in\{x,y,z\}} \sigma_a \otimes M_p^a$ gives

$$\begin{aligned} M_e = I_2 \otimes \left(\sum_{p\in\mathcal{P}_e} \omega_p^{(0)} M_p^0 \right) \\ + \sum_{a\in\{x,y,z\}} \sigma_a \otimes \left(\sum_{p\in\mathcal{P}_e} \omega_p^{(v)} M_p^a \right). \end{aligned} \tag{50}$$

Here, M_p^0 and M_p^a are the Pauli components of the assembled path matrix. The three spin-vector components share $\omega_p^{(v)}$ to preserve their joint transformation under rotations.

B.2.6 Operator Decoding and Residual Updates

After the aggregated operator M_e is re-encoded into irreducible coefficients by $\mathcal{E}_{\mathrm{orb}}$, an element-pair conditioning network generates channel-wise scales:

$$\mathbf{g}_e^{\mathrm{elem}} = 1 + \tanh\left[f_{\mathrm{elem}}\left(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_i \odot \mathbf{e}_j\right)\right]. \tag{51}$$

Here, $\mathbf{e}_i$ and $\mathbf{e}_j$ are element embeddings, and $\odot$ denotes elementwise multiplication. All magnetic components within an irreducible channel share the same scale.

The output mapping $\mathcal{F}_{\mathrm{out}}$ in Section 2.3 combines element-pair conditioning, an equivariant tensor product, scalar gating, and residual scaling. Together with $\mathcal{E}_{\mathrm{orb}}$, it forms the operator decoder D_{op}:

$$\begin{aligned} \Delta\mathbf{z}_e^{(t)} &= D_{\mathrm{op}}(M_e, \mathbf{Y}_e) \\ &= \eta_{\mathrm{MP}}\ \mathrm{Gate}\left[\mathrm{TP}_{\mathrm{out}}\left(\mathbf{g}_e^{\mathrm{elem}} \odot \mathcal{E}_{\mathrm{orb}}(M_e), \mathbf{Y}_e\right); \widetilde{\mathbf{z}}_e^{(t),0e}\right], \\ \mathbf{z}_e^{\mathrm{MP},(t)} &= \mathbf{z}_e^{(t)} + \Delta\mathbf{z}_e^{(t)}. \end{aligned} \tag{52}$$

The operator Gate uses the even-parity scalar components of the normalized pair features for gating. The residual coefficient η_{MP} is obtained by applying $\tanh$ to a learnable parameter. This coefficient is included in $\Delta\mathbf{z}_e^{(t)}$, so the residual update follows the convention in Section 2.3. No additional normalization is applied at the end of the matrix-path module. The updated representation is passed directly to the pair transition.

B.3 Hamiltonian Decoder and Training Objectives

B.3.1 noSOC Decoder

For each target pair $e \in \mathcal{E}_H$, the query features include the radial expansion with its cutoff envelope, the cutoff envelope itself, normalized distance, logarithmic distance, and a zero-displacement indicator. Together with the endpoint element embeddings, these features generate the scalar context $\mathbf{q}_e$ and query representation $\mathbf{v}_e$.

The decoder combines the final endpoint atomic representations, target geometry, and available pair representation to generate irreducible coefficients:

$$\begin{aligned} \mathbf{c}_e^{\mathrm{raw}} = \mathrm{TP}_{\mathrm{end}}\left(\mathbf{s}_i^{(L)} \oplus \mathbf{s}_j^{(L)}, \mathbf{Y}_e\right) \\ + \mathrm{TP}_{\mathrm{query}}\left(\mathbf{v}_e, \mathbf{Y}_e\right) + \chi_e\, \mathrm{TP}_{\mathrm{pair}}\left(\mathbf{z}_e^{(L)}, \mathbf{Y}_e\right), \end{aligned} \tag{53}$$

where $\oplus$ denotes feature concatenation and $\chi_e = \mathbb{I}[e \in \mathcal{E}_{\mathrm{msg}}]$ indicates whether the target pair belongs to the message-passing graph. When $\chi_e = 0$, the third term is omitted, and the target pair is decoded using only its endpoint representations and query features.

For output irreducible channel c, the descriptor $\mathbf{o}_c$ encodes the endpoint orbital angular momenta, output angular momentum, and channel identity. The corresponding channel gate is

$$g_{e,c}^{\mathrm{orb}} = 1 + 0.1 \tanh\left[\frac{(W_q \mathbf{q}_e)^{\mathsf{T}} \mathbf{o}_c}{\sqrt{d_{\mathrm{orb}}}}\right], \tag{54}$$

where d_{orb} is the descriptor dimension. All magnetic components within an irreducible channel share the same gating coefficient. The gated coefficients and decoded orbital matrix are

$$\mathbf{c}_e = \mathbf{g}_e^{\mathrm{orb}} \odot \mathbf{c}_e^{\mathrm{raw}}, \qquad \widetilde{H}_e^{\mathrm{noSOC}} = \mathcal{D}_H^{\mathrm{noSOC}}(\mathbf{c}_e). \tag{55}$$

Here, $\mathcal{D}_H^{\mathrm{noSOC}}$ converts irreducible coefficients into Hamiltonian matrices in the orbital basis. On-site and off-site blocks use the same decoder architecture.

For the reverse pair $\bar{e}$, Hermitian projection is defined as

$$[\mathcal{P}_{\mathrm{Herm}}(A)]_e = \frac{1}{2}\left(A_e + A_{\bar{e}}^{\dagger}\right). \tag{56}$$

The final noSOC prediction is

$$\widehat{H}_e^{\mathrm{noSOC}} = \left[\mathcal{P}_{\mathrm{Herm}}\left(\widetilde{H}^{\mathrm{noSOC}}\right)\right]_e. \tag{57}$$

This projection constrains the relationship between forward and reverse periodic pairs. For an on-site block, it reduces to averaging the matrix with its Hermitian conjugate.

B.3.2 SOC Branch Decoders

We use $r \in \{0, v\}$ to denote the spin-independent and spin-dependent correction branches, respectively. Each branch has independent feature transformations before decoding, a standard decoder, and an on-site correction head. The standard decoder and correction heads together constitute the Hamiltonian decoder of the corresponding branch.

Each branch generates correction coefficients through two paths. The first transforms the branch features and produces the standard coefficients. The second passes the branch representations through one additional sparse equivariant backbone layer before an off-site joint decoding head. Both additional backbone layers enable orbital-state updates. The spin-dependent branch also includes a direct joint decoding head acting on its transformed features.

Before standard decoding, each branch applies residual transformations to its atomic and pair representations:

$$\begin{aligned} \mathbf{s}_i^{r,\mathrm{dec}} &= \mathbf{s}_i^r + \eta_s^r W_s^r \,\mathrm{Norm}(\mathbf{s}_i^r), \\ \mathbf{z}_e^{r,\mathrm{dec}} &= \mathbf{z}_e^r + \eta_z^r W_z^r \,\mathrm{Norm}(\mathbf{z}_e^r). \end{aligned} \tag{58}$$

Here, $\mathbf{s}_i^r$ and $\mathbf{z}_e^r$ are the representations produced by the corresponding branch. The query context and query representation are also updated:

$$\begin{aligned} \mathbf{q}_e^r &= \mathbf{q}_e + f_q^r\left[\mathrm{LN}(\mathbf{q}_e)\right], \\ \mathbf{v}_e^r &= \mathbf{v}_e + W_v^r \,\mathrm{LN}(\mathbf{q}_e). \end{aligned} \tag{59}$$

The standard decoder uses the same three-term construction as Section B.3.1, with $\mathbf{s}^{r,\mathrm{dec}}$, $\mathbf{z}^{r,\mathrm{dec}}$, and $\mathbf{v}_e^r$, and independently parameterized mappings for each branch. The branch-specific context $\mathbf{q}_e^r$ conditions the orbital-channel gates, on-site correction heads, and off-site joint decoding heads.

For an on-site pair $e = (i, i, \mathbf{0})$, second-order features are constructed through a tensor product of atomic features:

$$\mathbf{u}_i^{r,\mathrm{on}} = \mathrm{TP}_{\mathrm{on}}^r\left(W_L^r \widetilde{\mathbf{s}}_i^{r,\mathrm{dec}}, W_R^r \widetilde{\mathbf{s}}_i^{r,\mathrm{dec}}\right). \tag{60}$$

The corresponding on-site correction is

$$\delta \mathbf{c}_i^{r,\mathrm{on}} = \alpha_{\mathrm{on}}^r \,\mathrm{Gate}\left[\mathbf{u}_i^{r,\mathrm{on}}; \left((\widetilde{\mathbf{s}}_i^{r,\mathrm{dec}})^{0e}, \mathbf{q}_e^r\right)\right], \tag{61}$$

where α_{on}^r is a learnable scalar. This correction is added to the raw coefficients of the standard decoder before orbital-channel gating and branch-specific output-channel selection. We denote the resulting coefficients by $\mathbf{c}_e^{r,\mathrm{dec}}$.

The three off-site joint decoding heads share the same architecture but have independent parameters. They comprise the direct joint decoding head of the spin-dependent branch and the refinement heads of both branches. The direct head uses $\mathbf{s}^{v,\mathrm{dec}}$ and $\mathbf{z}^{v,\mathrm{dec}}$. The two refinement heads instead use the outputs of their additional backbone layers, without applying the standard decoder's feature transformations to those outputs. Each head uses the query context $\mathbf{q}_e^r$ of its corresponding branch.

For any joint decoding head, let $\mathbf{s}_i, \mathbf{s}_j, \mathbf{z}_e$ denote its input features. These features are first projected to reduced channel dimensions:

$$\mathbf{a}_i = W_i \,\mathrm{Norm}(\mathbf{s}_i), \qquad \mathbf{a}_j = W_j \,\mathrm{Norm}(\mathbf{s}_j), \qquad \mathbf{p}_e^{\mathrm{joint}} = W_p \,\mathrm{Norm}(\mathbf{z}_e). \tag{62}$$

The corresponding invariant context is

$$\boldsymbol{\zeta}_e = \mathbf{q}_e^r \oplus \mathbf{a}_i^{0e} \oplus \mathbf{a}_j^{0e} \oplus (\mathbf{p}_e^{\mathrm{joint}})^{0e}. \tag{63}$$

The endpoint features are first combined through a tensor product:

$$\mathbf{u}_e^{\mathrm{joint}} = \mathrm{Gate}_{2\sigma}\left[\mathrm{Norm}\left(\mathrm{TP}_{aa}\left(\mathbf{a}_i, \mathbf{a}_j\right)\right); \boldsymbol{\zeta}_e\right]. \tag{64}$$

The resulting endpoint feature is then fused with the pair feature:

$$\mathbf{j}_e = \mathrm{Gate}_{2\sigma}\left[\mathrm{Norm}\left(\frac{W_p' \mathbf{p}_e^{\mathrm{joint}} + W_u' \mathbf{u}_e^{\mathrm{joint}} + \mathrm{TP}_{pu}\left(\mathbf{p}_e^{\mathrm{joint}}, \mathbf{u}_e^{\mathrm{joint}}\right)}{\sqrt{3}}\right); \boldsymbol{\zeta}_e\right]. \tag{65}$$

Here, $\mathrm{Gate}_{2\sigma}$ uses $2\,\mathrm{sigmoid}$ channel weights generated from the invariant context. The output correction coefficients are

$$\delta \mathbf{c}_e = \mathrm{TP}_{\mathrm{dec}}\left(\mathbf{j}_e, \mathbf{Y}_e\right), \tag{66}$$

followed by selection of the irreducible channels required by the corresponding correction branch.

Off-site corrections are applied only to off-site target pairs within the message-passing graph. The corresponding mask is

$$m_e^{\mathrm{off}} = \chi_e \left(1 - \mathbb{I}[i = j,\ \mathbf{R} = \mathbf{0}]\right). \tag{67}$$

The final coefficients of the two SOC correction branches are

$$\begin{aligned} \mathbf{c}_e^0 &= \mathbf{c}_e^{0,\mathrm{dec}} + m_e^{\mathrm{off}}\, \delta \mathbf{c}_e^{0,\mathrm{refine}}, \\ \mathbf{c}_e^v &= \mathbf{c}_e^{v,\mathrm{dec}} + m_e^{\mathrm{off}} \left(\delta \mathbf{c}_e^{v,\mathrm{joint}} + \delta \mathbf{c}_e^{v,\mathrm{refine}}\right). \end{aligned} \tag{68}$$

Here, joint denotes the direct joint decoding head of the spin-dependent branch. The superscript refine denotes a joint decoding head following an additional backbone layer.

The summed coefficients are decoded into the spin-independent correction ΔH_e^0 and spin-dependent components H_e^a, where $a \in \{x, y, z\}$. Pauli assembly, Hermitian projection, and addition to the frozen noSOC baseline follow Section 2.2, yielding the final prediction $\widehat{H}_e^{\mathrm{SOC}}$.

B.3.3 Training Objectives

Training losses are evaluated over valid matrix elements after decoding and Hermitian projection. Invalid orbital channels introduced by padding are excluded. To distinguish reference data from frozen-model predictions, we denote the reference blocks by $H_e^{\mathrm{noSOC,ref}}$ and $H_e^{\mathrm{SOC,ref}}$.

The noSOC model uses mean squared error:

$$\mathcal{L}_{\mathrm{noSOC}} = \left\langle \left| \widehat{H}_e^{\mathrm{noSOC}} - H_e^{\mathrm{noSOC,ref}} \right|^2 \right\rangle, \tag{69}$$

where $\langle \cdot \rangle$ denotes the average over valid matrix elements across all target pairs in the current batch.

The SOC loss is computed from the Pauli components of the final Hamiltonian. Using the spin-major basis ordering in Section 2.2, the reference and predicted matrices are written as

$$\begin{aligned} H_e^{\mathrm{SOC,ref}} &= \sum_{a \in \{0,x,y,z\}} \sigma_a \otimes H_e^{\mathrm{SOC},a,\mathrm{ref}}, \\ \widehat{H}_e^{\mathrm{SOC}} &= \sum_{a \in \{0,x,y,z\}} \sigma_a \otimes \widehat{H}_e^{\mathrm{SOC},a}, \qquad \sigma_0 = I_2. \end{aligned} \tag{70}$$

Here, $\widehat{H}_e^{\mathrm{SOC},a}$ denotes a component of the final prediction after assembly and Hermitian projection, distinct from the unprojected branch output H_e^a. In particular, the scalar component $\widehat{H}_e^{\mathrm{SOC},0}$ includes the frozen noSOC baseline and the spin-independent correction.

The SOC loss applies fixed scale normalization and weighting across on-site/off-site categories and Pauli channels:

$$\mathcal{L}_{\mathrm{SOC}} = \frac{\sum_{b\in\{\mathrm{on},\mathrm{off}\}} \sum_{a\in\{0,x,y,z\}} \lambda_b\lambda_a\, s_{b,a}^{-2} \left\langle \left| \widehat{H}_e^{\mathrm{SOC},a} - H_e^{\mathrm{SOC},a,\mathrm{ref}} \right|^2 \right\rangle_b}{\left(\sum_b \lambda_b\right)\left(\sum_a \lambda_a\right)}. \tag{71}$$

The category $b = \mathrm{on}$ contains on-site pairs $e = (i, i, \mathbf{0})$, whereas $b = \mathrm{off}$ contains the remaining off-site pairs. The average $\langle\cdot\rangle_b$ includes only valid matrix elements in the corresponding category of the current batch. The quantities $s_{b,a}$ are category- and channel-specific scales that remain fixed during training. The coefficients λ_b and λ_a are the category and channel weights, respectively. The three spin-vector channels use equal weights, $\lambda_x = \lambda_y = \lambda_z$.

For complex-valued matrix elements, errors are measured by squared magnitudes. Defining $\delta H_e^a = \widehat{H}_e^{\mathrm{SOC},a} - H_e^{\mathrm{SOC},a,\mathrm{ref}}$, we have

$$|\delta H_e^a|^2 = \left[\mathrm{Re}(\delta H_e^a)\right]^2 + \left[\mathrm{Im}(\delta H_e^a)\right]^2. \tag{72}$$

C $E(3)$ Equivariance Proof

We adopt the representation conventions in Sections 2.1–2.3 and Appendix B, keep the unit-cell indexing convention fixed, and transform the lattice and atomic coordinates jointly. The noSOC outputs transform in orbital space, whereas the SOC outputs transform in joint spin–orbital space.

C.1 Group Action and Output Transformation Laws

An arbitrary Euclidean transformation is written as $g = (Q, \mathbf{t})$, where $Q \in O(3)$ is an orthogonal matrix and $\mathbf{t} \in \mathbb{R}^3$ is a translation. Its action on a periodic structure is

$$\mathbf{L}' = Q\mathbf{L}, \qquad \mathbf{r}_i' = Q\mathbf{r}_i + \mathbf{t}, \qquad Z_i' = Z_i. \tag{73}$$

For a lattice translation $\mathbf{R} = \mathbf{L}\mathbf{n}$, the integer cell index $\mathbf{n}$ remains fixed, giving

$$\mathbf{R}' = Q\mathbf{R} = \mathbf{L}'\mathbf{n}. \tag{74}$$

Thus, the periodic pair $e = (i, j, \mathbf{R})$ corresponds to $e' = (i, j, Q\mathbf{R})$, and the target pair set transforms as

$$g\mathcal{E}_H = \{(i, j, Q\mathbf{R}) : (i, j, \mathbf{R}) \in \mathcal{E}_H\}. \tag{75}$$

We suppress the target-set argument in the output formulas below while retaining this correspondence between e and e'.

The representation matrix on the valid orbital space of atom i is

$$D_i(Q) = \bigoplus_{a\in\mathcal{S}_i} D^{(l_a,(-1)^{l_a})}(Q), \tag{76}$$

where $D^{(l,p)}$ denotes the irreducible representation with angular momentum l and parity p. Distinct radial shells with the same angular momentum correspond to separate representation copies. The required noSOC transformation law is

$$\widehat{H}_{e'}^{\mathrm{noSOC}}(g\mathcal{X}) = D_i(Q)\, \widehat{H}_e^{\mathrm{noSOC}}(\mathcal{X})\, D_j(Q)^\dagger. \tag{77}$$

For the SOC model, define $Q_{\mathrm{ax}} = (\det Q)Q \in SO(3)$. Let $S(Q)$ be a spin-$\frac{1}{2}$ rotation matrix corresponding to Q_{ax}, satisfying

$$S(Q)\sigma_a S(Q)^\dagger = \sum_{b\in\{x,y,z\}} (Q_{\mathrm{ax}})_{ba}\sigma_b. \tag{78}$$

This accounts for the axial-vector transformation of spin. With spin-major basis ordering, the joint representation is

$$U_i(Q) = S(Q) \otimes D_i(Q). \tag{79}$$

The overall sign ambiguity of the spin rotation cancels in the two-sided operator transformation. The required SOC transformation law is therefore

$$\widehat{H}^{\mathrm{SOC}}_{e'}(g\mathcal{X}) = U_i(Q)\, \widehat{H}^{\mathrm{SOC}}_{e}(\mathcal{X})\, U_j(Q)^{\dagger}. \tag{80}$$

Below, $V_i(Q)$ denotes the transformation matrix on the relevant valid basis space. We use $V_i = D_i$ in orbital space and $V_i = U_i$ in joint spin–orbital space.

C.2 Geometric Inputs and Orbital Encoding

Under the joint transformation of the lattice and atomic coordinates, the relative displacement transforms as

$$\begin{aligned} \mathbf{d}_{e'} &= \mathbf{r}'_j + \mathbf{R}' - \mathbf{r}'_i \\ &= Q(\mathbf{r}_j + \mathbf{R} - \mathbf{r}_i) = Q\mathbf{d}_e. \end{aligned} \tag{81}$$

The global translation therefore cancels. Orthogonality of Q preserves distances and inner products:

$$\|\mathbf{d}_{e'}\| = \|\mathbf{d}_e\|, \qquad (Q\mathbf{d}_1)^{\top}(Q\mathbf{d}_2) = \mathbf{d}_1^{\top}\mathbf{d}_2. \tag{82}$$

Consequently, radial basis functions, cutoff envelopes, and zero-displacement indicators are invariant. The cutoff graph also preserves its connectivity under the correspondence $e \mapsto e'$.

The fractional-coordinate displacement $\boldsymbol{\delta}_e = \mathbf{L}^{-1}\mathbf{d}_e$ satisfies

$$\boldsymbol{\delta}_{e'} = (\mathbf{L}')^{-1}\mathbf{d}_{e'} = (Q\mathbf{L})^{-1}Q\mathbf{d}_e = \boldsymbol{\delta}_e. \tag{83}$$

Its Fourier encodings are therefore invariant. Element identities, shell types, edge types, and valid-shell masks remain unchanged. Thus, the geometric context $\mathbf{g}_e$ and its symmetrized form $\bar{\mathbf{g}}_e$, defined in Appendix B.1, are invariant.

Let $\mathbf{Y}_{e,\ell}$ denote the degree-ℓ component of the spherical-harmonic features $\mathbf{Y}_e$. These features obey

$$\mathbf{Y}_{e',\ell} = D^{(\ell,(-1)^{\ell})}(Q)\mathbf{Y}_{e,\ell}. \tag{84}$$

For zero displacement, only the $\ell = 0$ component is retained and all higher-order components are set to zero, which also satisfies this transformation law.

The base atomic features consist of the normalized atomic number $Z_i/100$, a constant term, the element fraction in the structure, and a learnable element embedding. These quantities are invariant and populate only the $0e$ channels; the remaining irreducible channels are initialized to zero. The base pair features are constructed from invariant atomic, radial, and geometric inputs through equivariant tensor products with $\mathbf{Y}_e$. Hence,

$$\begin{aligned} \left(\mathbf{s}_i^{\mathrm{base}}\right)' &= \rho_s(Q)\mathbf{s}_i^{\mathrm{base}}, \\ \left(\mathbf{z}_{e'}^{\mathrm{base}}\right)' &= \rho_z(Q)\mathbf{z}_e^{\mathrm{base}}, \end{aligned} \tag{85}$$

where ρ_s and ρ_z are the direct sums of irreducible representations carried by the atomic and pair features, respectively.

The orbital-shell tokens in Section 2.1 are initialized from element identities, shell indices, and angular-momentum labels. These labels specify shell types rather than magnetic components, so the token features are invariant scalars. Intra-atomic self-attention, residual feed-forward updates, and attention pooling therefore preserve invariance and produce an invariant atomic encoding $\mathbf{u}_i$.

Likewise, shell-pair tokens combine invariant endpoint shell features with the invariant geometric context $\mathbf{g}_e$. Their feed-forward updates and multi-head attention pooling produce an invariant pair encoding $\mathbf{u}_e$. The map ι_{0e} injects the atomic encoding into even-parity scalar channels, while the equivariant tensor product of $\mathbf{u}_e$ and $\mathbf{Y}_e$ generates features with the required pair representations.

The learnable residual coefficients are scalars, and each residual contribution transforms in the same representation as its corresponding base feature. The initial representations therefore satisfy

$$\begin{aligned} \mathbf{s}_i^{(0)\prime} &= \rho_s(Q)\mathbf{s}_i^{(0)}, \\ \mathbf{z}_{e'}^{(0)\prime} &= \rho_z(Q)\mathbf{z}_e^{(0)}. \end{aligned} \tag{86}$$

C.3 Equivariant Operations and Backbone Updates

An equivariant linear map W satisfies

$$W\rho_{\text{in}}(Q) = \rho_{\text{out}}(Q)W, \tag{87}$$

where ρ_{in} and ρ_{out} denote the input and output representations, respectively. A tensor product constructed through Clebsch–Gordan coupling satisfies

$$\text{TP}\big(\rho_1(Q)\mathbf{x}, \rho_2(Q)\mathbf{y}\big) = \rho_{\text{out}}(Q)\,\text{TP}(\mathbf{x}, \mathbf{y}). \tag{88}$$

This identity also holds when the tensor-product weights are generated from invariant features.

For features belonging to the same irreducible representation (ℓ, p), orthogonality gives

$$\sum_{c,m} \left| x_{cm}^{(\ell,p)\prime} \right|^2 = \sum_{c,m} \left| x_{cm}^{(\ell,p)} \right|^2, \tag{89}$$

where c indexes channels and m indexes magnetic components. RMS normalization grouped by irreducible representation constructs its normalization factor from these squared magnitudes, so the factor is invariant. The learnable scale of each channel is shared across all magnetic components, yielding

$$\text{Norm}\big(\rho(Q)\mathbf{x}\big) = \rho(Q)\,\text{Norm}(\mathbf{x}). \tag{90}$$

Gating coefficients are generated from $0e$ features and shared across all magnetic components of each irreducible channel. For an invariant context $\mathbf{c}$ and its scalar gating coefficient $\gamma(\mathbf{c})$,

$$\gamma(\mathbf{c}')\rho(Q)\mathbf{x} = \rho(Q)\big[\gamma(\mathbf{c})\mathbf{x}\big], \qquad \mathbf{c}' = \mathbf{c}. \tag{91}$$

Ordinary scalar nonlinearities act on $0e$ channels, whereas nonscalar channels are updated through invariant gates. Odd-parity scalars are not subjected to ordinary scalar activations that would violate their parity. These gates and feature transitions therefore preserve equivariance.

Sparse atomic attention exchanges atomic information within the neighborhoods of the computational graph. Queries, keys, and values are generated from the current features through equivariant mappings, with queries and keys in each attention head sharing the same representation structure. Their inner product satisfies

$$\left\langle \rho_h(Q)\mathbf{q}_i^{(h)}, \rho_h(Q)\mathbf{k}_j^{(h)} \right\rangle = \left\langle \mathbf{q}_i^{(h)}, \mathbf{k}_j^{(h)} \right\rangle. \tag{92}$$

Attention scores combine this inner product with a bias generated from the $0e$ components of the pair features. Since neighborhoods retain their correspondence under the group action, attention scores and their softmax weights are invariant. Aggregating equivariant value features with these weights therefore produces equivariant atomic updates.

Local message aggregation constructs neighborhood messages through equivariant mappings and tensor products, followed by summation over neighbors. Global slots aggregate atomic information across the structure and redistribute it to individual atoms. These slots are latent feature collections, and their indices do not represent spatial directions. Both the atom-to-slot assignment weights and the slot-to-atom return weights are generated from invariant scalar features. Slot aggregation and redistribution are therefore invariant weighted sums of equivariant features and preserve the corresponding transformation laws. Local–global fusion combines the two updates through equivariant mappings and invariant gates, with residual additions between channels of the same representation type, and thus preserves equivariance.

Node-to-pair feedback maps endpoint atomic features to pair updates through equivariant mappings and tensor products. Its scalar conditioning information is invariant, and its output has the same representation type as the current pair features. Both the feedback update and the residual addition therefore preserve equivariance.

In layers with orbital-state updates, shell and shell-pair tokens are updated only from existing invariant tokens and the $0e$ components of the current features. The updated tokens therefore remain invariant. Subsequent attention pooling, injection into even-parity scalar channels, and equivariant tensor products with spherical-harmonic features preserve the corresponding feature transformation laws.

C.4 Equivariance of Orbital-Operator Matrix Paths

C.4.1 Operator Construction and Normalization

The fixed orbital basis transformations $\mathcal{E}_{\mathrm{orb}}$ and $\mathcal{D}_{\mathrm{orb}}$ in Appendix B.2 convert between operator matrices and their irreducible coefficients. Their intertwining properties follow from the fixed Clebsch–Gordan basis transformations. For a shell pair (a, b), column-wise vectorization gives

$$\mathrm{vec}\big(D_a(Q)B_{ab}D_b(Q)^\dagger\big) = \big[D_b(Q)^* \otimes D_a(Q)\big]\,\mathrm{vec}(B_{ab}). \tag{93}$$

The fixed basis transformation Q_{ab} satisfies

$$Q_{ab}\big[D_b(Q)^* \otimes D_a(Q)\big] = \rho^{\mathrm{op}}_{ab}(Q)Q_{ab}. \tag{94}$$

Thus, the two-sided transformation of an orbital matrix corresponds to the transformation of its irreducible coefficients. SOC encoding and decoding additionally couple the orbital representation to the axial-vector spin representation through fixed spin-coupling matrices, which satisfy the corresponding intertwining identities.

Let ρ^{op}_e denote the representation carried by the operator coefficients for pair e. On the valid spaces, orbital decoding satisfies

$$\mathcal{D}_{\mathrm{orb}}\big(\rho^{\mathrm{op}}_e(Q)\mathbf{c}\big) = V_i(Q)\mathcal{D}_{\mathrm{orb}}(\mathbf{c})V_j(Q)^\dagger, \tag{95}$$

and orbital encoding satisfies

$$\mathcal{E}_{\mathrm{orb}}\big(V_i(Q)AV_j(Q)^\dagger\big) = \rho^{\mathrm{op}}_e(Q)\mathcal{E}_{\mathrm{orb}}(A). \tag{96}$$

For padded matrices, we extend the representation to the complete shell space and denote it by $\overline{V}_i(Q)$. Consistent with the matrix-element definitions in Appendix B.2.1, each orbital index identifies both a radial shell and a magnetic component within that shell. Spatial transformations mix the magnetic components within each shell through the corresponding representation matrix, while distinct radial shells remain separate representation copies. The valid-orbital mask P_i retains complete irreducible representation copies and therefore satisfies

$$P_i\overline{V}_i(Q) = \overline{V}_i(Q)P_i. \tag{97}$$

In joint spin–orbital space, the same orbital selection is applied to both spin components, so the mask obeys the same commutation relation. We use $\overline{V}_i$ for padded operators below and restrict to valid subblocks when needed.

For the atom-pair operator encoder defined in Appendix B.2.1, the normalized pair features and spherical-harmonic features are combined through an equivariant tensor product. The resulting operator coefficients therefore satisfy

$$\mathbf{c}^{B\prime}_{e'} = \rho^{\mathrm{op}}_e(Q)\mathbf{c}^B_e. \tag{98}$$

Applying the fixed orbital decoding map gives

$$\widetilde{B}'_{e'} = \overline{V}_i(Q)\widetilde{B}_e\overline{V}_j(Q)^\dagger. \tag{99}$$

Since the valid-orbital masks commute with the representation matrices, the masked operator satisfies

$$\begin{aligned} B'_{e'} &= P_i\widetilde{B}'_{e'}P_j \\ &= P_i\overline{V}_i(Q)\widetilde{B}_e\overline{V}_j(Q)^\dagger P_j \\ &= \overline{V}_i(Q)B_e\overline{V}_j(Q)^\dagger. \end{aligned} \tag{100}$$

For the atomic operator encoder, the equivariant linear maps W_L and W_R produce projected features $\mathbf{a}^L_k$ and $\mathbf{a}^R_k$ that satisfy

$$\mathbf{a}^{L\prime}_k = \rho_L(Q)\mathbf{a}^L_k, \qquad \mathbf{a}^{R\prime}_k = \rho_R(Q)\mathbf{a}^R_k, \tag{101}$$

where ρ_L and ρ_R are the representations carried by the two projected features. Their equivariant tensor product produces operator coefficients satisfying

$$\begin{aligned} \mathbf{c}^{G\prime}_k &= \mathrm{TP}_G\left(\rho_L(Q)\mathbf{a}^L_k, \rho_R(Q)\mathbf{a}^R_k\right) \\ &= \rho^{\mathrm{op}}_{kk}(Q)\mathbf{c}^G_k. \end{aligned} \tag{102}$$

Here, ρ_{kk}^{op} denotes the representation carried by the atomic operator coefficients. The fixed orbital decoding map therefore yields

$$\widetilde{A}'_k = \overline{V}_k(Q)\widetilde{A}_k\overline{V}_k(Q)^\dagger. \tag{103}$$

The unnormalized propagation matrix is then

$$A_k = P_k\widetilde{A}_kP_k + \gamma_k I_k. \tag{104}$$

Since γ_k is generated from $0e$ features, $\gamma'_k = \gamma_k$. The identity on the valid subspace satisfies

$$\overline{V}_k(Q)I_k\overline{V}_k(Q)^\dagger = I_k. \tag{105}$$

Consequently,

$$A'_k = \overline{V}_k(Q)A_k\overline{V}_k(Q)^\dagger. \tag{106}$$

Magnitude normalization uses the Frobenius norm, which is invariant under unitary transformations:

$$\left\|\overline{V}_k(Q)A_k\overline{V}_k(Q)^\dagger\right\|_F^2 = \|A_k\|_F^2. \tag{107}$$

The padded dimension and normalization constant remain unchanged under spatial transformations. The scalar magnitude factor therefore satisfies $\rho(A'_k) = \rho(A_k)$, yielding

$$G'_k = \frac{A'_k}{1+\rho(A'_k)} = \overline{V}_k(Q)G_k\overline{V}_k(Q)^\dagger. \tag{108}$$

C.4.2 Periodic Closure and Path Composition

Periodic closure is preserved under orthogonal transformations:

$$\mathbf{R}_1 + \mathbf{R}_2 = \mathbf{R} \quad\Longrightarrow\quad Q\mathbf{R}_1 + Q\mathbf{R}_2 = Q\mathbf{R}. \tag{109}$$

Together with the correspondence of computational-graph edges, this gives a bijection between the path sets $\mathcal{P}_e$ and $\mathcal{P}_{e'}$, with $p = (k, \mathbf{R}_1, \mathbf{R}_2)$ corresponding to $p' = (k, Q\mathbf{R}_1, Q\mathbf{R}_2)$.

For brevity, we suppress the argument Q of the representation matrices. The path operator transforms as

$$\begin{aligned} M'_{p'} &= B'_{ik}G'_kB'_{kj} \\ &= \left(\overline{V}_iB_{ik}\overline{V}_k^\dagger\right)\left(\overline{V}_kG_k\overline{V}_k^\dagger\right)\left(\overline{V}_kB_{kj}\overline{V}_j^\dagger\right) \\ &= \overline{V}_iB_{ik}G_kB_{kj}\overline{V}_j^\dagger \\ &= \overline{V}_iM_p\overline{V}_j^\dagger. \end{aligned} \tag{110}$$

The intermediate representation matrices cancel through $\overline{V}_k^\dagger\overline{V}_k = I$. The resulting operator therefore has the two-sided transformation law required for the target pair. Since the group action preserves the valid subspaces, the same relation holds for the valid subblocks.

C.4.3 Invariant Path Scoring and Aggregation

The path-scoring context comprises $0e$ atomic and atom-pair features, element embeddings, distances, directional inner products, operator magnitudes, and path counts. These quantities are invariant. Pooled orbital tokens included when orbital-state updates are enabled are also invariant. Hence,

$$\boldsymbol{\xi}'_{p'} = \boldsymbol{\xi}_p, \qquad a_{p'}^{(h)\prime} = a_p^{(h)}. \tag{111}$$

Softmax normalization over the corresponding complete path sets preserves this invariance:

$$\omega_{p'}^{(h)\prime} = \omega_p^{(h)}. \tag{112}$$

For single-head aggregation,

$$\begin{aligned} M'_{e'} &= \sum_{p'\in\mathcal{P}_{e'}} \omega'_{p'}M'_{p'} \\ &= \sum_{p\in\mathcal{P}_e} \omega_p\overline{V}_iM_p\overline{V}_j^\dagger \\ &= \overline{V}_iM_e\overline{V}_j^\dagger. \end{aligned} \tag{113}$$

For two-head SOC aggregation, we distinguish the scalar and spin-vector components. The following expressions are written on valid subblocks; the padded-space derivation is identical. The Pauli decomposition is

$$M_p = I_2 \otimes M_p^0 + \sum_{a\in\{x,y,z\}} \sigma_a \otimes M_p^a. \tag{114}$$

The joint spin–orbital transformation gives

$$M_{p'}^{0\prime} = D_i M_p^0 D_j^\dagger, \tag{115}$$

and

$$M_{p'}^{b\prime} = \sum_{a\in\{x,y,z\}} (Q_{\mathrm{ax}})_{ba} D_i M_p^a D_j^\dagger. \tag{116}$$

Because the three spin-vector components share the same invariant path weights,

$$\sum_{p'\in\mathcal{P}_{e'}} \omega_{p'}^{(v)\prime} M_{p'}^{b\prime} = \sum_{a\in\{x,y,z\}} (Q_{\mathrm{ax}})_{ba} D_i \left(\sum_{p\in\mathcal{P}_e} \omega_p^{(v)} M_p^a \right) D_j^\dagger. \tag{117}$$

Spin-vector aggregation therefore commutes with the spatial transformation. The scalar component likewise preserves its transformation law under its separate invariant weights $\omega_p^{(0)}$. Reassembling the aggregated components gives

$$M'_{e'} = U_i M_e U_j^\dagger. \tag{118}$$

C.4.4 Operator Decoding and Residual Updates

The operator decoder applies orbital re-encoding $\mathcal{E}_{\mathrm{orb}}$, followed by the output mapping $\mathcal{F}_{\mathrm{out}}$. Orbital re-encoding converts the two-sided operator transformation into the corresponding irreducible feature transformation. Element-pair conditioning and scalar gates are generated from invariant inputs, and their coefficients are shared across the magnetic components of each irreducible channel. The tensor product with $\mathbf{Y}_e$ is equivariant. The resulting pair increment therefore satisfies

$$\Delta\mathbf{z}_{e'}^{(t)\prime} = \rho_z(Q)\Delta\mathbf{z}_e^{(t)}. \tag{119}$$

The learnable residual coefficient η_{MP}, included in the increment as defined in Appendix B.2, is a scalar and does not alter this transformation law.

Since the current pair features and the increment transform under the same representation, their residual sum satisfies

$$\begin{aligned} \mathbf{z}_{e'}^{\mathrm{MP},(t)\prime} &= \mathbf{z}_{e'}^{(t)\prime} + \Delta\mathbf{z}_{e'}^{(t)\prime} \\ &= \rho_z(Q)\left(\mathbf{z}_e^{(t)} + \Delta\mathbf{z}_e^{(t)}\right) \\ &= \rho_z(Q)\mathbf{z}_e^{\mathrm{MP},(t)}. \end{aligned} \tag{120}$$

Thus, the matrix-path module preserves the equivariance of the pair representations.

C.5 Hamiltonian Decoding and Hermitian Projection

The initial representations are equivariant, and Sections C.3 and C.4 establish that the backbone operations and matrix-path module preserve this property. Induction over the layers therefore establishes the corresponding feature transformation laws for the backbone and both correction branches.

The decoding queries in Appendix B.3 are constructed from elemental and geometric invariants. Tensor products involving endpoint atomic features, pair features, and spherical-harmonic features preserve equivariance. Pair-feature availability masks and on-site/off-site masks depend on graph indices and zero-displacement conditions and are invariant. Orbital-channel gates share their coefficients across all magnetic components of each irreducible channel. The decoded noSOC orbital blocks therefore satisfy

$$\widetilde{H}_{e'}^{\mathrm{noSOC}\prime} = D_i(Q)\, \widetilde{H}_e^{\mathrm{noSOC}}\, D_j(Q)^\dagger. \tag{121}$$

The SOC branch decoders, on-site corrections, and additional off-site refinement likewise consist of equivariant mappings, tensor products, normalization, and invariant gating. The spin-independent

correction follows the orbital-matrix transformation law, whereas the three spin-dependent components undergo both two-sided orbital transformations and mixing under the axial-vector spin representation. Following the Pauli-component transformation laws in Section C.4, the assembled correction

$$\Delta H_e = I_2 \otimes \Delta H_e^0 + \sum_{a\in\{x,y,z\}} \sigma_a \otimes H_e^a \tag{122}$$

satisfies

$$\Delta H'_{e'} = U_i(Q)\, \Delta H_e\, U_j(Q)^\dagger. \tag{123}$$

This joint spin–orbital transformation law applies to the assembled matrix ΔH_e, rather than to an individual orbital-matrix component ΔH_e^0 or H_e^a.

Hermitian projection acts on a prediction set in which both forward and reverse pair blocks are defined. For the reverse pair $\bar{e} = (j, i, -\mathbf{R})$, suppose that a collection of matrix blocks A satisfies

$$\begin{aligned} A'_{e'} &= V_i(Q) A_e V_j(Q)^\dagger, \\ A'_{\bar{e}'} &= V_j(Q) A_{\bar{e}} V_i(Q)^\dagger, \end{aligned} \tag{124}$$

where $\bar{e}' = (j, i, -Q\mathbf{R})$. Then

$$\begin{aligned} [\mathcal{P}_{\mathrm{Herm}}(A')]_{e'} &= \frac{1}{2}\left(A'_{e'} + (A'_{\bar{e}'})^\dagger\right) \\ &= \frac{1}{2} V_i(Q) \left(A_e + A_{\bar{e}}^\dagger\right) V_j(Q)^\dagger \\ &= V_i(Q) [\mathcal{P}_{\mathrm{Herm}}(A)]_e V_j(Q)^\dagger. \end{aligned} \tag{125}$$

Hermitian projection therefore commutes with the group action and preserves equivariance. We use $V_i = D_i$ for the noSOC output and $V_i = U_i$ for the assembled SOC correction.

The baseline H_e^{noSOC} generated by the frozen noSOC model obeys the orbital-space equivariance law. Replicating it in the two spin-diagonal blocks gives

$$\begin{aligned} &U_i(Q) \left(I_2 \otimes H_e^{\mathrm{noSOC}}\right) U_j(Q)^\dagger \\ &\quad = I_2 \otimes \left(D_i(Q) H_e^{\mathrm{noSOC}} D_j(Q)^\dagger\right), \end{aligned} \tag{126}$$

where $S(Q) I_2 S(Q)^\dagger = I_2$. Freezing the model parameters does not alter this transformation law. The baseline and the projected correction therefore transform under the same joint spin–orbital representation, so their sum

$$\widehat{H}_e^{\mathrm{SOC}} = I_2 \otimes H_e^{\mathrm{noSOC}} + [\mathcal{P}_{\mathrm{Herm}}(\Delta H)]_e \tag{127}$$

remains equivariant.

Consequently, the final predictions satisfy

$$\widehat{H}_{e'}^{\mathrm{noSOC}}(g\mathcal{X}) = D_i(Q)\, \widehat{H}_e^{\mathrm{noSOC}}(\mathcal{X})\, D_j(Q)^\dagger \tag{128}$$

and

$$\widehat{H}_{e'}^{\mathrm{SOC}}(g\mathcal{X}) = U_i(Q)\, \widehat{H}_e^{\mathrm{SOC}}(\mathcal{X})\, U_j(Q)^\dagger. \tag{129}$$

Global translations cancel in relative displacements. Spatial inversion, $Q = -I_3$, is represented by the orbital parity factors, while $Q_{\mathrm{ax}} = I_3$, consistent with the axial-vector nature of spin. Pairformer-H therefore satisfies $E(3)$ equivariance under the adopted representation conventions.

D Datasets and Reference Calculations

D.1 Public Materials Benchmarks

We use datasets obtained from the public DeepH-E3 benchmarks or generated through additional calculations on their atomic configurations as two-dimensional materials benchmarks. These include monolayer graphene, monolayer MoS_2, bilayer graphene, bilayer bismuthene, bilayer Bi_2Se_3, and bilayer Bi_2Te_3. These systems span different elemental compositions, orbital bases, and interlayer

Table 3: Two-dimensional materials benchmarks and their local orbital bases.

Material	Configurations	Atoms	Orbital basis
Monolayer graphene	450	72	C: $2s2p1d$
Monolayer MoS_2	500	75	Mo: $3s2p2d$; S: $2s2p1d$
Bilayer graphene	300	64	C: $2s2p1d$
Bilayer bismuthene	576	36	Bi: $3s2p2d$
Bilayer Bi_2Se_3	576	90	Bi: $3s2p2d$; Se: $3s2p1d$
Bilayer Bi_2Te_3	256	90	Bi, Te: $3s2p2d$

couplings, enabling evaluation across different local environments and orbital-coupling patterns. The public datasets and their original computational procedures are described by Gong et al. (2023).

Table 3 summarizes the number of configurations, the number of atoms per configuration, and the local orbital basis for each material. The coefficients preceding s, p, and d indicate the number of radial functions for the corresponding angular momentum and exclude spin degrees of freedom.

The noSOC experiments on monolayer graphene, monolayer MoS_2, bilayer graphene, and bilayer Bi_2Te_3 use the publicly released Hamiltonians. For bilayer bismuthene and bilayer Bi_2Se_3, we retain the lattice vectors and atomic coordinates of the public configurations and perform new self-consistent OpenMX calculations with both spin polarization and SOC disabled to obtain the corresponding noSOC Hamiltonians. The SOC experiments on bilayer Bi_2Te_3 and bilayer Bi_2Se_3 use the publicly released SOC data.

D.2 Construction of the Bulk T_d-WTe_2 Dataset

To extend the evaluation to three-dimensional periodic crystals, we construct a noSOC dataset for bulk T_d-WTe_2. The initial structure is obtained from ICSD entry 653169 hosted by AFLOW (Curtarolo et al., 2012; Zagorac et al., 2019). It belongs to space group $Pmn2_1$ (No. 31) and contains 12 atoms per crystallographic unit cell. We construct a $2 \times 2 \times 1$ supercell containing 48 atoms.

The dataset contains 256 configurations, comprising one unperturbed supercell and 255 configurations generated by combining lattice strain with atomic displacements. Each Cartesian component of the atomic displacements is sampled from a zero-mean Gaussian distribution with a standard deviation selected from

$$\sigma_r \in \{0.015,\ 0.030,\ 0.045,\ 0.060\}\ \text{Å}. \tag{130}$$

The mean displacement over all atoms is subtracted from each configuration to remove rigid translation.

Lattice strain is constructed from a random matrix $\mathbf{A}$ whose entries are independently sampled from a zero-mean Gaussian distribution with a standard deviation selected from

$$\sigma_\varepsilon \in \{0.004,\ 0.008,\ 0.012,\ 0.016\}. \tag{131}$$

We first symmetrize the matrix as $\mathbf{E} = (\mathbf{A} + \mathbf{A}^\mathsf{T})/2$ and clip its diagonal entries to $[-0.025, 0.025]$, yielding $\overline{\mathbf{E}}$. The applied strain is then

$$\boldsymbol{\varepsilon} = \overline{\mathbf{E}} - \frac{\mathrm{tr}(\overline{\mathbf{E}})}{6}\mathbf{I}. \tag{132}$$

With lattice vectors stored as columns, the lattice matrix is updated according to

$$\mathbf{L}' = (\mathbf{I} + \boldsymbol{\varepsilon})\mathbf{L}. \tag{133}$$

Atomic coordinates are first transformed with the lattice, then perturbed by the sampled displacements and wrapped into the periodic cell. Configurations with a minimum periodic interatomic distance below 1.85 Å are resampled.

Each sampling attempt generates the strain matrix before the atomic displacements. The displacement standard deviation cycles through the listed values for successive configurations, while the strain standard deviation changes every four configurations. Consequently, all 16 combinations are traversed once per cycle.

Table 4: Basis sets, pseudopotentials, and k-point grids used for the additional noSOC reference calculations.

Material	Basis set	Pseudopotential	k-point grid
Bilayer bismuthene	Bi8.0-$s3p2d2$	Bi_PBE19	$3 \times 3 \times 1$
Bilayer Bi_2Se_3	Bi8.0-$s3p2d2$	Bi_PBE19	$3 \times 3 \times 1$
	Se7.0-$s3p2d1$	Se_PBE19	
Bulk T_d-WTe_2	W7.0-$s3p2d2$	W_PBE19	$3 \times 2 \times 2$
	Te7.0-$s3p2d2$	Te_PBE19	

Table 5: Shared settings for the additional noSOC DFT calculations.

Parameter	Setting
Real-space integration-grid energy cutoff	300 Ry
Electronic temperature	300 K
Self-consistency convergence threshold	10^{-7} Hartree
Maximum number of self-consistent iterations	400
Charge-mixing method	Rmm-DIISK
Partial core correction	Enabled

D.3 DFT Reference Calculations

For the publicly released Hamiltonians, we retain the reference data generated using the original DFT settings. The additional noSOC Hamiltonians for bilayer bismuthene and bilayer Bi_2Se_3, together with those for the newly constructed bulk T_d-WTe_2 dataset, are generated using OpenMX 3.9 with localized pseudoatomic orbital bases (Ozaki, 2003). All additional calculations are self-consistent single-point calculations with fixed input lattice vectors and atomic coordinates.

For these three materials, we use the PBE exchange–correlation functional (Perdew et al., 1996) and pseudopotentials from the OpenMX DFT_DATA19 database, with both spin polarization and SOC disabled. The complete basis specifications, pseudopotentials, and k-point grids are listed in Table 4.

The `s3p2d2` and `s3p2d1` bases contain 19 and 14 spatial orbitals per atom, respectively. The three materials share the principal self-consistent calculation settings summarized in Table 5.

The electronic temperature controls the smearing of electronic occupations. The self-consistency convergence threshold corresponds to the OpenMX parameter `scf.criterion`, which measures the change in eigenvalue energy between successive self-consistent iterations.

The resulting real-space Hamiltonians are organized into matrix blocks indexed by atom pairs and lattice translations and serve as reference targets for model prediction. Consistent basis definitions and orbital ordering are used across configurations of the same material to maintain a common correspondence between matrix elements and orbital channels.

E Training Configurations and Evaluation Protocol

E.1 Dataset Splits

Following the dataset-splitting procedure of DeepH-E3 (Gong et al., 2023), we partition each dataset into training, validation, and test sets at the level of complete atomic configurations, for parameter optimization, model selection, and final evaluation, respectively. All atomic pairs and Hamiltonian blocks belonging to the same configuration are assigned to the same subset. The partitions are determined using a fixed random permutation and predefined subset sizes and remain unchanged throughout the corresponding experiments.

Table 6: Dataset partitions. Counts refer to independent atomic configurations.

Dataset	Setting	Total	Train	Val.	Test	Unused
Monolayer graphene	noSOC	450	270	90	90	0
Monolayer MoS_2	noSOC	500	300	100	100	0
Bilayer graphene	noSOC	300	180	60	60	0
Bilayer bismuthene	noSOC	576	231	113	113	119
Bilayer Bi_2Se_3	noSOC	576	231	113	113	119
Bilayer Bi_2Te_3	noSOC	256	204	38	12	2
Bulk T_d-WTe_2	noSOC	256	204	38	12	2
Bilayer Bi_2Se_3	SOC	576	231	113	113	119
Bilayer Bi_2Te_3	SOC	256	204	38	12	2

Table 7: Cutoff radii and initial learning rates for the Pairformer-H noSOC models.

Dataset	Cutoff radius r_c (Å)	Initial learning rate
Monolayer graphene	7.2	8×10^{-3}
Monolayer MoS_2	7.2	8×10^{-3}
Bilayer graphene	7.2	8×10^{-3}
Bilayer bismuthene	7.2	8×10^{-3}
Bilayer Bi_2Se_3	7.2	8×10^{-3}
Bilayer Bi_2Te_3	9.0	8×10^{-3}
Bulk T_d-WTe_2	9.0	8×10^{-3}

E.2 noSOC Model and Optimization Settings

The noSOC model contains $L = 3$ backbone layers. Atomic and pair representations use the same hidden irreducible representations:

$$64 \times 0e + 32 \times 1o + 16 \times 2e + 8 \times 3o + 8 \times 4e. \tag{134}$$

The maximum angular momentum of the spherical-harmonic features is $l_{\max} = 5$, and the number of radial basis functions is 48. Orbital tokens have dimension 32 and use four attention heads. The path-scoring network has hidden dimension 128. Matrix paths use single-head aggregation, with softmax normalization over paths connecting the same target pair.

The cutoff radii and initial learning rates are listed in Table 7.

The cutoff radius determines the feature-update graph with edge set $\mathcal{E}_{\mathrm{msg}}$. The target pair set $\mathcal{E}_H$ for Hamiltonian prediction is determined by the reference matrix blocks.

Training uses the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$, without weight decay. The gradient norm is clipped at 1.0. Training runs on two GPUs with a per-device batch size of 1 and a global batch size of 2, using 32-bit floating-point precision. The maximum number of training epochs is set to 3000.

The learning rate is controlled by a ReduceLROnPlateau scheduler applied to an exponential moving average of the validation loss. The smoothing coefficient is 0.1, the decay factor is 0.5, the patience is 180, the cooldown is 40, the relative improvement threshold is 0.05, and the minimum learning rate is 3×10^{-5}. If the validation loss exceeds twice its historical minimum for 20 consecutive validation evaluations, the best model and its optimizer state are restored. The restored learning rate is then multiplied by 0.8.

E.3 SOC Initialization and Fine-Tuning

The SOC model initializes transferable parameters from the corresponding best noSOC checkpoint and contains $L_s = 2$ shared backbone layers. These are followed by independently parameterized spin-independent and spin-dependent correction branches. Each branch contains one backbone layer, one additional off-site refinement layer, and its corresponding decoder.

Table 8: Parameter-group optimization settings for SOC fine-tuning.

Parameter group	Initial learning rate
Trainable shared backbone	5×10^{-5}
Spin-independent correction branch	1×10^{-4}
Spin-dependent correction branch	2.5×10^{-4}
Off-site refinement layers	5×10^{-4}
Off-site decoders	1×10^{-3}

During fine-tuning, the last shared backbone layer, both correction branches, the off-site refinement layers, and the decoders are updated jointly. The first shared backbone layer and the input encoders remain frozen. A separate noSOC model used to generate the baseline remains frozen throughout training.

The hidden irreducible representations, spherical-harmonic features, radial basis functions, and orbital-token settings are identical to those of the noSOC model. The cutoff radius is 7.2 Å for bilayer Bi_2Se_3 and 9.0 Å for bilayer Bi_2Te_3. Shared layers and the spin-independent correction branch use single-head path aggregation. The spin-dependent branch uses two aggregation heads for the scalar and spin-vector components, respectively, with all three spin-vector components sharing the same path weights. Orbital-shell and shell-pair states are updated only in the spin-dependent branch and the off-site refinement layers of both branches. Gradients propagated from the spin-dependent branch to the shared input are scaled by 0.1.

SOC training uses Adam with the same momentum parameters, numerical stability constant, weight decay, and gradient clipping as the noSOC model. Parameter-group learning rates are listed in Table 8.

Training begins with five epochs of linear learning-rate warmup. Each parameter group subsequently adjusts its learning rate independently according to its monitored validation metric. The decay factor is 0.5, the minimum improvement threshold is 0.001 meV, and the minimum learning rate is 10^{-5}. Training begins with 100 epochs of fine-tuning, followed by 2900 epochs of regular training.

The training objective is evaluated on the final predicted Pauli components. On-site and off-site terms receive weights of 0.25 and 1.0, respectively. The scalar-component loss and the mean loss over the three spin-vector components receive equal weights, and the three spin-vector channels are weighted equally. Complex errors are measured by squared magnitudes.

Normalization scales are estimated exclusively from the training set and remain fixed during training. Separate scales are estimated for on-site and off-site terms. For the scalar component, the scale is the standard deviation of the difference between the reference component and the frozen noSOC baseline. For the spin-vector components, scales are the standard deviations of the corresponding reference components. Validation and test data are not used to estimate these scales.

E.4 Baseline Configurations

E.4.1 DeepH-E3

DeepH-E3 uses three equivariant message-passing modules. The input embedding is $64 \times 0e$, and the intermediate hidden irreducible representations are

$$64 \times 0e + 32 \times 1o + 16 \times 2e + 8 \times 3o + 8 \times 4e + 4 \times 5o. \tag{135}$$

The maximum angular momentum of the spherical-harmonic features is $l_{\max} = 5$, with parity explicitly distinguished. The cutoff radius is 7.2 Å for monolayer graphene, monolayer MoS_2, bilayer graphene, bilayer bismuthene, and bilayer Bi_2Se_3, and 9.0 Å for bilayer Bi_2Te_3 and bulk T_d-WTe_2. The SOC experiments on bilayer Bi_2Se_3 and bilayer Bi_2Te_3 use the same hidden representations and their respective cutoff radii, with outputs corresponding to Hamiltonian blocks that explicitly include spin degrees of freedom. The model uses the Adam optimizer.

Table 9: Initial learning rates of the baseline models.

Dataset	Setting	DeepH-E3	MACE-H	SLEM
Monolayer graphene	noSOC	8×10^{-3}	8×10^{-3}	8×10^{-3}
Monolayer MoS_2	noSOC	8×10^{-3}	8×10^{-3}	8×10^{-3}
Bilayer graphene	noSOC	8×10^{-3}	8×10^{-3}	8×10^{-3}
Bilayer bismuthene	noSOC	8×10^{-3}	8×10^{-3}	8×10^{-3}
Bilayer Bi_2Se_3	noSOC	8×10^{-3}	8×10^{-3}	8×10^{-3}
Bilayer Bi_2Te_3	noSOC	8×10^{-3}	8×10^{-3}	8×10^{-3}
Bulk T_{d}-WTe_2	noSOC	8×10^{-3}	8×10^{-3}	8×10^{-3}
Bilayer Bi_2Se_3	SOC	8×10^{-3}	8×10^{-3}	—
Bilayer Bi_2Te_3	SOC	8×10^{-3}	8×10^{-3}	—

E.4.2 MACE-H

MACE-H uses three interaction layers, correlation order 3, and a maximum spherical-harmonic angular momentum of $l_{\max} = 5$. The radial representation uses 128 learnable Gaussian basis functions, and the hidden layers of the radial network have width 64. The average-neighbor parameter is set to 50, and equivariant layer normalization is used.

The noSOC experiments on monolayer graphene, monolayer MoS_2, bilayer graphene, bilayer bismuthene, and bilayer Bi_2Se_3, together with the SOC experiments on bilayer Bi_2Se_3 and bilayer Bi_2Te_3, use hidden irreducible representations

$$64 \times 0e + 64 \times 1o + 64 \times 2e + 64 \times 3o + 64 \times 4e + 64 \times 5o. \tag{136}$$

The noSOC experiments on bilayer Bi_2Te_3 and bulk T_{d}-WTe_2 use

$$64 \times 0e + 64 \times 1o + 64 \times 2e. \tag{137}$$

Dataset-specific cutoff radii are identical to those of DeepH-E3. The SOC experiments predict Hamiltonian blocks that explicitly include spin degrees of freedom. The model uses the Adam optimizer.

E.4.3 SLEM

SLEM uses four network layers with hidden irreducible representations

$$64 \times 0e + 64 \times 1o + 32 \times 2e + 16 \times 3o + 16 \times 4e. \tag{138}$$

The number of radial basis functions is 18, the environment embedding multiplicity is 32, and the latent dimension is 128. Both the tensor-product radial network and the latent network contain two hidden layers of width 128. Outputs are converted into Hamiltonian matrices using the e3tb equivariant orbital decoder.

For monolayer graphene, monolayer MoS_2, bilayer graphene, bilayer bismuthene, and bilayer Bi_2Se_3, the cutoff radius is 7.2 Å and the average-neighbor parameter is 50. For bilayer Bi_2Te_3 and bulk T_{d}-WTe_2, these values are 9.0 Å and 80, respectively. The model uses the Adam optimizer, with a maximum of 3000 training epochs.

E.4.4 Baseline Optimization Settings

All three baselines use Adam momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Their initial learning rates are listed in Table 9.

The baseline models use the same learning-rate scheduler and scheduling settings as the Pairformer-H noSOC model.

E.5 Hamiltonian Error and Evaluation Settings

Hamiltonian prediction accuracy is measured by the mean absolute error over valid matrix elements, reported in meV:

$$\mathrm{MAE} = \frac{1}{N_{\mathrm{el}}} \sum_{\mathcal{X}\in\mathcal{T}} \sum_{e\in\mathcal{E}_H(\mathcal{X})} \sum_{(\alpha,\beta)\in\mathcal{V}_e} \left| [\widehat{H}_e]_{\alpha\beta} - [H_e^{\mathrm{ref}}]_{\alpha\beta} \right|, \tag{139}$$

where $\mathcal{T}$ is the set of test configurations, $\mathcal{V}_e$ contains the valid matrix-element indices for pair e, and N_{el} is the total number of evaluated elements. In the noSOC setting, α and β index orbitals. In the SOC setting, they index the joint orbital–spin basis, and complex errors are measured by their moduli.

Evaluation uses the final predictions after Hermitian projection. SOC predictions include both the frozen noSOC baseline and the learned corrections. Padded, invalid orbital channels are excluded. Absolute errors are summed over all valid matrix elements in all test configurations and divided by the total element count, without first averaging separately within each configuration.

On-site and off-site elements contribute according to their actual counts. The category weights and normalization scales used in the training loss are not applied during evaluation. On-site terms are defined by $e = (i, i, \mathbf{0})$; all other terms are classified as off-site.

F Ablation Studies and Results Across Random Seeds

F.1 Ablation Studies

We conduct ablation studies on the bilayer Bi_2Te_3 noSOC dataset to evaluate orbital-operator matrix paths and orbital encoding. All configurations use random seed 22 and identical training, validation, and test sets. Each model is trained from scratch for 3000 epochs. We report the mean absolute error (MAE) and root mean squared error (RMSE) over valid Hamiltonian matrix elements in the test set. Apart from the ablated modules, all configurations share the same remaining network settings, optimizer, learning-rate schedule, and training objective. Detailed training configurations are provided in Appendix E.

We compare the following three configurations:

- **Full model.** Both orbital encoding and orbital-operator matrix paths are enabled. Orbital encoding injects shell information into the initial atomic and pair representations. Matrix paths compose path operators as $M_p = B_{ik}G_kB_{kj}$, followed by adaptive weighted aggregation and operator decoding to generate residual pair updates.
- **Without matrix paths.** Orbital encoding is retained, while the entire matrix-path module is disabled. This removes the construction of atomic-pair operators and intermediate-atom propagation operators, path matrix multiplication, path scoring and softmax-weighted aggregation, and the residual updates generated by operator decoding. Adaptive weighting and element-pair conditioning within the matrix-path module are also disabled.
- **Without matrix paths and orbital encoding.** In addition to removing matrix paths, this configuration disables orbital encoding, including shell-token initialization and attention updates, atomic pooling, shell-pair token construction and pooling, and the residual mappings that inject orbital encodings into the initial representations. The initial representations are therefore $\mathbf{s}_i^{(0)} = \mathbf{s}_i^{\mathrm{base}}$ and $\mathbf{z}_e^{(0)} = \mathbf{z}_e^{\mathrm{base}}$.

All three configurations retain the base elemental and geometric features, pair transitions, atomic updates, endpoint atomic feedback, the Hamiltonian decoder, and Hermitian projection. Disabling orbital encoding does not change the orbital basis used for the output Hamiltonian matrices. The ablation configurations and test results are summarized in Table 10.

With orbital encoding retained, removing matrix paths increases the test MAE from 0.2370 to 0.5874 meV and the RMSE from 0.5253 to 1.1690 meV. Relative to the configuration without matrix paths, the full model reduces the MAE by 59.65%, demonstrating the importance of orbital-operator matrix paths for prediction accuracy under this experimental setting. This comparison evaluates the

Table 10: Ablation results on the bilayer Bi_2Te_3 noSOC dataset. All configurations use the same data split and random seed. Errors are reported in meV.

Orbital encoding	Matrix paths	Test MAE	Test RMSE
On	On	**0.2370**	**0.5253**
On	Off	0.5874	1.1690
Off	Off	0.5934	1.1926

Table 11: Test results for the full model on the bilayer Bi_2Te_3 noSOC dataset with different random seeds. Errors are reported in meV.

Random seed	Test structures	Test MAE	Test RMSE
11	12	0.2670	0.6230
22	12	0.2370	0.5253
42	12	0.2430	0.5451

contribution of the complete module, including the combined effects of operator composition, path aggregation, and the corresponding residual updates.

F.2 Results Across Random Seeds

We further report test results for the full model using random seeds 11, 22 and 42 on the bilayer Bi_2Te_3 noSOC dataset. All three runs use the same model architecture, training hyperparameters, and training, validation, and test set sizes. The random seed affects both data splitting and training, so this comparison reflects their combined variation. The results are reported in Table 11.

G Computational Efficiency Comparison

To evaluate the computational efficiency of Pairformer-H at inference, we compare the time required to generate Hamiltonians using Pairformer-H and self-consistent DFT calculations on 12 test structures from the bilayer Bi_2Te_3 noSOC dataset. Each structure contains 90 atoms, and both methods use the same atomic structures.

DFT calculations are performed using OpenMX 3.9 with the PBE exchange–correlation functional, Bi8.0-s3p2d2 and Te7.0-s3p2d2 localized orbital basis sets, a cutoff energy of 300 Ry, and a $3 \times 3 \times 1$ k-point mesh. The electronic temperature is 300 K, and the SCF convergence threshold is 10^{-7} Hartree.

The CPU comparison is conducted on the same compute node equipped with Intel Xeon Gold 6330 processors, with four physical CPU cores allocated to each method. DFT uses four MPI processes with one thread per process, whereas Pairformer-H uses one process with four computational threads. We additionally benchmark GPU inference on a separate server equipped with Intel Xeon Platinum 8481C processors and an NVIDIA GeForce RTX 4090 D GPU. On this server, GPU inference uses one GPU.

DFT timing covers the interval from launching OpenMX to process termination, including input reading, SCF iterations, and native output writing. Pairformer-H timing includes reading structural files, constructing the periodic graph, preparing and transferring data, performing the model forward pass, converting the predicted Hamiltonian matrices, and writing the HDF5 output. Timed predictions do not read reference Hamiltonians, overlap matrices, or precomputed edge lists, and the model's static geometry cache is disabled. Model initialization and weight loading are treated as one-time setup steps and excluded from per-structure inference times. The model runs in evaluation mode with gradient computation disabled and a batch size of one. After two warm-up predictions, all test structures are evaluated over five timing rounds. GPU operations are synchronized at the corresponding timing boundaries. DFT is run once per structure. Reported times are obtained by first averaging repeated measurements for each structure and then averaging over the 12 structures.

Table 12: Average Hamiltonian calculation times for the bilayer Bi_2Te_3 noSOC test structures. All times are reported in seconds per structure.

Method	Hardware	Full pipeline	Forward pass
DFT (OpenMX)	Xeon Gold 6330	1937.43	–
Pairformer-H (CPU)	Xeon Gold 6330	38.10	33.84
Pairformer-H (GPU)	RTX 4090 D	3.22	0.72

On the same CPU node with the same number of allocated cores, Pairformer-H achieves a $50.85\times$ speedup over DFT, defined as the ratio of their mean full-pipeline times. On the supplementary benchmark server, GPU inference provides a further $11.83\times$ speedup over CPU inference.

H Electronic-Structure and Optical-Response Calculations

H.1 Computational Settings

We use the Hamiltonians predicted by Pairformer-H to calculate band structures, densities of states, and interband optical conductivities for one test sample from each of the seven noSOC datasets and two SOC datasets, and compare the results with DFT references. For each configuration, both calculations use the same atomic structure, orbital basis, and DFT overlap matrix. Only the Hamiltonian is replaced to assess how Hamiltonian prediction errors affect downstream property calculations.

The real-space Hamiltonian and overlap matrices are transformed to reciprocal space through Fourier sums over lattice translations:

$$H_{ij}(\mathbf{k}) = \sum_{\mathbf{R}} e^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} H_{ij}^{\mathbf{R}}, \qquad S_{ij}(\mathbf{k}) = \sum_{\mathbf{R}} e^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} S_{ij}^{\mathbf{R}}, \tag{140}$$

where $\mathbf{R}$ is a lattice translation vector, and $H_{ij}^{\mathbf{R}}$ and $S_{ij}^{\mathbf{R}}$ are the Hamiltonian and overlap blocks associated with the corresponding periodic atomic pair. After assembling the full matrices, we solve the generalized eigenvalue problem:

$$H(\mathbf{k})\mathbf{c}_{n\mathbf{k}} = \varepsilon_{n\mathbf{k}} S(\mathbf{k})\mathbf{c}_{n\mathbf{k}}, \tag{141}$$

where $\varepsilon_{n\mathbf{k}}$ and $\mathbf{c}_{n\mathbf{k}}$ are the band eigenvalue and corresponding eigenvector, respectively.

Predicted and reference results use identical k-point paths, integration grids, and broadening parameters. For band-structure and density-of-states plots, both results use the DFT Fermi level of the corresponding configuration as their common energy reference.

H.2 Band Structures and Densities of States

The two-dimensional systems use the Γ–M–K–Γ band path. For bulk T_d-WTe_2, the path comprises four segments: X–Γ–Y, L–Γ–Z, N–Γ–M, and R–Γ. Each complete path contains 120 sampling points. Band curves are not connected across discontinuous path segments.

Density-of-states calculations use uniform $10\times10\times1$ k-point grids for the two-dimensional systems and a $10\times10\times10$ grid for bulk T_d-WTe_2. The sampled eigenvalues are broadened using a Gaussian with a standard deviation of 0.05 eV. The energy grid contains 1200 points spanning 5 eV below and above the corresponding DFT Fermi level. The density of states is normalized by the number of k points and reported in states/eV.

Figures 5–7 present the noSOC band structures and densities of states of monolayer graphene, bilayer graphene, monolayer MoS_2, bilayer bismuthene, bilayer Bi_2Se_3, and bulk T_d-WTe_2. Figure 8 presents the band structure and density of states of bilayer Bi_2Se_3 in the SOC setting. For the samples shown, the band structures reconstructed using Pairformer-H closely agree with the DFT references and reproduce the main band dispersions along the high-symmetry paths. The corresponding densities of states also show good agreement in their principal peak positions and overall profiles, indicating that the predicted Hamiltonians preserve the main electronic-structure features near the Fermi level in both noSOC and SOC settings.

Table 13: Mean absolute errors of band eigenvalues for representative test samples, reported in meV.

Material	Spin setting	Band MAE
Monolayer graphene	noSOC	0.2568
Bilayer graphene	noSOC	2.3775
Monolayer MoS_2	noSOC	0.3912
Bilayer bismuthene	noSOC	1.0255
Bilayer Bi_2Se_3	noSOC	0.6649
Bilayer Bi_2Te_3	noSOC	1.5803
Bulk T_d-WTe_2	noSOC	2.6291
Bilayer Bi_2Te_3	SOC	4.4544
Bilayer Bi_2Se_3	SOC	6.0960

To quantify the accuracy of band-structure reconstruction, we calculate the mean absolute error of band eigenvalues for representative test samples:

$$\mathrm{MAE}_{\mathrm{band}} = \frac{1}{|\mathcal{W}|} \sum_{(n,\mathbf{k})\in\mathcal{W}} \left| \varepsilon_{n\mathbf{k}}^{\mathrm{pred}} - \varepsilon_{n\mathbf{k}}^{\mathrm{DFT}} \right|. \tag{142}$$

Here, $\mathcal{W}$ contains the reference states along the band path satisfying $\left|\varepsilon_{n\mathbf{k}}^{\mathrm{DFT}} - E_{\mathrm{F}}^{\mathrm{DFT}}\right| \leq 1.5\,\mathrm{eV}$. At each $\mathbf{k}$ point, predicted and reference eigenvalues are sorted in ascending energy and matched by index. Both use the DFT Fermi level as their common energy reference.

H.3 Interband Optical Conductivity

We further calculate the longitudinal interband optical conductivity $\sigma_{xx}(\omega)$ for representative test samples from each dataset, comparing its real and imaginary components obtained using predicted and DFT Hamiltonians.

Both calculations share the overlap matrix, Fermi level, temperature, k-point grid, and broadening parameters. Uniform $10 \times 10 \times 1$ k-point grids are used for the two-dimensional systems, and a $10 \times 10 \times 10$ grid is used for bulk T_d-WTe_2. Electronic occupations follow the Fermi–Dirac distribution at 300 K. The noSOC calculations include a spin-degeneracy factor of two. The SOC calculations explicitly include spin degrees of freedom and use a spin-degeneracy factor of one.

Bands included in the transition sums lie within 12 eV below and above the DFT Fermi level. The photon-energy range is 0.05–10 eV, sampled at 401 points. The Lorentzian broadening parameter is 0.05 eV, and the transition-energy bin width is 0.005 eV. Only interband contributions are included.

In a local orbital basis, evaluating velocity matrix elements requires accounting for basis nonorthogonality and contributions from position-operator matrix elements (Lee et al., 2018; Esteve-Paredes & Palacios, 2023). Velocity matrix elements are evaluated using a localized-orbital expression that includes derivatives of the overlap matrix:

$$v_{nm}^{x}(\mathbf{k}) = \frac{1}{\hbar}\mathbf{c}_{n\mathbf{k}}^{\dagger}\left[\mathcal{D}_x H(\mathbf{k}) - \frac{\varepsilon_{n\mathbf{k}} + \varepsilon_{m\mathbf{k}}}{2}\mathcal{D}_x S(\mathbf{k})\right]\mathbf{c}_{m\mathbf{k}}. \tag{143}$$

Here, $\mathcal{D}_x$ denotes the covariant derivative used in the calculation. Its real-space construction includes the x component of the atomic-pair displacement $\mathbf{R} + \mathbf{r}_j - \mathbf{r}_i$, thereby accounting for the intracell orbital-center positions. The optical conductivity is obtained by summing interband transitions using the velocity matrix elements, energy differences, and occupation differences (Kubo, 1957; Calderín et al., 2017; Lee et al., 2020).

The conductivity is normalized by the full simulation-cell volume and plotted in units of $10^6\,\mathrm{S\,m^{-1}}$. For two-dimensional systems containing vacuum, this value depends on the adopted cell-height convention. The corresponding sheet conductance is

$$G_{xx}(\omega) = L_z \sigma_{xx}(\omega), \tag{144}$$

where L_z is the cell height along the layer normal.

Table 14: Integrated relative errors of the interband optical conductivity for representative test samples from each dataset.

Material	Spin setting	Real part	Imaginary part
Monolayer graphene	noSOC	0.2932%	0.4375%
Bilayer graphene	noSOC	0.3271%	0.4758%
Monolayer MoS_2	noSOC	0.0789%	0.1313%
Bilayer bismuthene	noSOC	0.1978%	0.3138%
Bilayer Bi_2Se_3	noSOC	0.0976%	0.1388%
Bilayer Bi_2Te_3	noSOC	0.1524%	0.1680%
Bulk T_d-WTe_2	noSOC	0.0778%	0.1429%
Bilayer Bi_2Te_3	SOC	0.5679%	1.0991%
Bilayer Bi_2Se_3	SOC	0.6353%	0.9240%

To quantify agreement between predicted and reference spectra, we calculate the integrated relative error separately for the real and imaginary components:

$$\epsilon_{\mathrm{rel}} = \frac{\int_{E_{\min}}^{E_{\max}} |f_{\mathrm{pred}}(E) - f_{\mathrm{DFT}}(E)| \, \mathrm{d}E}{\int_{E_{\min}}^{E_{\max}} |f_{\mathrm{DFT}}(E)| \, \mathrm{d}E}. \tag{145}$$

Here, $E = \hbar\omega$, and f denotes either $\mathrm{Re}\,\sigma_{xx}$ or $\mathrm{Im}\,\sigma_{xx}$. The integration range is 0.05–10 eV.

As shown in Figures 9–16, the predicted optical conductivities closely match the DFT references in both noSOC and SOC settings. Table 14 reports integrated relative errors below 0.5% for both components of all seven noSOC samples and at most 1.0991% across the two SOC samples.

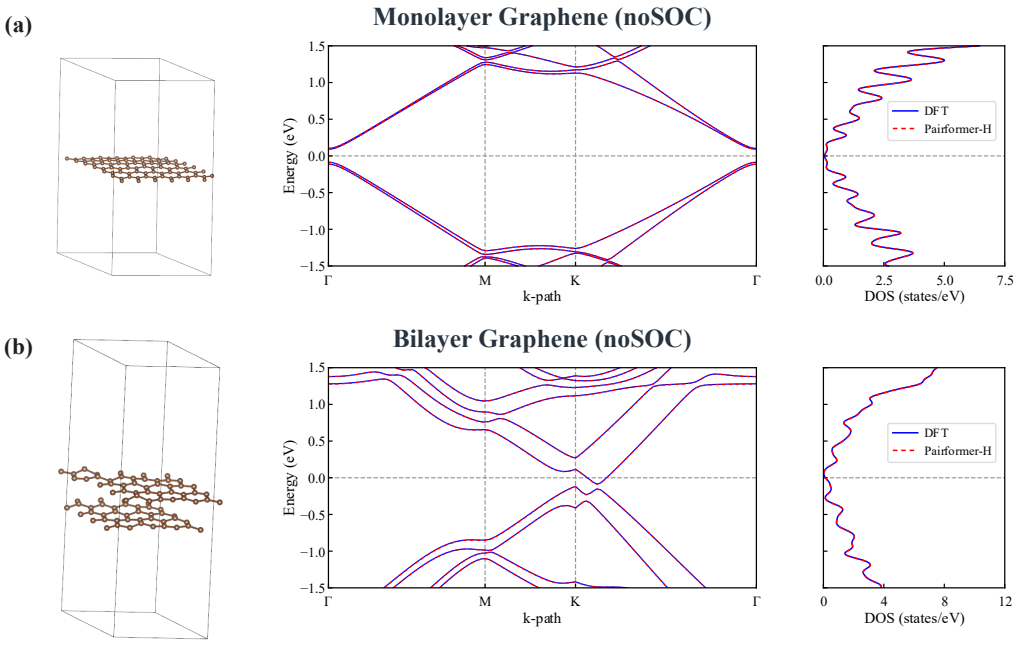


Figure 5: Band structures and densities of states reconstructed from Hamiltonians calculated by DFT and predicted by Pairformer-H in the noSOC setting: (a) monolayer graphene and (b) bilayer graphene. From left to right: crystal-structure illustrations, band structures, and densities of states.

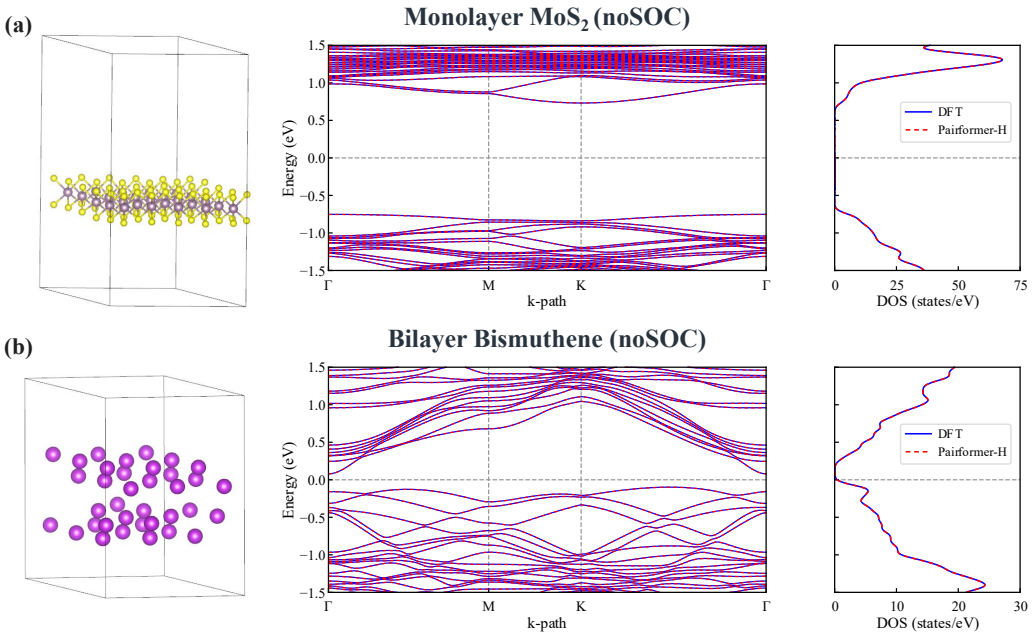


Figure 6: Band structures and densities of states reconstructed from Hamiltonians calculated by DFT and predicted by Pairformer-H in the noSOC setting: (a) monolayer MoS_2 and (b) bilayer bismuthene. From left to right: crystal-structure illustrations, band structures, and densities of states.

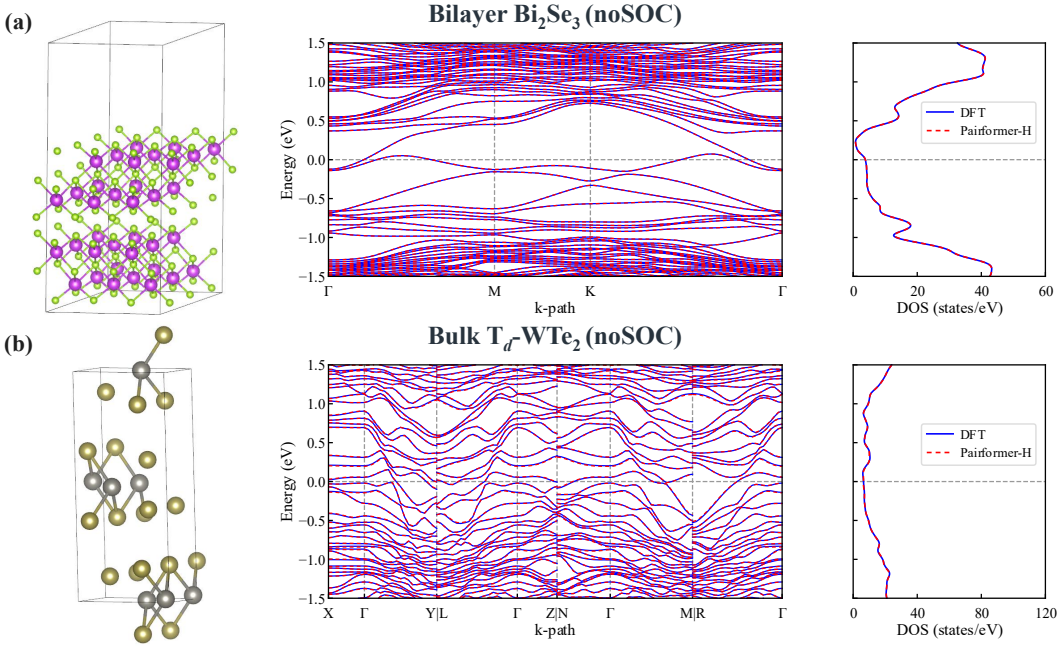


Figure 7: Band structures and densities of states reconstructed from Hamiltonians calculated by DFT and predicted by Pairformer-H in the noSOC setting: (a) bilayer Bi_2Se_3 and (b) bulk T_d-WTe_2. From left to right: crystal-structure illustrations, band structures, and densities of states.

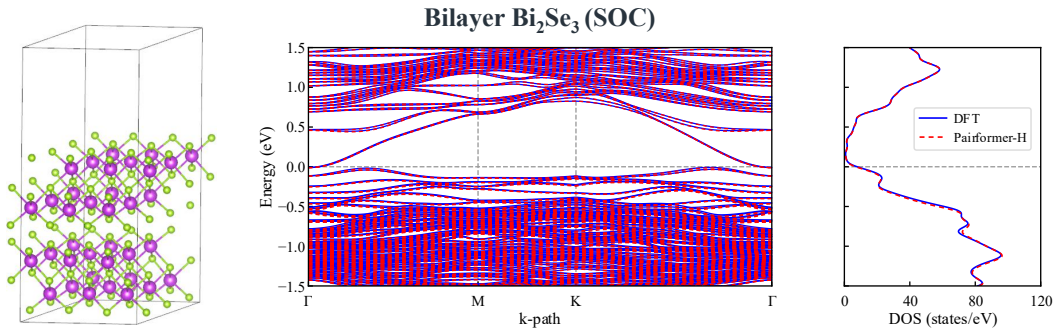


Figure 8: Band structure and density of states reconstructed from Hamiltonians calculated by DFT and predicted by Pairformer-H for bilayer Bi_2Se_3 in the SOC setting. From left to right: crystal-structure illustration, band structure, and density of states.

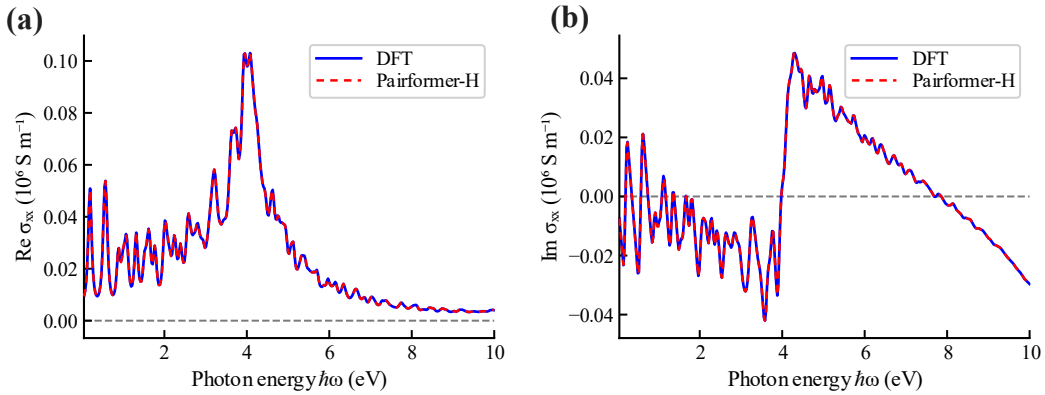


Figure 9: Interband optical conductivity of monolayer graphene in the noSOC setting, calculated from Hamiltonians obtained using DFT and predicted by Pairformer-H: (a) real and (b) imaginary components of σ_{xx}.

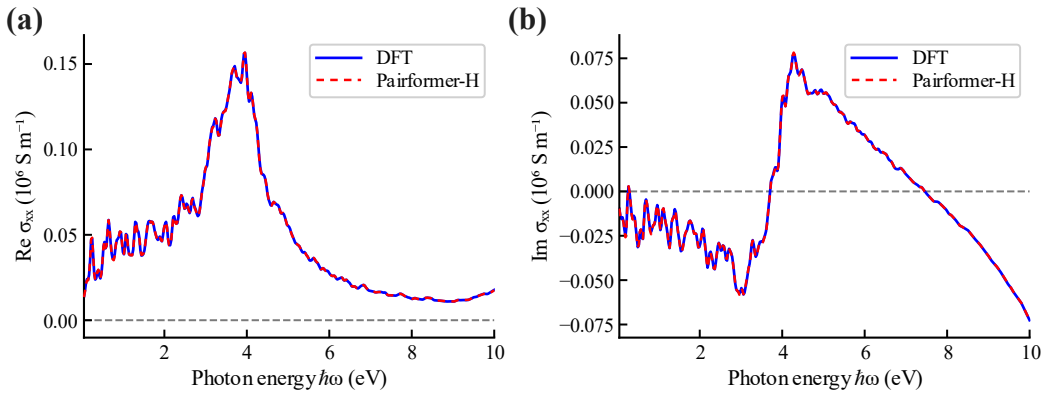


Figure 10: Interband optical conductivity of bilayer graphene in the noSOC setting, calculated from Hamiltonians obtained using DFT and predicted by Pairformer-H: (a) real and (b) imaginary components of σ_{xx}.

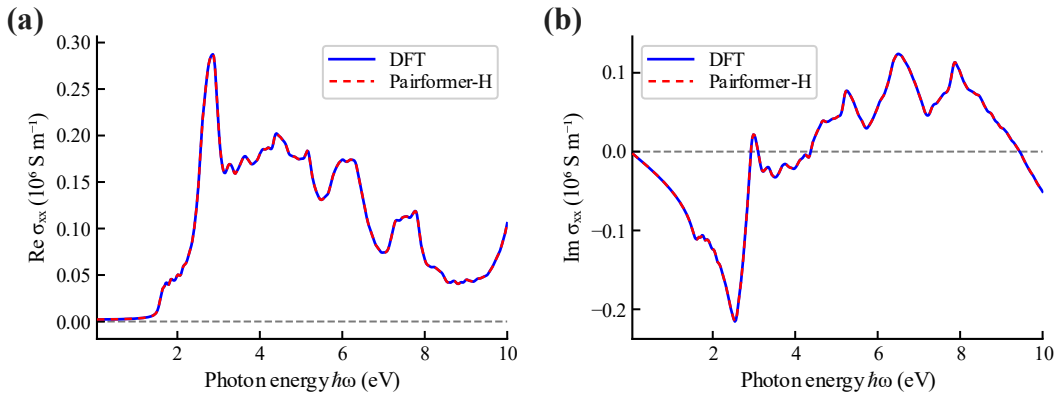


Figure 11: Interband optical conductivity of monolayer MoS_2 in the noSOC setting, calculated from Hamiltonians obtained using DFT and predicted by Pairformer-H: (a) real and (b) imaginary components of σ_{xx}.

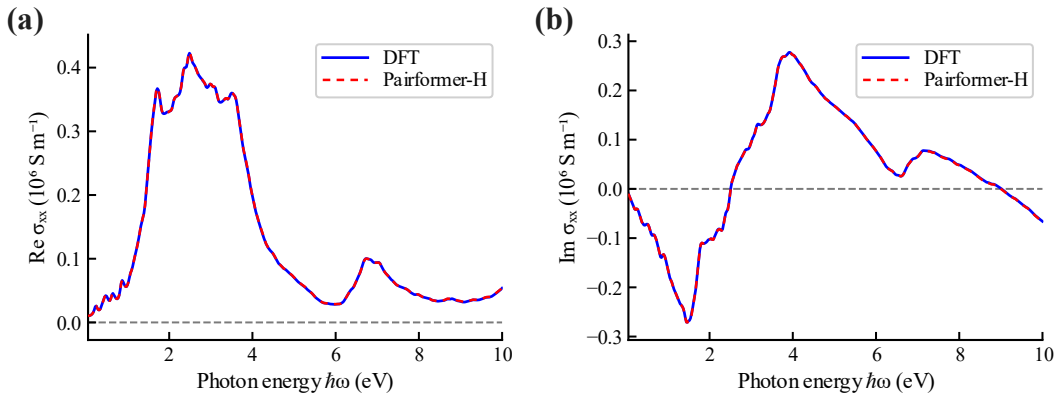


Figure 12: Interband optical conductivity of bilayer bismuthene in the noSOC setting, calculated from Hamiltonians obtained using DFT and predicted by Pairformer-H: (a) real and (b) imaginary components of σ_{xx}.

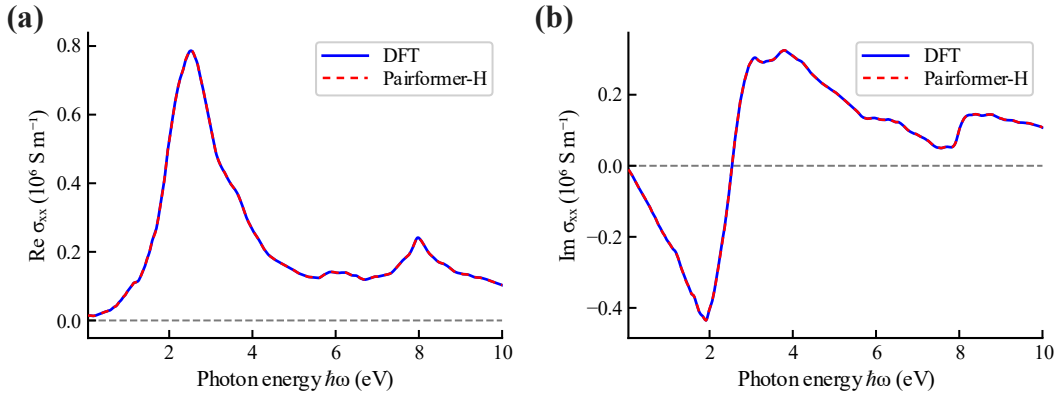


Figure 13: Interband optical conductivity of bilayer Bi_2Se_3 in the noSOC setting, calculated from Hamiltonians obtained using DFT and predicted by Pairformer-H: (a) real and (b) imaginary components of σ_{xx}.

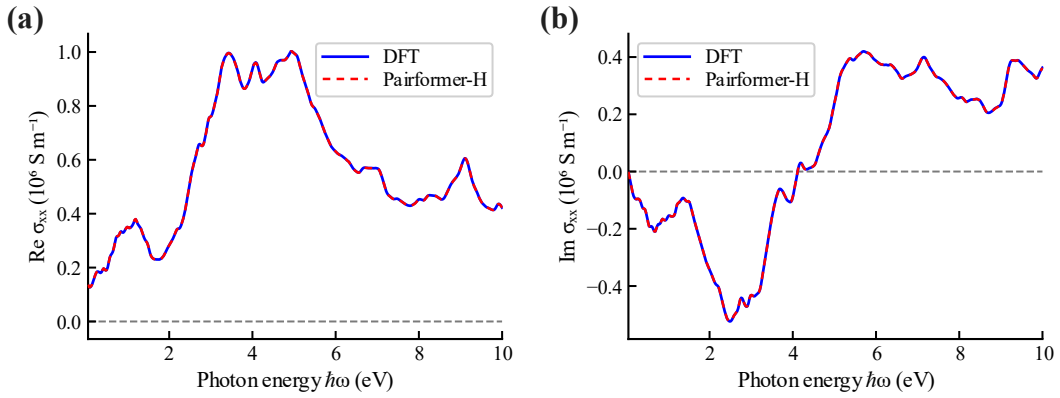


Figure 14: Interband optical conductivity of bulk T_{d}-WTe_2 in the noSOC setting, calculated from Hamiltonians obtained using DFT and predicted by Pairformer-H: (a) real and (b) imaginary components of σ_{xx}.

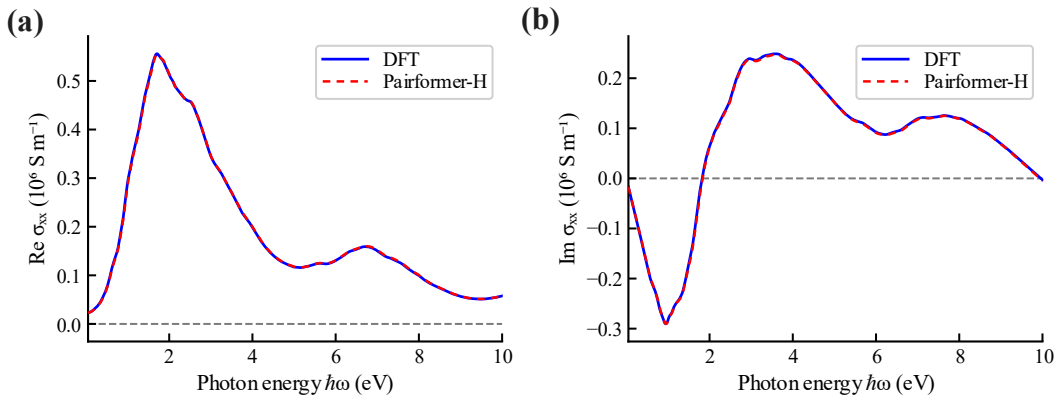


Figure 15: Interband optical conductivity of bilayer Bi_2Te_3 in the SOC setting, calculated from Hamiltonians obtained using DFT and predicted by Pairformer-H: (a) real and (b) imaginary components of σ_{xx}.

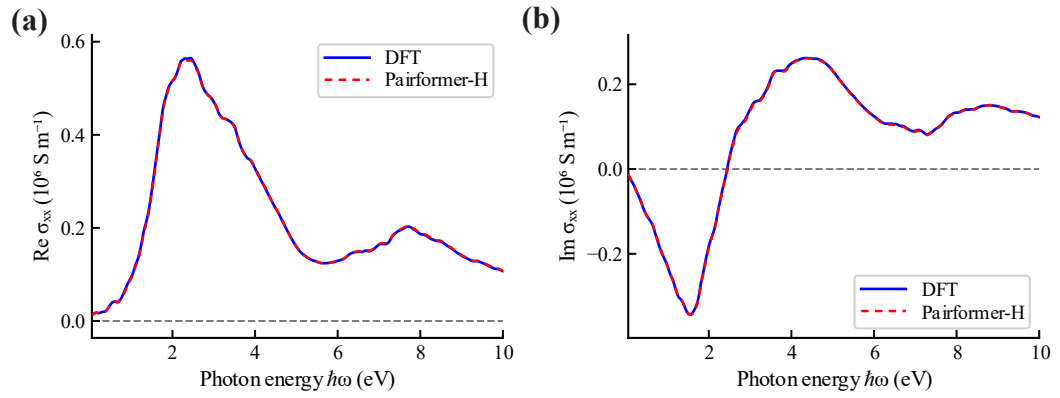


Figure 16: Interband optical conductivity of bilayer Bi_2Se_3 in the SOC setting, calculated from Hamiltonians obtained using DFT and predicted by Pairformer-H: (a) real and (b) imaginary components of σ_{xx}.