Adaptive Multi-Frame Learning for Expressive and Stable Atomic Representations

Jun Wang[1], **Yifan Zeng**[1,2], **Bo Han**[3], **Fengwang Li**[2,*], **Aoni Xu**[2,*], **Tongliang Liu**[1*]
[1]Sydney AI Centre, School of Computer Science, The University of Sydney
[2]School of Chemical and Biomolecular Engineering and ARC Centre of Excellence for Green Electrochemical Transformation of Carbon Dioxide, The University of Sydney
[3]Department of Computer Science, Hong Kong Baptist University
{fengwang.li, aoni.xu, tongliang.liu}@sydney.edu.au

Abstract

Reliable prediction of material properties from atomistic structures requires machine learning models that respect SE(3) symmetry. Frame-based methods address this challenge by constructing equivariant coordinate systems that align SE(3)-equivalent structures into unified representations. While attractive for their flexibility and efficiency, existing frame-based methods face two major limitations: a single frame type is often insufficient for heterogeneous atomic environments, and frame constructions can become unstable near highly symmetric or degenerate configurations. To address these issues, we propose the Multi-Frame Adaptive Network (MFAN), which leverages multiple frame types and introduces a local adaptive mechanism to combine them according to atomic environments, enabling different geometric references across atoms and complementary local-global information for each atom. Building on this multi-frame architecture, we further incorporate a frame-quality-aware weighting scheme that downweights unreliable frames before degeneration, thereby improving the continuity of the resulting representation. Experiments on crystal property prediction benchmarks demonstrate the leading performance of MFAN, with consistent gains from adaptive multi-frame learning. Additional force prediction experiments show improved stability in continuity-sensitive settings.

1 Introduction

A fundamental question in chemistry and materials science is the relationship between structure and properties [3]. While density functional theory (DFT) provides quantum-mechanical predictions via iterative electronic structure calculations, the high computational cost restricts their broader applicability [12, 31]. Recently, machine learning has emerged as a promising alternative for predicting material properties directly from atomic structures [8, 13, 2, 19]. In many applications, these methods achieve near-DFT accuracy with orders-of-magnitude higher efficiency, enabling the exploration of larger materials spaces and more complex dynamics [54, 41, 66, 75, 72].

Symmetry plays a fundamental role in physical systems, as exemplified by Noether's theorem [23]. For machine learning models of materials, respecting geometric symmetries is likewise essential for physical consistency and reliable generalization [19]. In general, a basic requirement is invariance or equivariance to SE(3) transformations (3D translations and rotations) of the input geometry. To incorporate these symmetries, a wide range of approaches have been developed, from SE(3)-invariant features [4, 61, 8] to equivariant neural networks [22, 2, 19]. Among them, frame-based methods have attracted growing attention [15, 17, 71, 29, 26]. These methods derive equivariant coordinate systems

*Corresponding authors.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

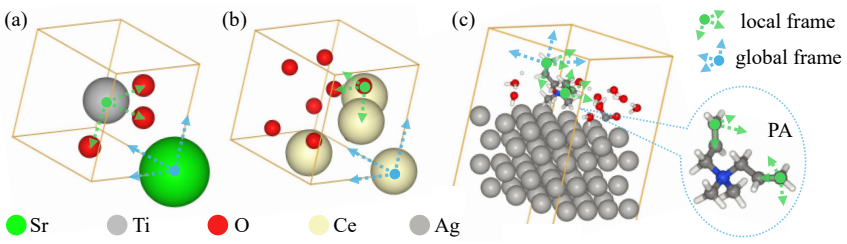

Figure 1: Illustration of diverse frame requirements. In strontium titanate (a), Sr atoms occupy regular crystallographic positions, making the global lattice frame natural, whereas Ti-O interactions depend on local coordination and are better captured by local frames. In cerium dioxide (b), the ordered Ce pattern is also aligned with the lattice, whereas O-related vacancy formation and migration are governed by local environments. In polymeric ammonium (PA)-modulated formate adsorption (c), the global frame is dominated by the Ag surface, requiring local frames to describe PA. Yet local frames cannot distinguish the symmetric termini of PA, so global information remains necessary.

from atomic structures, thereby mapping SE(3)-equivalent configurations to consistent canonical representations. This enables them to retain richer directional, multi-body geometric information than invariant descriptors. By scalarizing geometric quantities through equivariant frames, they are also typically more lightweight than equivariant networks, avoiding explicit equivariant feature propagation and costly tensor-product operations [15].

Despite this promise, currently the practical use of frame-based methods still faces two major challenges. First, existing methods typically impose a single, uniform frame construction across the entire system. This leaves a gap between frame construction and representation needs: the former is usually fixed at the method level, whereas the latter can vary across atoms. As illustrated in Figures 1(a) and (b), even within a single material, different atoms may favor fundamentally different types of frames because of differences in local environment and coordination. Moreover, as shown in Figure 1(c), a single atom may still require multiple frame types: local frames capture fine-grained geometry, while global frames provide complementary context to distinguish locally similar substructures [26]. Second, frame constructions are intrinsically vulnerable to discontinuity [18]. When the input approaches a highly symmetric or degenerate configuration, even small perturbations can induce abrupt changes in the constructed frames. This issue is especially problematic in applications such as force prediction and transition-state searching, where discontinuities in the energy surface can produce spurious forces and destabilize optimization trajectories.

In this work, we introduce the Multi-Frame Adaptive Network (MFAN) to address both limitations from a unified perspective. Rather than committing to a single frame construction principle, MFAN treats frame usage as an adaptive combination problem. It leverages multiple frame types and aggregates them with adaptive weights determined by each atom's local environment, enabling atoms to favor distinct geometric references and integrate complementary local and global information when needed. This adaptive combination further incorporates frame-quality awareness by dynamically downweighting near-degenerate or unstable frames, so that their contributions smoothly decay to zero as they approach degeneracy, thereby improving the continuity and stability of the resulting geometric representations. We evaluate MFAN on two benchmark groups: crystal property prediction on JARVIS-DFT and Materials Project (MP) datasets, and structure-to-energy-and-forces prediction on MPTrj-light, a subset sampled from MP relaxation trajectories, together with Cu/O and our self-collected CeO_2/Pt datasets for application-oriented surface scenarios. The results collectively validate the effectiveness of the multi-frame framework.

Our contributions are summarized as follows:

- We propose adaptive multi-frame learning, moving beyond fixed frame construction and usage strategies through adaptive combination to better handle diverse chemical environments.
- We instantiate this idea as MFAN, a flexible framework for integrating heterogeneous frames that decouples model architecture from frame construction and readily supports new frames.

- We evaluate MFAN on diverse benchmarks, showing that adaptive multi-frame learning consistently improves accuracy and stability over single-frame variants and highlighting the model's broad potential across crystalline, surface, and application-driven settings.

2 Preliminaries

In this section, we first describe the prediction tasks and symmetry requirements considered in this work, and then briefly introduce the basic concepts of frames.

2.1 Material structure representations and symmetry requirements

Following previous works [1, 70, 13], we represent a periodic cell with N atoms by triplet $(\boldsymbol{A}, \boldsymbol{P}, \boldsymbol{L})$, where $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_N] \in \mathbb{R}^{p\times N}$ represents atomic features, $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, ..., \boldsymbol{p}_N] \in \mathbb{R}^{3\times N}$ contains Cartesian coordinates, and $\boldsymbol{L} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3] \in \mathbb{R}^{3\times 3}$ defines the lattice vectors. Repetition along the lattice vectors gives the infinite periodic material structure $(\hat{\boldsymbol{A}}, \hat{\boldsymbol{P}})$, where $\hat{\boldsymbol{A}} = \{\hat{\boldsymbol{a}}_{i,\boldsymbol{k}} \mid \hat{\boldsymbol{a}}_{i,\boldsymbol{k}} = \boldsymbol{a}_i, \boldsymbol{k} \in \mathbb{Z}^3, i = 1, \ldots, N\}$ and $\hat{\boldsymbol{P}} = \{\hat{\boldsymbol{p}}_{i,\boldsymbol{k}} \mid \hat{\boldsymbol{p}}_{i,\boldsymbol{k}} = \boldsymbol{p}_i + \boldsymbol{L}\boldsymbol{k}, \boldsymbol{k} \in \mathbb{Z}^3, i = 1, \ldots, N\}$.

For crystalline materials, this representation is natural because the periodic cell directly corresponding to the crystal unit cell. It is also commonly used for amorphous materials in first-principles calculations and atomistic simulations [34, 35, 73]. For example, in surface modeling, two lattice vectors define the periodic surface, while the third is surface-normal and includes a sufficiently large vacuum layer to avoid interactions between repeated slabs, as shown in Figure 1(c) [41, 72].

We consider three symmetry requirements for material property prediction. The first is *permutation invariance*, requiring predictions to be invariant to reordering of atomic indices, which is naturally satisfied by most graph neural network (GNN) or transformer-based models. The second is *periodic invariance*, requiring predictions to remain unchanged under equivalent representations of the same periodic cell [70]. This makes the use of global information, including global frames, more delicate, whereas local models often satisfy this requirement more naturally. The third and most important one is SE(3) symmetry, which requires predictions to be invariant or equivariant under translations and rotations. Formally, let $f_\theta(\boldsymbol{A}, \boldsymbol{P}, \boldsymbol{L})$ be a prediction model, we say f_θ is *SE(3)-invariant* if, for any 3D rotation matrix $\boldsymbol{Q} \in \mathbb{R}^{3\times 3}$ with $\boldsymbol{Q}^\mathsf{T}\boldsymbol{Q} = \boldsymbol{I}$ and $\det(\boldsymbol{Q}) = 1$, and for any translation vector $\boldsymbol{b} \in \mathbb{R}^3$, we have

$$f_\theta(\boldsymbol{A}, \boldsymbol{P}, \boldsymbol{L}) = f_\theta(\boldsymbol{A}, \boldsymbol{Q}\boldsymbol{P} + \boldsymbol{b}, \boldsymbol{Q}\boldsymbol{L}). \tag{1}$$

Similarly, we say f_θ is *SE(3)-equivariant* if

$$\mathcal{M}(\boldsymbol{Q}) f_\theta(\boldsymbol{A}, \boldsymbol{P}, \boldsymbol{L}) = f_\theta(\boldsymbol{A}, \boldsymbol{Q}\boldsymbol{P} + \boldsymbol{b}, \boldsymbol{Q}\boldsymbol{L}), \tag{2}$$

where $\mathcal{M}(\boldsymbol{Q})$ denotes the matrix representation of $\boldsymbol{Q}$ acting on the codomain of f_θ. Both SE(3)-invariant and SE(3)-equivariant prediction tasks are common in practice. In this work, we primarily focus on SE(3)-invariant tasks.

Specifically, Structure-to-Energy-and-Forces (S2EF) prediction involves estimating the energy of a chemical structure and the forces acting on its atoms, which is essential for molecular dynamics [7]. While forces are SE(3)-equivariant, they are also expected to be conservative, so we compute them as the negative gradients of the predicted energy with respect to atomic coordinates to meet both requirements [1, 13, 73, 19], i.e., $\hat{\boldsymbol{F}}_i = -\nabla_{\boldsymbol{p}_i}\hat{E}$. Thus, the core objective remains the prediction of an SE(3)-invariant energy, whose smoothness is critical for accurate and physically meaningful forces.

2.2 Frame-based models and typical frames

For SE(3)-invariance, translational invariance can be naturally handled by using relative coordinates, so the main challenge is rotation invariance, i.e., *SO(3)-invariance*. Frame-based methods address this by constructing a frame $\mathcal{F}$, namely one or a set of equivariant rotation matrices from $\boldsymbol{P}$ and $\boldsymbol{L}$. When $\mathcal{F}$ is a single rotation matrix, Eq. (2) gives, for an arbitrary rotation $\boldsymbol{Q}$, we have

$$\mathcal{F}(\boldsymbol{Q}\boldsymbol{P}, \boldsymbol{Q}\boldsymbol{L})^\mathsf{T}\boldsymbol{Q}\boldsymbol{P} = (\boldsymbol{Q}\mathcal{F}(\boldsymbol{P}, \boldsymbol{L}))^\mathsf{T}\boldsymbol{Q}\boldsymbol{P} = \mathcal{F}(\boldsymbol{P}, \boldsymbol{L})^\mathsf{T}\boldsymbol{Q}^\mathsf{T}\boldsymbol{Q}\boldsymbol{P} = \mathcal{F}(\boldsymbol{P}, \boldsymbol{L})^\mathsf{T}\boldsymbol{P}. \tag{3}$$

Therefore, it defines a canonical rotation (or coordinate system) that maps all rotationally equivalent inputs to a common representation. As a result, $f_\theta(\boldsymbol{A}, \mathcal{F}_{\boldsymbol{P},\boldsymbol{L}}^\mathsf{T}\boldsymbol{P}, \mathcal{F}_{\boldsymbol{P},\boldsymbol{L}}^\mathsf{T}\boldsymbol{L})$ is rotationally invariant,

which is known as *canonicalization* [18]. When $\mathcal{F}$ is a set of frames, invariant predictions can still be achieved by *frame averaging* [59], $\bar{f}_\theta = \sum_{F \in \mathcal{F}(\boldsymbol{P}, \boldsymbol{L})} f_\theta(\boldsymbol{A}, F^\mathsf{T}\boldsymbol{P}, F^\mathsf{T}\boldsymbol{L})/|\mathcal{F}(\boldsymbol{P}, \boldsymbol{L})|$. Since canonicalization is the main usage in this work, we treat $\mathcal{F}$ as a matrix whenever unambiguous. When $\mathcal{F}$ denotes a set, we view it as a uniform distribution and approximate the average by randomly sampling one frame during computation [17, 18]. For material property prediction, atom-wise representations can be computed using either a shared frame for all atoms or atom-specific frames constructed from each atom's local environment; we refer to these as *global frames* and *local frames*, respectively [26].

Several frame construction methods have been proposed, including

- *Neighbor-pair frame*, which enumerates all pairs of neighboring atoms for each central atom and uses the two resulting bond directions together with their orthogonal cross-product direction to construct a local frame set [58, 18].
- *Lattice frame*, which constructs a global frame from the three shortest non-coplanar vectors in the lattice spanned by integer combinations of the lattice vectors [71].
- *PCA frame*, which is obtained by applying principal component analysis to the centered atomic positions and selecting the eigenvectors in the order of their eigenvalues [59, 17].
- *QR frame*, which first selects a full-rank, non-degenerate set of vectors from atomic positions, then applies generalized QR decomposition, and uses Q to define the frame [48].

Among them, the latter two can be used for both local and global frames. A key challenge for frame construction is discontinuity. For example, PCA frames may change abruptly when eigenvalues are close. More generally, all unweighted finite frames for SO(3) group are unlikely to be continuous [18]. This poses a major limitation to model stability and optimization, particularly when force prediction requires taking derivatives.

3 Methodology

In this section, we present the Multi-Frame Adaptive Network (MFAN), a GNN-based model built around edge-level multi-frame geometric feature extraction. We first introduce our multi-frame design for richer geometric references, then describe the continuity-preserving mechanism for stabilizing the geometric representation, and finally present the full model.

3.1 Multi-frame geometric feature extraction

Revisiting Eq. (3), a frame $\mathcal{F}$ in frame-based models can be viewed as a geometric encoder that transforms SO(3)-equivariant vectors $\boldsymbol{V}$ into rotation-invariant scalar features through the projection $\mathcal{F}^\mathsf{T}\boldsymbol{V}$. These invariant features can then be processed directly by standard neural networks without breaking symmetry, which forms the basis of many frame-based SE(3)-invariant models [17, 58, 71, 29]. Moreover, when the frame is non-degenerate, this encoding is invertible and thus information-preserving in principle, supporting the effectiveness of frame-based methods [15, 16].

Despite this theoretical completeness, different frame constructions can exhibit substantially different learning efficiency across tasks and datasets in practice [67, 29, 26]. As illustrated in Figure 1, a global frame may provide useful holistic context but lose local resolution, while local frames capture fine geometry but may miss global distinctions. Different frame construction strategies, such as PCA and neighbor-pair, also differ in stability and geometric meaning. This makes it difficult for a single-frame encoding to consistently achieve strong performance, and motivates the use of multiple frames: since frame-based scalarization and vectorization are linear operations, they naturally extend to a set of frames. For a frame set $\{\mathcal{F}^{(1)}, \ldots, \mathcal{F}^{(K)}\}$ with K frames, the concatenated projection $[\mathcal{F}^{(1)}, \ldots, \mathcal{F}^{(K)}]^\mathsf{T}\boldsymbol{V}$ remains SO(3)-invariant. Multiple frames can therefore serve as complementary geometric encoders of the same structure, providing a richer representation space.

Building on this idea, MFAN develops a geometric feature extraction module that integrates multiple candidate frames and adaptively combines the most informative and stable geometric references through a lightweight routing network. The resulting module converts geometric information into SE(3)-invariant representations for subsequent neural network processing. Specifically, to preserve translational invariance, for each neighboring atom pair (i, j), we use the relative displacement

$\boldsymbol{d}_{ij} = \boldsymbol{p}_j - \boldsymbol{p}_i$ as the geometric feature.[2] Following [71, 29], we decompose $\boldsymbol{d}_{ij}$ into an SE(3)-invariant distance $r_{ij} = \|\boldsymbol{d}_{ij}\|$ and an SO(3)-equivariant direction $\hat{\boldsymbol{d}}_{ij} = \boldsymbol{d}_{ij}/r_{ij}$. The distance r_{ij} is further expanded by a set of uniformly spaced radial basis functions, $\mathrm{RBF}_k(s) = \exp(-\gamma_k(s-\mu_k)^2)$, to obtain a richer representation, and modulated by an envelope function $u(r)$ that smoothly decays to zero near the neighboring cutoff [21]. This yields the final vectorized distance feature $\boldsymbol{r}_{ij}$.

The multi-frame design is used in directional feature processing. For each atom i, we first construct its frame set $\{\mathcal{F}_i^{(1)}, \ldots, \mathcal{F}_i^{(K)}\}$, where global frames are shared across atoms and local frames are constructed from its local environment in a consistent manner. Given each frame $\mathcal{F}_i^{(k)}$, we express the direction $\hat{\boldsymbol{d}}_{ij}$ in its coordinate system as $\boldsymbol{c}_{ij}^{(k)} = (\mathcal{F}_i^{(k)})^\mathsf{T} \hat{\boldsymbol{d}}_{ij}$, which gives an SE(3)-invariant directional descriptor. Each $\boldsymbol{c}_{ij}^{(k)}$ is then processed by a lightweight embedding and alignment network $\phi^{(k)}(\cdot)$. To aggregate information from different frames, a routing network predicts an importance score $w_{ij}^{(k)}$ from the local environment, encoded by the SE(3)-invariant node representations $\boldsymbol{x}_i$, $\boldsymbol{x}_j$, and SE(3)-invariant edge representation $\boldsymbol{e}_{ij}$. Meanwhile, each frame is assigned a quality score $\mathcal{Q}_i^{(k)}$ to characterize its continuity and stability, as introduced in the next subsection. Finally, we normalize the product of $w_{ij}^{(k)}$ and $\mathcal{Q}_i^{(k)}$ across frames to obtain the overall routing weight $\bar{w}_{ij}^{(k)}$, and compute the aggregated directional feature as $\boldsymbol{\theta}_{ij} = \sum_k \bar{w}_{ij}^{(k)} \phi^{(k)}\left(\boldsymbol{c}_{ij}^{(k)}\right)$.

In MFAN, the geometric feature extraction module produces $\boldsymbol{r}_{ij}$ and $\boldsymbol{\theta}_{ij}$ for the network backbone. Since r_{ij} and $\boldsymbol{c}_{ij}^{(k)}$ are SE(3)-invariant, it suffices for $\mathcal{Q}_i^{(k)}$ to satisfy the same property so that all geometric inputs respect the symmetry constraint, thereby ensuring SE(3)-invariant predictions. The completeness of single-frame representations has been well studied in prior works [15, 71], and our multi-frame design does not compromise this property. In this work, we combine at least one global frame with at least one local frame in MFAN to capture both global orientation and local geometric details, alleviating the limitation of purely local frames in distinguishing locally equivalent atoms, such as those at the same Wyckoff position in crystals [26]. Proposition 1 formalizes this property by showing that the inclusion of a non-degenerate shared global frame ensures that the multi-frame geometric representation can distinguish rotationally equivalent local environments. Details and proofs of all propositions are provided in Appendix A.

Proposition 1. *Let p and q be two atoms with neighbor sets $\{p_i\}_{i=1}^m$ and $\{q_i\}_{i=1}^m$, respectively. Assume that $\{\boldsymbol{d}_{qq_i}\}_{i=1}^m \neq \{\boldsymbol{d}_{pp_i}\}_{i=1}^m$, but $\{\boldsymbol{d}_{qq_i}\}_{i=1}^m = \{\boldsymbol{Q}\boldsymbol{d}_{pp_i}\}_{i=1}^m$ for some non-identity rotation $\boldsymbol{Q} \in SO(3)$. Consider their frame sets $\{\mathcal{F}_p^{(1)}, \ldots, \mathcal{F}_p^{(K)}\}$ and $\{\mathcal{F}_q^{(1)}, \ldots, \mathcal{F}_q^{(K)}\}$, if there exists k such that the k-th frame is a non-degenerate shared global frame, i.e., $\mathcal{F}_p^{(k)} = \mathcal{F}_q^{(k)}$, then $\{[r_{pp_i}, \boldsymbol{c}_{pp_i}^{(1)}, \ldots, \boldsymbol{c}_{pp_i}^{(K)}]\}_{i=1}^m \neq \{[r_{qq_i}, \boldsymbol{c}_{qq_i}^{(1)}, \ldots, \boldsymbol{c}_{qq_i}^{(K)}]\}_{i=1}^m$. Consequently, the resulting multi-frame representation can distinguish p and q.*

3.2 Continuity preserving mechanism

While frame-based methods provide an efficient route to SO(3)-invariant representations, discontinuities in frame construction can harm stability and derivative-based predictions such as forces. As discussed in the preliminaries, unweighted finite frames for SO(3) group are generally not continuous [18]. A natural remedy is to introduce frame weighting so that near-degenerate frames are smoothly suppressed. However, existing weighted frame-averaging methods typically require enumerating all neighboring pairs [58, 18], which is computationally prohibitive in practice.

To address this issue, we adopt a different strategy via our multi-frame architecture. Instead of exhaustively enumerating all neighboring pairs, we combine several computationally efficient frame constructions and assign each of them a quality-aware score, so that unstable frames are suppressed before degeneracy. As shown in Propositions 2 and 3, for common frame constructions, the onset of discontinuity can often be characterized by simple conditions, and the corresponding degenerate configurations are typically confined to measure-zero subsets of the configuration space. Therefore,

[2] In MFAN, two atoms are treated as neighbors and connected by a GNN edge if their distance is below a predefined cutoff. With periodic boundary conditions, an atom pair (i, j) may yield multiple edges from different periodic images, each with displacement $\boldsymbol{d}_{ij(\boldsymbol{m})} = \boldsymbol{p}_j - \boldsymbol{p}_i + \boldsymbol{L}\boldsymbol{m}$ for $\boldsymbol{m} \in \mathbb{Z}^3$. When the periodic image index need not be specified explicitly, we write ij for a valid edge $ij(\boldsymbol{m})$ to simplify notation.

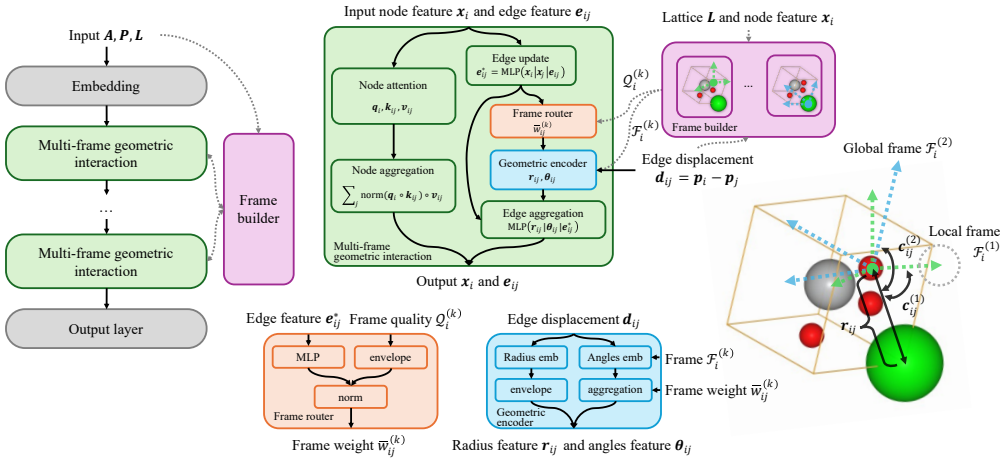

Figure 2: Illustration of the proposed MFAN architecture. MFAN follows a GNN pipeline with stacked interaction layers for iterative node and edge updates. In each layer, node representations are updated by attention-based aggregation of neighborhood information, while edge representations are enriched by a geometric encoder. This encoder uses the proposed multi-frame mechanism to extract directional information under multiple local and global frames, and a quality-aware frame router to adaptively combine them for improved geometric expressiveness and stability.

by combining heterogeneous frames and reweighting them according to their stability or continuity, the model can adaptively switch across candidate frames as configurations vary, favoring the more stable ones to maintain overall stability at lower computational cost.

Proposition 2. *Let p be an atom with neighbor set $\{p_1, \ldots, p_m\}$. If there exists constants $d_{min}, d_{max}, \varepsilon > 0$ such that $d_{min} < \|\boldsymbol{d}_{pp_i}\| < d_{max}$ and $\|\boldsymbol{d}_{pp_i} \times \boldsymbol{d}_{pp_j}\| \geq \varepsilon$, $\forall\, 1 \leq i < j \leq m$, then the neighbor-pair frame of p depends continuously on the atomic positions.*

Proposition 3. *Let p be an atom with neighbor set $\{p_1, \ldots, p_m\}$, and let $\Sigma_p = \sum_{i=1}^{m} \boldsymbol{d}_{pp_i} \boldsymbol{d}_{pp_i}^{\mathsf{T}}$ be the covariance matrix constructed from its neighbors. Suppose that there exists a constant $\delta > 0$ such that the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of Σ_p satisfy $\min_{i \neq j} |\lambda_i - \lambda_j| \geq \delta$. Then the (sign-invariant) PCA frame of p depends continuously on the atomic positions.*

Moreover, as illustrated by the above two propositions, the quantities that characterize frame degeneracy, such as inter-neighbor angles or eigenvalue gaps, are often SE(3)-invariant. These quantities can therefore be directly used to construct the frame quality score $\mathcal{Q}_i^{(k)} \in [0, +\infty)$, where values close to zero indicate that the frame is approaching degeneracy. Proposition 4 formalizes how quality scores can be used to preserve continuity: if the scores indicate impending frame degeneracy and induce continuous weights that vanish in neighborhoods of the corresponding singular sets, the weighted aggregation of multi-frame representations remain continuous even when individual frames are discontinuous.

Proposition 4. *Let $\mathcal{F}^{(k)}(\boldsymbol{P})$ be a frame constructed from the configuration $\boldsymbol{P}$ for $k = 1, \ldots, K$, and $g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$ be the SO(3)-invariant (or equivariant) representation induced by the k-th frame, where g_k is continuous in both arguments.*[3] *Let $w_k(\boldsymbol{P})$ be continuous, SO(3)-invariant weights, and define $H(\boldsymbol{P}) = \sum_{k=1}^{K} w_k(\boldsymbol{P})\, g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$. Suppose that each frame $\mathcal{F}^{(k)}(\boldsymbol{P})$ is continuous outside a closed singular set Σ_k, and $w_k(\boldsymbol{P})$ vanishes continuously in an ε_k-neighborhood of Σ_k for some $\varepsilon_k > 0$. Then $H(\boldsymbol{P})$ is continuous and SO(3)-invariant (or equivariant).*

In MFAN, we incorporate an additional envelope function when constructing the frame quality score $\mathcal{Q}_i^{(k)}$, smoothly truncating it to zero before the raw score reaches zero, i.e., the frame becomes discontinuous, to meet the conditions of Proposition 4. Multiplying it with the importance score $w_{ij}^{(k)}$ gives the overall routing weight of each frame, thereby promoting continuity of the aggregated

[3] In Proposition 4, $\boldsymbol{P}$ denotes a simplified geometric configuration variable, with the full dependence on atomic features and lattice variables omitted for clarity.

multi-frame representation. In particular, for S2EF tasks, the lattice frame is independent of atomic positions and remains unchanged under position perturbations or differentiation. We therefore fix its quality score to 1, ensuring that directional features do not vanish simultaneously.

3.3 Overall MFAN architecture

As illustrated in Figure 2, MFAN adopts an attention-enhanced graph neural network (GNN) as its backbone, where each atom is treated as a node and edges are constructed between atoms within a predefined cutoff. In the embedding layer, atomic features $\boldsymbol{a}_i$ and interatomic distances r_{ij} are embedded as the initial node and edge representations, denoted by $\boldsymbol{x}_i$ and $\boldsymbol{e}_{ij}$, respectively. The node and edge representations are then iteratively updated through interaction layers. For node updates, we follow the iComformer architecture [71]. For a central atom i and its neighbor j, the query $\boldsymbol{q}_i$ is computed from $\boldsymbol{x}_i$, while the key $\boldsymbol{k}_{ij}$ and value $\boldsymbol{v}_{ij}$ are obtained from the concatenated feature $[\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{e}_{ij}]$. Neighbor messages are then aggregated to produce the updated node representation as $\boldsymbol{x}_i^{\text{new}} = \text{softplus}\left(\boldsymbol{x}_i + \sum_{j\in\mathcal{N}(i)} \text{sigmoid}(\text{norm}(\boldsymbol{q}_i \circ \boldsymbol{k}_{ij})) \circ \boldsymbol{v}_{ij}\right)$, where $\circ$ and $\text{norm}(\cdot)$ denote the Hadamard product and element-wise normalization. For edge updates, we first compute an intermediate representation $\boldsymbol{e}_{ij}^*$ from $[\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{e}_{ij}]$. This representation is used to estimate the frame importance score $w_{ij}^{(k)}$ in the geometric feature extraction module, and is further combined with the resulting geometric features $\boldsymbol{r}_{ij}$ and $\boldsymbol{\theta}_{ij}$ to produce the updated edge representation $\boldsymbol{e}_{ij}^{\text{new}}$. Finally, average pooling over node representations gives a structure-level representation for prediction. When needed, forces can be computed as the negative gradient of the predicted total energy with respect to atomic positions. The detailed architecture of MFAN are provided in Appendix B.

4 Related Works

Existing methods for modeling materials and chemical structures under SE(3) symmetry can be broadly grouped into three categories. The first builds on SE(3)-invariant geometric features, such as interatomic distances and angles. Starting from BPNN [4], which relies on manually designed descriptors, subsequent works has explored new architectures, including GNNs and transformers [61, 9, 51, 70, 47, 64], as well as more informative features and training targets, such as magnetic moments and dihedral angles [56, 10, 8, 25, 13]. These models are efficient, and recent theoretical studies suggest that fully connected GNNs relying only on distance information can still be nearly E(3)-complete [40]. However, under practical sparse and moderate-cutoff settings, their reliance on invariant features can limit the modeling of complex geometric relationships.

A second line of work develops SE(3)-equivariant neural networks, which operate on equivariant representations directly, often using spherical harmonics and tensor-product operations [65, 22]. This framework has been advanced by improved nonlinearities [6], efficient tensor products [37, 44], attention mechanisms [20, 42], and support for higher-order equivariant predictions [74]. Application-oriented models such as MACE [1, 2] and eSEN [19] further demonstrate strong empirical performance. Overall, these models provide strong geometric modeling power, but explicit equivariance can constrain neural network architectures and increase computational cost.

Thirdly, frame-based methods provide an alternative by introducing equivariant reference frames that canonicalize geometric data without explicit tensor algebra [59, 15]. They support both SE(3)-invariant and SE(3)-equivariant predictions [58, 33, 52]. Among them, global frames construct a single coordinate system for the entire structure, with developments from PCA-based designs to stochastic variants, minimal-frame methods, and lattice frames for crystalline systems [59, 17, 48, 71, 27]. In contrast, local frames build atom-specific coordinate systems from local geometry, offering greater expressivity but potentially losing relative orientation information [67, 26]. Early local frame designs relied on bond directions or average displacements [15, 16, 58, 73], while recent work introduces dynamic variants [29].

Currently, few methods combine local and global information in frame-based representations. Among them, [26] composes a learned invariant local transformation with an equivariant global frame; [28] uses local-frame edge encoding and a global-frame feed-forward module; and [49] exchanges tensorial messages between local frames and maps outputs back to the global coordinate system. These methods integrate frames through prescribed compositions or processing pipelines. In contrast, MFAN provides a general multi-frame framework that processes independently constructed frames in

parallel and learns atom-dependent weights to adaptively adjust their contributions. Continuity is another key concern in frame-based models: finite unweighted frames that preserve continuity are conjectured not to exist for SO(3) [18], motivating weighted frame ensembles [58, 18]. However, these approaches typically generate a large amount of frames from a single construction principle, leading to high computational cost and limited flexibility for heterogeneous local environments.

To improve computational efficiency, some recent studies on material property prediction relax strict SE(3) symmetry requirements by using data augmentation or regularization [55, 60, 53]. In force-field modeling, some methods also predict forces directly rather than deriving them from energy [43, 55]. These approaches reflect different trade-offs between computational efficiency and physical consistency. In this work, we focus on SE(3)-invariant energy modeling and conservative force fields.

5 Experiments

To comprehensively evaluate MFAN, we use crystal property prediction as the primary benchmark for assessing predictive accuracy, with force prediction serving as a supplementary probe of model stability. Additional ablation studies, stability assessments, results for alternative frame combinations, and analyses of computational complexity and efficiency are presented in Appendix C. Our code is publicly available at `https://github.com/IngramWang/mfan`.

5.1 Crystal property prediction

We evaluate the effectiveness of our multi-frame architecture on crystal property prediction tasks using two commonly used benchmarks, *JARVIS-DFT* [11] and *Materials Project (MP)* [30]. Following prior works, we use the same data splits for direct comparison with a wide range of baselines, from GNN models with invariant features such as CGCNN [69] and SchNet [61] to recent frame-based transformers like iComFormer [71], CrystalFramer [29] and SPFrame-based eComFormer [26]. To validate MFAN's applicability to different frame constructions and assess their effects, we evaluate five MFAN variants: lattice frame, PCA frame, QR frame, lattice + QR frames, and lattice + PCA frames. For the lattice + PCA setting, we additionally include a lattice-frame variant that incorporates neighbor directions, since lattice vectors alone may not provide a unique orientation in highly symmetric crystals such as FCC and BCC structures. Detailed frame implementations are presented in Appendix B.5. Experimental configurations, including dataset information and training and model hyperparameters, are provided in Appendix D.

Results on JARVIS-DFT are reported in Table 1. Even with a single frame, MFAN improves over existing methods on most tasks, suggesting that the proposed geometric feature extraction module can effectively incorporate frame-based information. More importantly, the multi-frame variants consistently outperform their single-frame counterparts. Overall, MFAN variants occupy the best and second-best positions on nearly all JARVIS-DFT tasks. Table 2 reports the MP results, where the multi-frame variants again outperform the single-frame ones and achieve competitive performance against strong baselines. The results of additional frame combinations, including pairs of local frames such as QR + PCA, are reported in Appendix C.4, which demonstrate that the benefits of the multi-frame architecture extend beyond global-local frame combinations. The complexity analysis is provided in Appendix C.5. Taking advantage of our efficient edge update design, with the same number of interaction blocks, MFAN uses fewer parameters than iComFormer. Since the multi-frame structure is used only for geometric feature extraction, adding more frames incurs only moderate computational overhead.

5.2 Stability and force prediction

We further examine the performance of the multi-frame architecture with frame-quality-aware weighting using S2EF tasks across three datasets as a complementary evaluation. Among them, *MPTrj-light* is sampled from MP trajectories and focuses on light-to-mid elements in inorganic materials [13]. We also include two datasets from realistic applications: *Cu/O*, which studies oxygen restructuring on Cu catalyst surfaces [41], and *CeO_2/Pt*, our self-collected dataset for Pt-cluster reconstruction on CeO_2 surfaces. The latter is particularly challenging due to localized Ce $4f$ electrons and the associated multiple electronic configurations. Since most baseline models in the previous subsection do not support force prediction, here we include two representative S2EF models

Table 1: Crystal property prediction results on the JARVIS-DFT dataset. Results are reported in mean absolute error (MAE). The training/validation/test set sizes and units are listed under each objective. Bold and underline indicate the best and second-best results, respectively.

Method	Form. energy 44578/5572/5572 eV/atom	Total energy 44578/5572/5572 eV/atom	Bandgap (OPT) 44578/5572/5572 eV	Bandgap (MBJ) 14537/1817/1817 eV	E hull 44296/5537/5537 eV
CGCNN [69]	0.063	0.078	0.20	0.41	0.17
SchNet [61]	0.045	0.047	0.19	0.43	0.14
MEGNet [9]	0.047	0.058	0.145	0.34	0.084
GATGNN [51]	0.047	0.056	0.17	0.51	0.12
ALIGNN [25]	0.0331	0.037	0.142	0.31	0.076
Matformer [70]	0.0325	0.035	0.137	0.30	0.064
PotNet [47]	0.0294	0.032	0.127	0.27	0.055
iComFormer [71]	0.0272	0.0288	0.122	0.26	0.047
eComFormer [71]	0.0284	0.0315	0.124	0.283	0.044
Crystalformer [64]	0.0306	0.0320	0.128	0.30	0.046
Crystalframer [29]	0.0263	0.0279	0.117	0.242	0.047
eComFormer with SPFrame [26]	0.0261	0.0276	**0.107**	0.239	0.042
MFAN (ours)					
- `lattice only`	0.0257	0.0270	0.124	0.258	0.040
- `PCA only`	0.0262	0.0281	0.130	0.260	0.039
- `QR only`	<u>0.0254</u>	0.0271	0.126	0.243	0.042
- `lattice + PCA`	<u>0.0254</u>	<u>0.0266</u>	<u>0.115</u>	**0.233**	**0.037**
- `lattice + QR`	**0.0250**	**0.0264**	0.118	<u>0.234</u>	<u>0.038</u>

Table 2: Crystal property prediction results on the MP dataset. Results are reported in mean absolute error (MAE). The training/validation/test set sizes and units are listed under each objective. Bold and underline indicate the best and second-best results, respectively.

Method	Form. energy 60000/5000/4239 eV/atom	Bandgap 60000/5000/4239 eV	Bulk modulus 4664/393/393 log(GPa)	Shear modulus 4664/392/393 log(GPa)
CGCNN [32]	0.031	0.292	0.047	0.077
SchNet [61]	0.033	0.345	0.066	0.099
MEGNet [9]	0.030	0.307	0.060	0.099
GATGNN [51]	0.033	0.280	0.045	0.075
ALIGNN [25]	0.022	0.218	0.051	0.078
Matformer [70]	0.021	0.211	0.043	0.073
PotNet [47]	0.0188	0.204	0.040	0.065
iComFormer [71]	0.0183	0.193	0.0380	<u>0.0637</u>
eComFormer [71]	0.0182	0.202	0.0417	0.0729
Crystalformer [64]	0.0186	0.198	0.0377	0.0689
Crystalframer [29]	0.0172	0.185	**0.0338**	0.0677
eComFormer with SPFrame [26]	0.0171	<u>0.181</u>	0.0371	0.0672
MFAN (ours)				
- `lattice only`	0.0167	0.189	0.0373	0.0674
- `PCA only`	0.0167	0.185	0.0386	0.0681
- `QR only`	<u>0.0163</u>	0.185	0.0383	0.0641
- `lattice + PCA`	0.0166	**0.179**	0.0355	0.0653
- `lattice + QR`	**0.0161**	0.184	<u>0.0340</u>	**0.0635**

of comparable scale as additional baselines: MACE [1, 2], an equivariant neural network, and AlphaNet [73], a frame-based model. We use these comparisons together with MFAN variants to evaluate the accuracy-stability trade-off of adaptive multi-frame learning. Besides energy and force errors, we report *%Force Outlier*, defined as the percentage of force components exceeding 10 eV/Å after applying Gaussian noise with $\sigma = 0.05$Å to atomic positions. This noise level is comparable to femtosecond-scale molecular-dynamics displacements, and the threshold conservatively captures anomalously large forces that are most likely to destabilize a simulation trajectory.

For force prediction, PCA frames are excluded because differentiating through them can cause numerical issues. As shown in Table 3, the lattice frame performs better on the crystalline *MPTrj-light* dataset, while the QR frame is more effective on the two surface datasets. In general, the multi-frame variant also outperforms its single-frame counterparts on accuracy. For %Force Outlier, all models perform relatively poorly on *MPTrj-light*, consistent with previous observations that current universal datasets may still insufficiently cover perturbed non-equilibrium structures [14]. On the two task-specific datasets, lattice-frame-based MFAN shows strong stability, while the lattice

+ QR variant maintains comparable stability and substantially improves accuracy, highlighting the advantage of multi-frame learning. More stability assessments including other force outlier thresholds and controlled perturbation tests can be found in Appendix C.2 and C.3. These results reveal potential failures of individual frames in symmetric structures and show that the quality-aware multi-frame architecture consistently maintains stable predictions.

Table 3: Results on S2EF tasks. Energy MAE is reported for total energy in *eV/atom*, and force MAE is reported for per-coordinate force components in *eV/Å*. %Force Outlier denotes the percentage of predicted force components exceeding 10 eV/Å after Gaussian position perturbations with $\sigma = 0.05$ Å. We repeat the random perturbation procedure five times and report the mean %Force Outlier. Bold and underline indicate the best and second-best results, respectively.

Method	MPTrj-light			Cu/O			CeO_2/Pt		
	Energy MAE	Force MAE	%Force Outlier	Energy MAE	Force MAE	%Force Outlier	Energy MAE	Force MAE	%Force Outlier
AlphaNet [73]	0.0167	0.0764	0.2169	0.0012	0.0362	0.0036	0.0060	0.1184	0.0093
MACE [1]	0.0341	0.0704	0.2058	0.0012	0.0272	0.0009	0.0117	0.0978	**0.0057**
MFAN (ours)									
- `lattice only`	**0.0125**	<u>0.0645</u>	**0.1970**	0.0017	0.0430	**0.0002**	<u>0.0057</u>	0.0974	<u>0.0060</u>
- `QR only`	0.0131	0.0649	0.2000	<u>0.0011</u>	<u>0.0256</u>	0.0006	0.0058	<u>0.0960</u>	0.0091
- `lattice + QR`	<u>0.0127</u>	**0.0631**	**0.1970**	**0.0010**	**0.0251**	<u>0.0003</u>	**0.0055**	**0.0944**	0.0064

6 Discussion

This work studies two key issues in frame-based material property prediction models: heterogeneous frame requirements and discontinuities near degenerate structures. To this end, the Multi-Frame Adaptive Network (MFAN) formulates frame usage as an adaptive combination problem. By dynamically weighting heterogeneous local and global frames conditioned on atomic environments, MFAN captures informative geometric references and complementary structural cues. Its frame-quality-aware mechanism further suppresses unreliable frames and promotes stable geometric representations near singular configurations. Experiments on crystal benchmarks demonstrate MFAN's strong predictive accuracy, broad applicability across frame constructions, and consistent gains from adaptive multi-frame learning, while continuity-sensitive force prediction tasks indicate improved stability.

MFAN still leaves room for improvement. While its SE(3)-invariant design based on distances and directions is effective, equivariant vector or tensor representations may be better suited to complex tasks. The lattice frame is stable for fixed lattice vectors, but NPT simulations require robust frame designs under varying lattice vectors. Beyond benchmark evaluation, MFAN can serve as a deployable component in atomistic simulation workflows, opening up opportunities for scientific discovery. Further limitations and future work are discussed in Appendix E.

Acknowledgments and Disclosure of Funding

Tongliang Liu is partially supported by the following Australian Research Council projects: DP260102466, FT220100318, DP220102121, LP220100527, LP220200949. Bo Han is supported by RGC Research Fellow Scheme No. RFS2627-2S02, RGC Young Collaborative Research Grant No. C2005-24Y, RGC General Research Fund No. 12202026, RGC General Research Fund No. 12200725, NSFC Major Research Plan No. 92570109, NSFC General Program No. 62376235, HKBU Faculty Niche Research Areas No. RC-FNRA-IG/25-26/SCI/05, and HKBU CSD Departmental Incentive Scheme. Jun Wang is partially supported by the JD Technology Scholarship for Postgraduate Research in Artificial Intelligence and the OpenAI Researcher Access Program. This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI Australia), an NCRIS enabled capability supported by the Australian Government. The authors acknowledge the use of the Sydney GPU Cluster, a service provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney. The authors also thank Dr. Suqin Yuan for providing additional computational resources.

References

[1] Ilyes Batatia, David Peter Kovacs, Gregor N. C. Simm, Christoph Ortner, and Gabor Csanyi. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*, 2022.

[2] Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M Elena, Dávid P Kovács, Janosh Riebesell, Xavier R Advincula, Mark Asta, Matthew Avaylon, William J Baldwin, et al. A foundation model for atomistic materials chemistry. *The Journal of chemical physics*, 163(18), 2025.

[3] Rohit Batra, Le Song, and Rampi Ramprasad. Emerging materials intelligence ecosystems propelled by machine learning. *Nature Reviews Materials*, 6(8):655–678, 2021.

[4] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters*, 98(14):146401, 2007.

[5] Peter E Blöchl. Projector augmented-wave method. *Physical review B*, 50(24):17953, 1994.

[6] Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J Bekkers, and Max Welling. Geometric and physical quantities improve E(3) equivariant message passing. *arXiv preprint arXiv:2110.02905*, 2021.

[7] Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, et al. Open catalyst 2020 (OC20) dataset and community challenges. *ACS Catalysis*, 11(10):6059–6072, 2021.

[8] Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2(11):718–728, 2022.

[9] Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31 (9):3564–3572, 2019.

[10] Kamal Choudhary and Brian DeCost. Atomistic line graph neural network for improved materials property predictions. *npj Computational Materials*, 7(1):185, 2021.

[11] Kamal Choudhary, Kevin F Garrity, Andrew CE Reid, Brian DeCost, Adam J Biacchi, Angela R Hight Walker, Zachary Trautt, Jason Hattrick-Simpers, A Gilad Kusne, Andrea Centrone, et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *npj computational materials*, 6(1):173, 2020.

[12] Sambit Das, Phani Motamarri, Vikram Gavini, Bruno Turcksin, Ying Wai Li, and Brent Leback. Fast, scalable and accurate finite-element based ab initio calculations using mixed precision computing: 46 PFLOPS simulation of a metallic dislocation system. In *International conference for high performance computing, networking, storage and analysis*, pages 1–11, 2019.

[13] Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J Bartel, and Gerbrand Ceder. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5(9):1031–1041, 2023.

[14] Bowen Deng, Yunyeong Choi, Peichen Zhong, Janosh Riebesell, Shashwat Anand, Zhuohan Li, KyuJung Jun, Kristin A Persson, and Gerbrand Ceder. Systematic softening in universal machine learning interatomic potentials. *npj Computational Materials*, 11(1):9, 2025.

[15] Weitao Du, He Zhang, Yuanqi Du, Qi Meng, Wei Chen, Nanning Zheng, Bin Shao, and Tie-Yan Liu. SE(3) equivariant graph neural networks with complete local frames. In *International Conference on Machine Learning*, pages 5583–5608, 2022.

[16] Yuanqi Du, Limei Wang, Dieqiao Feng, Guifeng Wang, Shuiwang Ji, Carla P Gomes, Zhi-Ming Ma, et al. A new perspective on building efficient and expressive 3D equivariant graph neural networks. *Advances in Neural Information Processing Systems*, 36:66647–66674, 2023.

[17] Alexandre Agm Duval, Victor Schmidt, Alex Hernández-Garcıa, Santiago Miret, Fragkiskos D Malliaros, Yoshua Bengio, and David Rolnick. FAENet: Frame averaging equivariant gnn for materials modeling. In *International Conference on Machine Learning*, pages 9013–9033, 2023.

[18] Nadav Dym, Hannah Lawrence, and Jonathan W Siegel. Equivariant frames and the impossibility of continuous canonicalization. In *International Conference on Machine Learning*, pages 12228–12267, 2024.

[19] Xiang Fu, Brandon M Wood, Luis Barroso-Luque, Daniel S Levine, Meng Gao, Misko Dzamba, and C Lawrence Zitnick. Learning smooth and expressive interatomic potentials for physical property prediction. In *International Conference on Machine Learning*, pages 17875–17893, 2025.

[20] Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. SE(3)-transformers: 3D roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems*, 33:1970–1981, 2020.

[21] Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020.

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

[23] Sophie Hermann and Matthias Schmidt. Why Noether's theorem applies to statistical mechanics. *Journal of Physics: Condensed Matter*, 34(21):213001, 2022.

[24] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves, Bjørk Hammer, Cory Hargus, et al. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.

[25] Tim Hsu, Tuan Anh Pham, Nathan Keilbart, Stephen Weitzner, James Chapman, Penghao Xiao, S Roger Qiu, Xiao Chen, and Brandon C Wood. Efficient and interpretable graph network representation for angle-dependent properties applied to optical spectroscopy. *npj Computational Materials*, 8(1):151, 2022.

[26] Haowei Hua and Wanyu Lin. Local-global associative frames for symmetry-preserving crystal structure modeling. *arXiv preprint arXiv:2505.15315*, 2025.

[27] Haowei Hua, Jingwen Yang, Wanyu Lin, and Pan Zhou. Revisiting the canonicalization for fast and accurate crystal tensor property prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 417–425, 2026.

[28] Tinglin Huang, Zhenqiao Song, Rex Ying, and Wengong Jin. Protein-nucleic acid complex modeling with frame averaging transformer. *Advances in Neural Information Processing Systems*, 37:127154–127180, 2024.

[29] Yusei Ito, Tatsunori Taniai, Ryo Igarashi, Yoshitaka Ushiku, and Kanta Ono. Rethinking the role of frames for SE(3)-invariant crystal structure modeling. In *International Conference on Learning Representations*, 2025.

[30] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

[31] Weile Jia, Han Wang, Mohan Chen, Denghui Lu, Lin Lin, Roberto Car, E Weinan, and Linfeng Zhang. Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning. In *International conference for high performance computing, networking, storage and analysis*, pages 1–14, 2020.

[32] Oumar Kaba and Siamak Ravanbakhsh. Equivariant networks for crystal structures. *Advances in Neural Information Processing Systems*, 35:4150–4164, 2022.

[33] Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, pages 15546–15566, 2023.

[34] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B*, 54(16):11169, 1996.

[35] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science*, 6(1):15–50, 1996.

[36] Georg Kresse and Daniel Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review b*, 59(3):1758, 1999.

[37] Seung Yul Lee, Hojoon Kim, Yutack Park, Dawoon Jeong, Seungwu Han, Yeonhong Park, and Jae W Lee. FlashTP: Fused, sparsity-aware tensor product for machine learning interatomic potentials. In *International Conference on Machine Learning*, 2025.

[38] Xiu-Chuan Li, Xiaobo Xia, Fei Zhu, Tongliang Liu, Xu-Yao Zhang, and Cheng-Lin Liu. Dynamics-aware loss for learning with label noise. *Pattern Recognition*, 144:109835, 2023.

[39] Xiuchuan Li and Tongliang Liu. Efficient and trustworthy causal discovery with latent variables and complex relations. In *International Conference on Learning Representations*, volume 2025, pages 34349–34383, 2025.

[40] Zian Li, Xiyuan Wang, Shijia Kang, and Muhan Zhang. On the completeness of invariant geometric deep learning models. *arXiv preprint arXiv:2402.04836*, 2024.

[41] Zan Lian, Federico Dattila, and Núria López. Stability and lifetime of diffusion-trapped oxygen in oxide-derived copper CO2 reduction electrocatalysts. *Nature Catalysis*, 7(4):401–411, 2024.

[42] Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3D atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.

[43] Yi-Lun Liao, Brandon Wood, Abhishek Das, and Tess Smidt. EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.

[44] Yi-Lun Liao, Alexander J Hoffman, Sabrina C Shen, Alexandre Duval, Sam Walton Norwood, and Tess Smidt. EquiformerV3: Scaling efficient, expressive, and general SE(3)-equivariant graph attention transformers. *arXiv preprint arXiv:2604.09130*, 2026.

[45] Yexiong Lin, Yu Yao, Zhaoqing Wang, Xu Shen, Jun Yu, Bo Han, and Tongliang Liu. Improving the instance-dependent transition matrix estimation by exploiting self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[46] Yexiong Lin, Yu Yao, Yuxuan Du, Jun Yu, Bo Han, Mingming Gong, and Tongliang Liu. Do we need to penalize variance of losses for learning with label noise? *Journal of Machine Learning Research*, 27(57):1–38, 2026.

[47] Yuchao Lin, Keqiang Yan, Youzhi Luo, Yi Liu, Xiaoning Qian, and Shuiwang Ji. Efficient approximations of complete interatomic potentials for crystal property prediction. In *International Conference on Machine Learning*, pages 21260–21287, 2023.

[48] Yuchao Lin, Jacob Helwig, Shurui Gui, and Shuiwang Ji. Equivariance via minimal frame averaging for more symmetries and efficiency. *arXiv preprint arXiv:2406.07598*, 2024.

[49] Peter Lippmann, Gerrit Gerhartz, Roman Remme, and Fred A Hamprecht. Beyond canonicalization: How tensorial messages improve equivariant message passing. In *International Conference on Learning Representations*, volume 2025, pages 88067–88087, 2025.

[50] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[51] Steph-Yves Louis, Yong Zhao, Alireza Nasiri, Xiran Wang, Yuqi Song, Fei Liu, and Jianjun Hu. Graph convolutional neural networks with global attention for improved materials property prediction. *Physical Chemistry Chemical Physics*, 22(32):18141–18148, 2020.

[52] George Ma, Yifei Wang, Derek Lim, Stefanie Jegelka, and Yisen Wang. A canonicalization perspective on invariant and equivariant learning. *Advances in Neural Information Processing Systems*, 37:60936–60979, 2024.

[53] Arslan Mazitov, Filippo Bigi, Matthias Kellner, Paolo Pegolo, Davide Tisi, Guillaume Fraux, Sergey Pozdnyakov, Philip Loche, and Michele Ceriotti. PET-MAD as a lightweight universal interatomic potential for advanced materials modeling. *Nature Communications*, 16(1):10653, 2025.

[54] Amil Merchant, Simon Batzner, Samuel S Schoenholz, Muratahan Aykol, Gowoon Cheon, and Ekin Dogus Cubuk. Scaling deep learning for materials discovery. *Nature*, 624(7990):80–85, 2023.

[55] Mark Neumann, James Gin, Benjamin Rhodes, Steven Bennett, Zhiyi Li, Hitarth Choubisa, Arthur Hussey, and Jonathan Godwin. Orb: A fast, scalable neural network potential. *arXiv preprint arXiv:2410.22570*, 2024.

[56] Cheol Woo Park and Chris Wolverton. Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Physical Review Materials*, 4(6): 063801, 2020.

[57] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[58] Sergey Pozdnyakov and Michele Ceriotti. Smooth, exact rotational symmetrization for deep learning on point clouds. *Advances in Neural Information Processing Systems*, 36:79469–79501, 2023.

[59] Omri Puny, Matan Atzmon, Heli Ben-Hamu, Ishan Misra, Aditya Grover, Edward J Smith, and Yaron Lipman. Frame averaging for invariant and equivariant network design. *arXiv preprint arXiv:2110.03336*, 2021.

[60] Benjamin Rhodes, Sander Vandenhaute, Vaidotas Šimkus, James Gin, Jonathan Godwin, Tim Duignan, and Mark Neumann. Orb-v3: atomistic simulation at scale. *arXiv preprint arXiv:2504.06231*, 2025.

[61] Kristof T Schütt, Huziel E Sauceda, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. SchNet–a deep learning architecture for molecules and materials. *The Journal of chemical physics*, 148(24), 2018.

[62] Leslie N Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, volume 11006, pages 369–386, 2019.

[63] Gilbert W. Stewart and Ji-guang Sun. *Matrix Perturbation Theory*. Academic Press, Boston, 1990. ISBN 0126702306.

[64] Tatsunori Taniai, Ryo Igarashi, Yuta Suzuki, Naoya Chiba, Kotaro Saito, Yoshitaka Ushiku, and Kanta Ono. Crystalformer: Infinitely connected attention for periodic structure encoding. *arXiv preprint arXiv:2403.11686*, 2024.

[65] 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.

[66] Feng Wang, Zebing Ma, and Jun Cheng. Accelerating computation of acidity constants and redox potentials for aqueous organic redox flow batteries by machine learning potential-based molecular dynamics. *Journal of the American Chemical Society*, 146(21):14566–14575, 2024.

[67] Xiyuan Wang and Muhan Zhang. Graph neural network with local frame for molecular potential energy surface. In *Learning on Graphs Conference*, pages 19–1. PMLR, 2022.

[68] Junfan Xia, Yaolong Zhang, and Bin Jiang. The evolution of machine learning potentials for molecules, reactions and materials. *Chemical Society Reviews*, 54(10):4790–4821, 2025.

[69] Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.

[70] Keqiang Yan, Yi Liu, Yuchao Lin, and Shuiwang Ji. Periodic graph transformers for crystal material property prediction. *Advances in Neural Information Processing Systems*, 35:15066–15080, 2022.

[71] Keqiang Yan, Cong Fu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Complete and efficient graph transformers for crystal material property prediction. *arXiv preprint arXiv:2403.11857*, 2024.

[72] Yu Yang, Jun Wang, Yaohui Shi, Xue Han, Yuhang Wang, Biswaranjan Mohanty, Yongxiang Liang, Mianqi Xue, Jie Zeng, Tongliang Liu, et al. Decoupling product selectivity in electrocatalytic CO2 reduction by steering the interfacial water structure. *Journal of the American Chemical Society*, 2026.

[73] Bangchen Yin, Jiaao Wang, Weitao Du, Pengbo Wang, Penghua Ying, Haojun Jia, Zisheng Zhang, Yuanqi Du, Carla Gomes, Chenru Duan, et al. AlphaNet: scaling up local-frame-based neural network interatomic potentials. *npj Computational Materials*, 11(1):332, 2025.

[74] Haiyang Yu, Zhao Xu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Efficient and equivariant graph networks for predicting quantum hamiltonian. In *International Conference on Machine Learning*, pages 40412–40424, 2023.

[75] Claudio Zeni, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Zilong Wang, Aliaksandra Shysheya, Jonathan Crabbé, Shoko Ueda, et al. A generative model for inorganic materials design. *Nature*, 639(8055):624–632, 2025.

A Proofs

In this section, we provide the proofs of all propositions in this article. For clarity, we restate each proposition before its proof, together with the necessary definitions and details. We also summarize the notation used throughout the paper in Table 4.

Symbol	Description
N	Number of atoms in a periodic cell.
$\boldsymbol{a}_i$	Atomic feature of atom i.
$\boldsymbol{A}$	Matrix of atomic features, $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N]$.
$\boldsymbol{p}_i$	Cartesian coordinates of atom i.
$\boldsymbol{P}$	Matrix of cartesian coordinates $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N]$.
$\boldsymbol{l}_i$	The i-th lattice vector of the periodic cell.
$\boldsymbol{L}$	Matrix of lattice vectors as $\boldsymbol{L} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3]$.
$\mathrm{SE}(3)$	The 3D special Euclidean group, representing all rotations and translations.
$\mathrm{SO}(3)$	The 3D special orthogonal group, representing all rotations
K	Number of frames.
$\mathcal{F}$	Frame (a single equivariant rotation matrix or a set thereof). Throughout this work, $\mathcal{F}$ is treated as a matrix; if it denotes a set, a matrix is randomly sampled from it.
$\mathcal{F}_i^{(k)}$	The k-th frame associated with atom i. It may be either a local frame constructed from the neighborhood of atom i, or a shared global frame.
$\mathcal{Q}_i^{(k)}$	The k-th frame quality score of atom i. It takes values in $[0, \infty)$; smaller values indicate that the frame is closer to discontinuity.
$\boldsymbol{d}_{ij}$	Displacement vector from atom i to j, $\boldsymbol{d}_{ij} = \boldsymbol{p}_j - \boldsymbol{p}_i$.
r_{ij}	Length of the displacement vector, $r_{ij} = \|\boldsymbol{d}_{ij}\|$.
$\hat{\boldsymbol{d}}_{ij}$	Normalized displacement vector, $\hat{\boldsymbol{d}}_{ij} = \boldsymbol{d}_{ij}/r_{ij}$.
$\boldsymbol{c}_{ij}^{(k)}$	Direction $\hat{\boldsymbol{d}}_{ij}$ expressed in frame $\mathcal{F}_i^{(k)}$, $\boldsymbol{c}_{ij}^{(k)} = (\mathcal{F}_i^{(k)})^\mathsf{T} \hat{\boldsymbol{d}}_{ij}$.
$w_{ij}^{(k)}$	Importance score of frame $\mathcal{F}_i^{(k)}$ for edge ij.
$\bar{w}_{ij}^{(k)}$	Normalized weight of frame $\mathcal{F}_i^{(k)}$ for edge ij according to $w_{ij}^{(k)}$ and $\mathcal{Q}_i^{(k)}$.
$\boldsymbol{r}_{ij}$	Length feature of edge ij from r_{ij}.
$\boldsymbol{\theta}_{ij}$	Aggregated directional feature of edge ij from all $\bar{w}_{ij}^{(k)}$ and $\boldsymbol{c}_{ij}^{(k)}$.
$\boldsymbol{x}_i$	SE(3)-invariant representation of atom i from GNN.
$\boldsymbol{e}_{ij}$	SE(3)-invariant representation of edge ij from GNN.
$\{p_i\}_{i=1}^m$	Neighbor set of atom p.
$\boldsymbol{Q} \in SO(3)$	Rotation matrix.
Σ_p	Covariance matrix constructed from neighbors of atom p.

Table 4: Summary of main symbols used in this work.

A.1 Proposition 1

Proposition. *Let p and q be two atoms with neighbor sets $\{p_i\}_{i=1}^m$ and $\{q_i\}_{i=1}^m$, respectively. Assume that*

$$\{\boldsymbol{d}_{qq_i}\}_{i=1}^m = \{\boldsymbol{Q}\boldsymbol{d}_{pp_i}\}_{i=1}^m$$

for some non-identity rotation $\boldsymbol{Q} \in SO(3)$, while

$$\{\boldsymbol{d}_{qq_i}\}_{i=1}^m \neq \{\boldsymbol{d}_{pp_i}\}_{i=1}^m.$$

Consider their frame sets $\{\mathcal{F}_p^{(1)}, \ldots, \mathcal{F}_p^{(K)}\}$ and $\{\mathcal{F}_q^{(1)}, \ldots, \mathcal{F}_q^{(K)}\}$. If the k-th frame is a permutation-invariant SE(3)-equivariant local frame, then

$$\{(r_{pp_i}, \boldsymbol{c}_{pp_i}^{(k)})\}_{i=1}^m = \{(r_{qq_i}, \boldsymbol{c}_{qq_i}^{(k)})\}_{i=1}^m,$$

i.e., such a local frame cannot distinguish p and q.

However, if there exists k such that the k-th frame is a non-degenerate shared global frame, i.e., $\mathcal{F}_p^{(k)} = \mathcal{F}_q^{(k)}$, then

$$\{[r_{pp_i}, \boldsymbol{c}_{pp_i}^{(1)}, \ldots, \boldsymbol{c}_{pp_i}^{(K)}]\}_{i=1}^m \neq \{[r_{qq_i}, \boldsymbol{c}_{qq_i}^{(1)}, \ldots, \boldsymbol{c}_{qq_i}^{(K)}]\}_{i=1}^m,$$

and thus the multi-frame representation can distinguish p and q.

Proof. We first show that a local frame cannot distinguish p and q. Let the k-th frame be a local frame constructed by a permutation-invariant SO(3)-equivariant rule. Since

$$\{\boldsymbol{d}_{qq_i}\}_{i=1}^m = \{\boldsymbol{Q}\boldsymbol{d}_{pp_i}\}_{i=1}^m,$$

the two local neighborhoods differ only by the rotation $\boldsymbol{Q}$. We may choose a corresponding ordering of the neighbors of q, i.e., a permutation π of $\{1, \ldots, m\}$, such that

$$\boldsymbol{d}_{qq_{\pi(i)}} = \boldsymbol{Q}\boldsymbol{d}_{pp_i}, \qquad i = 1, \ldots, m.$$

Accordingly, letting

$$\boldsymbol{D}_p = [\boldsymbol{d}_{pp_1}, \ldots, \boldsymbol{d}_{pp_m}], \qquad \boldsymbol{D}_q = [\boldsymbol{d}_{qq_{\pi(1)}}, \ldots, \boldsymbol{d}_{qq_{\pi(m)}}],$$

we have

$$\boldsymbol{D}_q = \boldsymbol{Q}\boldsymbol{D}_p.$$

Notice that for a permutation-invariant frame construction, the resulting frame is independent of the specific ordering. Let the k-th local frame be generated by an SE(3)-equivariant frame construction rule applied to the local neighborhood matrix. We have

$$\mathcal{F}_q^{(k)} = \mathcal{F}^{(k)}(\boldsymbol{D}_q) = \mathcal{F}^{(k)}(\boldsymbol{Q}\boldsymbol{D}_p) = \boldsymbol{Q}\mathcal{F}^{(k)}(\boldsymbol{D}_p) = \boldsymbol{Q}\mathcal{F}_p^{(k)}.$$

Then for each neighbor,

$$r_{qq_{\pi(i)}} = \|\boldsymbol{d}_{qq_{\pi(i)}}\| = \|\boldsymbol{Q}\boldsymbol{d}_{pp_i}\| = \|\boldsymbol{d}_{pp_i}\| = r_{pp_i},$$

$$\boldsymbol{c}_{qq_{\pi(i)}}^{(k)} = (\mathcal{F}_q^{(k)})^\mathsf{T}\hat{\boldsymbol{d}}_{qq_{\pi(i)}} = (\boldsymbol{Q}\mathcal{F}_p^{(k)})^\mathsf{T}(\boldsymbol{Q}\hat{\boldsymbol{d}}_{pp_i}) = (\mathcal{F}_p^{(k)})^\mathsf{T}\hat{\boldsymbol{d}}_{pp_i} = \boldsymbol{c}_{pp_i}^{(k)}.$$

Therefore,

$$\{(r_{pp_i}, \boldsymbol{c}_{pp_i}^{(k)})\}_{i=1}^m = \{(r_{qq_i}, \boldsymbol{c}_{qq_i}^{(k)})\}_{i=1}^m.$$

Hence, such a local frame cannot distinguish p and q.

We now consider the case where there exists k such that the k-th frame is a shared global frame, i.e.,

$$\mathcal{F}_p^{(k)} = \mathcal{F}_q^{(k)}.$$

Since this frame is non-degenerate (equivalently, $\mathcal{F}_p^{(k)}$ is full-rank), the mapping

$$\boldsymbol{d} \mapsto (r, \boldsymbol{c}^{(k)}) = \left(\|\boldsymbol{d}\|, (\mathcal{F}_p^{(k)})^\mathsf{T}\frac{\boldsymbol{d}}{\|\boldsymbol{d}\|}\right)$$

is invertible. Therefore,

$$\{\boldsymbol{d}_{qq_i}\}_{i=1}^m \neq \{\boldsymbol{d}_{pp_i}\}_{i=1}^m \quad \Longrightarrow \quad \{(r_{qq_i}, \boldsymbol{c}_{qq_i}^{(k)})\}_{i=1}^m \neq \{(r_{pp_i}, \boldsymbol{c}_{pp_i}^{(k)})\}_{i=1}^m.$$

That is, the coordinate sets under the shared global frame are different.

Thus, since a part of the concatenated multi-frame coordinates is already different, we further have

$$\{[r_{pp_i}, \boldsymbol{c}_{pp_i}^{(1)}, \ldots, \boldsymbol{c}_{pp_i}^{(K)}]\}_{i=1}^m \neq \{[r_{qq_i}, \boldsymbol{c}_{qq_i}^{(1)}, \ldots, \boldsymbol{c}_{qq_i}^{(K)}]\}_{i=1}^m.$$

Therefore, the multi-frame representation can distinguish p and q. □

Remarks. The above proposition highlights a limitation of local frames: they cannot distinguish neighborhoods that are related by a rotation. In principle, a GNN may still differentiate such atoms through message passing if their environments differ beyond the local neighborhood. However, this ability is fundamentally limited by the exponentially decaying influence of distant nodes in practical architectures. Moreover, when p and q are related by a global symmetry of the structure—e.g., occupying the same Wyckoff position in a crystal—even global message passing cannot break this equivalence, and local-frame-based representations remain indistinguishable [26].

A.2 Proposition 2

Proposition. *Let p be an atom with neighbor set $\{p_1, \dots, p_m\}$. If there exist constants $d_{\min}, d_{\max}, \varepsilon > 0$ such that*

$$d_{\min} < \|\boldsymbol{d}_{pp_i}\| < d_{\max}, \qquad \forall\, 1 \le i \le m,$$

and

$$\|\boldsymbol{d}_{pp_i} \times \boldsymbol{d}_{pp_j}\| \ge \varepsilon, \qquad \forall\, 1 \le i \ne j \le m,$$

then the neighbor-pair frame of p, defined as

$$\mathcal{F}_p = \left\{ \mathcal{F}_p^{ij} \;\middle|\; \mathcal{F}_p^{ij} = [\boldsymbol{e}_p^{ij,1}, \boldsymbol{e}_p^{ij,2}, \boldsymbol{e}_p^{ij,3}],\ 1 \le i \ne j \le m \right\},$$

where

$$\hat{\boldsymbol{d}}_{pp_i} = \frac{\boldsymbol{d}_{pp_i}}{\|\boldsymbol{d}_{pp_i}\|},$$

$$\boldsymbol{e}_p^{ij,1} = \hat{\boldsymbol{d}}_{pp_i}, \qquad \boldsymbol{e}_p^{ij,2} = \frac{\hat{\boldsymbol{d}}_{pp_j} - (\hat{\boldsymbol{d}}_{pp_j}^{\mathsf{T}} \hat{\boldsymbol{d}}_{pp_i})\hat{\boldsymbol{d}}_{pp_i}}{\|\hat{\boldsymbol{d}}_{pp_j} - (\hat{\boldsymbol{d}}_{pp_j}^{\mathsf{T}} \hat{\boldsymbol{d}}_{pp_i})\hat{\boldsymbol{d}}_{pp_i}\|}, \qquad \boldsymbol{e}_p^{ij,3} = \boldsymbol{e}_p^{ij,1} \times \boldsymbol{e}_p^{ij,2},$$

depends continuously on the atomic positions.

Proof. Notice that $\boldsymbol{d}_{pp_i}$ depends continuously on the atomic positions for each i. Since $\|\boldsymbol{d}_{pp_i}\| \ge d_{\min} > 0$, the normalization

$$\boldsymbol{d}_{pp_i} \mapsto \hat{\boldsymbol{d}}_{pp_i} = \frac{\boldsymbol{d}_{pp_i}}{\|\boldsymbol{d}_{pp_i}\|}$$

is also continuous. Hence each unit direction $\hat{\boldsymbol{d}}_{pp_i}$ depends continuously on the atomic positions.

Fix any pair $1 \le i \ne j \le m$. By assumption,

$$\|\boldsymbol{d}_{pp_i} \times \boldsymbol{d}_{pp_j}\| \ge \varepsilon > 0.$$

Therefore,

$$\|\hat{\boldsymbol{d}}_{pp_i} \times \hat{\boldsymbol{d}}_{pp_j}\| = \frac{\|\boldsymbol{d}_{pp_i} \times \boldsymbol{d}_{pp_j}\|}{\|\boldsymbol{d}_{pp_i}\|\,\|\boldsymbol{d}_{pp_j}\|} \ge \frac{\varepsilon}{d_{\max}^2}.$$

The vector $\hat{\boldsymbol{d}}_{pp_j} - (\hat{\boldsymbol{d}}_{pp_j}^{\mathsf{T}} \boldsymbol{e}_p^{ij,1})\boldsymbol{e}_p^{ij,1}$ is the component of $\hat{\boldsymbol{d}}_{pp_j}$ orthogonal to $\hat{\boldsymbol{d}}_{pp_i}$. Hence,

$$\left\| \hat{\boldsymbol{d}}_{pp_j} - (\hat{\boldsymbol{d}}_{pp_j}^{\mathsf{T}} \boldsymbol{e}_p^{ij,1})\boldsymbol{e}_p^{ij,1} \right\|^2 = \|\hat{\boldsymbol{d}}_{pp_j}\|^2 - (\hat{\boldsymbol{d}}_{pp_j}^{\mathsf{T}} \hat{\boldsymbol{d}}_{pp_i})^2 = 1 - (\hat{\boldsymbol{d}}_{pp_j}^{\mathsf{T}} \hat{\boldsymbol{d}}_{pp_i})^2.$$

On the other hand, since both $\hat{\boldsymbol{d}}_{pp_i}$ and $\hat{\boldsymbol{d}}_{pp_j}$ are unit vectors,

$$\|\hat{\boldsymbol{d}}_{pp_i} \times \hat{\boldsymbol{d}}_{pp_j}\|^2 = \|\hat{\boldsymbol{d}}_{pp_i}\|^2 \|\hat{\boldsymbol{d}}_{pp_j}\|^2 - (\hat{\boldsymbol{d}}_{pp_i}^{\mathsf{T}} \hat{\boldsymbol{d}}_{pp_j})^2 = 1 - (\hat{\boldsymbol{d}}_{pp_i}^{\mathsf{T}} \hat{\boldsymbol{d}}_{pp_j})^2.$$

Therefore, we have

$$\left\| \hat{\boldsymbol{d}}_{pp_j} - (\hat{\boldsymbol{d}}_{pp_j}^{\mathsf{T}} \boldsymbol{e}_p^{ij,1})\boldsymbol{e}_p^{ij,1} \right\| = \|\hat{\boldsymbol{d}}_{pp_i} \times \hat{\boldsymbol{d}}_{pp_j}\|,$$

the denominator in the definition of $\boldsymbol{e}_p^{ij,2}$ is bounded away from zero by $\varepsilon / d_{\max}^2$. As a result, $\boldsymbol{e}_p^{ij,2}$ is obtained from continuous operations (subtraction, inner product, and normalization away from zero), and is thus continuous.

Finally, $\boldsymbol{e}_p^{ij,3} = \boldsymbol{e}_p^{ij,1} \times \boldsymbol{e}_p^{ij,2}$ is also continuous, since the cross product is continuous. Therefore, for every ordered pair $i \ne j$, the frame $\mathcal{F}_p^{ij}$ depends continuously on the atomic positions. Since $\mathcal{F}_p$ consists of finitely many such frames, the entire neighbor-pair frame set $\mathcal{F}_p$ also depends continuously on the atomic positions. □

Remarks. (1) Here, the continuity of a frame set $\mathcal{F}$ is understood through frame averaging: for any (fixed) continuous function f, the quantity

$$\frac{1}{|\mathcal{F}|}\sum_{F\in\mathcal{F}} f(\boldsymbol{A}, F^{\mathsf{T}}\boldsymbol{P}, F^{\mathsf{T}}\boldsymbol{L})$$

is continuous with respect to the atomic positions and lattice vectors. In this proposition, since the cardinality of $\mathcal{F}_p$ is fixed, it reduces to the continuity of each individual frame $\mathcal{F}_p^{ij}$.

(2) The above proposition is stated for one specific construction of the neighbor-pair frame. In practice, other definitions are also possible. For example, one may define the second or third axis using the cross product of two neighbor directions, such as $\boldsymbol{d}_{pp_i} \times \boldsymbol{d}_{pp_j}$. Nevertheless, the underlying principle is the same: as long as the two selected neighbor vectors span a non-degenerate plane, the resulting frame is well-defined and depends continuously on the atomic positions.

(3) In this proposition, we assume that the neighbor set of p is fixed. If the neighbor set is not fixed but determined by a cutoff radius r_c, the same conclusion still holds locally provided no atom lies in the annulus $[r_c - \delta,\ r_c + \delta]$ for some $\delta > 0$. In that case, the neighbor set remains unchanged under sufficiently small perturbations, and the above argument applies.

A.3 Proposition 3

Proposition. *Let p be an atom with a fixed neighbor set $\{p_1, \ldots, p_m\}$, and let*

$$\Sigma_p = \sum_{i=1}^{m} \boldsymbol{d}_{pp_i}\boldsymbol{d}_{pp_i}^{\mathsf{T}} \in \mathbb{R}^{3\times 3}$$

be the covariance matrix constructed from the displacement vectors. Let

$$\lambda_1 > \lambda_2 > \lambda_3$$

be the eigenvalues of Σ_p, with corresponding unit eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$, and assume that there exists a constant $\delta > 0$ such that

$$\min_{i\neq j} |\lambda_i - \lambda_j| \geq \delta.$$

Define the (sign-invariant) PCA frame of p as

$$\mathcal{F}_p = \{[\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \hat{\boldsymbol{v}}_3] \mid \hat{\boldsymbol{v}}_i \in \{\boldsymbol{v}_i, -\boldsymbol{v}_i\},\ i = 1, 2, 3\}\,.$$

Then $\mathcal{F}_p$ depends continuously on the atomic positions.

Proof. Since $\boldsymbol{d}_{pp_i}$ depends continuously on the atomic positions for each i, the covariance matrix

$$\Sigma_p = \sum_{i=1}^{m} \boldsymbol{d}_{pp_i}\boldsymbol{d}_{pp_i}^{\mathsf{T}}$$

also depends continuously on the atomic positions. Consider a perturbation of the atomic positions, and let

$$\tilde{\Sigma}_p = \Sigma_p + E$$

be the corresponding perturbed covariance matrix. Since Σ_p depends continuously on the atomic positions, we have $\|E\| \to 0$ as the perturbation of the atomic positions tends to zero. By assumption,

$$\min_{i\neq j} |\lambda_i - \lambda_j| \geq \delta > 0,$$

so the three eigenvalues are distinct. Since Σ_p is a real symmetric matrix, its three distinct eigenvalues correspond to three mutually orthogonal one-dimensional eigenspaces. Moreover, according to the Weyl's inequality, the eigengap condition ensures that under sufficiently small perturbations, these eigenspaces remain uniquely associated with the eigenvalues near $\lambda_1, \lambda_2, \lambda_3$.

By the Davis-Kahan theorem [63], for each $i \in \{1, 2, 3\}$, if $\tilde{\boldsymbol{v}}_i$ is a unit eigenvector of $\tilde{\Sigma}_p$ associated with the perturbed eigenvalue near λ_i, then

$$\sin\angle\big(\mathrm{span}(\boldsymbol{v}_i), \mathrm{span}(\tilde{\boldsymbol{v}}_i)\big) \leq \frac{\|E\|}{\delta_i'},$$

where δ_i' denotes the spectral separation between λ_i and the perturbed eigenvalues associated with the other eigenspaces, i.e.,

$$\delta_i' = \min_{j \neq i} |\lambda_i - \tilde{\lambda}_j|.$$

By Weyl's inequality, after ordering the perturbed eigenvalues $\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3$ according to the eigenvalues they perturb from, we have

$$|\tilde{\lambda}_j - \lambda_j| \leq \|E\|.$$

Therefore, for $j \neq i$,

$$|\lambda_i - \tilde{\lambda}_j| \geq |\lambda_i - \lambda_j| - |\lambda_j - \tilde{\lambda}_j| \geq \delta - \|E\|.$$

For sufficiently small perturbations such that $\|E\| < \delta/2$, we have

$$\delta_i' \geq \delta/2.$$

Consequently,

$$\sin \angle\big(\mathrm{span}(\boldsymbol{v}_i), \mathrm{span}(\tilde{\boldsymbol{v}}_i)\big) \leq \frac{2\|E\|}{\delta}.$$

Since δ is fixed and $\|E\| \to 0$, it follows that

$$\sin \angle\big(\mathrm{span}(\boldsymbol{v}_i), \mathrm{span}(\tilde{\boldsymbol{v}}_i)\big) \to 0.$$

as the perturbation of the atomic positions tends to zero. Hence each eigenspace $\mathrm{span}(\boldsymbol{v}_i)$ depends continuously on the atomic positions.

The remaining ambiguity is the sign of each eigenvector, since $\boldsymbol{v}_i$ and $-\boldsymbol{v}_i$ span the same one-dimensional eigenspace. As a result, a specific ordered basis $[\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3]$ may exhibit discontinuous sign changes, even though the underlying eigenspaces vary continuously. However, the sign-invariant PCA frame

$$\mathcal{F}_p = \{[\hat{\boldsymbol{v}}_1, \hat{\boldsymbol{v}}_2, \hat{\boldsymbol{v}}_3] \mid \hat{\boldsymbol{v}}_i \in \{\boldsymbol{v}_i, -\boldsymbol{v}_i\},\ i = 1, 2, 3\}$$

depends only on these three one-dimensional eigenspaces and not on a particular sign choice of the eigenvectors. Therefore, sign flips of individual eigenvectors do not change the frame itself. Thus, since the eigenspaces depend continuously on the atomic positions, and $\mathcal{F}_p$ is uniquely determined by them, it follows that the PCA frame $\mathcal{F}_p$ also depends continuously on the atomic positions. □

Remarks. For notational simplicity, the PCA frame above is defined up to independent sign choices of the three eigenvectors, and is therefore $E(3)$-equivariant rather than strictly $SE(3)$-equivariant, since it does not distinguish reflections. If an $SE(3)$-equivariant frame is desired, one can fix the handedness by adjusting the sign of the third axis so that the resulting matrix has determinant one. After this modification, we enumerate the signs of the first two frame vectors, while the sign of the third frame vector is determined by the first two. Therefore, the resulting frame set remains invariant to the arbitrary signs returned by the PCA decomposition, and the continuity argument is preserved.

A.4 Proposition 4

Proposition. *Let $\mathcal{F}^{(k)}(\boldsymbol{P})$ be a frame constructed from the configuration $\boldsymbol{P}$ for $k = 1, \ldots, K$, and*

$$g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$$

be the representation induced by the k-th frame, where g_k is continuous in both arguments. Assume that each $g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$ is SO(3)-invariant (or equivariant), and let $w_k(\boldsymbol{P})$ be continuous, SO(3)-invariant scalar weights. Define

$$H(\boldsymbol{P}) = \sum_{k=1}^{K} w_k(\boldsymbol{P})\, g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P})).$$

For each k, suppose that:

- *$\mathcal{F}^{(k)}(\boldsymbol{P})$ is continuous outside a closed singular set Σ_k;*
- *there exists $\varepsilon_k > 0$ such that*

$$w_k(\boldsymbol{P}) = 0 \qquad \textit{whenever } \mathrm{dist}(\boldsymbol{P}, \Sigma_k) \leq \varepsilon_k.$$

Then $H(\boldsymbol{P})$ is continuous. Moreover, if each $g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$ is SO(3)-invariant, then $H(\boldsymbol{P})$ is SO(3)-invariant; if each $g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$ is SO(3)-equivariant, then $H(\boldsymbol{P})$ is SO(3)-equivariant.

Proof. We first prove continuity. Fix $k \in \{1, \dots, K\}$ and define

$$h_k(\boldsymbol{P}) = w_k(\boldsymbol{P})\, g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P})).$$

Then

$$H(\boldsymbol{P}) = \sum_{k=1}^{K} h_k(\boldsymbol{P}).$$

It therefore suffices to show that each h_k is continuous. Define

$$U_k := \{\boldsymbol{P} : \operatorname{dist}(\boldsymbol{P}, \Sigma_k) > \varepsilon_k\}.$$

Since Σ_k is closed, U_k is open. Moreover, every $\boldsymbol{P} \in U_k$ lies outside Σ_k, so $\mathcal{F}^{(k)}(\boldsymbol{P})$ is continuous on U_k. By assumption, g_k is continuous in both arguments, hence

$$g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$$

is continuous on U_k. Since $w_k(\boldsymbol{P})$ is also continuous, $h_k(\boldsymbol{P})$ is continuous on U_k.

Now consider the closed set

$$N_k := \{\boldsymbol{P} : \operatorname{dist}(\boldsymbol{P}, \Sigma_k) \le \varepsilon_k\}.$$

By assumption, $w_k(\boldsymbol{P}) = 0$ on N_k, so

$$h_k(\boldsymbol{P}) = 0 \qquad \text{for all } \boldsymbol{P} \in N_k.$$

Hence h_k is continuous on N_k.

It remains to check continuity at boundary points satisfying

$$\operatorname{dist}(\boldsymbol{P}_0, \Sigma_k) = \varepsilon_k.$$

At such a point, $\boldsymbol{P}_0 \notin \Sigma_k$ because $\varepsilon_k > 0$, so $\mathcal{F}^{(k)}$ is continuous at $\boldsymbol{P}_0$, and therefore

$$g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$$

is continuous, hence locally bounded, near $\boldsymbol{P}_0$. Meanwhile, $w_k(\boldsymbol{P}_0) = 0$, and w_k is continuous, so $w_k(\boldsymbol{P}) \to 0$ as $\boldsymbol{P} \to \boldsymbol{P}_0$. Therefore,

$$h_k(\boldsymbol{P}) = w_k(\boldsymbol{P})\, g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P})) \to 0 = h_k(\boldsymbol{P}_0).$$

Thus h_k is continuous at every boundary point, and hence continuous on the whole configuration space. Since H is a finite sum of continuous functions h_k, it follows that H is continuous.

We next prove SO(3)-invariance or equivariance. If each

$$g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$$

is SO(3)-invariant, then for any $\boldsymbol{Q} \in \mathrm{SO}(3)$,

$$H(\boldsymbol{Q}\boldsymbol{P}) = \sum_{k=1}^{K} w_k(\boldsymbol{Q}\boldsymbol{P})\, g_k(\boldsymbol{Q}\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{Q}\boldsymbol{P})) = \sum_{k=1}^{K} w_k(\boldsymbol{P})\, g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P})) = H(\boldsymbol{P}),$$

where we used the SO(3)-invariance of both w_k and g_k.

If instead each

$$g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$$

is SO(3)-equivariant, then

$$\begin{aligned} H(\boldsymbol{Q}\boldsymbol{P}) &= \sum_{k=1}^{K} w_k(\boldsymbol{Q}\boldsymbol{P})\, g_k(\boldsymbol{Q}\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{Q}\boldsymbol{P})) \\ &= \sum_{k=1}^{K} w_k(\boldsymbol{P})\, \mathcal{M}(\boldsymbol{Q}) g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P})) \\ &= \mathcal{M}(\boldsymbol{Q}) H(\boldsymbol{P}), \end{aligned}$$

so H is SO(3)-equivariant.

This completes the proof. □

Remarks. (1) For simplicity, the notation here is slightly different from the full form $f(\boldsymbol{A}, \boldsymbol{P}, \boldsymbol{L})$. Since the atomic features $\boldsymbol{A}$ are irrelevant to the present continuity and symmetry arguments, they may be absorbed into g_k. The lattice vectors $\boldsymbol{L}$ may either be absorbed into $\boldsymbol{P}$, so that continuity is understood jointly with respect to atomic coordinates and lattice vectors, or be absorbed into g_k and $\mathcal{F}^{(k)}$. The latter corresponds to the usual construction in periodic systems, where one first computes relative displacement vectors from the atomic coordinates and lattice vectors, and then performs the subsequent frame construction and representation computation based on these displacements.

(2) Strictly speaking, $g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P}))$ need not be evaluated on Σ_k, where the frame may be undefined or multi-valued. Correspondingly, we interpret each weighted frame-induced term by its continuous extension:

$$h_k(\boldsymbol{P}) = \begin{cases} w_k(\boldsymbol{P}) g_k(\boldsymbol{P}, \mathcal{F}^{(k)}(\boldsymbol{P})), & \boldsymbol{P} \notin N_k, \\ 0, & \boldsymbol{P} \in N_k, \end{cases}$$

where N_k is the neighborhood of Σ_k on which $w_k(\boldsymbol{P}) = 0$. Under this interpretation, the singular frame-induced representation is never evaluated inside N_k, and h_k is locally constant there. Therefore, its derivative also vanishes inside N_k, which in principle suppresses not only discontinuities in the forward representation but also anomalous derivative contributions near singular frames. In practical automatic differentiation frameworks, however, this ideal short-circuit evaluation is not always realized. Even when the outer quality weight is zero, the internal frame-construction branch may still be evaluated during backpropagation. If this branch contains numerically ill-conditioned operations near singular configurations, unstable gradients can still arise as an implementation-level issue. Nevertheless, quality-aware weighting remains an effective mechanism for suppressing unreliable frame contributions and their derivative effects, provided that the underlying frame construction can be evaluated stably.

(3) This proposition shows that a possible discontinuity in a frame-induced representation can be avoided by suppressing its contribution before the configuration reaches the singular set. Requiring w_k to vanish on an ε_k-neighborhood of Σ_k, rather than only on Σ_k itself, prevents any uncontrolled behavior of $\mathcal{F}^{(k)}$ from surviving in the weighted aggregation. In practice, the singular sets Σ_k are typically induced by geometric degeneracies, such as collinearity, rank deficiency, or repeated eigenvalues, and are usually of measure zero.

B Details of the MFAN architecture

In this section, we provide a detailed description of the MFAN architecture. Following the main text, we first describe the processing of geometric features, then introduce the backbone network, including node and edge updates, and finally present the output layer. At the end of this section, we also provide implementation details of the frame constructions used in our experiments, along with their corresponding quality scores.

B.1 Input embedding and geometric features

MFAN is implemented on top of a graph neural network, where atoms are treated as nodes and neighboring atom pairs within a cutoff are connected by edges. The initial node representation $\boldsymbol{x}_i$ is obtained by a linear embedding of the input atomic feature $\boldsymbol{A}$. In our implementation, this feature may be either a pre-computed vector or a one-hot encoding of the atomic number, followed by a linear projection to the hidden dimension d.

For edges, given an atom pair (i, j), we first enumerate all periodic images of atom j that fall within the cutoff radius of atom i using lattice vectors. For each such image $\boldsymbol{m} \in \mathbb{Z}^3$, we construct an edge and compute the relative displacement

$$\boldsymbol{d}_{ij(\boldsymbol{m})} = \boldsymbol{p}_j - \boldsymbol{p}_i + \boldsymbol{L}\boldsymbol{m}.$$

For notational simplicity, although multiple edges may exist between i and j, we use ij to denote these edges collectively and $\boldsymbol{d}_{ij}$ to represent their displacement vectors. Note that message passing in the network is bidirectional; therefore, if ij is an edge, then ji is also included as an edge. The radial feature is constructed in two steps. Let $r_{ij} = \|\boldsymbol{d}_{ij}\|$. First, the scalar $-0.75/r_{ij}$ is expanded by a Gaussian radial basis layer

$$\mathrm{RBF}_k(r) = \exp\big(-\gamma_k(r - \mu_k)^2\big),$$

and we denote the resulting vector as

$$\mathrm{RBF}(r) = \big[\mathrm{RBF}_1(r), \ldots, \mathrm{RBF}_K(r)\big]^\mathsf{T},$$

where each component corresponds to a different basis function indexed by k with uniformly spaced centers. This is followed by a linear projection and a $\mathrm{Softplus}$ activation function:

$$\boldsymbol{r}^*_{ij} = \mathrm{Softplus}(\boldsymbol{W}_r\,\mathrm{RBF}(-0.75/r_{ij}) + \boldsymbol{b}_r)\,.$$

Second, when distance smoothing is enabled, the resulting feature is multiplied by the polynomial envelope [21] with $p = 6$:

$$u(d) = 1 - \frac{(p+1)(p+2)}{2}d^p + p(p+2)d^{p+1} - \frac{p(p+1)}{2}d^{p+2},$$

where d is the rescaled distance clipped to $[0, 1]$, so that the radial feature smoothly decays to zero near the cutoff. The final radial feature is $\boldsymbol{r}_{ij} = u(r_{ij})\boldsymbol{r}^*_{ij}$.

For directional features, we first normalize the displacement to $\hat{\boldsymbol{d}}_{ij} = \boldsymbol{d}_{ij}/\|\boldsymbol{d}_{ij}\|$. For each frame $\mathcal{F}^{(k)}_i$, we first express the direction $\hat{\boldsymbol{d}}_{ij}$ in its coordinate system as

$$\boldsymbol{c}^{(k)}_{ij} = (\mathcal{F}^{(k)}_i)^\mathsf{T}\hat{\boldsymbol{d}}_{ij}$$

Then, we expand each element in $\boldsymbol{c}^{(k)}_{ij}$ by another Gaussian basis layer on $[-1, 1]$, followed by a linear layer and $\mathrm{Softplus}$. This gives a directional embedding $\boldsymbol{\Theta}^{(k)}_{ij}$ for each frame.

B.2 Node update

The node update follows the Comformer-style interaction block [71]. Let $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ be the current node representations and let $\boldsymbol{e}_{ij}$ be the current edge representation. For each edge, MFAN computes

$$\boldsymbol{q}_i = W_q\boldsymbol{x}_i, \qquad \boldsymbol{k}_i = W_k\boldsymbol{x}_i, \qquad \boldsymbol{k}_j = W_k\boldsymbol{x}_j, \qquad \boldsymbol{v}_i = W_v\boldsymbol{x}_i, \qquad \boldsymbol{v}_j = W_v\boldsymbol{x}_j,$$

and linearly projects the edge feature as $\tilde{\boldsymbol{e}}_{ij} = W_e\boldsymbol{e}_{ij}$. The projected edge feature is fused with the key and value through two-layer MLP,

$$\tilde{\boldsymbol{k}}_{ij} = \psi_k([\boldsymbol{k}_i, \boldsymbol{k}_j, \tilde{\boldsymbol{e}}_{ij}]), \qquad \tilde{\boldsymbol{v}}_{ij} = \psi_v([\boldsymbol{v}_i, \boldsymbol{v}_j, \tilde{\boldsymbol{e}}_{ij}]),$$

where ψ_k, ψ_v consists of $\mathrm{Linear\text{-}SiLU\text{-}Linear}$. Then the message along edge ij is then formed as

$$\boldsymbol{m}_{ij} = \tilde{\boldsymbol{v}}_{ij} \circ \sigma\Big(\mathrm{norm}\Big(\boldsymbol{q}_i \circ \tilde{\boldsymbol{k}}_{ij}\Big)\Big)\,,$$

where $\circ$ denotes the Hadamard product, σ denotes the sigmoid gate and $\mathrm{norm}(\cdot)$ is implemented as either batch normalization or layer normalization, depending on the configuration. Layer normalization is used when the batch size is small or in S2EF tasks, where batch normalization may introduce non-physical correlations. Messages are summed over neighbors and added back to the input through a residual $\mathrm{Softplus}$ update. This block is repeated over multiple interaction layers.

$$\boldsymbol{x}^{\mathrm{new}}_i = \mathrm{Softplus}\Big(\boldsymbol{x}_i + \mathrm{Agg}_{j\in\mathcal{N}(i)}\boldsymbol{m}_{ij}\Big)\,.$$

In our implementation, the aggregation is a simple summation over neighboring edges.

B.3 Edge update and multi-frame aggregation

For edge updates, we first form an intermediate edge representation

$$\boldsymbol{e}^*_{ij} = \phi_e([\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{e}_{ij}]),$$

where ϕ_e is implemented by a linear layer followed by SiLU. This intermediate representation is used both as the routing context for multi-frame aggregation and as part of the final edge update.

Let $\boldsymbol{\Theta}^{(k)}_{ij} \in \mathbb{R}^{3\times d}$ be the directional embedding extracted from the k-th frame. MFAN supports two frame aggregation networks $\phi^{(k)}$ used in this work. For relatively complex tasks (S2EF tasks in experiments), each of the three directional channels is first processed by a shared two-layer

MLP. The resulting features are modulated by frame-specific FiLM parameters $(\gamma^{(k)}, \beta^{(k)})$, summed over the three frame axes, and then passed through a frame-specific two-layer MLP. For relatively simple tasks (crystal property prediction tasks in experiments), we additionally introduce a learnable alignment matrix $A^{(k)} \in \mathbb{R}^{3\times3}$ for each frame and first rotate the three directional channels by $A^{(k)}$. The transformed channels are processed by a shared MLP with frame-specific FiLM modulation, concatenated, and projected by a shared linear layer.

For routing, the model predicts one logit for each frame from $\boldsymbol{e}^*_{ij}$. In complex tasks, the router may use independent sigmoid gates for different frames, allowing information from each frame to be better preserved. For simpler tasks, it can alternatively use a softmax over frames. When frame-quality-aware weighting is enabled, the frame quality scores are first mapped through a smooth cutoff function. Specifically, we reuse the envelope function introduced above, mapping $\mathcal{Q}_i^{(k)} \le \varepsilon_0$ to $r = 1$ and $\mathcal{Q}_i^{(k)} \ge \varepsilon_1$ to $r = 0$, with $\varepsilon_1 = 5\varepsilon_0$ in our experiments.. The resulting values are then multiplied with the router weights before normalization. The final aggregated directional feature is

$$\boldsymbol{\theta}_{ij} = \sum_k \bar{w}_{ij}^{(k)}\, \phi^{(k)}(\boldsymbol{\Theta}_{ij}^{(k)}),$$

where $\bar{w}_{ij}^{(k)}$ combines the environment-dependent router weight and the frame-quality weight. In addition, an optional layer normalization can be applied to θ_{ij} to further improve stability and reduce model complexity.

The updated edge representation is finally computed from the concatenation of the intermediate edge representation, the radial feature, and the aggregated directional feature:

$$\boldsymbol{e}_{ij}^{\text{new}} = \text{Softplus}\big(\boldsymbol{e}_{ij} + \psi_e([\boldsymbol{e}^*_{ij}, \boldsymbol{r}_{ij}, \boldsymbol{\theta}_{ij}])\big)\,,$$

where ψ_e is a small MLP followed by an output projection and optional normalization.

B.4 Pooling and output layers

After the interaction blocks, MFAN uses mean pooling over atoms in each structure to obtain a graph-level representation. In the default setting used in this work, node representations are first pooled,

$$\bar{\boldsymbol{x}} = \frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}} \boldsymbol{x}_i,$$

and then passed through a prediction head composed of one or more fully connected layers with SiLU activation, followed by a linear output layer. An alternative implementation applies the prediction head to each atom first and then mean-pools the per-atom outputs. This design is commonly adopted in S2EF tasks [4, 13, 73], as it yields more physically meaningful predictions by modeling per-atom energy contributions.

For invariant property prediction, the pooled output is used directly. For S2EF tasks, MFAN predicts the energy and computes atomic forces by automatic differentiation,

$$\hat{\boldsymbol{F}}_i = -\nabla_{\boldsymbol{p}_i}\hat{E},$$

thereby guaranteeing that the force field is conservative. In the implementation, the predicted per-atom energy is rescaled by the number of atoms in the structure before differentiation, so that the gradient is taken with respect to the corresponding total energy.

B.5 Frames and frame-quality scores

MFAN is a flexible framework that can accommodate a wide range of frame constructions. Below, we detail the specific frame constructions used in our experiments, along with their corresponding quality scores. Current there are two commonly used types of frames, namely E(3) frames and SE(3) frames. The former does not distinguish chirality, while the latter enforces a consistent orientation, corresponding to $\det(\mathcal{F}) = \pm1$ and $\det(\mathcal{F}) = 1$, respectively. Any E(3) frame can be converted to an SE(3) frame by flipping the third frame vector if needed to enforce $\det(\mathcal{F}) = 1$. We omit this distinction in the following for brevity.

PCA frame. We implement our PCA frame as a local frame. For each atom i, We first aggregate neighbor displacements around it to form

$$\boldsymbol{a}_i = \sum_{j \in \mathcal{N}(i)} \boldsymbol{d}_{ij}, \qquad \Sigma_i = \sum_{j \in \mathcal{N}(i)} \boldsymbol{d}_{ij} \boldsymbol{d}_{ij}^{\mathsf{T}}.$$

We then compute the eigenvectors of Σ_i and use the three principal directions as the frame axes. Unlike PCA-based frames constructed via random sampling, we instead use $\boldsymbol{a}_i$ to deterministically fix the orientation of the frame vectors, reducing computational cost while avoiding discontinuities induced by randomness. Specifically, for each principal direction, we select the eigenvector whose angle with $\boldsymbol{a}_i$ is less than 90° as the frame vector. The quality score is designed to reflect both spectral degeneracy and directional ambiguity. Specifically, we use the minimum of the square root of two eigenvalue gaps and the dot products between $\boldsymbol{a}_i$ and three eigenvectors. Therefore, the score becomes small when two principal directions are hard to distinguish or when the sign of an eigenvector is ambiguous.

QR frame. The original QR-based frame is SE(3)-equivariant but not permutation-invariant [48]. We therefore introduce several modifications to make it more compatible with our setting. Specifically, we use the QR frame as a local frame. For each atom, we first sort neighboring vectors by distance, and then select the first nonzero vector, the first vector that is not collinear with it, and the first vector that is not coplanar with the first two. These three vectors are stacked into a 3×3 matrix and orthonormalized by batched QR decomposition, with the resulting Q matrix defining the frame. Its quality score contains two components. The first is a selection-stability term δ_g, defined as the minimum distance gap between each selected vector and the next valid competitor in the sorted neighbor list. This term becomes small when multiple candidate neighbors have nearly identical distances, where the distance-based selection may become ambiguous. The second is a QR-stability term δ_r, defined as the minimum absolute value of the three diagonal elements of R, which becomes small when the selected vectors are nearly linearly dependent. The overall quality combines these two effects as

$$\mathcal{Q} = \min(\alpha \delta_g, \delta_r),$$

where $\alpha > 1$ downweights the selection-stability term relative to the QR-stability term. This choice reflects that selection ambiguity mainly affects the ordering of selected frame vectors, whose influence is partly mitigated by our symmetric axis aggregation, whereas QR instability can lead to ill-conditioned orthonormalization and more severe numerical artifacts.

Lattice frame. We use the lattice frame as its original version [71]. Specifically, we first select lattice points $e = n_1\boldsymbol{l}_1 + n_2\boldsymbol{l}_2 + n_3\boldsymbol{l}_3$ with the smallest non-zero norms, using the smallest as $\boldsymbol{e}_1$ and the next two (ensuring linear independence) as $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$. The signs are adjusted so that the angles between $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, and between $\boldsymbol{e}_1$ and $\boldsymbol{e}_3$, are acute. The lattice frame is a global frame that depends only on the lattice vectors and is independent of atomic positions. In S2EF tasks, we focus on the continuity of predictions with respect to atomic positions. Therefore, in our experiments, we assign a constant quality score of 1 to the lattice frame. This reflects its role as a stable geometric reference in the fixed-lattice setting considered in this work.

Lattice-neighbor frame. The lattice-neighbor frame is an extension of the lattice frame designed for situations where multiple lattice directions play nearly equivalent roles. Instead of treating all three lattice-frame axes symmetrically, we retain a single canonical lattice axis—the longest lattice axis $\boldsymbol{e}_3$—and complement it with local neighbor information. This choice is motivated by the anisotropic nature of many structures (e.g., HCP crystals or layered materials), where one lattice direction often plays a distinct structural role compared to the others. Concretely, we first aggregate the neighbor displacements around atom i into

$$\boldsymbol{a}_i = \sum_{j \in \mathcal{N}(i)} \boldsymbol{d}_{ij}.$$

We then construct the second axis as $\boldsymbol{b}_i = \boldsymbol{e}_3 \times \boldsymbol{a}_i$, and use $[\boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{e}_3]$ as the frame after normalization. In this way, the frame preserves a stable lattice reference while allowing local geometry to break the near-equivalence among lattice directions. Its quality score is set to $\|\boldsymbol{b}_i\|$, which captures both the magnitude of $\boldsymbol{a}_i$ and its alignment with $\boldsymbol{e}_3$, reflecting whether $\boldsymbol{a}_i$ is close to being collinear with $\boldsymbol{e}_3$.

Actually, we follow the relaxed definition used in the original lattice frame construction for lattice-neighbor frame, where the frame vectors are treated as equivariant directional axes rather than an orthonormal rotation matrix. Therefore, they are not explicitly orthogonalized, which helps preserve the original directional information from the lattice and neighboring atoms. To avoid scale effects, however, each frame vector is normalized when computing the directional features. Thus, the directional features depend only on the orientations of the frame vectors, not on their magnitudes.

In our experiments, MFAN variants labeled as `lattice only`, `PCA only`, and `QR only` use only the corresponding single frame. The variant labeled as `lattice + QR` employs both the lattice frame and the QR frame, while `lattice + PCA` uses the original lattice frame, its variant lattice-neighbor frame, and the PCA frame. Through the flexible interface of MFAN, additional frames can be readily incorporated. Beyond atomic positions $\boldsymbol{P}$ and lattice vectors $\boldsymbol{L}$, even SO(3)-invariant atomic or edge representations $\boldsymbol{x}_i$ and $\boldsymbol{e}_{ij}$ can be used for frame construction, enabling dynamic frames [29].

C Supplemental experiments

In this section, we present supplemental experiments, including ablation studies, more force outlier thresholds, controlled perturbation tests, results for alternative frame combinations, and analyses of computational complexity and efficiency.

C.1 Ablation studies

We first conduct ablation studies to evaluate the effects of the frame routing gate and frame-quality-aware weighting in MFAN, focusing on both prediction accuracy and stability. The previous experiments have shown that incorporating multiple frames provides richer geometric information and helps the model better capture atomic environments. Building on this design, the frame routing gate generates frame importance scores $w_{ij}^{(k)}$ according to the local atomic environment, while frame-quality-aware weighting assigns quality scores $\mathcal{Q}_i^{(k)}$ according to the stability of each frame. Together, these two scores form the adaptive mechanism for integrating information from different frames used in MFAN.

We conduct ablation studies on the CeO_2/Pt dataset, with the results shown in Fig. 5. Here, `lattice + QR w/o routing gate` fixes the $w_{ij}^{(k)}$ to 0.5, while $\mathcal{Q}_i^{(k)}$ is fixed to 1 for `lattice + QR w/o frame quality`. We can find that adding the routing network improves both prediction accuracy and force stability. This highlights the importance of adaptively weighting frames according to local atomic environments. The effect of frame-quality-aware weighting is more nuanced. On the one hand, it substantially reduces anomalous force predictions, suggesting that suppressing unstable frames improves the empirical stability of frame-based representations. On the other hand, it slightly degrades prediction accuracy, possibly because the current quality score is relatively conservative. For the QR frame, it penalizes configurations where the directions of axes may change or where the ordering of axes becomes ambiguous. In practice, however, not every axis contributes equally to the model output, and the model may not be highly sensitive to the ordering of coordinate axes. Therefore, the design of frame quality scores reflects a trade-off between representation stability and fitting capacity. Developing less conservative but still continuity-aware quality scores is a promising direction for future work.

Table 5: Ablation study on the CeO_2/Pt dataset. Energy MAE is computed for total energy in *eV/atom*, and force MAE for per-coordinate force components in *eV/Å*. %Force Outlier denotes the percentage of predicted force components exceeding 10 eV/Å after Gaussian position perturbations with $\sigma = 0.05$ Å. Bold and underline indicate the best and second-best results, respectively.

Method	Energy MAE	Force MAE	%Force Outlier
MFAN (ours)			
- `lattice only`	0.0057	0.0974	**0.0060**
- `QR only`	0.0058	0.0960	0.0091
- `lattice + QR`	<u>0.0055</u>	<u>0.0944</u>	<u>0.0064</u>
- `lattice + QR w/o routing gate`	0.0058	0.0958	0.0072
- `lattice + QR w/o frame quality`	**0.0054**	**0.0930**	0.0080

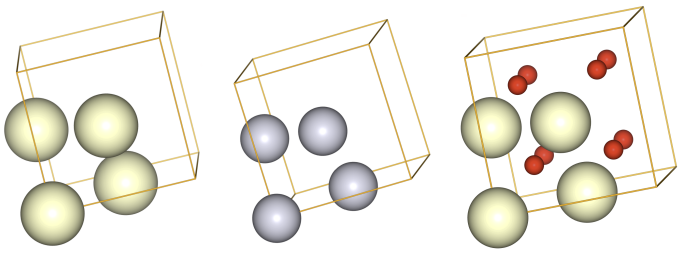

Figure 3: Atomic structures of crystalline Ce, Pt, and CeO_2, where Ce and Pt adopt FCC structures, and the Ce atoms in CeO_2 also form an FCC sublattice, with eight O atoms symmetrically distributed around each Ce atom.

C.2 Force outlier rates at a 5 eV/Å Threshold

In the main results, we adopt a conservative force outlier threshold of $10\ \mathrm{eV/\AA}$ to identify severe force spikes that can destabilize simulation trajectories. For the three S2EF datasets, more than 0.8% of the force components exceed $2\ \mathrm{eV/\AA}$ even without perturbations, and the fractions exceeding $5\ \mathrm{eV/\AA}$ are 0.19%, 0.04%, and 0.06%, respectively. Perturbed configurations can exhibit broader force distributions than the original structures sampled from MD or relaxation trajectories. Consequently, lower thresholds may capture physically plausible high-force configurations.

Table 6: Percentage of force outlier in different thresholds. Gaussian position perturbations with $\sigma = 0.05$ Å are applied to the atomic positions, and the force components exceeding 5 eV/Å and 10 eV/Å are reported. Bold and underline indicate the best and second-best results, respectively.

Method	MPTrj-light		Cu/O		CeO_2/Pt	
	5 eV/Å	10 eV/Å	5 eV/Å	10 eV/Å	5 eV/Å	10 eV/Å
AlphaNet [73]	1.3524	0.2169	0.0715	0.0036	0.1291	0.0093
MACE [1]	1.4277	0.2058	0.0708	0.0009	**0.1058**	**0.0057**
MFAN (ours)						
- `lattice only`	1.3354	**0.1970**	**0.0567**	**0.0002**	0.1225	<u>0.0060</u>
- `QR only`	<u>1.3321</u>	0.2000	0.0625	0.0006	0.1320	0.0091
- `lattice + QR`	**1.3308**	**0.1970**	<u>0.0610</u>	<u>0.0003</u>	<u>0.1172</u>	0.0064

In Table 6, we also report the percentage of force outlier at a lower threshold of $5\ \mathrm{eV/\AA}$. At this threshold, the measured outlier rates show greater variation, and the overall model ranking becomes more sensitive to the underlying force distributions. Particularly high rates are observed on MPTrj-light, where independent atomic perturbations can produce short interatomic contacts in dense bulk structures. Nevertheless, MFAN (lattice + QR) yields lower outlier rates than MFAN (QR) across all three datasets, supporting our main conclusion that combining a stable lattice reference with the local QR frame reduces anomalous force predictions.

C.3 Controlled perturbation tests

To further verify the stability of multi-frame structure and MFAN, we introduce small continuous perturbations around symmetric structures and track the frame variation and force variation relative to the displacement. We selected three ideal crystals: Ce (MP-12) and Pt (MP-126) adopt ideal FCC structures, where each atom has 12 equally distant nearest neighbors and the three principal directions are equivalent. CeO_2 (MP-20194) is also highly symmetric: the Ce atoms form an FCC sublattice, with eight O atoms symmetrically distributed around each Ce atom. The atomic structures of Ce, Pt, CeO_2 are shown in Figure 3. For each ideal symmetric structure $\boldsymbol{P}$, we independently perturb every atom with random displacements $\boldsymbol{u}$ whose magnitudes t range from 10^{-4} to 10^{-1} Å. At each perturbation scale, we evaluate three quantities: the paired discrepancy between the frames constructed, the corresponding frame-quality scores, and the effective stiffness

$$K_{\mathrm{eff}}(\boldsymbol{u}) = \frac{|\mathbf{F}(\boldsymbol{P}+\boldsymbol{u}) - \mathbf{F}(\boldsymbol{P}-\boldsymbol{u})|}{2|\boldsymbol{u}|}$$

which measures force sensitivity to structural perturbations. It approaches the root mean square magnitude of the Hessian-vector product, linking it to local energy-surface curvature relevant to phonon calculations and Hessian-based transition-state searches. For each perturbation magnitude t, we repeat the experiment using 100 independently sampled perturbations and report the mean. We compare three representative frame constructions: the *PCA* and *QR* frames used as candidate frames in MFAN, and the frame construction adopted by AlphaNet, referred to here as *edge-neighbor*.

Table 7: Results of controlled perturbation tests. We introduce small perturbations whose magnitudes t range from 10^{-4} to 10^{-1}Å around the ideal Ce, Pt, and CeO_2 structures and track the frame change, frame quality, and effective stiffness. For each perturbation magnitude, we repeat the experiment using 100 independently sampled perturbations and report the mean.

Metric	Frame / Model	Structure	$t = 0.1$	$t = 0.01$	$t = 0.001$	$t = 0.0001$
Frame change	Edge-neighbor	CeO_2	0.5627	0.5414	0.5458	0.5434
		Ce	0.5519	0.5390	0.5479	0.5392
		Pt	0.5512	0.5433	0.5453	0.5382
	QR	CeO_2	0.8178	0.8125	0.8131	0.8165
		Ce	0.8152	0.8110	0.8170	0.8157
		Pt	0.8116	0.8190	0.8119	0.8163
	PCA	CeO_2	0.8151	0.8142	0.8132	0.8144
		Ce	0.8140	0.8101	0.8174	0.8112
		Pt	0.8153	0.8180	0.8169	0.8173
Frame quality	Edge-neighbor	CeO_2	0.9993	0.0362	0.0036	0.0004
		Ce	0.3880	0.0389	0.0039	0.0004
		Pt	0.3539	0.0354	0.0035	0.0004
	QR	CeO_2	1.1797	0.1839	0.0185	0.0018
		Ce	0.7259	0.0770	0.0079	0.0008
		Pt	0.6977	0.0748	0.0076	0.0007
	PCA	CeO_2	3.1886	0.1134	0.0113	0.0011
		Ce	0.7876	0.0776	0.0076	0.0008
		Pt	0.8352	0.0848	0.0084	0.0008
Effective stiffness	MACE	CeO_2	8.1193	7.8498	7.9026	7.7972
		Ce	2.5303	2.4549	2.4883	2.4827
		Pt	8.2480	7.9473	7.8768	7.9054
	AlphaNet	CeO_2	8.4218	8.9323	43.4493	432.9285
		Ce	2.6098	5.1818	40.5845	404.0738
		Pt	7.6861	9.6389	49.6207	477.8717
	MFAN (lattice only)	CeO_2	8.3153	7.9652	8.0236	8.5011
		Ce	2.5506	2.4779	2.4565	2.4163
		Pt	7.8214	7.9300	7.9795	7.8964
	MFAN (QR only)	CeO_2	8.3442	7.8999	47.1726	3136.5235
		Ce	2.4424	4.4007	61.9293	4577.9513
		Pt	8.0124	8.1978	43.4336	8128.9348
	MFAN (lattice and QR)	CeO_2	7.4792	7.3751	7.9791	8.4725
		Ce	1.6069	2.9943	2.4182	2.4130
		Pt	7.7323	7.2834	7.8558	7.8738

As shown in Table 7, the root mean square difference between paired frame vectors remains nearly unchanged as the perturbation magnitude decreases, rather than converging to zero. This behavior indicates that the edge-neighbor, QR, and PCA frames are discontinuous near the degenerate configurations, where even small atomic displacements can induce substantial changes in frame orientation. In contrast, the quality scores of all three frame constructions consistently approach zero as the perturbation magnitude decreases, validating frame quality as an indicator of proximity to degeneracy. Accordingly, we can assign zero weight to any frame whose quality falls below a prescribed threshold, preventing unreliable frames from contributing to the prediction. The effective-stiffness results further demonstrate that frame discontinuities can propagate to predicted forces. Both AlphaNet and MFAN (QR-only) exhibit rapidly increasing effective stiffness as $\|\boldsymbol{u}\|$ decreases, with no indication of convergence. This divergent behavior is consistent with a force discontinuity or non-differentiable force response near the degenerate structure. In contrast, MFAN (lattice+QR) produces effective stiffness comparable to that of MACE and MFAN (lattice-only), both of which remain well behaved in this limit. This comparison shows that MFAN's quality-aware weighting successfully suppresses the degenerating QR branch and prevents its discontinuity from propagating to the force prediction.

C.4 Results for alternative frame combinations

Our main results focus primarily on global–local frame combinations, while the MFAN formulation is not restricted to a fixed number or type of frames. Here, we evaluate two additional combinations, *PCA + QR* and *lattice + PCA + QR*, on the five JARVIS-DFT targets using the same training protocol.

Table 8: Crystal property prediction results on the JARVIS-DFT dataset. Results are reported in mean absolute error (MAE). The training/validation/test set sizes and units are listed under each objective. Bold and underline indicate the best and second-best results, respectively.

Method	Form. energy 44578/5572/5572 eV/atom	Total energy 44578/5572/5572 eV/atom	Bandgap (OPT) 44578/5572/5572 eV	Bandgap (MBJ) 14537/1817/1817 eV	E hull 44296/5537/5537 eV
CGCNN [69]	0.063	0.078	0.20	0.41	0.17
SchNet [61]	0.045	0.047	0.19	0.43	0.14
MEGNet [9]	0.047	0.058	0.145	0.34	0.084
GATGNN [51]	0.047	0.056	0.17	0.51	0.12
ALIGNN [25]	0.0331	0.037	0.142	0.31	0.076
Matformer [70]	0.0325	0.035	0.137	0.30	0.064
PotNet [47]	0.0294	0.032	0.127	0.27	0.055
iComFormer [71]	0.0272	0.0288	0.122	0.26	0.047
eComFormer [71]	0.0284	0.0315	0.124	0.283	0.044
Crystalformer [64]	0.0306	0.0320	0.128	0.30	0.046
Crystalframer [29]	0.0263	0.0279	0.117	0.242	0.047
eComFormer with SPFrame [26]	0.0261	0.0276	**0.107**	0.239	0.042
MFAN (ours)					
- `lattice only`	0.0257	0.0270	0.124	0.258	0.040
- `PCA only`	0.0262	0.0281	0.130	0.260	0.039
- `QR only`	0.0254	0.0271	0.126	0.243	0.042
- `QR + PCA`	0.0253	0.0265	0.119	0.239	0.039
- `lattice + PCA`	0.0254	0.0266	<u>0.115</u>	**0.233**	**0.037**
- `lattice + QR`	**0.0250**	<u>0.0264</u>	0.118	<u>0.234</u>	0.038
- `lattice + QR + PCA`	<u>0.0251</u>	**0.0259**	0.118	0.236	**0.037**

As shown in Table 8, MFAN (PCA+QR) achieves lower MAE than both MFAN (PCA-only) and MFAN (QR-only) on four of the five targets; on energy above hull, it matches PCA-only and outperforms QR-only. This result shows that MFAN can also combine two local frames and that its gains are not restricted to combining global and local information. When lattice, PCA, and QR are used together, the resulting model improves over both lattice+PCA and lattice+QR on total energy, while remaining comparable to them on the other four targets. We attribute the diminishing marginal gain to partial redundancy: PCA and QR are different constructions, but both summarize the same local atomic neighborhood. Once the lattice frame supplies complementary global information, adding both local frames provides less additional information than combining either local frame with the lattice frame.

Thus, MFAN is not limited to global-local frame combinations; its benefit primarily depends on whether the incorporated frames provide complementary geometric information. Incorporating learned dynamic frames or other frames based on more distinct geometric information may provide greater gains than simply increasing the number of related local frames.

C.5 Complexity and efficiency

In this subsection, we evaluate model complexity and efficiency on JARVIS DFT dataset for the formation energy prediction, where the training and validation sets contain 44,578 and 5,572 samples, respectively. All experiments are conducted on a machine equipped with an Intel i9-13900F CPU, 32 GB RAM, and an NVIDIA RTX 4090 GPU (24 GB memory). Training time is reported in seconds per epoch (s/epoch), and inference time in milliseconds per sample (ms/sample). The batch size is set to 64 for both training and evaluation. To obtain stable measurements, we discard the first three epochs and report the average over epochs 4 to 8.

As shown in Table 9, MFAN adopts an efficient edge update design, which leads to a lower parameter count than iComFormer when the number of interaction blocks is fixed. This indicates that, at the block level, MFAN provides a more parameter-efficient formulation. In practice, however, MFAN typically employs a larger number of interaction blocks (e.g., 5 node and 2 edge blocks) to fully exploit the multi-frame mechanism. As a result, the overall parameter count and runtime of MFAN

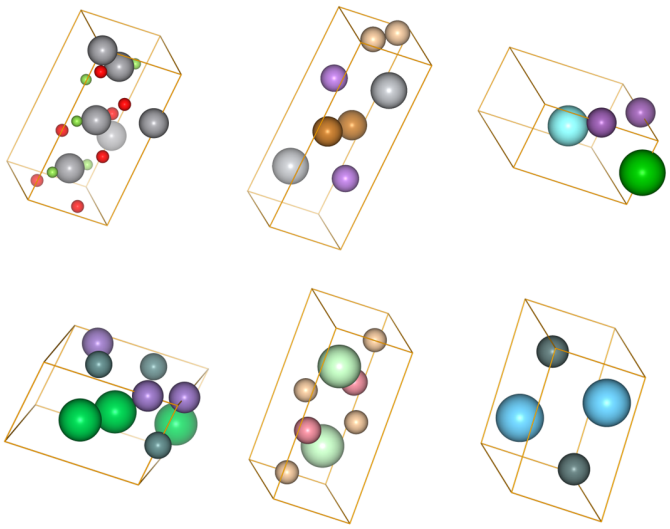

Figure 4: Visualization of representative structures from the JARVIS-DFT (top row) and Materials Project (MP, bottom row) datasets.

can be slightly higher than those of iComFormer (e.g., 4 node and 1 edge blocks). Nevertheless, MFAN remains more efficient and lightweight than typical eComFormer in terms of parameter count and computational cost. Compared with the SPFrame-based eComFormer, MFAN also avoids the additional SPFrame construction overhead reported in [26]. Notably, the multi-frame mechanism is only applied in the geometric feature extraction module within the edge update blocks, so the additional complexity introduced by multiple frames is minimal (on the order of only a few thousand extra parameters). However, the computational cost can vary depending on the choice of frames, and more complex frame constructions may incur higher runtime overhead compared to their impact on model complexity.

Table 9: Model Complexity and Efficiency. We evaluate model complexity and efficiency on the JARVIS DFT dataset for the formation energy prediction. In the `# Blocks` column, the notation "xNyE" denotes x node update blocks and y edge update blocks. Training time is reported in seconds per epoch *(s/epoch)* and evaluation time in milliseconds per sample *(ms/sample)*. All experiments are conducted on a machine with an Intel i9-13900F CPU and an NVIDIA RTX 4090 GPU.

Model	# Blocks	Frame	# Parameters	Training time	Evaluation time
iComformer	4N1E	lattice	4.9M	44.2	0.58
iComformer	5N2E	lattice	7.0M	64.9	0.77
eComformer-half	3N1E	-	5.6M	43.9	0.45
eComformer	3N1E	-	12.4M	98.1	0.64
MFAN	4N1E	lattice	4.4M	34.7	0.52
MFAN	4N1E	lattice + PCA	4.4M	40.4	0.58
MFAN	4N1E	lattice + QR	4.4M	47.5	0.75
MFAN	5N2E	lattice	6.1M	46.4	0.65
MFAN	5N2E	lattice + PCA	6.1M	56.0	0.75
MFAN	5N2E	lattice + QR	6.1M	63.2	0.91
MFAN	5N2E	lattice + PCA + lattice-nei	6.1M	65.2	0.84
MFAN	5N2E	lattice + PCA + QR	6.1M	72.8	1.01

D Details of experiment configurations

In this section, we provide details of the datasets and training configurations used in our experiments.

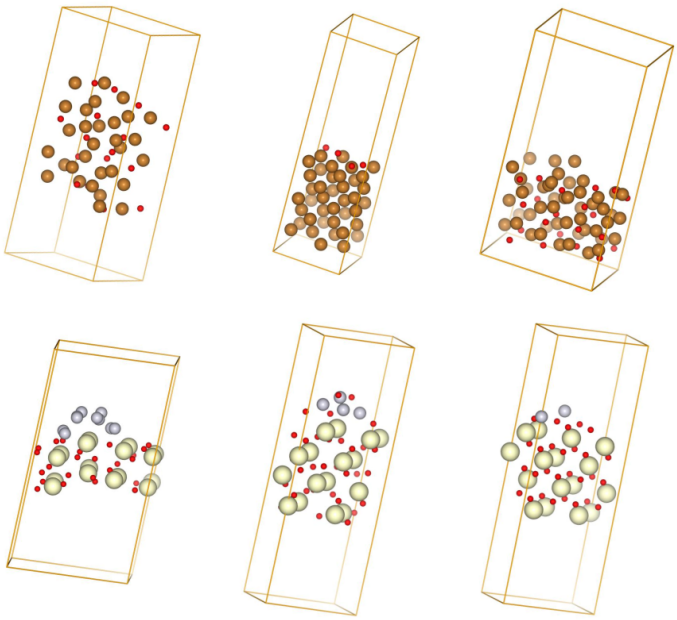

Figure 5: Visualization of representative structures from the Cu/O (top row) and CeO_2/Pt (bottom row) datasets.

D.1 Dataset information

The *JARVIS-DFT* 3D dataset contains 55,723 crystal materials with DFT-calculated properties obtained using OptB88vdW (OPT) and TBmBJ (MBJ) [11]. It is available as *dft_3d_2021* via the *jarvis-tools* package. We follow the data splits of prior work for direct comparison [70, 71], and use formation energy (formation_energy_peratom), total energy (optb88vdw_total_energy), bandgap (optb88vdw_bandgap and mbj_bandgap), and energy above hull (ehull) as prediction targets.

The Materials Project (MP) dataset is an online public repository of materials and their DFT-computed properties [30]. We use a snapshot containing 69,239 crystal materials for our experiments, which is available as *megnet* via *jarvis-tools* [9]. We follow the data splits of prior work as well and use formation energy (e_form), bandgap (gap_pbe), bulk modulus (bulk_modulus), and shear modulus (shear_modulus) as prediction targets. Representative samples from the JARVIS-DFT and MP datasets are shown in Figure 4.

The *MPTrj-light* dataset is constructed from the Materials Project Trajectory (MPtrj) dataset, which contains relaxation trajectories collected from the Materials Project [13]. We restrict our study to structures composed only of elements with atomic numbers from 6 to 35, excluding hydrogen to avoid the over-representation of hydrogen-containing configurations while focusing on light-to-mid elements commonly found in inorganic materials. The resulting subset contains 290,536 configurations with DFT-calculated total energies and atomic forces, and is split into training, validation, and test sets with a ratio of 2:1:1.

The *Cu/O* dataset is derived from the OD-Cu dataset developed for studying oxide-derived copper catalysts in CO_2 reduction electrocatalysis [41]. This original dataset was constructed from DFT calculations and active learning to model the reduction, oxygen diffusion, and restructuring of Cu-O systems under electrochemical conditions. In this work, we remove bulk structures and oxygen-free configurations, retaining surface configurations that contain oxygen. The resulting dataset contains 15,626 samples and is split into training, validation, and test sets with a ratio of 2:1:1. Unlike generic benchmark datasets, Cu/O originates from a realistic chemical application and captures the complex surface reconstruction and oxygen dynamics of oxide-derived Cu catalysts.

For MPTraj-light and Cu/O, we adopt a configuration-level random split rather than enforcing trajectory-wise separation, following common practice in application-oriented settings where tra-

jectory sampling is expensive and strict separation may lead to sparse and high-variance test sets. Besides, the dataset is divided into training, validation, and test sets with a ratio of 2:1:1. This relatively balanced split ensures that only a sparse subset of frames from each trajectory is included in the training set, reducing redundancy caused by highly correlated neighboring configurations, while reserving a substantial portion of frames for validation and testing. As a result, the evaluation remains both stable and sufficiently challenging.

We collect the CeO_2/Pt dataset based on DFT calculations to study the restructuring of Pt clusters on CeO_2 surfaces. It contains 6,178 configurations, primarily consisting of Pt atoms on CeO_2 surfaces, along with a small number of bulk Ce and Pt samples. This is an important heterogeneous catalysis system, since the metal-support interaction between Pt and CeO_2 gives rise to unique electronic effects that are critical for reactions such as methane activation, while also being challenging due to the complex electronic structure of cerium and the dynamic nature of Pt aggregation on oxide surfaces. The dataset is split into training, validation, and test sets with a ratio of 80%/5%/15%. All DFT calculations are performed using Vienna Ab Initio Simulation Package (VASP) with the PBE functional, where a Hubbard +U value of 5.5 eV is applied to Ce atoms to account for localized $4f$ states [34, 35, 57]. The PAW method is used with a plane-wave cutoff of 400 eV [5, 36], and a Γ-centered Monkhorst–Pack scheme with a k-point spacing of $0.03 \times 2\pi$ Å^{-1} is adopted. Representative samples from the Cu/O and CeO_2/Pt datasets are shown in Figure 5.

D.2 Training settings

In this subsection, we present the training and model configurations used in experiments.

All MFAN models are trained using the AdamW optimizer with a OneCycle learning rate scheduler [50, 62]. The number of epochs, learning rate, and the final_div_factor in the OneCycle schedule are adjusted for each task. For crystal property prediction, we use the smooth L1 loss with $\beta = 0.01$. For S2EF, we use smooth L1 loss with $\beta = 0.001$ for energy and $\beta = 0.01$ for forces. As indicated by the ablation study in Appendix C.1, the quality score mainly controls the trade-off between stability and accuracy. Since crystal property prediction tasks are less sensitive to stability issues, we set ϵ sufficiently small in this setting, so that the quality score is equal to one in almost all cases. In contrast, for S2EF tasks, we use frame-quality weighting with task-specific values of ϵ to improve derivative stability near frame degeneracies. Note that, consistent with prior work, labels are normalized for crystal property prediction, whereas raw (unnormalized) labels are used for S2EF. The batch size is set to 64 for crystal property prediction, and to 8 or 6 for S2EF depending on GPU memory constraints. Other hyperparameters of MFAN follow those of iComFormer [71].

The overall network architecture has been described in Section B. Except for the choice of frames, which is directly reflected in the experimental results, the models used across different tasks differ only in the number of node and edge update blocks, the dimensionality of node and edge representations, and the optional use of layer normalization in the end of directional feature processing. Following iComFormer, for crystal property prediction we further limit the maximum number of neighbors per atom after applying the distance cutoff, which reduces computational cost. For S2EF, no such neighbor limit is imposed to preserve continuity.

Specifically, for *JARVIS-DFT formation energy, total energy, and ehull*, we use a model with 5 node update layers and 2 edge update layers, and a representation dimension of 256. The maximum number of neighbors is set to 25. The model is trained for 500 epochs with a learning rate of 0.001, and the final_div_factor in the OneCycle schedule is set to 10000.

For *JARVIS-DFT Bandgap (OPT)*, we use a model with 5 node update layers and **1** edge update layers, and a representation dimension of 256. The maximum number of neighbors is set to **20**. The model is trained for 500 epochs with a learning rate of **0.002**, and the final_div_factor in the OneCycle schedule is set to 10000.

For *JARVIS-DFT Bandgap (MBJ)*, we use a model with 5 node update layers and **1** edge update layers, and a representation dimension of 256. The maximum number of neighbors is set to **20**. The model is trained for 500 epochs with a learning rate of 0.001, and the final_div_factor in the OneCycle schedule is set to **20**.

For *Material Project formation energy*, we use a model with 5 node update layers and 2 edge update layers, and a representation dimension of 256. The maximum number of neighbors is set to 25.

The model is trained for 500 epochs with a learning rate of 0.001, and the final_div_factor in the OneCycle schedule is set to 10000.

For *Material Project Bandgap*, we use a model with 5 node update layers and 2 edge update layers, and a representation dimension of 256. The maximum number of neighbors is set to 25. The model is trained for 500 epochs with a learning rate of **0.0005**, and the final_div_factor in the OneCycle schedule is set to 10000.

For *Material Project bulk modulus*, we use a model with **3** node update layers and 2 edge update layers, and a representation dimension of **96**. An additional layer normalization is applied to the directional features before the edge update. The maximum number of neighbors is set to **18**. The model is trained for 500 epochs with a learning rate of **0.0005**, and the final_div_factor in the OneCycle schedule is set to **50**.

For *Material Project shear modulus*, we use a model with **3** node update layers and **1** edge update layers, and a representation dimension of 256. An additional layer normalization is applied to the directional features before the edge update. The maximum number of neighbors is set to **12**. The model is trained for 500 epochs with a learning rate of 0.001, and the final_div_factor in the OneCycle schedule is set to **60**.

For all S2EF tasks, we use MFAN models with 5 node update layers and 2 edge update layers, with a representation dimension of 256. We adopt a cutoff radius of 6 Å. Following related works, we set the weight decay to 1×10^{-9} to avoid overly regularizing the learned potential energy surface under dense force supervision. We use a learning rate of 0.0001 for all S2EF experiments. For *MPTrj-light*, we train the model for 50 epochs with a final division factor of 10^4, with energy and force loss weights of 1.0 and 0.5, respectively, and frame quality weighting parameter $\epsilon = 0.1$. For *Cu/O*, we train the model for 100 epochs with a final division factor of 10^4, with energy and force loss weights of 1.0 and 4.0, respectively, and frame quality weighting parameter $\epsilon = 0.01$. For *CeO_2/Pt*, we train the model for 300 epochs with a final division factor of 10^8, with energy and force loss weights of 1.0 and 2.0, respectively, and frame quality weighting parameter $\epsilon = 0.01$. We do not include PCA-based frames for S2EF experiments, since the force target requires differentiating the predicted energy with respect to atomic positions, and therefore backpropagating through the frame construction. Although quality-aware weighting can suppress unreliable PCA-frame contributions and their derivative effects in principle, practical automatic differentiation may still evaluate the eigendecomposition branch before the outer quality weight is applied. Near spectral degeneracies, this can lead to unstable gradients. We therefore use lattice- and QR-based frames for more stable force training.

For the S2EF baselines, we use AlphaNet v0.0.1 with the architecture of its OMA-pretrained model, which contains 4 update layers and 176 hidden channels, resulting in 4.8M parameters. We also use MACE v0.3.14 with its default architecture, which contains 128 equivariant messages and 2 layers, each with correlation order 3. The parameter count of MACE depends on the number of elements in the training set and is therefore dataset-dependent. For example, on MPTrj-light, which includes 18 distinct elements, the model contains 3.4M parameters. For each dataset, AlphaNet is trained with the same learning rate and number of epochs as MFAN, while MACE uses its default learning rate and the same number of epochs as MFAN. For AlphaNet, we use the official cosine learning-rate schedule. For MACE, whose default ReduceLROnPlateau scheduler may not decay the learning rate within our limited epoch budget, we additionally test shortened scheduler settings with `lr_factor` $= 0.5$ and `scheduler_patience` $\in \{3, 10\}$. We choose the energy and force loss weights from $\{1, 2, 5, 10\}$ to improve overall validation accuracy while maintaining a balanced trade-off between energy and force prediction.

All experiments are conducted on single local servers equipped with an Intel i9-13900F CPU, 32GB or 64GB RAM, local SSD storage, and an NVIDIA RTX 4090 or 5090 GPU. On a server with 64GB RAM and an RTX 5090 GPU, each JARVIS-DFT, Cu/O, and CeO_2/Pt run can be completed within one day, while each MP and MPTrj-light run typically requires one to two days. RTX 4090 GPUs support the JARVIS-DFT, Cu/O, and CeO_2/Pt experiments, but MP and MPTrj-light require smaller batch sizes due to their higher memory demand. Including preliminary tuning, ablations, and failed runs, the full project required approximately 100–200 single-GPU days.

E Limitations, future directions, and broader context

MFAN is designed to explore the effectiveness of multi-frame learning. Therefore, we primarily focus on well-established crystal property prediction tasks and adopt the iComFormer backbone for a controlled comparison. Extending MFAN to broader applications in chemistry and materials science still leaves several aspects to be improved.

First, the current model design relies on SE(3)-invariant representations constructed from distances and directional projections, which are effective for crystal prediction but remain based on scalarized geometric features. In addition, our directional features are aggregated symmetrically over the frame axes, which improves robustness to axis-ordering ambiguities but may discard part of the ordering information in ordered frames. Incorporating more equivariant representations through iterative frame-based scalarization-vectorization mechanisms may provide a more flexible representation and further improve performance for more complex tasks [15, 16].

Second, the stability of the current lattice frame relies on fixed lattice vectors. This assumption becomes restrictive in settings where the lattice evolves, such as NPT molecular dynamics (where N, P, and T denote the controlled particle number, pressure, and temperature, respectively) and stress-strain prediction, which requires derivatives of the energy with respect to lattice vectors. Designing frame constructions or combinations that remain stable under simultaneous changes in atomic positions and lattice vectors is therefore an important direction for future work.

Third, MFAN currently considers a limited set of predefined frame constructions. A natural extension is to introduce more adaptive frames, for example by dynamically constructing frames from learned node or edge representations [29]. Such data-driven frames, combined with the multi-frame weighting mechanism, may better capture heterogeneous local environments. In addition, in the current design, frames are primarily used to construct directional features, while they themselves are not explicitly treated as features. However, especially for local and data-driven frames, they may encode useful information about the atomic environment. Incorporating frame representations directly as additional geometric features may provide complementary information, particularly in the multi-frame setting.

From a broader perspective, machine learning has become an important tool in chemistry and materials science for understanding chemical mechanisms and exploring materials space. Although universal models trained on large-scale datasets have made significant progress [2, 73, 19], their performance in specific applications can still differ from density functional theory (DFT) due to the complexity of chemical environments (e.g., adsorbates, electric fields, and solvation) and the diversity of DFT configurations (e.g., exchange-correlation functionals, cutoff energies, and various corrections) [68], especially for non-equilibrium structures like transition states [14]. Consequently, many practical workflows still rely on task-specific data collection and model training [41, 66, 72].

In this context, MFAN aims to provide an efficient and flexible model for academic settings with limited computational resources. While equivariant networks perform well in universal datasets, practical applications often involve simpler environments where their expressivity gains are less pronounced relative to their computational cost. Frame-based representations offer a balance by capturing richer geometry than distances and angles at moderate cost, which MFAN further enhances through its multi-frame architecture to improve flexibility and stability. For practical applications, stochasticity in configuration sampling and convergence issues in DFT calculations inevitably introduce a small fraction of noisy or unreliable samples, and incorporating noise-robust and trustworthy learning techniques may therefore be beneficial, particularly when training data are limited [38, 39, 45, 46]. We have also implemented an interface between MFAN and the Atomic Simulation Environment (ASE) [24], enabling its direct integration into molecular dynamics simulation pipelines. Since S2EF prediction underlies many downstream tasks such as structure relaxation and reaction energetics, this creates a practical path for deploying MFAN in broader chemistry and materials applications.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the main contributions and scope of the paper, including the proposed method and its intended setting. These claims are consistent with the methodology and are supported by the experimental results.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations are discussed in the main paper (Discussion section) and further elaborated in Appendix E, where we provide a more detailed analysis along with directions for future work. We examine the current limitations of the model and discuss potential extensions required for broader settings and applications. The proposed method does not introduce new assumptions beyond standard settings, and it does not involve issues related to fairness or privacy.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All theoretical results in the paper are stated with their corresponding assumptions, which are clearly specified in the propositions. Complete and formal proofs are provided in the Appendix A, while the main paper includes intuitive explanations to aid understanding. All theorems and formulas are properly numbered and cross-referenced, and relevant prior results are cited where needed.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All information necessary to reproduce the main experimental results is provided in the paper and supplementary material. We detail the model architecture in Appendix B, and the training procedures and evaluation protocols—including dataset descriptions, hyperparameters, and implementation details—in Appendix D. Upon publication, we will release the code as well as the self-collected dataset (CeO_2/Pt) used in this work, ensuring that the results can be reliably reproduced.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Our code and self-collected dataset (CeO_2/Pt) used in this work is publicly available at `https://github.com/IngramWang/mfan`. In the meantime, we provide detailed descriptions of the model architecture, training procedures, and experimental settings in the main article, Appendix B, and Appendix D.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies the training and evaluation details necessary to understand the results. The main text outlines the experimental setup, while full details—including data splits, hyperparameters, optimizer settings, and implementation details—are provided in Appendix D).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: The paper does not report error bars or formal statistical significance measures. This is consistent with common practice in related works and benchmarks considered in this paper, where results are typically reported as single-run metrics.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We report the experimental hardware setup in Appendix D. Appendix C also provides training and evaluation time for the efficiency analysis. These details allow the required computational resources to be estimated for reproducing the experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research presented in this paper complies with the NeurIPS Code of Ethics. It involves no human subjects, sensitive data, or ethically concerning applications, and all experiments are conducted using publicly available or properly sourced datasets.

   Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: This work focuses on machine learning methods for materials modeling and is primarily intended for scientific research. It does not involve human subjects, sensitive data, or applications with immediate societal consequences. While improved materials modeling could have long-term positive impacts (e.g., accelerating materials discovery), we do not identify direct negative societal impacts associated with this work.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This work does not involve models or datasets with a high risk of misuse. The proposed method is designed for scientific research in materials modeling and does not generate content or involve user-facing applications. Therefore, no specific safeguards are required.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All datasets and codebases used in this work are publicly available and properly cited. We follow standard usage practices and respect their licenses and terms of use to the best of our knowledge. The self-collected CeO_2/Pt dataset is labeled using VASP, for which our research group holds a valid license.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets (e.g., code or datasets) at submission time. The implementation and self-collected dataset will be released upon publication, and thus no accompanying documentation is provided at this stage.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This work does not involve crowdsourcing or research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This work does not involve research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core methods in this work do not involve large language models (LLMs).

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.