DR-Label: Label Deconstruction and Reconstruction of GNN Models for Catalysis Systems

Bowen Wang[1*], Chen Liang[2*], Jiaze Wang[1], Jiezhong Qiu[3], Furui Liu[3], Shaogang Hao[4], Dong Li[5], Guangyong Chen[3 †], Xiaolong Zou[2], Pheng-Ann Heng[1,3]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong,
[2] Shenzhen Geim Graphene Center, Institute of Materials Research, Tsinghua Shenzhen International Graduate School, Tsinghua University,
[3] Zhejiang Lab,
[4] Tencent,
[5] Huawei Noah's Ark Lab,
bowenwang@link.cuhk.edu.hk

Abstract

Attaining the equilibrium geometry of a catalyst-adsorbate system is key to fundamentally assessing its effective properties, such as adsorption energy. While machine learning methods with advanced representation or supervision strategies have been applied to boost and guide the relaxation processes of catalysis systems, existing methods that produce linearly aggregated geometry predictions are susceptible to edge representations ambiguity, and are therefore vulnerable to graph variations. In this paper, we present a novel graph neural network (GNN) supervision and prediction strategy **DR-Label**. Our approach mitigates the multiplicity of solutions in edge representation and encourages model predictions that are independent of graph structural variations. DR-Label first **D**econstructs finer-grained equilibrium state information to the model by projecting the node-level supervision signal to each edge. Reversely, the model **R**econstructs a more robust equilibrium state prediction by converting edge-level predictions to node-level via a sphere-fitting algorithm. When applied to three fundamentally different models, DR-Label consistently enhanced performance. Leveraging the graph structure invariance of the DR-Label strategy, we further propose DRFormer, which applied explicit intermediate positional update and achieves a new state-of-the-art performance on the Open Catalyst 2020 (**OC20**) dataset and the Cu-based single-atom alloys CO adsorption (**SAA**) dataset. We expect our work to highlight vital principles for advancing geometric GNN models for catalysis systems and beyond. Our code is available at https://github.com/bowenwang77/DR-Label

Introduction

Discovering efficient catalysts is a pressing issue in contemporary research (Newell et al. 2020; Liu et al. 2017; Appel et al. 2013). The effectiveness of a catalyst for optimizing a chemical reaction is closely related to its adsorption energy towards specific adsorbates participating in the reaction, a property inherent to the equilibrium state of the catalyst-adsorbate system (Chanussot et al. 2021; Tran and Ulissi 2018; Peterson and Nørskov 2012). Density functional theory (DFT) is a widely used first-principles calculation method of calculating adsorption energies, uncovering potentially high-performance catalysts for further experimental research(Sholl and Steckel 2022; Kresse and Joubert 1999; Kresse and Furthmüller 1996). However, the high computational cost renders exploring the entire configuration space impractical (Matera et al. 2019; Medford et al. 2018). With their fast inference capabilities and the growing availability of databases, machine learning methods such as graph neural network (GNN) models have gained continuing popularity for assessing adsorption energies of catalysis systems. By constructing a graph that treating atoms as nodes and interatomic relation as edges, GNNs excel at automatically extracting interatomic interactions and formulating effective features for property prediction using a physics-motivated multi-step message aggregation.

Researchers have increasingly focused on developing more elaborate message passing and representations during a model's forward pass to preserve physical symmetries. Some incorporate additional neighborhood information in each message-passing step (Schütt et al. 2017; Gasteiger, Groß, and Günnemann 2020; Gasteiger, Becker, and Günnemann 2021), while others employ advanced representations to maintain group equivariance/invariance properties (Thomas et al. 2018; Zitnick et al. 2022; Jing et al. 2021). In contrast to early efforts that directly predict the targeted graph-level properties for atomic systems (Gilmer et al. 2017; Xie and Grossman 2018), another trend involves adding auxiliary geometric supervision signals at the equilibrium states to improve model performance. More specifically, (Godwin et al. 2022; Liao and Smidt 2023; Satorras, Hoogeboom, and Welling 2021) incorporate atomic level supervision signals of equilibrium configurations, while (Liang et al. 2022; Méndez-Lucio et al. 2021; Yuan et al. 2023) extend supervision signals to the interatomic level. However, the former commonly employ linearly aggregated node predictions from edges. As shown in Fig. 1 (c) and (d), linear aggregation leads to a multiplic-

*These authors contributed equally.
†Corresponding author: Guangyong Chen

ity of solutions, causing ambiguities in edge representation. The latter ones simply separate edge-level and node-level geometric predictions into different network branches, leaving the underlying inter-coupling mechanism obscure. As further discussed in the Method section and shown in Fig. 2 (c) and (d), ambiguity in edge representations caused by the aggregation and update schemes is further exacerbated by the non-unique nature of graph construction algorithms (Gasteiger et al. 2022; Ying et al. 2021; Gilmer et al. 2017), rendering estimated atomic positions highly susceptible to graph structural changes. Consequently, GNN models must carefully consider the densely coupled nature of node and edge representations for geometry prediction.

In this work, we address the ambiguity of edge representation for GNN models on atomic systems and the non-unique nature of graph construction algorithms by developing **DR-Label**. DR-Label introduces principles for E(3)-equivariant label deconstruction, expanding the atom-wise equilibrium geometric state signal to the inter-atomic level, and thereby providing finer-grained supervision signals to models. During inference, the DR-Label schema first generates an inter-atomic level prediction and then reconstructs the node-level prediction through an inverse operation of the deconstruction process. This approach allows GNN models to bypass the multiplicity of edge representations, which is common in existing linearly aggregated node predictions. Due to the clear definition of edge-level representations, DR-Label integrated models can generate node-wise geometric predictions that are graph structural variations independent. We present a simple-to-implement instantiation of the DR-Label strategy "project and fit", and integrate it into three fundamentally different cutting-edge models. We evaluate our methods using the Open Catalyst 2020 (OC20) (Chanussot et al. 2021) dataset and the Cu-based single-atom alloys CO adsorption (SAA) dataset (Liang et al. 2022), simulating the structure relaxation processes and predicting adsorption energies of catalysis systems. Each model demonstrates a consistent performance enhancement when applying DR-Label. Leveraging the graph structure independent property of DR-Label, we further develop DRFormer, which achieves state-of-the-art performance on both the OC20 and SAA datasets. These results indicate that DR-Label can serve as a critical component for catalyst adsorption energy prediction models, and has the potential for broader application to other N-body systems state-property prediction tasks.

Preliminaries and Related Work

Graph Neural Network for Equilibrium State Prediction for Atomic Systems

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be the graph representation of an N-body atomic system with N atoms, where each element in the node set $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ represents each atom. v_i contains two part of feature $\{h_i, p_i\}$: $h_i \in \mathbb{R}^d$ refers the E(3)-invariant atom properties, and $p_i \in \mathbb{R}^3$ is the atomic position in the Euclidean space. $\mathcal{E} = \{e_{ij} = (v_i, v_j) | v_i, v_j \in \mathcal{V}\}$ is the set of directed edges from sender node v_i to receiver node v_j. For each e_{ij}, we denote the edge embedding as r_{ij} and edge direction as $\mathbf{d}_{ij} = \frac{p_i - p_j}{|p_i - p_j|}$. $\mathcal{N}_i$ refers to the set of 1-step neighbor nodes of v_i.

For the relaxation process of an atomic system, the set of atom positions $\mathcal{P}^0 = \{p_1^0, p_2^0, ..., p_N^0\}$ are dynamically changed to the equilibrium state $\mathcal{P}^\star = \{p_1^\star, p_2^\star, ..., p_N^\star\}$ due to the inter-atomic forces in structure relaxation, where $\Delta\mathcal{P}^\star = \{\Delta p_i^\star = p_i^\star - p_i^0 | i \in 1, 2, ..., N\}$. In the task of equilibrium geometry and property prediction of the atomic system, researchers develop a model $\mathcal{F}_{GNN}$ such that $\{\hat{\mathcal{A}}, \hat{\mathcal{G}}\} = \mathcal{F}_{GNN}(\mathcal{G}^0)$, where $\hat{\mathcal{G}} = (\hat{\mathcal{V}}, \hat{\mathcal{E}})$ is the output graph with estimated node and edge representations, and $\hat{\mathcal{A}}$ is the system-level property at equilibrium state that minimizes $||\mathcal{A}^\star - \hat{\mathcal{A}}||$, with $\mathcal{A}^\star$ being the ground truth. The equilibrium geometry forecast branch is commonly deployed to help the model to learn the underlying physical rules, where the model produces a new set of atomic positions $\hat{\mathcal{P}} = \{\hat{p}_1, \hat{p}_2, ..., \hat{p}_N\}$ that minimizes $\sum_{i \in N} ||p_i^\star - \hat{p}_i||_2$.

Catalyst Adsorption Energy Prediction

Compared to earlier approaches that depended on expert-engineered features or custom graph constructions (Xie and Grossman 2018; Chen et al. 2019), contemporary GNNs for atomic systems commonly employ geometric graphs (Gasteiger, Groß, and Günnemann 2020) for the automatic extraction of essential features. In the context of Initial Structure to Relaxed Energy (IS2RE) for catalysis systems, the geometric forecast of the Relaxed State $\mathcal{G}^\star$ is of paramount importance, because the relaxed energy $\mathcal{A}^\star$ is derived based on the relaxed state of the system $\mathcal{A}^\star = \mathbf{f}_{ads}(\mathcal{G}^\star)$. Existing catalyst adsorption energy prediction methods can be divided into iterative relaxation or direct methods. Relaxation methods (Zitnick et al. 2022; Gasteiger, Becker, and Günnemann 2021; Sriram et al. 2022; Gasteiger et al. 2022; Shuaibi et al. 2021) typically employ a Structure to Energy and Forces (S2EF) model, which emulates DFT calculations to predict the transient energy and forces within atomic systems. Subsequently, this model is iteratively used to refine atomic positions. Despite their increased accuracy, training these methods easily demands thousands of GPU days. In contrast, direct methods predict the relaxed state from the initial state and energy without intermediate supervision, reducing the training cost to tens of GPU days (Ying et al. 2021; Shi et al. 2022). Recent advancements, such as the introduction of node-level noise-correcting loss (Godwin et al. 2022) and the integration of equivariant features and attention mechanisms (Liao and Smidt 2023), revealed that auxiliary supervision of node positional displacement $\Delta\mathcal{P}^\star$ towards the relaxed state significantly improves model performance. We integrate DR-Label into both direct methods and the direct variant of iterative relaxation methods, demonstrating its effectiveness and broad applicability.

Invariant and Equivariant GNNs

Machine learning models that effectively predict geometric states have applications beyond catalyst adsorption energy assessment, including drug discovery (Méndez-Lucio et al. 2021; Zhou et al. 2023; Shen et al. 2022), protein structure prediction (Jing et al. 2021; Jumper et al. 2021), and engineering simulation (Pfaff et al. 2020; Sanchez-Gonzalez

et al. 2020). Graph neural networks play a central role, with group invariance and equivariance being critical properties to consider during model development. An E(3)-equivariant model $\mathbf{f}_{E(3)}$ should satisfy the condition $\mathbf{f}_{E(3)}(\psi(\mathcal{P})) = \psi(\mathbf{f}_{E(3)}(\mathcal{P}))$ for an E(3) group operation ψ. Early models, such as (Gilmer et al. 2017), were unsuitable for geometric state forecasting due to their lack of group equivariance. Top-performing models preserve E(3)-equivariance through advanced message-passing strategies. SchNet (Schütt et al. 2017) uses continuous edge encoders to model local correlations. DimeNet (Gasteiger, Groß, and Günnemann 2020) introduces edge representations grounded in Schrödinger's equation, encoding inter-atom distance and angular information, achieving comparable performance to DFT calculations on QM9 dataset. EGNN (Satorras, Hoogeboom, and Welling 2021) linearly aggregates edge representation with edge directions for E(n)-equivariance. GemNet (Gasteiger, Becker, and Günnemann 2021) includes dihedral angle information for enhanced performance. Other methods focus on designing finer equivariant representations (Liao and Smidt 2023; Thomas et al. 2018; Batzner et al. 2022; Fuchs et al. 2020), commonly using spherical harmonics for equivariance. (Zitnick et al. 2022) embed nodes as C channels of functions on a unit sphere, each function parameterized by L degree of spherical harmonic basis. The proposed DR-Label preserves physical symmetries while addressing the limitations of using linear aggregated node geometry predictions, which are commonly deployed in previous methods.

Method

In this section, we stress the limitations of existing models that generate linearly aggregated geometric state forecasts. Then we propose principles of label deconstruction and reconstruction (DR-Label). Finally, we instantiate an algorithm named "project and fit" that implements the DR-Label principles.

Limitation of Linearly Aggregated Node Prediction

As illustrated by Fig. 1 (a) and (b), in order to generate an E(3)-equivariant node-level positional displacement prediction, a straightforward idea commonly deployed by existing models (Satorras, Hoogeboom, and Welling 2021; Liao and Smidt 2023; Yuan et al. 2023; Gasteiger et al. 2022; Gasteiger, Becker, and Günnemann 2021; Gasteiger et al. 2020) is to take advantage of the natural E(3)-equivariance of outward edge directions, and use edges as the local coordinate system. The framework can be abstracted as:

$$r_{ij} = \mathbf{f}_{emb}\left(\mathcal{G}^0\right)_{ij}, \tag{1}$$

$$\hat{\mathbf{x}}_{ij} = \mathbf{f}_{vec}\left(\mathbf{f}_{geom}\left(p_i^0, p_j^0\right), r_{ij}\right), \tag{2}$$

$$\hat{p}_i = p_i^0 + \Delta\hat{p}_i = p_i^0 + \mathbf{f}_{pos}(\sum_{j\neq i} \hat{\mathbf{x}}_{ij}). \tag{3}$$

The functions can be implemented through customized message updating models such that $\mathbf{f}_{geom} : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^{h_v}$, $\mathbf{f}_{vec} : \mathbb{R}^{h_v} \times \mathbb{R}^{h_e} \to \mathbb{R}^{h_p}$, and $\mathbf{f}_{pos} : \mathbb{R}^{h_p} \to \mathbb{R}^3$.

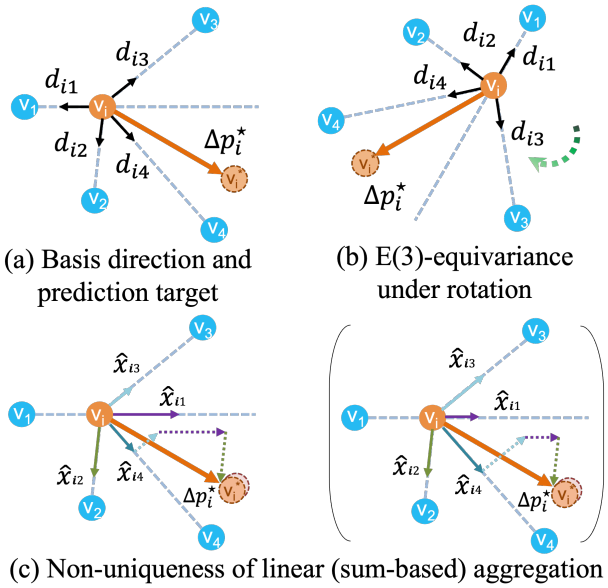
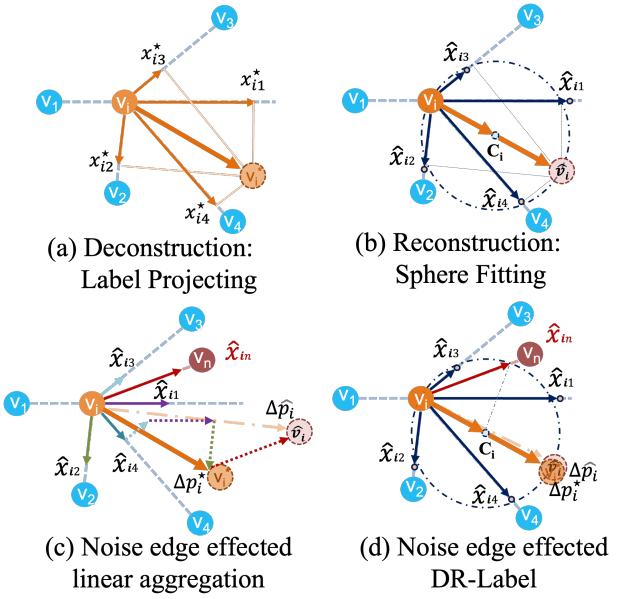


Figure 1: (a) and (b) depict the E(3)-equivariance demanded in geometric prediction models. (c) demonstrates linearly aggregated node predictions from edges can lead to non-uniqueness in edge representation.

$\mathbf{f}_{emb}$ serves as the whole backbone of neural architecture that can output edge features $r_{ij} \in \mathbb{R}^{h_e}$, which should be combined with E(3)-equivariant edge-direction encoding $\mathbf{f}_{geom}(p_i^0, p_j^0)$ for geometry-encoded edge embedding $\hat{\mathbf{x}}_{ij}$ by $\mathbf{f}_{vec}$. Then the positional displacement vector $\Delta\hat{p}_i$ is derived through a linear aggregation of incident $\hat{\mathbf{x}}_{ij}$, followed by another transformation $\mathbf{f}_{pos}$. For clearer elucidation, we simplify this process without sacrificing general applicability by abstracting it to a straightforward sum-aggregation of edge-wise predictions, shown in Fig. 1(c). In this example, $h_e = 1, h_v = 3, h_p = 3$. Therefore $r_{ij} \in \mathbb{R}$ and $\hat{\mathbf{x}}_{ij} \in \mathbb{R}^3$. We let $\mathbf{f}_{geom}(p_i, p_j) = \mathbf{d}_{ij} = \frac{p_i - p_j}{|p_i - p_j|}$, and $\hat{\mathbf{x}}_{ij} = \mathbf{f}_{vec}(\mathbf{d}_{ij}, r_{ij}) = r_{ij}\mathbf{d}_{ij}$. We also set $\mathbf{f}_{pos}$ as identity matrix and have $\Delta\hat{p}_i = \sum_{j\neq i} \hat{\mathbf{x}}_{ij}$.

The above linear aggregated framework is problematic because supervision signals for geometry state prediction lie within the node level. Therefore, shown in Fig. 1 (c), the physical significance of inter-atomic embeddings is unclear, leading to the multiplicity of ideal edge-embeddings, i.e., there exists more than one set of edge representations $\hat{\mathbf{x}}_{ij}, j \in \mathcal{N}_i$ that can generate the same $\Delta\hat{p}_i$. This can make the node prediction highly volatile under graph structure variations. As shown in Fig. 2 (c), suppose an additional noisy edge $e_{in} = (v_i, v_n)$ is introduced due to a change of the graph structure, then a different edge-wise prediction $\hat{\mathbf{x}}_{in}$ will be generated. For a linearly aggregated node displacement prediction, the results will be largely deviated from the original prediction due to the influence of $\hat{\mathbf{x}}_{in}$.

Even worse, variations in graph representations for atomic systems are ubiquitous and inevitable. First, the



(a) Deconstruction: Label Projecting

(b) Reconstruction: Sphere Fitting

(c) Noise edge effected linear aggregation

(d) Noise edge effected DR-Label

Figure 2: (a) and (b) gives an instantiation of DR-Label: the project and fit algorithm. (c) and (d) provide a comparative analysis of the impact of graph structural variations on node predictions, contrasting linearly aggregated node predictions with the DR-Label approach.

graph construction algorithm for atomic systems remains an open question that varies among different models. Prominent methods span cut-off distance-based graphs like Dimenet(Gasteiger, Groß, and Günnemann 2020), fixed nearest neighbor graphs such as GemNet-OC(Gasteiger et al. 2022), and fully connected graphs exemplified by Graphormer(Ying et al. 2021). Second, even under the same graph construction algorithm, diverse atomic instances bring about disparities in both geometric and topological graph statistics, making some graphs too dense while others disconnected. Notably, in fully connected graphs, the average node degree can fluctuate extensively since it corresponds to the system's atom count. Finally, even when both the construction algorithm and atomic system remain consistent, the geometric structure iteration process during relaxation can also alter inter-atomic distances, leading to recurrent edge set and geometry variations.

Under these conditions, the optimal model would be difficult to obtain. With non-unique ideal edge predictions and node predictions sensitive to fluctuating graph structures, the model's universality is compromised. Hence, it is imperative to attain clear and unique edge-wise supervision that can withstand structural changes in the graph. Conversely, node predictions aggregated from edges also need to be independent of graph structure variations.

Label Deconstruction and Reconstruction

In order to reduce the edge representation ambiguity caused by linearly aggregated node predictions and improve model robustness under graph structural variations, in this work we propose to address the problem by designing a function ϕ that constructs a unique set of edge-wise label $\{\mathbf{x}^\star_{ij}|e_{ij} \in \mathcal{E}\}$ from node label $\{\Delta p^\star_i|v_i \in \mathcal{V}\}$, to supervise the model in a finer-grained way. Reversely, the model shall generate a set of edge-wise prediction $\{\hat{\mathbf{x}}_{ij}|e_{ij} \in \mathcal{E}\}$ that can reconstruct the node-wise prediction $\{\Delta\hat{p}_i|v_i \in \mathcal{V}\}$. In this way, we are able to provide a concise label for each edge that is identical under different graph representations. At the same time, suppose there exists an oracle model $\mathbf{f}^\star$ such that as long as it can provide the optimal edge predictions $\hat{\mathbf{x}}_{ij} = \mathbf{x}^\star_{ij}$, we can always have $\Delta\hat{p}_i = \Delta p^\star_i$ under different graph structures. We name such a strategy as **DR-Label**, and summarize three principles to follow for implementing such strategy: uniqueness, reversibility, and E(3)-equivariance.

Uniqueness: When performing label decomposition, for each e_{ij}, $\mathbf{x}^\star_{ij}$ is identical under different graph representation $\mathcal{G}$. Moreover, for each $\{e_{ij}|j \in \mathcal{N}_i\}$, there does not exist another set of $\{\tilde{\mathbf{x}}^\star_{\mathbf{ij}} : j \in \mathcal{N}_i\}$ such that there exist $k \in \mathcal{N}_i$ and $\mathbf{x}^\star_{ik} \neq \tilde{\mathbf{x}}^\star_{\mathbf{ik}}$ while $\phi^{-1}(\{\tilde{\mathbf{x}}^\star_{\mathbf{ij}}\}) = \phi^{-1}(\{\mathbf{x}^\star_{ij}\}) = \Delta p^\star_i$

Reversibility: After deconstructing the label of node i,$\{\mathbf{x}^\star_{ij}|j \in \mathcal{N}_i\} = \phi(\Delta p^\star_i)$, there needs to exist a function ϕ^{-1}, s. t. $\Delta p^\star_i = \phi^{-1}(\{\mathbf{x}^\star_{ij}|j \in \mathcal{N}_i\})$

E(3)-equivariance: There exists a pair of group operations on nodes ψ_v and edges ψ_e, such that when an E(3) transformation ψ is performed to the N-body system, for both the deconstruction process and the reconstruction process, we have $\phi(\psi_v(\Delta p^\star_i)) = \psi_e(\phi(\Delta p^\star_i))$ and $\psi_v(\phi^{-1}(\{\mathbf{x}^\star_{ij}|j \in \mathcal{N}_i\})) = \phi^{-1}(\psi_e(\{\mathbf{x}^\star_{ij}|j \in \mathcal{N}_i\}))$.

Instantiation of DR-Label: Project and Fit

Based on the proposed principles that we discussed above, here we instantiate one example implementation of the DR-Label strategy, namely the 'Project and Fit' algorithm. Pseudocode of the algorithm is further detailed in the appendix.

As illustrated in Fig 2 (a) and (b), for the label deconstruction part, we project the positional shift vector $\Delta p^\star_i$ of node v_i on each outward edge direction by $\{\mathbf{x}^\star_{ij} = (\Delta p^\star_i \cdot \mathbf{d}_{ij})\mathbf{d}_{ij} \in \mathbb{R}^3\} = \phi(\Delta p^\star_i)$, and use the magnitude of projection $\{m^\star_{ij} = \Delta p^\star_i \cdot \mathbf{d}_{ij} \in \mathbb{R}\}$ as the label for edge ij, forming an asymmetric projection magnitude label matrix $\mathcal{M}^\star \in \mathbb{R}^{N\times N}$. The operator $\cdot$ denotes the dot product. For the label reconstruction, our model first generates the prediction of projection magnitude matrix $\hat{\mathcal{M}} \in \mathbb{R}^{N\times N}$. The prediction of positional displacement projection on edges will be thereby calculated by $\{\hat{\mathbf{x}}_{ij} = \hat{m}_{ij}\mathbf{d}_{ij} \in \mathbb{R}^3\}$. Then, to recover node-wise displacement $\Delta p^\star_i$ from its projects on different edges, we use a 3D sphere fitting algorithm to find a sphere that passes the origin p^0_i while having the lowest L^2 distance with each endpoint of projection vectors $\hat{\mathbf{x}}_{ij}$. Note that the sphere is unambiguously defined by sphere center C_i. According to Thales' theorem, for three distinct points p^0_i, $\hat{\mathbf{x}}_{ij}$ and $\hat{p}_i$ on a sphere, if $\overline{p^0_i\hat{p}_i}$ passes the sphere center, then $\angle p^0_i\hat{\mathbf{x}}_{ij}\hat{p}_i$ is a right angle. In this way, we are optimizing $\overline{\hat{p}_i\hat{\mathbf{x}}_{ij}}$ and $\overline{\hat{\mathbf{x}}_{ij}p^0_i}$ to be perpendicular as a reverse process of the displacement projection process, and consequently derive $\hat{p}_i = p^0_i + 2 * C_i$. Compared with the linear aggregated

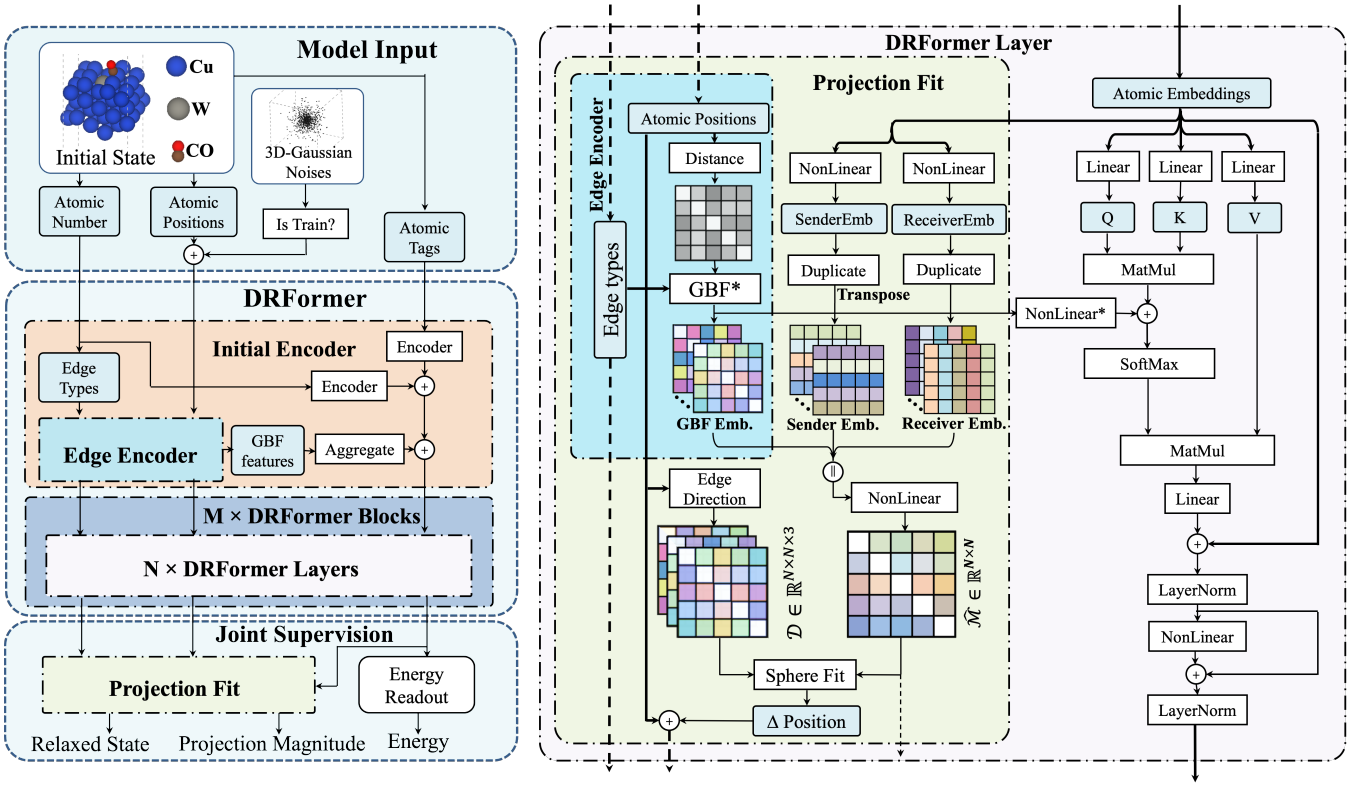


Figure 3: Left: the overall architecture of DR-Label integrated Graphormer and DRFormer. Right: breakdown of the key modules, including the DRFormer Layer, the Edge Encoder module, and the Projection Fit module. $||$ denotes concatenation and $+$ denotes addition. $\star$ indicates shared parameters across different layers in the model. Notably, the "Projection Fit" module inside the DRFormer layer is only activated at every F-th layer of InterPos integrated DRFormer.

node prediction shown in Eq. (3), our method proposes a non-linear node displacement prediction algorithm as

$$\hat{p}_i = p_i^0 + SphereFitting(\{\hat{\mathbf{x}}_{ij} | j \in \mathcal{N}_j\}). \quad (4)$$

The project and fit algorithm strictly fulfills the three principles, where the whole process are of E(3)-equivariance, and the two operations produce unique solutions regardless of graph structure while being reverse operations to each other. As Fig. 2(c) and (d) show and evidenced in our experiments, a model trained with the DR-Label can provide more robust results under the variation of the graph structure than linearly-aggregated node predictions.

Model Implementations

To evaluate the DR-Label schema's efficacy, in this section, we detailed DR-Label's integration with **3d-Graphormer**, a direct method. From this integration, we further developed **DRFormer** that achieved state-of-the-art performance on OC20 and SAA datasets. We also integrated DR-Label to the direct variant of two relaxation methods, **GemNet-OC** and **SCN**, which are further elaborated in the appendix.

DR-Label Integrated 3d-Graphormer

The DR-Label's integration into 3d-Graphormer involves adding a projection fit block after the final atomic embedding, as shown in Fig. 3. This block receives the initial atomic positions $\mathcal{P}^0$, edge type information from atomic endpoints, and 3d-Graphormer's atomic embeddings. In 3d-Graphormer, atomic positions help derive a symmetric $N \times N$ edge-level Euclidean distance matrix, where the ij element represents the distance between the i atom and the j atom. It is later combined with the edge type matrix to generate a Gaussian Basis Function (GBF) edge embedding.

Given the asymmetric nature of $\mathcal{M}^\star \in \mathbb{R}^{N \times N}$, we transform the final atomic embedding for each edge using two distinct nonlinear layers, producing node-level embeddings for sender and receiver nodes. The concluding edge-level embedding combines GBF edge embedding, sender node, and receiver node embeddings. This is then processed through another nonlinear layer to derive the projection magnitude matrix $\hat{\mathcal{M}} \in \mathbb{R}^{N \times N}$. Subsequently, we merge $\hat{\mathcal{M}}$ with the edge direction tensor $\mathcal{D} \in \mathbb{R}^{N \times N \times 3}$ to input the sphere-fitting algorithm, yielding an E(3)-equivariant vector prediction $\Delta\hat{p}_i$ for each node.

For label deconstruction, $\mathcal{M}^\star$ supervises the edge-wise

Model	MAE (eV) ↓					AEwT (%) ↑				
	ID	OOD Ads.	OOD Cat.	OOD Both	Average	ID	OOD Ads.	OOD Cat.	OOD Both	Average
3d-Graphormer	0.4329	0.5850	0.4441	0.5299	0.4980	–	–	–	–	–
3d-Graphormer+DR-Label	0.4574	0.5497	0.4686	0.4873	0.4907	6.34	3.95	6.02	4.34	5.13
GNS	0.54	0.65	0.55	0.59	0.5825	–	–	–	–	–
GNS+Noisy Nodes	0.47	0.51	0.48	0.46	0.4800	–	–	–	–	–
Equiformer	0.4222	0.5420	0.4231	0.4754	0.4657	7.23	3.77	7.13	4.10	5.56
Moleformer	0.413	0.523	0.432	0.473	0.4603	8.01	3.04	7.66	3.19	5.48
Equiformer+Noisy Nodes	0.4156	0.4976	**0.4165**	0.4344	0.4410	7.47	4.64	7.19	4.84	6.04
DRFormer w/o InterPos	**0.4085**	0.5149	0.4286	0.4635	0.4539	**9.08**	4.87	**8.36**	5.11	**6.86**
DRFormer-S	0.4187	0.4863	0.4321	0.4332	0.4425	8.39	**5.42**	8.12	5.44	6.84
DRFormer	0.4212	**0.4852**	0.4258	**0.4220**	**0.4386**	7.64	5.22	7.44	**5.47**	6.44

Table 1: OC20 IS2RE validation set results (direct method). Top result in bold and runner-ups underlined

Model	MAE (eV) ↓					AEwT (%) ↑				
	ID	OOD Ads.	OOD Cat.	OOD Both	Average	ID	OOD Ads.	OOD Cat.	OOD Both	Average
GemNet-OC	0.560	0.711	0.576	0.671	0.630	4.15	2.29	3.85	2.28	3.14
GemNet-OC + DR-Label	0.450	0.707	0.478	0.638	0.568	6.55	2.56	5.93	2.98	4.51
SCN	0.516	0.643	0.530	0.604	0.573	4.92	2.71	4.42	2.76	3.70
SCN + DR-Label	0.474	0.675	0.482	0.626	0.564	5.55	2.63	5.13	2.69	4.00
3d-Graphormer (Ensemble)	0.3976	0.5719	0.4166	0.5029	0.4722	8.97	3.45	8.18	3.79	6.1
GNS+Noisy Nodes	0.4219	0.5678	0.4366	0.4651	0.4728	9.12	4.25	8.01	4.64	6.5
Equiformer+Noisy Nodes	0.4171	0.5479	0.4248	0.4741	0.4660	7.71	3.70	7.15	4.07	5.66
Moleformer	0.4134	0.5346	0.4280	0.4581	0.4585	8.79	**4.67**	7.58	**4.87**	6.48
DRFormer w/o InterPos	0.3928	0.5523	0.4113	0.4766	0.4582	8.77	3.81	8.32	3.97	6.21
DRFormer-S	0.4148	0.5691	0.4365	0.4931	0.4783	8.07	4.09	7.31	4.30	5.94
DRFormer	0.4099	0.5387	0.4297	0.4647	0.4607	7.68	4.62	7.12	4.71	6.03
DRFormer-Average	**0.3789**	**0.5175**	**0.3966**	**0.4432**	**0.4340**	**9.35**	4.42	**8.48**	4.59	**6.71**

Table 2: OC20 IS2RE testing set results (direct method). Top result in bold and runner-ups underlined

magnitude matrix via the label projection algorithm. Node displacement prediction is also guided by $\Delta\mathcal{P}^{\star}$. We maintain 3d-Graphormer's energy readout for graph-level supervision and blend all supervision levels using $\mathcal{L} = \mathcal{L}_G + \lambda\mathcal{L}_V + \gamma\mathcal{L}_E$, where $\mathcal{L}$ represents total loss, and individual components denote graph, node, and edge-level losses. Parameters λ and γ adjust the supervision intensity.

The DRFormer Model

Taking advantage of the robustness of DR-Label under graph structural variations, we further advance our model, DRFormer, to include explicit intermediate atomic positional updates (**InterPos**) throughout the model. As shown in Fig. 3, we employ the projection fit block to update the atomic positions at every F-th layer of the DRFormer. GBF edge embedding is correspondingly modified for the graph attention component. In contrast to models that predict the relaxed geometry in a single positional update, DRFormer explicitly approximates the relaxation trace in a multi-step manner for better simulating a complex relaxation process.

To touch the limit of direct model performance, following **Noisy Nodes** (Godwin et al. 2022), DRFormer incorporates noise-augmented instances into each training batch. We also corrected the coordinate of atoms at the relaxed state in the training instances that passed the periodic boundary, with the similar procedure in Moleformer (Yuan et al. 2023). Further details can be found in the appendix.

Experiments and Results

Experimental Settings

Datasets. Our methodology and models are mainly evaluated on two datasets for catalysis systems. The primary dataset is OC20 (Chanussot et al. 2021), which comprises of 460k instances for IS2RE method training. It includes four distinct evaluation datasets: an in-distribution dataset (ID) and three out-of-distribution datasets featuring unseen struc-

tures in adsorbates (OOD-Ads), catalyst slabs (OOD-Cat), or both (OOD-Both). We adhere to the official validation and testing split to assess our model. The secondary dataset we employ is a smaller one from (Wang et al. 2021), predicting the CO adsorption energy of Cu-based single-atom alloy (SAA) catalysts. Composed of 41 doping species on different sites of 5 surfaces of ideal Cu crystals, the SAA dataset comprises of 3075 instances in total. CO molecules are adsorbed on various positions of the surfaces, leading to a diverse dataset. Each sample from both datasets includes the initial (unrelaxed) state and the corresponding DFT-calculated relaxed state and adsorption energy as the target. The two datasets are fundamentally differ since OC20 exclusively considers regular bulk catalyst surfaces.

Model Evaluation Settings. Following the OC20 settings, we use the mean absolute error of adsorption energy prediction (MAE, eV) and the percentage of absolute errors within the threshold of 0.02eV (AEwT,%) as our evaluation metrics. For experiments on the OC20 datasets, our test results are reported based on our submission to the official evaluation server on *eval.ai*(Yadav et al. 2019). In experiments on the SAA datasets, we follow the experimental setup from (Liang et al. 2022), performing 10 random splits with a train-validation-test split of 60%:20%:20% and report the mean and standard deviation of the test set results.

Catalyst Adsorption Energy Predictions

OC20 Results. Our results on the OC20 validation and testing sets are reported in Tab. 1 and 2, respectively. A "+ DR-Label" tag indicates that the DR-Label module is incorporated into the corresponding model; "w/o InterPos" suggests the absence of the intermediate positional update in the DRFormer model. "DRFomer-S" model employs simplified modifications of "DRFormer" that conserve 60.1% of computational resources during training. Hyperparameters are detailed in the appendix.

Firstly, we can observe that the integration of the DR-Label module consistently improves the performance of 3d-Graphormer, GemNet-OC, and SCN. For instance, adding the DR-Label module leads to a relative increase of 57.8% in AEwT value on the ID dataset and 54.0% on the OOD-Cat dataset for GemNet-OC. These results demonstrate that DR-Label effectively augments the supervision signal, guiding the model to learn the underlying physics of relaxation more accurately, thereby improving the precision of adsorption energy prediction.

Additionally, DRFormer models sets new state-of-the-art performance on the OC20 IS2RE validation set, particularly excelling in the AEwT metric. This indicates that our method significantly increases the proportion of accurate adsorption energy predictions. We further notice that though not achieves top performance on most of the metrics, DRFormer-S demands much smaller training resources than DRFormer, therefore providing a balanced model solution.

Finally, the DRFormer w/o InterPos model set a new single-model performance record on average MAE, while DRFormer demonstrates comparable performance on most metrics. Denoted as "DRFormer-Average", we achieve a new state-of-the-art direct method performance on the OC20 IS2RE test set by averaging the predictions of "DRFormer w/o InterPos", "DRFormer-S", and "DRFormer". Notably, while the 3d-Graphormer(Ensemble) requires 31 separately trained models (equating to 46.5 days of computation on an $8\times A100$ GPU machine) to achieve its final result, both "DRFormer w/o InterPos" and "DRFormer-S" need a maximum of 5 days of training under the same hardware conditions, while delivering comparable or superior results.

Model	MAE (meV) ↓	AEwT (%) ↑
Localized-cos + GBR	120 ± 18	29.9 ± 1.8
Gaussian-cos + GBR	130 ± 19	27.1 ± 2.1
Gaussian-tanh + GBR	132 ± 20	26.8 ± 1.8
SchNet	257 ± 113	10.3 ± 7.7
CGCNN	182 ± 89	14.7 ± 3.7
DimeNet	99 ± 16	32.9 ± 2.8
DimeNet++	94 ± 15	33.7 ± 1.5
MT-MD DimeNet++	87 ± 11	39.8 ± 2.5
3d-Graphormer	98 ± 16	33.0 ± 1.8
DRFormer-S	$\mathbf{83 \pm 14}$	$\mathbf{44.5 \pm 2.1}$
DRFormer-S-FT	$\mathbf{69 \pm 15}$	$\mathbf{56.0 \pm 2.3}$

Table 3: SAA dataset performance

Modules	Model (MAE (meV)↓)						
DR-Label	×	√	×	×	√	×	√
Noisy Nodes	×	×	√	×	√	√	√
InterPos	×	×	×	√	×	√	√
ID	693	688	673	*726*	664	720	**648**
OOD Ads.	753	733	722	*782*	705	771	**694**
OOD Cat.	665	678	652	*708*	646	694	**631**
OOD Both	669	656	651	*708*	635	700	**627**
Average	695	689	675	*731*	662	721	**650**

Table 4: Ablation study

SAA Dataset Results. We maintain the same model structure used in the OC20 experiments and apply it to the SAA dataset to further test our method. The results are presented in Tab. 3. By benchmarking the performance of the 3d-Graphormer and DRFormer-S model on the SAA dataset, we found that the former does not top the charts, while the latter achieves a new state-of-the-art performance on this dataset. In DRFormer-S-FT, we load the final OC20-trained checkpoint and fine-tune this pre-trained model on the SAA dataset. Compared with the previous best method, MT-MD DimeNet++, this procedure significantly enhances the final performance, leading to a 20.2% relative decrease in MAE and a 16.2% increase in AEwT.

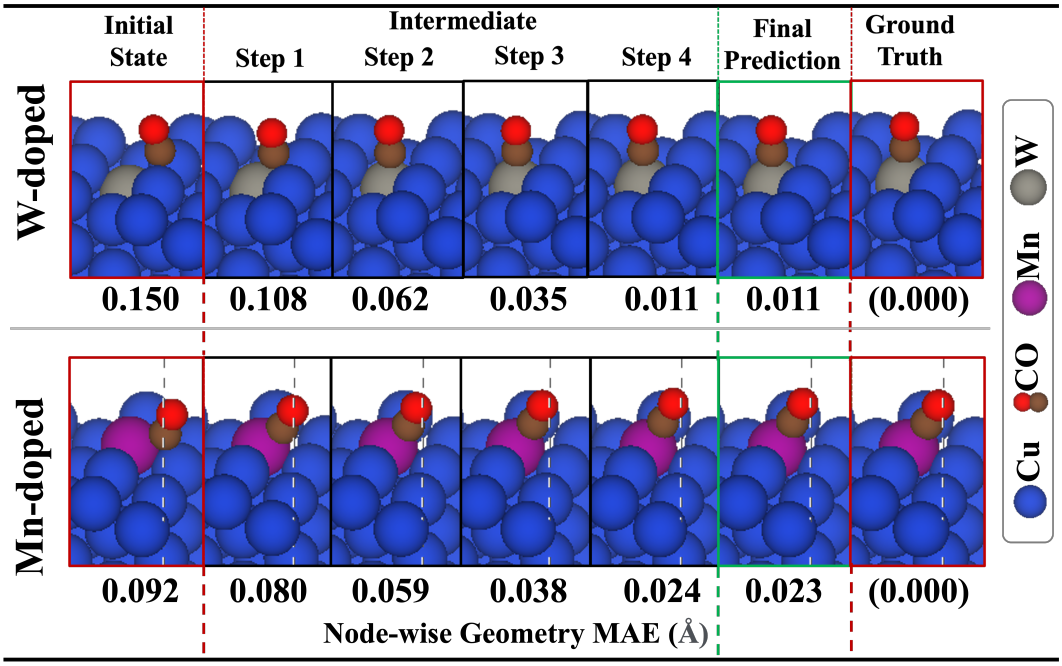


Figure 4: Visualization of intermediate geometries after each DRFormer block generated by the DRFormer model. (a) Adsorption of CO on W-doped Cu(441) surface; (b) Adsorption of CO on Mn-doped Cu(441) surface.

Ablation Studies

We further examine the effectiveness of individual model components by conducting an ablation study trained on 3D-Graphormer using the OC20 IS2RE 10k dataset, with results assessed on the OC20 IS2RE validation set. We utilized 3D-Graphormer's inherent geometry displacement prediction module (NodeTaskHead) to perform the intermediate positional update without DR-Label. The result is shown in Tab. 4. Firstly, in alignment with previous experimental observations, models incorporating the DR-Label module consistently outperform those without it. Secondly, the result shows that intermediate positional update alone worsened the result of 3d-Graphormer. Incorporating noisy nodes merely mitigates this decline. In contrast, the fusion of DR-Label with intermediate node position updates augments performance. This suggests that a multi-step approximation of the relaxation process enhances model efficacy. However, with every positional update, the graph's structure would need re-evaluation, potentially introducing disparities in edge geometries. DR-Label helps to reduce the accumulated prediction errors brought by this continuous restructuring, ensuring uniformity across multiple positional updates.

Intermediate Trace Visualization

While DRFormer has been designed to predict the relaxed state directly without additional intermediate supervision signals, we are interested in understanding whether our model can learn a physically reasonable trace that progressively approaches the equilibrium state. Given that DRFormer provides explicit updates of intermediate positions, we visualize the trace of these positions and assess the model's behavior during representation updating in Fig. 4.

For this experiment, we used DRFormer-S (evaluated on the SAA dataset test set) as an example. We extracted the intermediate positions after each DRFormer block, and visualized the results using VESTA. We present two examples in Fig. 4. The node-wise $L^2 - MAE(\mathring{A})$ of intermediate atomic positions towards the relaxed positions are indicated below each output. We observed the adsorbate navigating a geometric continuum between its initial position and its relaxed state. Concurrently, we noted that while the adsorbate-adsorbate distance remained relatively stable, the adsorbate-slab distance underwent a more noticeable change. This suggests our model autonomously learns authentic inter-atomic relationships and a spatially coherent dynamic process, even in the absence of auxiliary relaxation trace guidance.

Conclusion and Broader Impact

In this work, we introduce the DR-Label strategy, an effective method that successfully mitigates issues of ambiguity and non-uniqueness in edge representation. The integration of this module enhances three distinct models: 3d-Graphormer, GemNet-OC and SCN. We further propose the DRFormer architecture, which establishes a new state-of-the-art performance in adsorption energy prediction on both the OC20 and SAA datasets. Given its benefits, we foresee the potential of DR-Label as a universal tool for geometric GNN models predicting equilibrium states. We also expect DR-Label could stimulate advancements in edge-level auxiliary supervisions within general GNN models.

Acknowledgements

This work is supported by the National Key R&D Program of China (2022YFE0200700), the Hong Kong Innovation and Technology Fund (Project No. ITS/241/21), the National Natural Science Foundation of China (Project No. 6237073934, 3234101132 62006219), and the Natural Science Foundation of Guangdong Province (2022A1515011579).

References

Appel, A. M.; Bercaw, J. E.; Bocarsly, A. B.; Dobbek, H.; DuBois, D. L.; Dupuis, M.; Ferry, J. G.; Fujita, E.; Hille, R.; Kenis, P. J.; et al. 2013. Frontiers, opportunities, and challenges in biochemical and chemical catalysis of CO2 fixation. *Chemical reviews*, 113(8): 6621–6658.

Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; and Kozinsky, B. 2022. E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature communications*, 13(1): 2453.

Chanussot, L.; Das, A.; Goyal, S.; Lavril, T.; Shuaibi, M.; Riviere, M.; Tran, K.; Heras-Domingo, J.; Ho, C.; Hu, W.; et al. 2021. Open catalyst 2020 (OC20) dataset and community challenges. *ACS Catalysis*, 11(10): 6059–6072.

Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; and Ong, S. P. 2019. Graph networks as a universal machine learning framework for molecules and crystals. *Chemistry of Materials*, 31(9): 3564–3572.

Fuchs, F. B.; Worrall, D. E.; Fischer, V.; and Welling, M. 2020. SE(3)-transformers: 3D roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems*, 2020-Decem(3).

Gasteiger, J.; Becker, F.; and Günnemann, S. 2021. GemNet: Universal Directional Graph Neural Networks for Molecules. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Gasteiger, J.; Giri, S.; Margraf, J. T.; and Günnemann, S. 2020. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. In *Machine Learning for Molecules Workshop, NeurIPS*.

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

Gasteiger, J.; Shuaibi, M.; Sriram, A.; Günnemann, S.; Ulissi, Z.; Zitnick, C. L.; and Das, A. 2022. GemNet-OC: Developing Graph Neural Networks for Large and Diverse Molecular Simulation Datasets. *Transactions on Machine Learning Research (TMLR)*.

Gilmer, J.; et al. 2017. Neural message passing for quantum chemistry. In *ICML*, 1263–1272. PMLR.

Godwin, J.; Schaarschmidt, M.; Gaunt, A.; Sanchez-Gonzalez, A.; Rubanova, Y.; Veličković, P.; Kirkpatrick, J.; and Battaglia, P. 2022. Simple GNN regularisation for 3D molecular property prediction & beyond. In *International Conference on Learning Representations (ICLR)*.

Jing, B.; Eismann, S.; Suriana, P.; Townshend, R. J.; and Dror, R. 2021. Learning from Protein Structure with Geometric Vector Perceptrons. In *International Conference on Learning Representations*.

Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; and Hassabis, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873): 583–589.

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

Kresse, G.; and Joubert, D. 1999. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review b*, 59(3): 1758.

Liang, C.; Wang, B.; Hao, S.; Chen, G.; Heng, P.-A.; and Zou, X. 2022. Multi-Task Mixture Density Graph Neural Networks for Predicting Cu-based Single-Atom Alloy Catalysts for CO2 Reduction Reaction. *arXiv preprint arXiv:2209.07300*.

Liao, Y.-L.; and Smidt, T. 2023. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs. In *International Conference on Learning Representations*.

Liu, X.; Xiao, J.; Peng, H.; Hong, X.; Chan, K.; and Nørskov, J. K. 2017. Understanding trends in electrochemical carbon dioxide reduction rates. *Nature communications*, 8(1): 15438.

Matera, S.; Schneider, W. F.; Heyden, A.; and Savara, A. 2019. Progress in accurate chemical kinetic modeling, simulations, and parameter estimation for heterogeneous catalysis. *Acs Catalysis*, 9(8): 6624–6647.

Medford, A. J.; Kunz, M. R.; Ewing, S. M.; Borders, T.; and Fushimi, R. 2018. Extracting knowledge from data through catalysis informatics. *Acs Catalysis*, 8(8): 7403–7429.

Méndez-Lucio, O.; Ahmad, M.; del Rio-Chanona, E. A.; and Wegner, J. K. 2021. A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nature Machine Intelligence*, 3(12): 1033–1039.

Newell, R.; Raimi, D.; Villanueva, S.; Prest, B.; et al. 2020. Global Energy Outlook 2020: energy transition or energy addition. *Resources for the Future*.

Peterson, A. A.; and Nørskov, J. K. 2012. Activity descriptors for CO2 electroreduction to methane on transition-metal catalysts. *The Journal of Physical Chemistry Letters*, 3(2): 251–258.

Pfaff, T.; Fortunato, M.; Sanchez-Gonzalez, A.; and Battaglia, P. W. 2020. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations (ICLR)*.

Sanchez-Gonzalez, A.; Godwin, J.; Pfaff, T.; Ying, R.; Leskovec, J.; and Battaglia, P. 2020. Learning to simulate complex physics with graph networks. In *International conference on machine learning*, 8459–8468. PMLR.

Satorras, V. G.; Hoogeboom, E.; and Welling, M. 2021. E (n) equivariant graph neural networks. In *International conference on machine learning*, 9323–9332. PMLR.

Schütt, K.; Kindermans, P.-J.; Sauceda Felix, H. E.; Chmiela, S.; Tkatchenko, A.; and Müller, K.-R. 2017. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. *Advances in neural information processing systems*, 30.

Shen, C.; Zhang, X.; Deng, Y.; Gao, J.; Wang, D.; Xu, L.; Pan, P.; Hou, T.; and Kang, Y. 2022. Boosting Protein–Ligand Binding Pose Prediction and Virtual Screening Based on Residue–Atom Distance Likelihood Potential and Graph Transformer. *Journal of Medicinal Chemistry*, 65(15): 10691–10706.

Shi, Y.; Zheng, S.; Ke, G.; Shen, Y.; You, J.; He, J.; Luo, S.; Liu, C.; He, D.; and Liu, T.-Y. 2022. Benchmarking graphormer on large-scale molecular modeling datasets. *arXiv preprint arXiv:2203.04810*.

Sholl, D. S.; and Steckel, J. A. 2022. *Density functional theory: a practical introduction*. John Wiley & Sons.

Shuaibi, M.; Kolluru, A.; Das, A.; Grover, A.; Sriram, A.; Ulissi, Z.; and Zitnick, C. L. 2021. Rotation invariant graph neural networks using spin convolutions. *arXiv preprint arXiv:2106.09575*.

Sriram, A.; Das, A.; Wood, B. M.; Goyal, S.; and Zitnick, C. L. 2022. Towards Training Billion Parameter Graph Neural Networks for Atomic Simulations. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.

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

Tran, K.; and Ulissi, Z. W. 2018. Active learning across intermetallics to guide discovery of electrocatalysts for CO2 reduction and H2 evolution. *Nature Catalysis*, 1(9): 696–703.

Wang, D.; Cao, R.; Hao, S.; Liang, C.; Chen, G.; Chen, P.; Li, Y.; and Zou, X. 2021. Accelerated prediction of Cu-based single-atom alloy catalysts for CO2 reduction by machine learning. *Green Energy & Environment*.

Xie, T.; and Grossman, J. C. 2018. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14): 145301.

Yadav, D.; Jain, R.; Agrawal, H.; Chattopadhyay, P.; Singh, T.; Jain, A.; Singh, S. B.; Lee, S.; and Batra, D. 2019. Evalai: Towards better evaluation systems for ai agents. *arXiv preprint arXiv:1902.03570*.

Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; and Liu, T. Y. 2021. Do Transformers Really Perform Bad for Graph Representation? *Advances in Neural Information Processing Systems*, 34(NeurIPS): 28877–28888.

Yuan, Z.; Zhang, Y.; Tan, C.; Wang, W.; Huang, F.; and Huang, S. 2023. Molecular Geometry-aware Transformer for accurate 3D Atomic System modeling. *arXiv preprint arXiv:2302.00855*.

Zhou, G.; Gao, Z.; Ding, Q.; Zheng, H.; Xu, H.; Wei, Z.; Zhang, L.; and Ke, G. 2023. Uni-Mol: A Universal 3D Molecular Representation Learning Framework. In *International Conference on Learning Representations*.

Zitnick, C. L.; Das, A.; Kolluru, A.; Lan, J.; Shuaibi, M.; Sriram, A.; Ulissi, Z.; and Wood, B. 2022. Spherical Channels for Modeling Atomic Interactions. In *Advances in Neural Information Processing Systems (NeurIPS)*.