AlphaNet: Scaling Up Local Frame-based Atomistic Foundation Model

Bangchen Yin[*] Tsinghua University

Jiaao Wang[*†] UT Austin

Weitao Du DAMO Academy

Pengbo Wang HKUST

Penghua Ying Tel Aviv University

Haojun Jia Deep Principle, Inc.

Zisheng Zhang Stanford University

Yuanqi Du[†] Cornell University

Carla P. Gomes Cornell University

Chenru Duan[†] Deep Principle, Inc.

Hai Xiao[†] Tsinghua University

Graeme Henkelman[†] UT Austin

Abstract

We present AlphaNet, a local frame-based equivariant model designed to achieve both accurate and efficient simulations for atomistic systems. Recently, machine learning force fields (MLFFs) have gained prominence in molecular dynamics simulations due to their advantageous efficiency-accuracy balance compared to classical force fields and quantum mechanical calculations, alongside their transferability across various systems. Despite the advancements in improving model accuracy, the efficiency and scalability of MLFFs remain significant obstacles in practical applications. AlphaNet enhances computational efficiency and accuracy by leveraging the local geometric structures of atomic environments through the construction of equivariant local frames and learnable frame transitions. We substantiate the efficacy of AlphaNet across diverse datasets, including defected graphene, formate decomposition, zeolites, and surface reactions. AlphaNet consistently surpasses well-established models, such as NequIP and DeepPot, in terms of both energy and force prediction accuracy. Notably, AlphaNet offers one of the best trade-offs between computational efficiency and accuracy among existing models. Moreover, AlphaNet exhibits scalability across a broad spectrum of system and dataset sizes, affirming its versatility. Our code and data is available at `https://github.com/zmyybc/AlphaNet`.

1 Introduction

Molecular dynamics (MD) simulations have become essential for exploring and understanding complex phenomena in areas such as energy storage, catalysis, and biological systems [24, 3, 22]. While atomic forces necessary for these simulations can theoretically be derived from quantum mechanical approaches such as density functional theory (DFT), the computational cost associated with such first-principles methods severely limit their application to small systems and short time scales. As a result, many phenomena occurring over longer time frames and larger spatial scales remain inaccessible, even with the most powerful supercomputers.

Classical force fields, which rely on predefined mathematical forms, offer a computationally efficient alternative, allowing simulations of larger systems over extended periods [7, 23, 17]. However,

[*]These authors contributed equally to this work.

[†]Corresponding authors. Emails: `wangjiaao0720@utexas.edu`, `yuanqidu@cs.cornell.edu`, `duanchenru@deepprinciple.com`, `haixiao@tsinghua.edu.cn` and `henkelman@utexas.edu`.

Preprint. Under review.

the simplicity of these models often compromises their accuracy, leading to a trade-off between computational speed and the fidelity of the simulated dynamics.

The advent of machine learning has introduced a promising solution to this dilemma. By training models on data derived from ab initio calculations, machine learning forcefields (MLFF) can potentially achieve the accuracy of first-principles methods while maintaining the computational efficiency of classical force fields. Unlike traditional models that rely on explicit functional forms to describe bonded and non-bonded interactions, MLFFs are more flexible, learning to predict interactions based on the positions of atoms and their chemical identities.

As interest in applying machine learning techniques to large-scale atomistic systems continues to grow, the field has seen a surge in the development of various innovative models. One example among the others, DPA-1 [35], is a large-scale pre-trained model with improved attention architecture over DeepPot [32]. Similarly, JMP [28] leverages diverse molecular systems of different types as a joint pre-training model. MACE [1], on the other hand, introduces higher-order message passing as a complete basis of many-body atomic interactions. Collectively, these models represent significant strides in the evolution of machine learning-driven atomistic simulations [13, 2, 26, 29, 8].

However, the balance between computational efficiency and accuracy is key to the practical application of these models. On one side, efficient yet less expressive models are more suitable for tasks that weigh computation more than accuracy. On the other side, less efficient yet expressive models are more suitable for tasks which accuracy is more important.

Atomistic systems in 3D Euclidean space are invariant to Euclidean symmetry including rotation, translation and reflection. To reduce the unnecessary degree of freedoms, equivariant models are developed to build model that respects these symmetries [25, 31]. Among all the developed models, most of the expressive models by date [2, 1, 21] are based on spherical harmonics [15]. However, the computation to calculate the tensor product of irreducible representations impose expensive computational overheads. While another branch of work which achieve rotation equivariance through building equivariant frames which could either be local or global [18, 9, 33, 12]. The main benefit of frame-based approaches is that is eliminates the necessity to use tensor product which greatly improves the computational efficiency.

In this paper, we propose a local frame-based atomistic equivariant model (AlphaNet) for accurate yet highly efficient atomistic system simulations. Building on the success of frame-based MLFFs in small atomistic systems, we introduce an additional rotary positional embedding to enhance frame transition and temporal connection for multi-scale modeling. Extensive quantitative experiments demonstrate AlphaNet excels at accuracy, efficiency and scalability compared to state-of-the-art MLFF models on a variety of molecular systems, from defected graphene, formate decomposition, zeolites, to surface reactions.

2 Experiment

2.1 Dataset

The Defected Bilayer Graphene Dataset consists of reference structures specifically designed to train and validate machine learning potentials (MLPs). It includes three bilayer systems: V0V0 (pristine), V0V1 (single vacancy on the top layer), and V1V1 (single vacancy in both layers). The dataset comprises single-point DFT (PBE+MBD) energies and atomic forces calculated for various configurations, including different interlayer distances, stacking modes, and manually deformed structures. Additionally, snapshot configurations from MD simulations at different temperatures were included. The data were split into training, validation, and test sets, containing 3988, 4467, and 200 structures, respectively. Farthest point sampling (FPS) and principal component analysis (PCA) were used to ensure a representative and balanced division. [34]

The Formate Decomposition on Cu Dataset consists of configurations related to the decomposition process of formate on a Cu surface, specifically involving the cleavage of the C-H bond. The dataset includes initial, intermediate, and final states, such as monodentate and bidentate formate on Cu <110>, as well as the final state with an H ad-atom and desorbed CO2 in the gas phase. Nudged elastic band (NEB) method [16] was used to generate an initial reaction path for the C-H bond breaking, followed by 12 short ab initio molecular dynamics (AIMD) simulations. These simulations were performed using the CP2K code, resulting in a total of 6855 DFT structures with a time step of 0.5

fs and 500 steps per trajectory. AlphaNet was trained on 2500 uniformly sampled structures from the full dataset, with a validation set of 250 structures and the mean absolute error evaluated on the remaining structures. [2]

The Zeolite Dataset comprises essential reference structures designed for various applications such as catalysis, adsorption, and separation. Zeolites are highly valued porous materials that have been the subject of extensive scientific research due to their broad applicability. The dataset includes 16 distinct zeolites, with atomic-level trajectories obtained through *Ab Initio Molecular Dynamics* (AIMD) [4] simulations conducted at 2000 K using the Vienna *Ab Initio Simulation Package* (VASP) [19, 20]. For each zeolite, 80,000 snapshots were extracted, and the corresponding energies and atomic forces were calculated for every configuration. The dataset was randomly divided into training, validation, and test sets in a 6:2:2 ratio, comprising 48,000, 16,000, and 16,000 structures, respectively.

The OC2M Dataset is a large-scale subset of the OC20 Dataset, specifically designed for training and evaluating machine learning models in the context of interatomic potentials [5]. This dataset includes approximately 2 million atomic configurations in the training set, with validation and test sets containing 4 splits, which is In Domain (ID), out-of-domain adsorbate (OOD adsorbate), out-of-domain catalyst (OOD catalyst), and both of unseen adsorbate and unseen catalyst (OOD both). Each of them contains around 1 million structures. It encompasses configurations involving 56 different chemical elements, ensuring a broad and diverse representation of atomic environments. These configurations were generated using density functional theory (DFT) calculations, with a focus on a wide range of materials science applications, including bulk materials, surfaces, and defect structures. Originating from the comprehensive OC20 Dataset, the OC2M subset provides accurate total energies and atomic forces for each configuration. This extensive dataset is particularly valuable for developing and testing machine learning models that require a diverse and large-scale collection of training examples to achieve robust generalization across different types of atomic interactions.

2.2 Results

2.2.1 Accuracy

Our model demonstrates excellent precision across various experimental datasets. On datasets such as formate decomposition and defected graphene, as detailed in Table 1. We selected Nequip [2] as our baseline model due to its demonstrated superior performance in the research associated with the dataset. AlphaNet consistently outperforms the baseline model in terms of both energy and force accuracy.

The formate decomposition dataset, which serves as a representative example of catalytic surface reactions, specifically focuses on the dehydrogenation of formate ($HCOO^* \rightarrow H^* + CO_2$) on a Cu $\langle 110 \rangle$ surface. For this dataset, AlphaNet achieves a mean absolute error (MAE) of 45.5 meV/Å for force and 0.23 meV/atom for energy, compared to Nequip's 47.3 meV/Å and 0.50 meV/atom, respectively. These results underscore AlphaNet's ability to effectively capture the intricate nature of heterogeneous systems, which often involve both metallic and covalent bonding, along with complex charge transfer processes between the metal and adsorbed molecules. This high degree of accuracy highlights the model's suitability for simulating catalytic reactions with multiple interaction types.

Another challenge for machine learning force fields (MLFFs) is accurately modeling the sliding layer effect in materials like defected graphene, where interlayer interactions are critical. On the defected graphene test set, AlphaNet achieves a force MAE of 32.0 meV/Å and an energy MAE of 1.7 meV/atom, substantially outperforming Nequip, which records 60.2 meV/Å for force MAE and 1.9 meV/atom for energy MAE. Furthermore, AlphaNet successfully reproduces the binding energy curve for AB-stacked bilayer graphene as calculated by PBE+MBD, without requiring explicit long-range dispersion interaction corrections. The deviation in the shallow sliding potential energy landscape is less than 0.4 meV/atom, illustrating AlphaNet's capacity to handle systems with subtle interlayer forces and complex structural dynamics (as shown in Figure 3).

This ability to model both metallic and covalent bonding scenarios, as well as interlayer interactions, makes AlphaNet a versatile and powerful tool for materials design, particularly in fields requiring high accuracy for large-scale systems. These results suggest that AlphaNet can be effectively applied to a wide variety of systems, ranging from catalytic surfaces to layered materials, where precision in both energy and force calculations is critical for predicting material behavior.

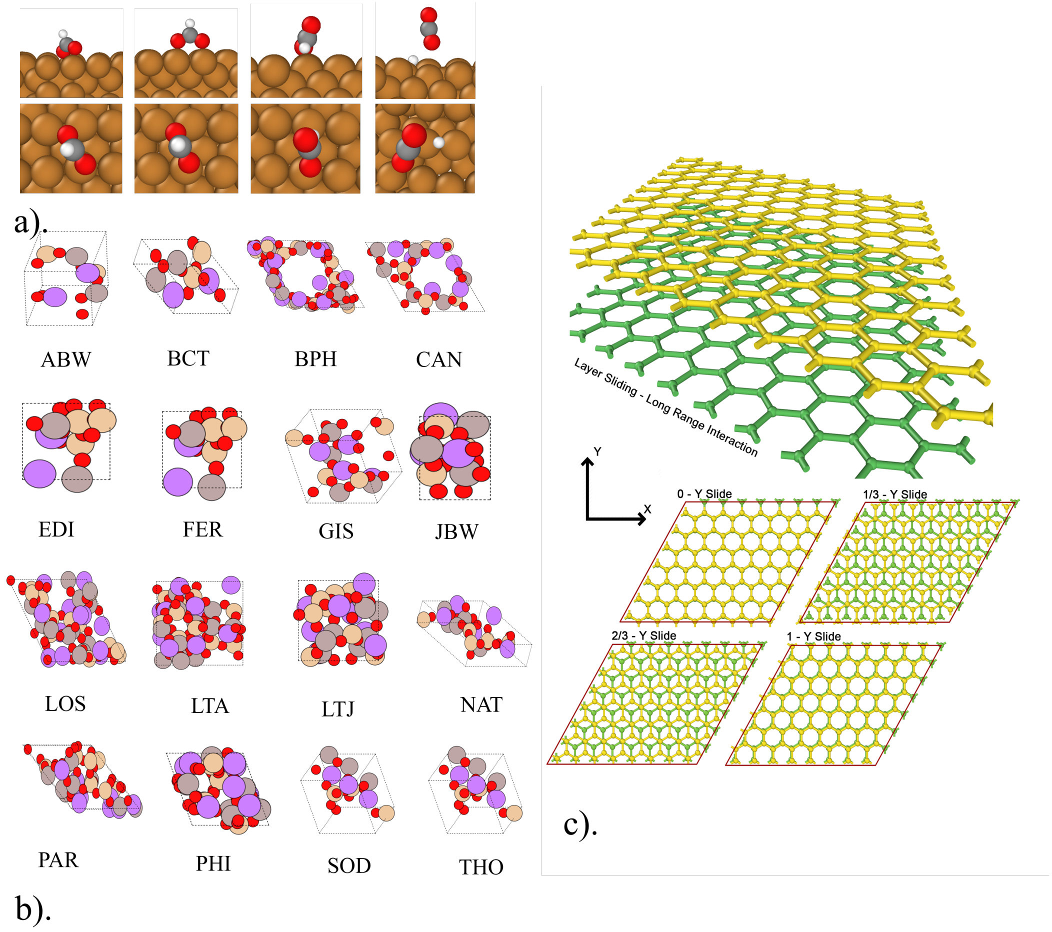


Figure 1: Illustration of Experimental Systems: a) The Formate Decomposition on Cu Dataset. Cu atoms are depicted in brown, C atoms in grey, O atoms in red, and H atoms in white. b) The Zeolite Dataset. Li atoms are shown in purple, Si atoms in grey, Al atoms in orange, and O atoms in red. c) The Defected Bilayer Graphene Dataset.

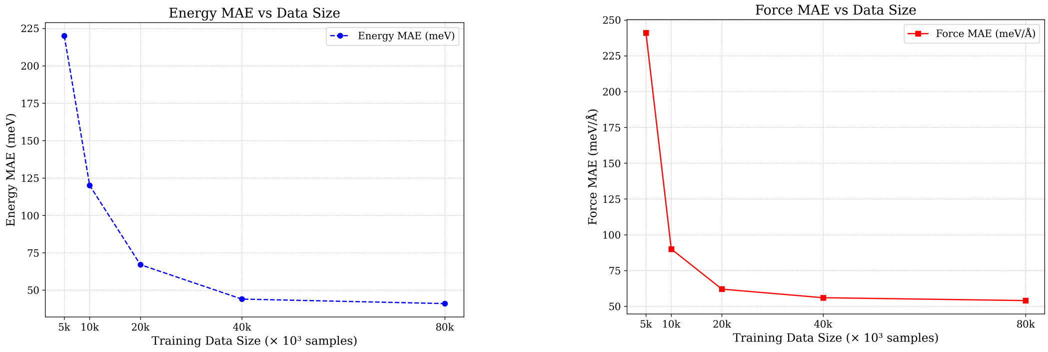
2.2.2 Scaling

In the context of machine learning force field (MLFF) models, performance across large and diverse datasets is crucial for validating model robustness and scalability. The zeolite dataset, comprising 16 types of zeolites and a total of 800,000 configurations, provides an excellent benchmark to evaluate the accuracy of these models. Here, we compare the performance of Deep Potential [32] and AlphaNet using two key metrics: Energy MAE (meV/atom) and Force MAE (meV/Å). A lower mean absolute error (MAE) indicates higher predictive accuracy.

Table 4 shows the results of Deep Pot and AlphaNet across different zeolite configurations. Across nearly all tested configurations, AlphaNet significantly outperforms Deep Pot in both energy and force predictions.

For instance, in the ABWopt configuration, AlphaNet achieves an Energy MAE of 41 meV, substantially lower than Deep Pot's 90 meV. Similarly, AlphaNet records a Force MAE of 54 meV/Å, outperforming Deep Pot's 90 meV/Å. This pattern of superiority is consistent across most configurations. Notable results include the CANopt structure, where AlphaNet shows an exceptional Force MAE of 10 meV/Å compared to Deep Pot's 90 meV/Å, indicating a marked improvement in force prediction accuracy. Additionally, in the FERopt structure, AlphaNet achieves an Energy MAE of 100 meV, significantly lower than Deep Pot's 290 meV, and a Force MAE of 18 meV/Å, much better than Deep Pot's 130 meV/Å.

The performance of the model in terms of Mean Absolute Error (MAE) for both energy (in meV) and force (in meV/Å) improves as the training data size increases from 5k to 80k samples. Specifically, as shown in Figure 2, the energy MAE decreases from 220 meV at 5k samples to 41 meV at 80k samples. Similarly, the force MAE decreases from 241 meV/Å at 5k samples to 54 meV/Å at 80k samples. This trend highlights the scalable performance of increasing the training dataset size in enhancing the accuracy of both energy and force predictions on the zeolite dataset.



(a) Energy MAE (meV) vs Data Size

(b) Force MAE (meV/Å) vs Data Size

Figure 2: MAE of Energy and Force predictions as a function of training data size on the zeolite dataset.

Even in cases where Deep Pot performs competitively, such as in the LOSopt configuration, where Deep Pot achieves an Energy MAE of 110 meV/atom compared to AlphaNet's 130 meV, AlphaNet still maintains an advantage in Force MAE with a value of 46 meV/Å compared to Deep Pot's 70 meV/Å. This suggests that while Deep Pot may occasionally perform well in energy prediction, AlphaNet is more consistent in providing better overall results, particularly in force predictions.

In terms of standout performance, AlphaNet demonstrates its predictive power in configurations like LTJopt, where it achieves an Energy MAE of just 1.2 meV compared to Deep Pot's 70 meV, and in the GISopt structure, where it delivers a Force MAE of only 3.1 meV/Å, far surpassing Deep Pot's 50 meV/Å.

These results indicate that AlphaNet is better equipped to handle complex zeolite configurations, providing both lower energy and force prediction errors across most datasets. This demonstrates its robustness and makes it a promising model for large-scale applications in materials design, where accuracy and scalability are critical.

The Open Catalyst Project (OCP) provides the OC20 dataset, a comprehensive collection of materials and structures designed to train machine learning models for catalysis research. At present, our work has focused on scaling our model to the OC2M S2EF task, with plans to extend the training to the full OC20 dataset in future research. We compared the performance of our model against established models such as EquiformerV2 [21], SchNet [26], and DimeNet++ [14].

We were able to compare our results with SchNet and DimeNet++ models trained on the full OC20 dataset (134M data points), as well as with EquiformerV2 trained on both the full dataset and the smaller 2M subset. Currently we report our results on the validation set as it is public available. Results on the test set would be done as soon as possible. It's important to note that each of these models was trained under different configurations, including training for both energy and force, training for energy only, and training for force only. The corresponding performance metrics are summarized in Table 2.

To control the training cost and make sure all of the elements are seen in the data, we first train our model on the full OC training set for 10,000 steps and then turns to 2M subset. We believe this would not provide additional benefits because we have tested that this procedure makes no difference in the validation ID set as there is no additional elements here. Our model trained on 2M subset for 2,000,000 steps demonstrates exceptional accuracy in energy prediction, achieving a mean absolute error (MAE) of 0.19 eV, which is significantly better than EquiformerV2's performance on both the full dataset (0.229 eV) and the 2M subset (0.285 eV). SchNet and DimeNet++, which were trained on the full dataset, exhibit MAE values above 0.35 eV. This trend holds consistently across different test sets, demonstrating the robustness of our model in energy prediction.

In terms of force prediction, while our model trained on the 2M subset does not surpass larger models like EquiformerV2 and DimeNet++-large, it still achieves competitive results relative to models of a similar scale trained on the full dataset. Table 3 provides additional insights, revealing that our model has substantially lower training costs and fewer parameters compared to EquiformerV2 and DimeNet++. This efficiency, combined with the model's high precision on this large dataset, indicates that our approach is computationally effective, requiring less data and resources while still delivering strong performance.

Given these promising results, we are confident that our model can be successfully scaled to the full OC20 dataset in future work, further enhancing its predictive accuracy for both energy and force calculations. This would allow us to continue improving the state-of-the-art in machine learning for catalysis research.

2.2.3 Speed

The primary advantage of machine learning force fields (MLFFs) over density functional theory (DFT) force fields lies in their significantly faster inference speed. While increasing the number of parameters in an ML model can improve its precision, it often results in slower inference times. To evaluate the speed of different MLFFs, we measured the average forward pass time across 2,000 zeolite structures with a batch size of 10. We also tested systems of varying sizes by expanding the unit cells. Although we aimed to maintain a consistent number of parameters across models, slight variations exist due to differences in architecture.

We compared several graph-based models, including Nequip, MACE, SchNet, and PaiNN, against our AlphaNet. As illustrated in Figure 4, AlphaNet shows a distinct speed advantage over other MLFFs, with a relatively slow increase in computational time as system size grows. This demonstrates that AlphaNet's speed scales efficiently with larger systems, making it a more suitable option for handling increasing system sizes.

Table 1: Performance of Models on Different Datasets

Dataset	Metric	NequIP	AlphaNet
Formate Decomposition	Force MAE (meV/Å)	47.3	**45.5**
	Energy MAE (meV/atom)	0.50	**0.23**
Defected Graphene	Force MAE (meV/Å)	60.2	**32.0**
	Energy MAE (meV/atom)	1.9	**1.7**

3 Discussion

Alphanet presents a significant advancement as a scalable machine learning potential, offering both high precision and fast computational speed. This success can be attributed to its local-frame-based architecture, which scalarizes geometric quantities over local frames and effectively aggregate local frames to construct local- and global-aware features. By utilizing this architecture, our model excels at capturing complex interactions across a broad spectrum of systems, including metallic, covalent, and ionic bonding, as well as long-range interactions.

One of the key advantages of Alphanet lies in its scalability. Specifically, our model is capable of maintaining high performance as the size of the dataset increases, which sets it apart from other large-scale atomistic models. Despite the growing complexity introduced by larger datasets, Alphanet operates with fewer parameters, making it more efficient while still delivering robust results. This ability to handle large datasets without sacrificing accuracy positions Alphanet as a promising solution for extensive simulations and real-world applications. In future work, we aim to extend the model's capabilities to explore additional types of interactions, such as hydrogen bonding, which are critical in a wide range of chemical and biological systems.

Table 2: Energy and Force MAE for different models

Model	ID	OOD-adsorbate	OOD-catalyst	OOD-both
Energy MAE (eV) ↓				
SchNet-full	0.43	0.49	0.53	0.72
SchNet-full-energy-only	0.38	0.45	0.55	0.70
SchNet-full-force-only	34.00	33.77	35.30	38.47
DimeNet++-full	0.48	0.47	0.53	0.65
DimeNet++-full-energy-only	0.35	0.40	0.51	0.65
DimeNet++-full-force-only	28.20	28.94	28.91	34.90
EquiformerV2-2M($\lambda_E = 2$)	0.28	-	-	-
EquiformerV2-full($\lambda_E = 2$)	0.22	-	-	-
AlphaNet-2M($\lambda_E = 4$)	<u>0.19</u>	**0.23**	**0.23**	**0.29**
AlphaNet-2M($\lambda_E = 20$)	**0.18**	<u>0.23</u>	<u>0.25</u>	<u>0.30</u>
Force MAE (eV/Å) ↓				
SchNet-full	0.050	0.054	0.052	0.064
SchNet-full-energy-only	0.485	0.496	0.485	0.547
SchNet-full-force-only	0.045	0.048	0.047	0.058
DimeNet++-full	0.045	0.047	0.045	0.055
DimeNet++-full-energy-only	0.341	0.339	0.341	0.362
DimeNet++-full-force-only	0.034	**0.035**	<u>0.035</u>	<u>0.041</u>
EquiformerV2($\lambda_E = 2$)	<u>0.022</u>	-	-	-
EquiformerV2-full($\lambda_E = 2$)	**0.016**	-	-	-
AlphaNet-2M($\lambda_E = 4$)	0.036	<u>0.037</u>	**0.035**	**0.040**
AlphaNet-2M($\lambda_E = 20$)	0.054	0.062	0.061	0.070

Note: The downward arrow (↓) indicates that lower values are preferable. The term 'full' in a model's name signifies that it was trained on the complete OC20 training set, '2M' indicates that the model was trained on OC2M sub set. λ_E refers to the weight of the energy loss relative to the force, with a value set to 100. Bold indicates the best performance, while <u>underline</u> represents the second best.

Table 3: Training Performance and Training Time for Different Models

Models	Energy MAE (eV) ↓	Force MAE (eV/Å) ↓	Training Time (GPU Days) ↓	Number of Parameters ↓
SchNet	0.549	0.0568	**194**	9.1M
DimeNet++	0.515	0.0328	1600	10.1M
EquiformerV2	0.236	**0.0157**	1368	153M
AlphaNet	**0.201**	0.0356	360	**6.1M**

Another avenue for further exploration is the interpretability of the model. While our current implementation achieves impressive results, gaining a deeper understanding of how the model interprets and processes interactions at the atomic level will enhance its applicability and trustworthiness in practical scenarios.

Despite these advancements, one challenge remains: the efficient use of GPU memory during training, particularly when dealing with large systems consisting of over 1,000 atoms (though we can parallelize it across multiple GPUs). This is a common limitation shared by most message-passing neural networks (MPNNs), and addressing it will be crucial for enabling the model to scale further in both size and complexity. Future work will focus on optimizing memory usage to overcome this bottleneck and ensure the model's applicability to even larger systems.

4 Methodology

4.1 Problem Formulation

Machine learning forcefields. A single state of a molecular system can be described by a set of atom types $h \in \mathbb{R}^{n\times d}$ and atomic positions $x \in \mathbb{R}^{n\times 3}$. In molecular dynamics, we are interested

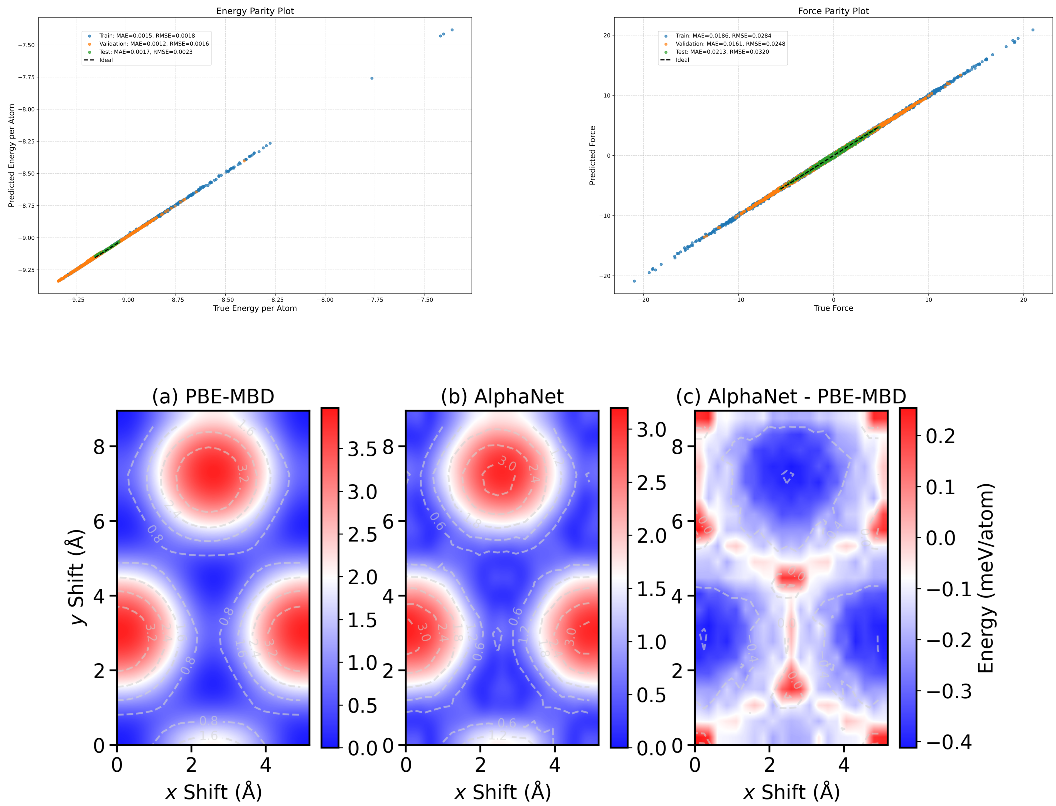


Figure 3: Comparisons between AlphaNet predictions and DFT calculations. Parity plots for (a) total energies and (b) atomic forces obtained for the training (blue), validation (orange), and test (green) data sets.(c) AlphaNet (middle) and DFT (left) sliding PES for bilayer graphene, and their difference (right).

in learning a function $f : \mathbb{R}^{n\times d} \times \mathbb{R}^{n\times 3} \rightarrow \mathbb{R}$ which predicts the energy of the current state of the molecular system readily to be used for simulation. The force can be further calculated by $-\nabla_x f(x)$, thus often referred as to machine learning force fields (MLFFs).

Message-passing neural networks. One common way to model molecular systems is through message-passing neural networks, which is a general neural architecture that leverages both permutation equivariance and locality in graph-structured data. Specifically, there are two essential operations in message-passing neural networks, message calculation and message update. The message calculation operation encodes structured messages from local environments $m_i = \oplus_{j\in\mathcal{N}(i)} f_m(f_h(x_i), h(x_j), e_{ij})$ where i and j denotes the index of different nodes, e_{ij} denotes optional edge features and f_m and f_h denotes neural networks that encode both edge and node features; The message update aggregates the incoming messages from the neighbors $m_i = f_u(m_i)$, where f_u is a neural network.

Equivariant message-passing neural networks. Equivariance is another crucial property to building MLFFs as molecular systems are invariant to the Euclidean group. A function f is said to be equivariant with respect to certain group G if $f \circ g(x) = g \circ f(x), \forall x \in X,\, g \in G$. One natural property of message-passing neural networks is the permutation equivariance such that changing the order of the input set of nodes do not result in difference output for the same node as it only depends on the neighbors which are unchanged under permutation. In addition to permutation, we need to consider the special Euclidean group in 3D, SE(3), which includes rotation and translation operations. We do so by leveraging local frames [10]. Specifically, we build a set of equivariant and complete frames and scalarize the geometric quantity which does not break SE(3)-equivariance under nonlinear neural network encodings.

Table 4: Performance of Deep Pot and AlphaNet on the Zeolite Dataset

Dataset	Metric	Deep Pot	AlphaNet
ABWopt	Energy MAE	90	**71**
	Force MAE	90	**54**
BCTopt	Energy MAE	110	**73**
	Force MAE	**50**	55
BPHopt	Energy MAE	210	**41**
	Force MAE	**60**	90
CANopt	Energy MAE	150	**62**
	Force MAE	90	**10**
EDIopt	Energy MAE	40	**5.4**
	Force MAE	50	**59**
FERopt	Energy MAE	290	**100**
	Force MAE	130	**18**
GISopt	Energy MAE	60	**56**
	Force MAE	50	**3.1**
JBWopt	Energy MAE	150	**18**
	Force MAE	70	**50**
LOSopt	Energy MAE	**110**	130
	Force MAE	70	**46**
LTAopt	Energy MAE	150	**90**
	Force MAE	64	**60**
LTJopt	Energy MAE	70	**1.2**
	Force MAE	110	**60**
NATopt	Energy MAE	210	**47**
	Force MAE	110	**9.9**
PARopt	Energy MAE	90	**6.5**
	Force MAE	70	**3.5**
PHIopt	Energy MAE	60	**16**
	Force MAE	120	**50**
SODopt	Energy MAE	200	**9.5**
	Force MAE	110	**21**
THOopt	Energy MAE	160	**70**
	Force MAE	60	**47**

To build equivariant and complete frames from atomic positions, we consider the following edge-based frames following [10]

$$\mathcal{F}_{ij} = \left(\hat{e}^1_{ij}, \hat{e}^2_{ij}, \hat{e}^3_{ij}\right) = \left(\frac{x_i - x_j}{\|x_i - x_j\|}, \frac{(x_i - \overline{x}) \times (x_j - \overline{x})}{\|(x_i - \overline{x}) \times (x_j - \overline{x})\|}, \frac{(x_i - x_j) \times ((x_i - \overline{x}) \times (x_j - \overline{x}))}{\|(x_i - x_j) \times ((x_i - \overline{x}) \times (x_j - \overline{x}))\|} \right)$$

To scalarize geometric quantity such as a vector v, we take the inner product between it and the frames which results in a set of invariant scalars and can be transformed back through a tensorization operation

$$\mathrm{Scalarize}(v, \mathcal{F}_{ij}) = (s^1_{ij}, s^2_{ij}, s^3_{ij}) = (e^1_{ij} \cdot v, e^2_{ij} \cdot v, e^3_{ij} \cdot v)$$

$$\mathrm{Tensorize}(s_{ij}, \mathcal{F}_{ij}) = s^1_{ij} e^1_{ij} + s^2_{ij} e^2_{ij} + s^3_{ij} e^3_{ij}$$

Efficient and expressive equivariant message-passing neural networks. Following [11], we leverage an efficient and expressive extension of the above equivariant message-passing neural networks based on local frames. We consider two modules to improve the expressiveness while

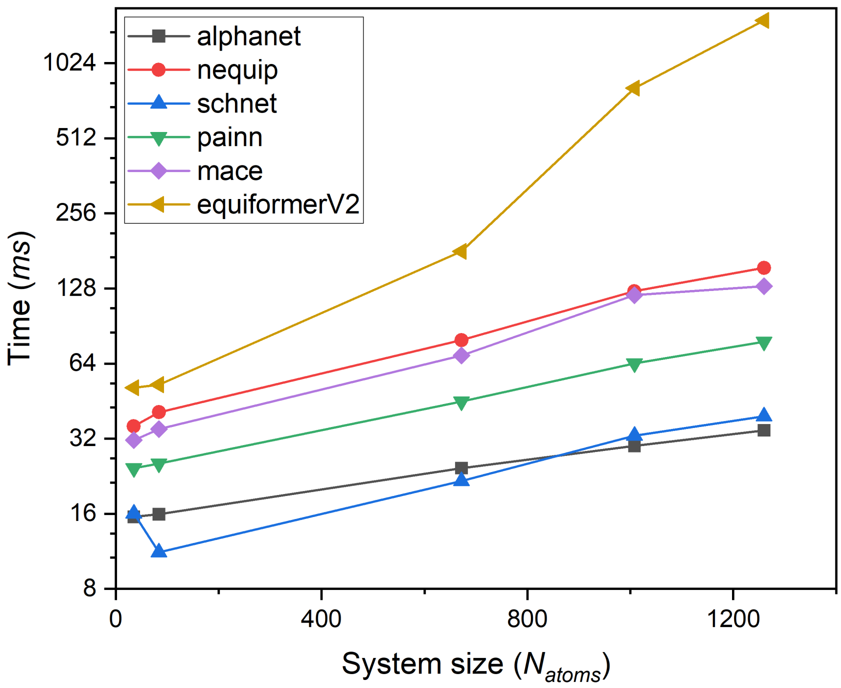


Figure 4: Comparison of inference speed for various machine learning force fields. The x-axis represents the number of atoms in the systems, while the y-axis shows the average time required to predict energy and forces for systems in batches of 10, across a total of 200 batches.

maintaining the efficiency of the network. Specifically, we leverage a local structure encoding module which scalarizes not only features from local neighbors but also from overlapping neighbors along each edge, denoted as $A_{ij} := f_l(\text{Scalarize}(S_{i-j}, \mathcal{F}_{ij}))$, where f_l denotes a neural network and S_{i-j} represents overlapping neighbors of node i and j. In addition, we use an additional frame transition module which encodes the alignment between each neighboring frame to improve the expressiveness. We realize this by incorporating vector-based message passing to amortize the cost where $\mathbf{m}_{ij}$ denotes equivariant edge features. This framework is highly efficient as it avoids expensive higher-body message-passing neural networks and higher-order tensor updates.

$$\{h_i, x_i\} = \{f_u(\oplus_{j\in\mathcal{N}(i)} f_m(h(x_i), h(x_j), A_{ij}, e_{ij})), \oplus_{j\in\mathcal{N}(i)} f_{\mathbf{m}}(h(x_i), h(x_j), A_{ij}, e_{ij})\mathbf{m}_{ij}\}$$

Rotary positional embedding as invariant frame transition. Positional embedding are known to improve and stabilize the training of transformers [27]. We introduce the elegant relationship between the commonly used Rotary Position Embedding (RoPE) and frame transition we use in our architecture [30]. In fact the frame transition can be considered as a generalized rotary positional embedding for 3D equivariant frames. Note that the spirit of rotary position embedding is to design an embedding scheme for relative position in sequential data. To build the rotary positional embedding, [30] utilizes the relative index between nodes x_i and x_j of a one-dimensional sequence to find a function $g(x_i, x_j, i-j)$ such that

$$\langle f_q(x_i, i), f_k(x_j, j)\rangle = g(x_i, x_j, i-j)$$

where f_q and f_k denote the neural network embeddings of query and key inside an attention layer. Fortunately, there is an explicit way of building g by multiplying $f_{\{q,k\}}(x_i, i)$ with the diagonal matrix expanded by rotation matrices [30]:

$$\begin{pmatrix} \cos\theta_n, & -\sin\theta_n \\ \sin\theta_n, & \cos\theta_n \end{pmatrix}$$

where $\theta_n = 10000^{-2\frac{(n-1)}{d}}, \quad n \in [1, 2, \dots, \frac{d}{2}]$, and d is the dimension of features. From a geometric point of view, the index i of x_i provides a one-dimensional position embedding such that the inner product $\langle f_q(x_i, i), f_k(x_j, j)\rangle$ only depend on the relative position $i - j$. In the three-dimensional world, we can utilize the frame transition matrices between x_i and x_j that plays a similar rule of the index difference $i - j$ in a one-dimensional sequence. In addition, we introduce the rotary positional embedding on invariant features h and represent the rotation by the multiplication of complex numbers. We predict an additional complex number from our invariant features h_i for each node i and multiply it by the scalarized features h_i. Similar to the equivariant message passing in each layer of the network, the learned rotary embedding is also applied in each layer.

Temporal connection. In addition to commonly adapted message passing in the spatial domain, we consider another dimension, temporal domain, which can be described by the depth of the neural network [6]. As message-passing neural networks are local, we believe it is beneficial to integrate multi-scale information. Specifically, for each consecutive layer, we learn a kernel that linear transforms and aggregates features. Given the invariant feature embeddings h^{l_1} at layer l_1 and $h^{(l_2)}$ at layer l_2, the kernel is of the matrix product form: $M^{(l_{3,1,2})}$ which transforms $h^{(l_1)}$ and $h^{(l_2)}$ to a new feature $h^{(l_3)}$ and added as a residual into the next layer:

$$h^{(l_3)} = \sum_{k_1=1}^{K_1} \left(\sum_{k_2=1}^{K_2} M^{(l_{3,1,2})}_{k_1,k_2} \cdot h^{(l_2)}_{k_2} \right) \cdot h^{(l_1)}_{k_1}$$

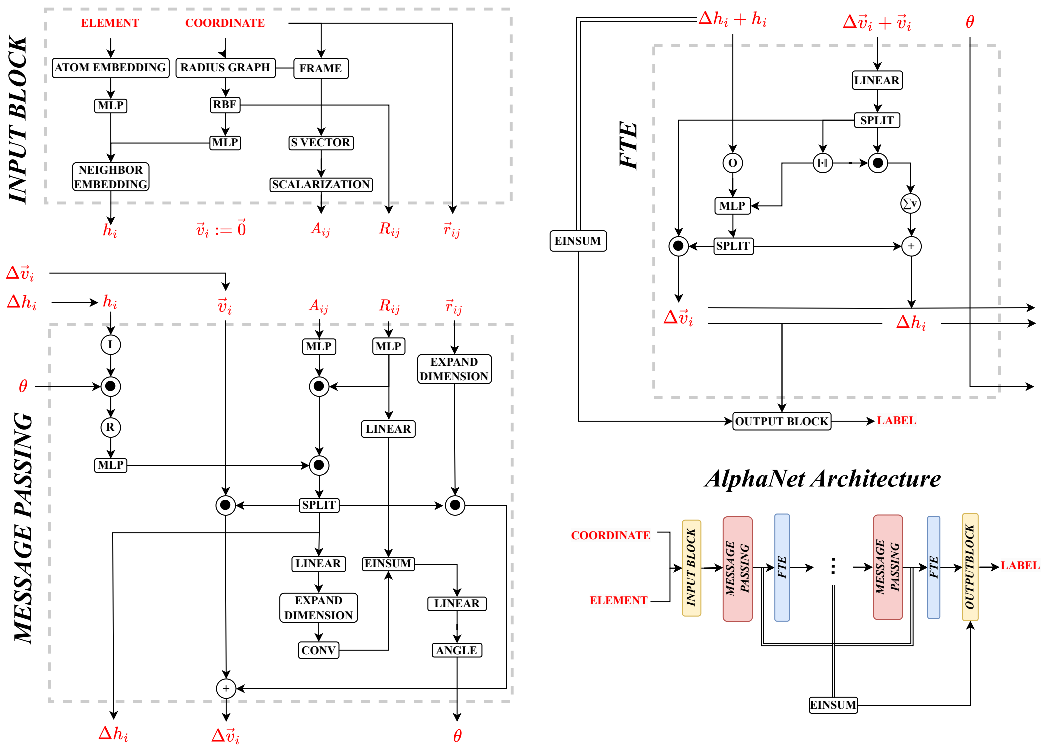


Figure 5: Overview of AlphaNet framework. We first process the input atomic types and coordinates to scalarized features and frames and pass them further to a loop of message passing and frame transition layers, followed by an output block with a temporal connection to the final output. I/R denotes imaginary and real number, $\|\bullet\|$ denotes norm, $\bullet$ denotes element-wise multiplication, and $\circ$ denotes vector scaling.

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 Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

[2] Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022.

[3] Mauro Boero, Michele Parrinello, and Kiyoyuki Terakura. First principles molecular dynamics study of ziegler natta heterogeneous catalysis. *Journal of the American Chemical Society*, 120(12):2746–2752, 1998.

[4] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.*, 55:2471–2474, Nov 1985.

[5] Lowik Chanussot, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, Javier Heras-Domingo, Caleb Ho, Weihua Hu, Aini Palizhati, Anuroop Sriram, Brandon Wood, Junwoong Yoon, Devi Parikh, C. Lawrence Zitnick, and Zachary Ulissi. Open catalyst 2020 (oc20) dataset and community challenges. *ACS Catalysis*, 11(10):6059–6072, May 2021.

[6] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

[7] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[8] B. Deng, P. Zhong, K. Jun, et al. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence*, 5:1031–1041, 2023.

[9] Weitao Du, He Zhang, Yuanqi Du, Qi Meng, Wei Chen, Bin Shao, and Tie-Yan Liu. Equivariant vector field network for many-body system modeling. *CoRR*, abs/2110.14811, 2021.

[10] 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. PMLR, 2022.

[11] 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, 2024.

[12] Alexandre Agm Duval, Victor Schmidt, Alex Hernández-García, Santiago Miret, Fragkiskos D. Malliaros, Yoshua Bengio, and David Rolnick. FAENet: Frame averaging equivariant GNN for materials modeling. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 9013–9033. PMLR, 23–29 Jul 2023.

[13] Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems*, 34:6790–6802, 2021.

[14] Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations (ICLR)*, 2020.

[15] Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks, 2022.

[16] Graeme Henkelman, Blas P. Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics*, 113(22):9901–9904, 12 2000.

[17] J. E. Jones and Sydney Chapman. On the determination of molecular fields. —ii. from the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 106(738):463–477, 1924.

[18] John Jumper, Richard Evans, Alexander Pritzel, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596:583–589, 2021.

[19] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, Oct 1996.

[20] G. Kresse and J. Hafner. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B*, 48:13115–13118, Nov 1993.

[21] Yi-Lun Liao, Brandon M Wood, Abhishek Das, and Tess Smidt. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. In *The Twelfth International Conference on Learning Representations*.

[22] Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, and David E. Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

[23] A. D. Jr. MacKerell, D. Bashford, M. Bellott, R. L. Jr. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998. PMID: 24889800.

[24] William Richards, Tatsuya Tsujimura, Lincoln Miara, et al. Design and synthesis of the superionic conductor na10snp2s12. *Nature Communications*, 7:11009, 2016.

[25] Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. *CoRR*, abs/2102.09844, 2021.

[26] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 03 2018.

[27] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. Association for Computational Linguistics, 2018.

[28] Nima Shoghi, Adeesh Kolluru, John R. Kitchin, Zachary W. Ulissi, C. Lawrence Zitnick, and Brandon M. Wood. From molecules to materials: Pre-training large generalizable models for atomic property prediction, 2024.

[29] Justin S Smith, Olexandr Isayev, and Adrian E Roitberg. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chemical science*, 8(4):3192–3203, 2017.

[30] Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

[31] Nathaniel Thomas, Tess E. 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. *CoRR*, abs/1802.08219, 2018.

[32] Han Wang, Linfeng Zhang, Jiequn Han, and E Weinan. Deepmd-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Computer Physics Communications*, 228:178–184, 2018.

[33] weitao Du, Yuanqi Du, Limei Wang, Dieqiao Feng, Guifeng Wang, Shuiwang Ji, Carla P Gomes, and Zhi-Ming Ma. A new perspective on building efficient and expressive 3d equivariant graph neural networks. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[34] Penghua Ying, Amir Natan, Oded Hod, and Michael Urbakh. Effect of interlayer bonding on superlubric sliding of graphene contacts: A machine-learning potential study. *ACS Nano*, 18(14):10133–10141, 2024. PMID: 38546136.

[35] D. Zhang, H. Bi, F. Z. Dai, et al. Pretraining of attention-based deep learning potential model for molecular simulation. *npj Computational Materials*, 10(94), 2024.