npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-025-01817-w

AlphaNet: scaling up local-frame-based neural network interatomic potentials

Bangchen Yin[1,10], Jiaao Wang[2,10] ✉, Weitao Du[3], Pengbo Wang[4], Penghua Ying[5], Haojun Jia[6], Zisheng Zhang[7,8], Yuanqi Du[9] ✉, Carla Gomes[9], Chenru Duan[6] ✉, Graeme Henkelman[2] ✉ & Hai Xiao[1] ✉

Molecular dynamics simulations demand an unprecedented combination of accuracy and scalability to tackle grand challenges in catalysis and materials design. To bridge this gap, we present AlphaNet, a local-frame-based equivariant model that simultaneously improves computational efficiency and predictive precision for interatomic interactions. By constructing equivariant local frames with learnable geometric transitions and enabling contractions through spatial domain and temporal domain, AlphaNet enhances the representational capacity of atomic environments, achieving state-of-the-art accuracy in energy and force predictions. Extensive benchmarks on large-scale datasets spanning molecular reactions, crystal stability, and surface catalysis (Matbench Discovery and OC2M) demonstrate its superior performance over existing neural network interatomic potentials while ensuring scalability across diverse system sizes with varying types of interatomic interactions. The synergy of accuracy, efficiency, and transferability positions AlphaNet as a transformative tool for modeling multiscale phenomena, decoding dynamics in catalysis and functional interfaces, with direct implications for accelerating the discovery of complex molecular systems and functional materials. Our code and data are available at https://github.com/zmyybc/AlphaNet.

Molecular dynamics (MD) simulations have become essential for investigating and elucidating complex phenomena across diverse fields, including biology, catalysis, and energy engineering[1–8]. While atomic forces necessary for MD simulations can be derived from quantum mechanical approaches such as density functional theory (DFT)[9], the computational cost associated with such first-principles methods severely limits their applications to small-sized systems (~10^3 atoms) and short timescales (~10^1 ps). Consequently, a wide spectrum of phenomena occurring over much larger 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 trajectories[10–15]. However, the simplicity of these models often compromises their accuracy, necessitating a delicate trade-off between computational efficiency and the fidelity of the simulated dynamics, especially for reactive chemistry[14,16–18].

The advent of machine learning techniques has introduced a promising solution to this dilemma. By training models on data derived from first-principles calculations, neural network interatomic potentials (NNIPs) can potentially achieve the accuracy of first-principles methods while maintaining the computational efficiency of classical force fields[19–25]. Unlike traditional models that rely on explicit functional forms to describe the bonded and non-bonded interactions, NNIPs are more flexible, learning to predict interactions based on the positions of atoms and their chemical identities[26].

As interest in applying NNIPs 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[27], is a large-scale pre-trained model with improved attention architecture over DeepPot[28]. Similarly, JMP[29] leverages diverse molecular systems of different types as a joint pre-training model. MACE[19], on the other hand, introduces higher-order message passing as a complete basis of many-body atomic

[1]Department of Chemistry, Tsinghua University, Beijing, China. [2]Department of Chemistry and the Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, USA. [3]DAMO Academy, Beijing, China. [4]Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong, China. [5]Department of Physical Chemistry, Tel Aviv University, Tel Aviv, Israel. [6]Deep Principle Inc., Cambridge, MA, USA. [7]Department of Chemical Engineering, Stanford University, Stanford, CA, USA. [8]SLAC National Accelerator Laboratory, Menlo Park, CA, USA. [9]Department of Computer Science, Cornell University, Ithaca, NY, USA. [10]These authors contributed equally: Bangchen Yin, Jiaao Wang. ✉e-mail: wangjiaao0720@utexas.edu; yd392@cornell.edu; duanchenru@deepprinciple.com; henkelman@utexas.edu; haixiao@tsinghua.edu.cn

interactions. Collectively, these models represent significant strides in the evolution of NNIP-driven atomistic simulations[20,30–37].

Nevertheless, the balance between computational efficiency and accuracy is key to practical applications of these models. On one side, efficient yet less expressive models are more suitable for tasks that weigh computational cost more than accuracy. On the other side, less efficient yet expressive models are more suitable for tasks where accuracy is more important. Atomistic systems in 3D Euclidean space are invariant to Euclidean symmetries including rotation, translation and reflection. Thus, equivariant models are developed to respect these symmetries[38,39], and most of these expressive models by date[19,20,31] are based on the spherical harmonics[40], but calculating the tensor product of irreducible representations imposes expensive computational overheads. While another branch of work achieves rotation equivariance through building equivariant frames that can be either local or global[41–44], and the main benefit of frame-based approaches is eliminating the necessity to involve the tensor product, thus greatly improving the computational efficiency.

Herein, we propose a local-frame-based equivariant atomistic model, named AlphaNet, for accurate yet highly efficient atomistic simulations. Building on the success of frame-based NNIPs in small atomistic systems, we introduce an additional rotary position embedding to enable multi-body message passing and temporal connection for multi-scale modeling. Extensive quantitative benchmarks demonstrate that AlphaNet excels at accuracy, efficiency and scalability compared with the state-of-the-art (SOTA) NNIP models on a variety of datasets, including formate decomposition, defected graphenes, zeolites, Open Catalyst 2020 (OC20), and Matbench discovery WBM test set, which cover various types of interatomic interactions across a rich set of systems and phenomena, ranging from molecules and bulks to surfaces, from gas-phase reactions and crystal stability to surface catalysis.

Results

We systematically assess the capabilities of AlphaNet by examining its prediction accuracy, scalability with respect to model and dataset sizes, as well as computational efficiency in terms of inference speed and memory usage. Initially, we validate atomic-level prediction accuracy across five diverse chemical systems, simultaneously demonstrating AlphaNet's scalability concerning dataset and system sizes. Subsequently, we quantify computational efficiency by benchmarking both the memory footprint and inference costs of the model. Comprehensive details regarding datasets, comparative methodologies, and hyperparameter configurations are provided in Supplementary Information.

High accuracy of AlphaNet across all datasets

We first consider 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 ⟨110⟩ surface. For this dataset, AlphaNet achieves a mean absolute error (MAE) of 42.5 meV/Åfor force and 0.23 meV/atom for energy, compared to NequIP[20] 47.3 meV/Åand 0.50 meV/atom, respectively, as shown in Table 1. These results highlight AlphaNet's capability to model heterogeneous systems involving metallic and covalent bonding, as well as complex charge transfer processes. This high degree of accuracy highlights the model's suitability for simulating catalytic reactions with multiple interaction types.

A further challenge for NNIPs is the accurate modeling of inter-layer sliding effects in layered materials such as defected graphene, where interlayer interactions are essential. On the defected graphene dataset, AlphaNet attains a force MAE of 19.4 meV/Å and an energy MAE of 1.2 meV/atom, significantly outperforming NequIP's 60.2 meV/Å and 1.9 meV/atom, respectively. Additionally, AlphaNet successfully reproduces the binding energy profile for AB-stacked bilayer graphene, benchmarked against PBE +MBD calculations. Deviations within the shallow sliding potential energy landscape are less than 0.4 meV/atom, highlighting AlphaNet's robust capabilities in modeling subtle interlayer forces and complex structural dynamics (see Fig. 1).

The third dataset evaluated is the zeolite dataset, consisting of 16 distinct zeolite types and comprising a total of 800,000 configurations. As summarized in Table 2, AlphaNet demonstrates significantly improved performance compared to Deep Potential[28], and a around 20% improvement on other equivariant models. The detailed performance on each system is demonstrated in Supplementary Table 3, where we can see our model shows best performance on 13 out of 16 systems. High temperature (4500K) molecular dynamics were also conducted to further validate the models' performance (Supplementary Fig. 1).

To address the next challenge, we consider the Open Catalyst Project OC20 dataset[45], a comprehensive collection of crystal structures specifically curated for training NNIP models in catalysis applications. Our study focuses on the OC2M subset and evaluates performance on the Structure-to-Energy-and-Force (S2EF) task. We benchmark our proposed model, AlphaNet-M, against established state-of-the-art methods, including EquiformerV2[31], EScAIP[30], eSCN[46], GemNet-OC[47], SchNet[33], and DimeNet ++[48]. As shown in Table 3 AlphaNet-M, trained on the 2M subset for 2,000,000 steps, achieves remarkable accuracy in energy prediction, yielding a mean absolute error (MAE) of 0.24 eV. This accuracy is on par with larger-scale models such as EquiformerV2 and EScAIP. In contrast, models like SchNet and DimeNet++, even when trained on the full dataset, demonstrate inferior performance, with MAE values exceeding 0.35 eV. However, AlphaNet-M does not surpass the larger-scale models (EquiformerV2 and EScAIP) in force prediction tasks. This discrepancy likely arises from the inherent conservativeness of the model architectures, which enforce consistency by computing forces strictly as negative gradients of the predicted energies ($-\nabla_x E(x)$).

The model's performance cannot be adequately assessed based solely on accuracy metrics[49]. We extended the application of AlphaNet-S to additional materials science tasks following the methodology of Chipsff[50]. Utilizing pretrained NNIPs, we evaluated multiple material properties including bulk relaxation, elastic constants, bulk modulus (Kv) prediction, and various system-level tasks across 104 diverse materials. As evidenced in Table 4, AlphaNet-S demonstrates slightly above-average performance overall. Notably, our analysis reveals that no single model dominates across all metrics: while ALIGNN-FF and ORB v2 excel in lattice relaxation predictions, AlphaNet-S, MACE-MP, and SevenNet-l3i5 show capabilities in property prediction tasks. To mitigate potential random errors in model assessment, future work should incorporate more testing samples.

Lastly, we validate the effectiveness of AlphaNet using the established Matbench discovery WBM test set[51]. We trained two variants of AlphaNet: a large model employing Gaussian radial basis functions (AlphaNet-L) and a compact model using Bessel radial basis functions

Table 1 | Results on the formate decomposition and defected graphene datasets

Dataset	Metric	NequIP	MACE	Mattersim	Grace	AlphaNet
Formate Decomposition	Force MAE (meV/Å) ↓	47.3	54.1	46.8	53.9	**42.5**
	Energy MAE (meV/atom) ↓	0.50	0.31	0.45	0.48	**0.23**
Defected Graphene	Force MAE (meV/Å) ↓	60.2	32.7	28.3	71.5	**19.4**
	Energy MAE (meV/atom) ↓	1.9	1.4	1.7	2.3	**1.2**

Boldface indicates the best performance.

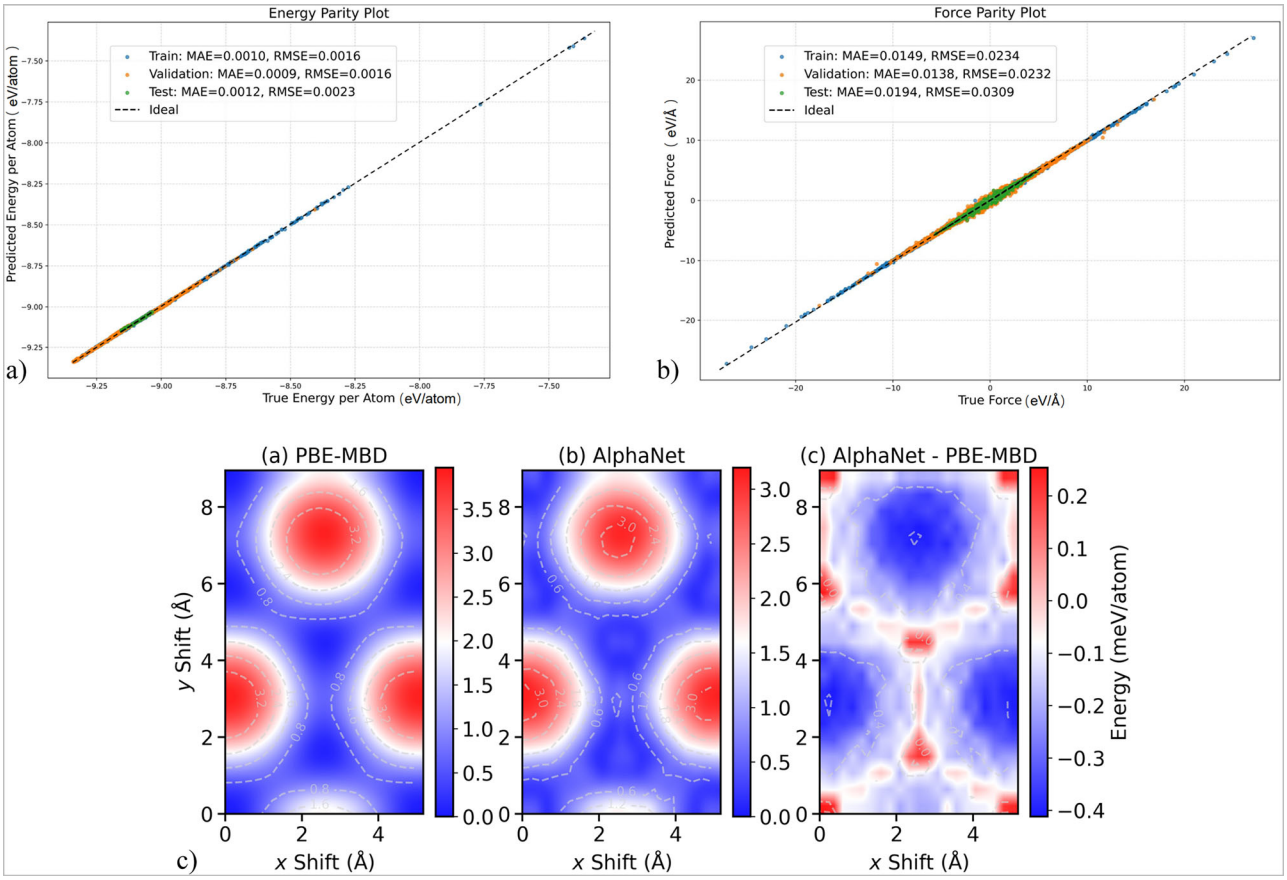

Fig. 1 | Comparisons between AlphaNet predictions and DFT calculations on the defected graphene dataset. a Parity plot for total energies (units: eV/Å) obtained for the training (blue), validation (orange), and test (green) data sets. **b** Parity plot for atomic forces (units: eV/Å) obtained for the training (blue), validation (orange), and test (green) data sets. **c** Sliding potential energy surface (PES) for bilayer graphene: DFT results (left), AlphaNet predictions (middle), and their difference (right).

Table 2 | Average performance across all Zeolite frameworks

Model	Energy MAE ↓	Force MAE ↓
AlphaNet	**17.1**	**15.9**
NequIP	19.2	17.1
MACE	20.9	18.3
MatterSim-5M	24.4	22.1
Deep Potential	134.8	81.5
GRACE-2L	25.6	23.4

Boldface indicates the best performance. (Unit for energy is meV and unit for force is meV/Å).

Table 3 | Results on the OC20 validation set

Model	Parameters ↓	Energy MAE (eV) ↓	Force MAE (eV/Å) ↓
SchNet (full)	9.1M	0.54	0.55
DimeNet++ (full)	10.1M	0.53	0.048
GemNet	38M	0.29	0.026
eSCN	51M	0.28	0.021
EScAIP-Medium	146M	0.25	0.019
EquiformerV2 ($\lambda_E = 2$)	146M	0.28	0.022
AlphaNet-M ($\lambda_E = 4$)	6.1M	0.25	0.040
AlphaNet-M ($\lambda_E = 20$)	6.1M	0.24	0.062

"full" denotes models trained on the full OC20 training set, others are trained on the 2M subset. λ_E refers to the weight of the energy loss, the weight of the force loss λ_F is set to 100. The top 3 performances are underlined.

(AlphaNet-S). As shown in Table 5, the Bessel function variant shows consistent improvements over the Gaussian-based model across multiple metrics. AlphaNet demonstrates a combination of computational efficiency and predictive performance. While models like eSEN-30M-MP achieve impressive accuracy benchmarks (F1 = 0.831, DAF = 5.260), AlphaNet delivers competitive results with significantly fewer parameters. Specifically, AlphaNet-S achieves an F1 score of 0.808 and DAF score of 4.915 with only 4.5 million parameters, approaching the performance of computationally intensive frameworks like eqV2 S DeNS (31.2M parameters). This architectural efficiency is particularly noteworthy when considering AlphaNet's R2 performance (0.745 for L-variant, 0.796 for S-variant), which outperforms similarly sized models while approaching the accuracy of much larger networks. In the table, E, F, and S denote energy, forces, and stress as training targets, respectively, while G/D indicates whether the forces and stress are conservative. AlphaNet maintains physics-informed design within a resource-efficient framework, achieving a balance between computational cost and predictive accuracy, making it highly scalable for large-scale materials discovery.

AlphaNet exhibits effective scaling with model, data, and system sizes

In addition to demonstrating robust performance on large datasets, we systematically evaluate model performance across varying architectures and data scales using the zeolite dataset. Our composite validation loss metric (4 × Energy MAE + 100 × Force MAE) reveals several key trends in Fig. 2: (1) Increasing training data from 50k to 800k samples consistently reduces

Table 4 | Model performance comparison on elastic properties prediction

Model Type	err_a (Å)	err_b (Å)	err_c (Å)	err_vol (Å³)	err_kv (GPa)	err_c11 (GPa)	err_c44 (GPa)
AlphaNet-S	0.032	0.032	0.103	2.41	95	41	40
MACE-MP	0.027	0.027	0.075	2.85	94	42	38
SevenNet-l3i5	0.030	0.031	0.098	3.58	104	43	37
ORB v2	0.018	0.019	0.045	2.11	118	81	74
CHGNET	0.036	0.038	0.072	3.29	89	59	46
ALIGNN-FF	0.011	0.011	0.014	0.42	135	196	74

Underlined values indicate performance in the top 3 for that metric.

Table 5 | Results on the Matbench Discovery WBM unique prototype dataset

Model	F1 ↑	DAF ↑	Prec ↑	Acc ↑	MAE ↓	R2 ↑	Size ↓	Type
CHGNet	0.613	3.361	0.514	0.851	0.063	0.689	413k	EFSGM
MACE-MP-0	0.669	3.777	0.577	0.878	0.057	0.697	4.69M	EFSG
GRACE-2L-MPtrj	0.691	4.163	0.636	0.896	0.052	0.741	15.3M	EFSG
SevenNet-0	0.724	4.252	0.65	0.904	0.048	0.75	842k	EFSG
SevenNet-l3i5	0.76	4.629	0.708	0.92	0.044	0.776	1.17M	EFSG
ORB MPtrj	0.765	4.702	0.719	0.922	0.045	0.756	25.2M	EFSD
DPA3-v1-MPtrj	0.765	4.654	0.711	0.921	0.042	0.798	3.37M	EFSG
eqV2 S DeNS	0.815	5.042	0.771	0.941	0.036	0.788	31.2M	EFSD
EScAIP	0.782	5.634	0.712	0.939	0.038	0.783	45M	EFSD
eSEN-30M-MP	0.831	5.260	0.804	0.946	0.033	0.822	30.1M	EFSG
AlphaNet-L	0.799	4.863	0.743	0.933	0.041	0.745	16.2M	EFSG
AlphaNet-S	0.808	4.915	0.751	0.935	0.037	0.796	4.5M	EFSG

Boldface indicates the best performance. Underlined values indicate performance in the top 3 for that metric.

validation loss from ~ 20 to < 3; (2) With 64 hidden channels in Fig. 2a, 3-layer models marginally outperform 4-layer (Δ = 0.05), while 5-layer models perform similarly to 2-layer architectures; (3) Expanding to 128–256 channels improves all models in Fig. 2b to 2d, with diminishing returns beyond 176 channels; (4) Deeper models (5-layer) show accelerated improvement, ultimately achieving slight superiority. These results suggest deeper architectures require larger hidden dimensions for optimal performance, though 3–4 layer models with 128 channels offer the best efficiency-performance tradeoff.

Our systematic ablation studies reveal Rotary Position Embedding (RoPE) as the most critical technical component, consistently enhancing performance across all model configurations-a finding further substantiated by gradient clipping analysis presented in Supplementary Table 4 and Supplementary Fig. 2. From Tables 6 and 7, we can see that the data demonstrates striking variation in RoPE's effectiveness depending on dataset characteristics and model architecture: While performance on the relatively simple Formate and Graphite systems shows limited sensitivity to model depth and hidden channel dimensions, the structurally diverse Zeolite dataset exhibits scaling behavior where increased model capacity (particularly beyond 3 layers and 128 hidden channels) yields substantial improvements, with RoPE amplifying these gains most significantly in complex systems. Specifically, RoPE's impact hierarchy (Zeo. > Gra. > For.) directly correlates with configuration variation, delivering up to 1.9 validation loss reduction in 5-layer Zeo. models. Specifically, configuration variation refers to the differences in the number of atoms per structure within the training dataset. In the For. dataset, all structures have the same number of atoms. In the Gra. dataset, the smallest structure contains only 4 atoms, while the largest has 98 atoms. In the Zeo. dataset, the number of atoms ranges from 64 to 256. This amplification effect becomes increasingly pronounced in larger architectures, as evidenced by the progressive benefit shown in deeper layers (0.6 to1.8 loss reduction from L2 to L5 in Zeo.) and wider channels (1.4 to1.9 improvement from 64 to 256 channels in 4-layer models), though diminishing returns emerge beyond 176 channels. Consequently, optimal performance-efficiency balance is achieved at 4 layers with 128 hidden channels, or more than 3 layers with 176 hidden channels, leveraging RoPE's architectural advantages without incurring excessive computational costs.

AlphaNet demonstrates an enhanced speed-accuracy trade-off

The primary advantage of NNIP models over density functional theory (DFT)-calculated energies and forces is their substantially faster inference speed. However, increasing the number of parameters in an NNIP model to enhance its predictive accuracy typically results in slower inference speeds. We conducted speed benchmarking tests among pretrained models on the Matbench Leaderboard using a set of 28 distinct zeolite structures. These structures were systematically expanded through a series of 9 expansion steps until reaching a maximum of approximately 10,000 atoms per structure. The maximum atom counts at each expansion step were: 672, 3360, 4704, 6048, 6720, 8064, 9072, 10,080 and 10,752. The evaluated models include AlphaNet, SevenNet[52], EquiformerV2[31], MACE[19] and Mattersim[37].

For each pretrained model, we measured the average inference speed per structure at each expansion step. The testing continued until the model encountered Out of Memory (OOM) errors for more than half of the structures at a given step, at which point we terminated the benchmarking for that model.

All tests were performed on the same hardware configuration: a single NVIDIA A100-SXM4-40GB GPU. For consistency, we used the ASE (Atomic Simulation Environment) interface for all predictions.

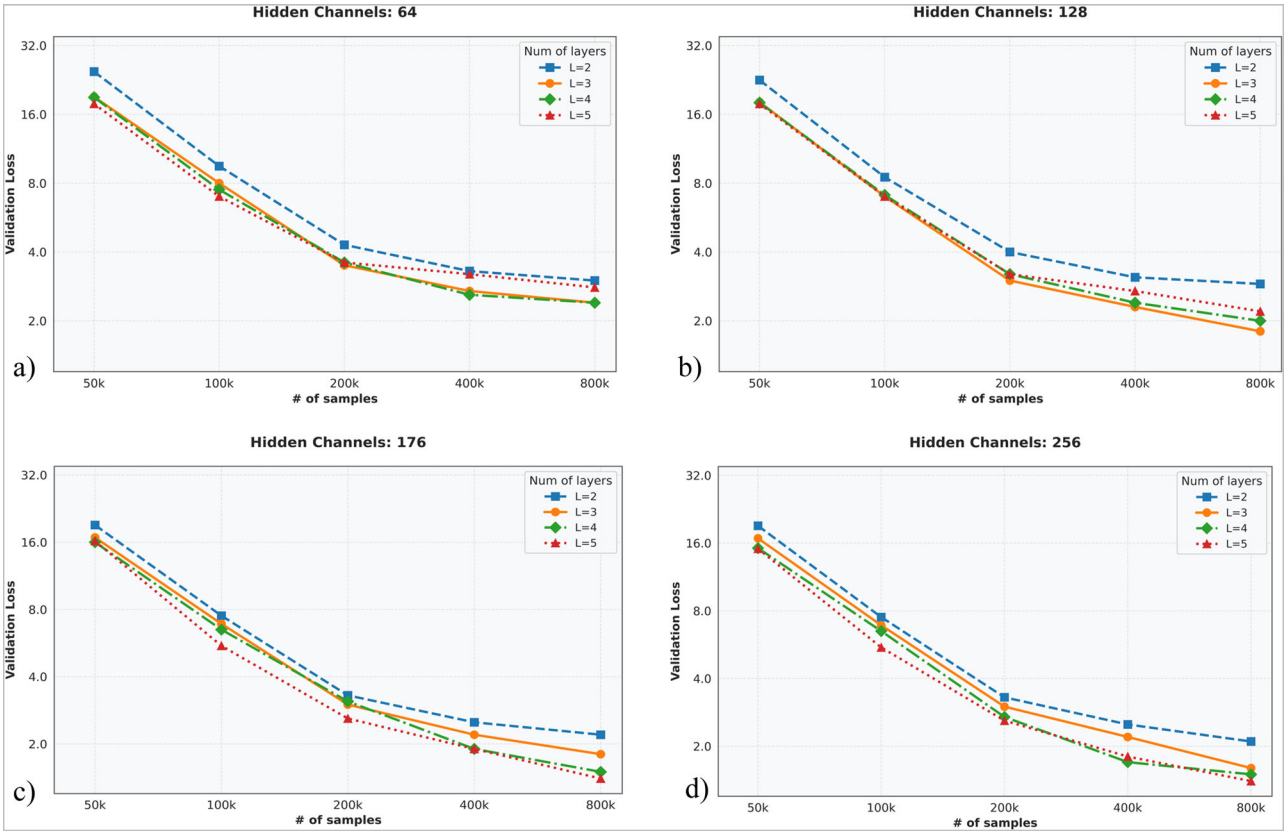

Fig. 2 | Comparative analysis of neural network performance across different hidden channel configurations. (Num of layers: 2: blue, 3: yellow, 4: green, 5: red). **a** Results for 64 hidden channels showing layer-wise performance. **b** Results for 128 hidden channels with layer differentiation. **c** Results for 176 hidden channels demonstrating layer-dependent variations. **d** Results for 256 hidden channels across model depths.

Table 6 | Ablation study on RoPE and layer depth (hidden channels = 128) across datasets (validation loss ↓)

Layers	RoPE	Gra.	For.	Zeo.
2	w/	3.4	4.4	3.3
	w/o	4.2	4.6	3.9
3	w/	3.2	4.2	1.9
	w/o	4.0	4.4	3.6
4	w/	3.2	4.2	2.0
	w/o	4.0	4.3	3.6
5	w/	3.3	4.4	2.3
	w/o	4.2	5.4	4.1
6	w/	3.3	4.5	-
	w/o	4.5	5.3	-

Dataset abbreviations: Gra. (Defected Graphene),For. (Formate Decomposition), Zeo.(Zeolite), also used in Table 7.

Table 7 | Ablation study on hidden channels of a 4-layer model across datasets (validation loss ↓)

Hidden Channels	RoPE	Gra.	For.	Zeo.
64	w/	3.3	4.5	2.2
	w/o	3.6	4.9	3.6
128	w/	3.2	4.2	2.0
	w/o	4.0	4.3	3.6
176	w/	3.2	4.2	1.5
	w/o	3.9	4.5	3.4
256	w/	3.3	4.2	1.3
	w/o	3.9	4.8	3.2

As shown in Fig. 3(a), AlphaNet-S demonstrates the fastest inference speed across all system sizes, maintaining this advantage until reaching its memory limit at approximately 10,000 atoms - a threshold shared by Mattersim-v1-5M. The larger parameter models - AlphaNet-L, EquiformerV2 Small, and ORB v2 - exhibit significantly higher computational demands, encountering Out-of-Memory (OOM) errors within the first three expansion steps (around 4000 atoms). These models show a more rapid increase in computational time relative to system size compared to other architectures. Notably, MACE-MP-0 emerges as the sole model capable of successfully processing systems exceeding 10,000 atoms, though its inference speed is more than twice as slow as AlphaNet-S for equivalent system sizes. Furthermore, we assess the inference speed of pre-trained models featured on the Matbench Discovery leaderboard[51]. To ensure a fair comparison, we select compliant, open-source models that are freely accessible without additional authorization. The average energy prediction speeds of these models are presented in Fig. 3b. Among these, GRACE implemented using TensorFlow demonstrates significantly faster inference compared to the other models, which are built on PyTorch, and our model, AlphaNet-S, comes in the second place. In particular, AlphaNet-L exhibits an inference speed comparable to models possessing approximately ten times fewer parameters, such as SevenNet-l3i5. This finding indicates that AlphaNet maintains competitive computational efficiency despite its relatively larger size, raising an interesting prospect for investigating its potential performance with reduced parameters in future studies.

The final efficiency evaluation is conducted using the OC20 dataset, with results summarized in Table 8. Notably, EScAIP demonstrates

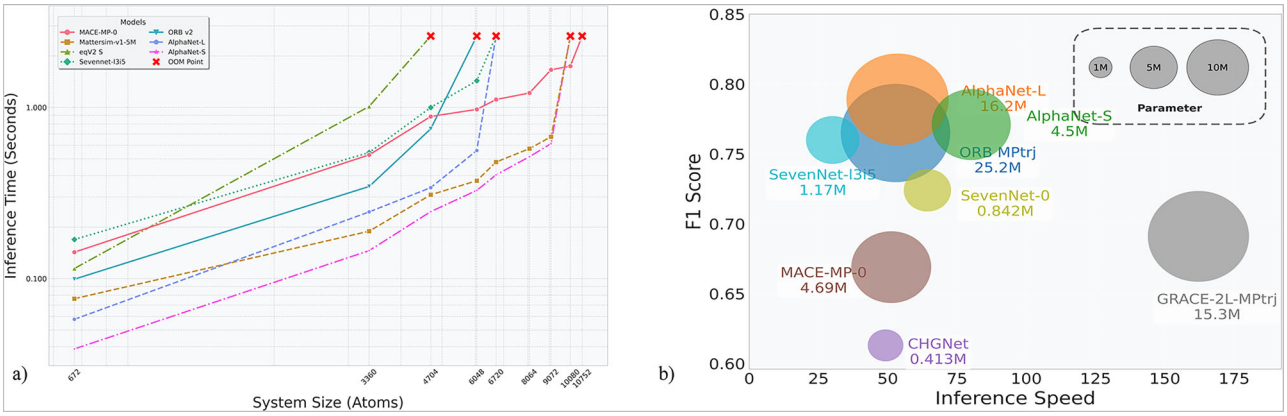

Fig. 3 | Computational performance benchmarking of neural network interatomic potentials (NNIPs). a Inference time versus system size scaling on zeolite structures. The x-axis represents the maximum system size (number of atoms) at each expansion step, while the y-axis shows the computational time (units: second) in energy evaluations per structure. **b** Speed-accuracy trade-off analysis on Matbench Discovery benchmark. All models comply with the official leaderboard requirements. Performance is measured in energy evaluations per second (x-axis) versus prediction accuracy (F1 score) (y-axis).

Table 8 | Efficiency evaluated on the OC20 dataset, using a single A100 40GB GPU

Models	Parameters ↓	Training Speed (Sample/Sec) ↑	Training Memory (GB/Sample) ↓	Inference Speed (Sample/Sec) ↑	Inference Memory (GB/Sample) ↓	Conserv.
SchNet	9.1M	39	2.5	112.1	1.7	✓
DimeNet++	10.1M	29	3.6	92	2.5	✓
GemNet-OC	38M	16.22	1.63	39.2	1.61	✓
eSCN	51M	8.5	1.75	25.2	1.74	×
EScAIP	83M	37.8	1.23	111.8	1.09	×
EquiformerV2	31M	15.7	1.63	36.72	1.61	×
AlphaNet	6.1M	41.2	2.2	120.1	1.5	✓

Conserv. denotes energy conservation. Underlined values indicate performance in the top 3 for that metric.

outstanding efficiency in both inference speed and memory usage, even given its substantial model size. AlphaNet exhibits a marginally faster inference speed; however, it incurs higher memory consumption compared to EScAIP. It is important to emphasize that models such as AlphaNet, SchNet, and DimeNet++ are conservative in their force calculations, inherently relying on PyTorch's autograd functionality, thus necessitating additional computational overhead and consequently increasing inference time and memory usage.

Discussion

AlphaNet represents a significant advance in scalable neural network interatomic potentials, achieving exceptional accuracy while maintaining computational efficiency. Its success stems from a novel local-frame-based architecture (as summarized in Fig. 4) that scalarizes geometric information within local frames and effectively aggregates these frames to produce features sensitive to both local and global chemical environments. Leveraging this architecture, AlphaNet adeptly captures complex atomic interactions across diverse chemical bonding types, including metallic, covalent, ionic, and long-range interactions, demonstrating versatility and broad applicability.

One of AlphaNet's key strengths is its scalability. The model consistently maintains high performance on a varying dataset size and complexity of atomic systems. Remarkably, AlphaNet achieves robust accuracy while utilizing fewer parameters compared to conventional approaches, enhancing its computational efficiency. This capability to effectively manage large-scale datasets without compromising performance positions AlphaNet as a highly promising framework for extensive simulations and practical real-world applications.

In future work, we plan to extend the model's capabilities to account for additional interaction types, notably hydrogen bonding, which plays a critical role across diverse chemical and biological systems. Another important direction is enhancing the interpretability of the proposed model. Although our current implementation demonstrates strong predictive performance, its applicability and reliability in practical scenarios can be significantly improved by achieving deeper insight into how the model identifies, processes, and leverages atomic-level interactions.

Additionally, addressing the efficient use of GPU memory during training presents a long-term challenge, particularly for large systems exceeding 10,000 atoms-even when leveraging parallelization across multiple GPUs. This limitation, common among most message-passing neural networks, must be overcome to facilitate scalability toward larger and more complex systems. Future research will prioritize memory optimization strategies to surmount this bottleneck, ensuring broader applicability and enabling analysis of increasingly intricate molecular systems.

Methods

Framework formulation and relevant concepts

Neural network interatomic potentials provide a way to describe a single state of a molecular system 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 in learning a function $f : \mathbb{R}^{n\times d} \times \mathbb{R}^{n\times 3} \to \mathbb{R}$ 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 neural network interatomic potentials (NNIPs). The subsequent paragraphs delineate the essential components of our proposed model.

One common way to model molecular systems is through message-passing neural networks, which are general neural architectures that leverage both permutation equivariance and locality in graph-structured data. Specifically, there are two essential operations in message-passing

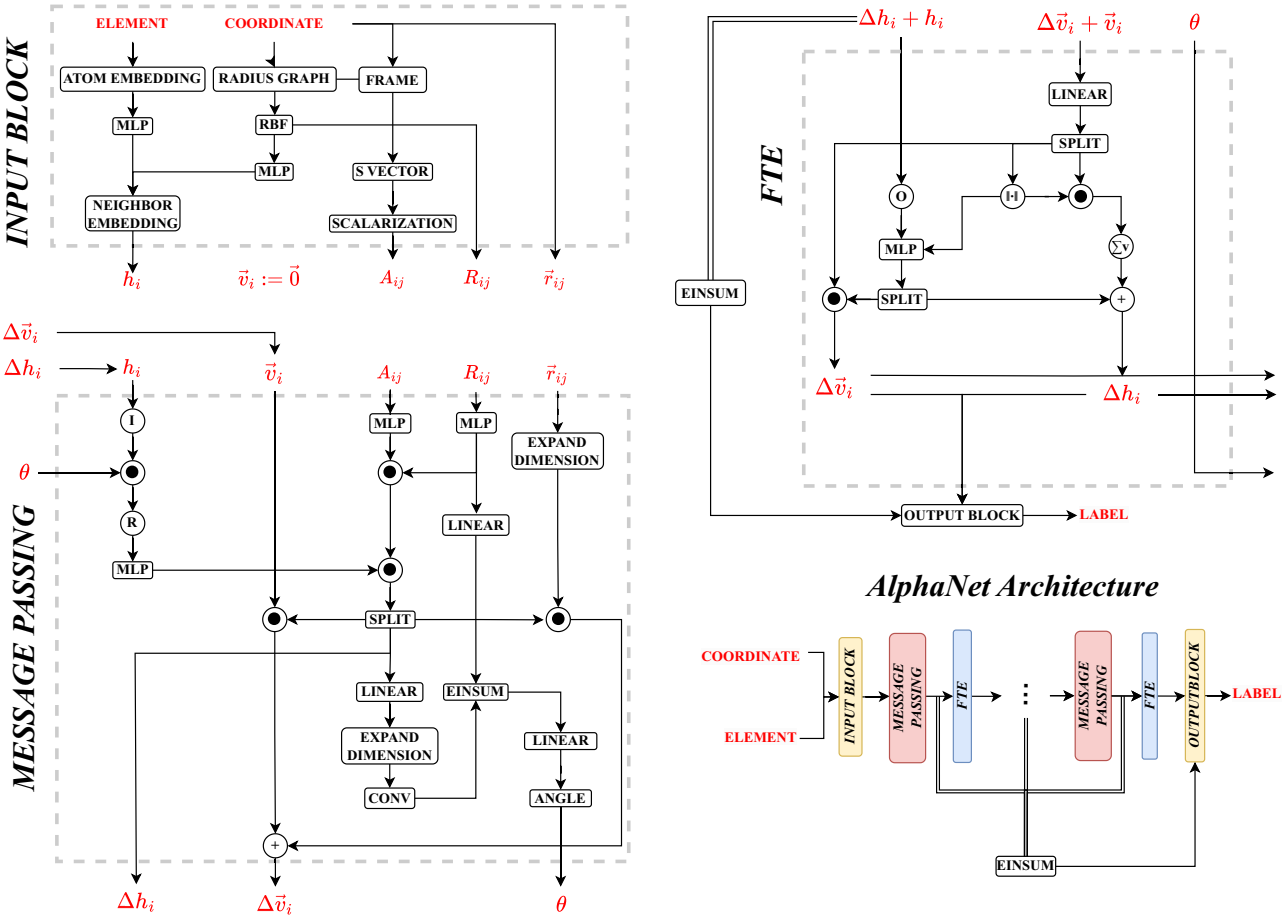

Fig. 4 | Overview of AlphaNet framework. We first process the input atomic types and coordinates to scalarized features and frames and pass them further to a loop of message passing and frame transition layers, followed by an output block with a temporal connection to the final output. *I*/*R* denotes imaginary and real number, $\|\bullet\|$ denotes norm, $\bullet$ denotes element-wise multiplication, and $\circ$ denotes vector scaling.

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 denote the index of different nodes, e_{ij} denotes optional edge features and f_m and f_h denote 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.

Equivariance is another crucial property in building MLFFs as molecular systems are invariant to the Euclidean group. A function f is said to be equivariant with respect to a 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 permutation equivariance such that changing the order of the input set of nodes does not result in a different output for the same node, since 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[42]. 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.

To build equivariant and complete frames from atomic positions, we consider the following edge-based frames following[42]:

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

To scalarize geometric quantities 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 inverted via a tensorization operation:

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

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

Following[53], we further 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 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.

$$\{h_i, x_i\} = \{f_u(\oplus_{j\in\mathcal{N}(i)} f_m(h(x_i), h(x_j), A_{ij}, e_{ij})), \oplus_{j\in\mathcal{N}(i)} f_{\mathbf{m}}(h(x_i), h(x_j), A_{ij}, e_{ij})\mathbf{m}_{ij}\} \tag{4}$$

Core innovations of AlphaNet

Rotary positional embedding is introduced as multi-body message passing contraction and invariant frame transition. Positional embeddings are known to improve and stabilize the training of transformers[54]. We establish the relationship between the commonly used Rotary Position Embedding (RoPE) and the frame transition used in our architecture[55]. 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,[55] 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

$$\langle f_q(x_i, i), f_k(x_j, j)\rangle = g(x_i, x_j, i-j) \tag{5}$$

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[55]:

$$\begin{pmatrix} \cos\theta_n, & -\sin\theta_n \\ \sin\theta_n, & \cos\theta_n \end{pmatrix} \tag{6}$$

where $\theta_n = 10000^{-2^{\frac{(n-1)}{d}}}, \quad n \in [1, 2, \ldots, \frac{d}{2}]$, 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 depends 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 play a similar role to the index difference $i-j$ in a one-dimensional sequence.

From another perspective, RoPE enables multi-body message passing. Common message passing typically relies on the summation of two-body interactions. In contrast, RoPE facilitates multi-body interactions through a mechanism reminiscent of matrix product state contraction[56]. Specifically, we employ a learnable kernel to perform convolutions over neighboring nodes.

The process begins by constructing an input sequence that combines a virtual node with projected quantum state features:

$$\mathbf{Q} = [\mathbf{a}, \mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_H] \tag{7}$$

where $\mathbf{a} \in \mathbb{C}^{B\times 1\times D}$ serves as a virtual node, while $\mathbf{q}_h \in \mathbb{C}^{B\times 1\times D}$ represents the complex feature projection of the h-th head.

The first convolution operation performs MPS-like tensor contraction:

$$\mathbf{C}_{i,k} = \sum_{j=1}^{H+1}\sum_{l=1}^{D} Q_{j,l}\cdot \mathcal{K}_{1,j,l,k} \tag{8}$$

where $\mathcal{K}_1 \in \mathbb{C}^{B\times(H+1)\times D\times \chi_2}$ is the MPS-inspired kernel tensor and $\mathbf{C} \in \mathbb{C}^{B\times\chi_2}$ contains the intermediate features capturing multi-body correlations.

The interaction is enhanced through a distance-aware diagonal term:

$$\mathbf{b} = \mathbf{a}\,\mathrm{diag} \otimes \mathbf{w}\,\mathrm{diag} + \mathbf{1} \tag{9}$$

where $\mathbf{a}\,\mathrm{diag} = \mathrm{SiLU}(f\,\mathrm{diag}(\mathbf{rbf}ij)) \in \mathbb{R}^{B\times\chi_2}$ incorporates activated distance information from radial basis functions, and $\mathbf{w}\,\mathrm{diag} \in \mathbb{R}^{\chi_1}$ provides learnable diagonal weighting.

A second convolution then transforms the features:

$$\mathbf{Z} = \mathbf{C}\cdot\mathbf{D} = \sum_{k=1}^{\chi_2} C_{i,k}\cdot D_{k,l} \tag{10}$$

where $\mathbf{D} = f_{\mathrm{linear}}(\mathbf{b}) \in \mathbb{C}^{\chi_2\times\chi_1}$ is the complex transformation matrix and $\mathbf{Z} \in \mathbb{C}^{B\times\chi_1}$ represents the final complex output.

Finally. the complex features undergo final processing through separate real and imaginary transformations followed by phase extraction, which is RoPE:

$$\mathrm{RoPE} = \angle(f\mathrm{fc}(\mathrm{Re}(\mathbf{Z})) + i\cdot f_{\mathrm{fc}}(\mathrm{Im}(\mathbf{Z}))) \tag{11}$$

To summarize, we introduced 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.

In addition to the spatial domain as in RoPE, we also consider the temporal domain, which can be described by the depth of the neural network[57]. As message-passing neural networks are local, it is beneficial to integrate multi-scale information. Specifically, for each consecutive layer, we learn a kernel that linearly 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 adds it as a residual into the next layer.

$$h^{(l_3)} = \sum_{k_1=1}^{K_1}\left(\sum_{k_2=1}^{K_2} M_{k_1,k_2}^{(l_{3,1,2})}\cdot h_{k_2}^{(l_2)}\right)\cdot h_{k_1}^{(l_1)} \tag{12}$$

Data availability

Our code and data are available at https://github.com/zmyybc/AlphaNet.

Received: 2 May 2025; Accepted: 27 September 2025;
Published online: 17 November 2025

References

1. Fermi, E., Pasta, J. R. & Ulam, S. M. Studies of nonlinear problems i https://api.semanticscholar.org/CorpusID:117035863 (1955).
2. Fermi, E.*Collected Papers* Vol. II (University of Chicago Press, Chicago, IL, 1965).
3. Alder, B. J. & Wainwright, T. E. Phase transition for a hard sphere system. *J. Chem. Phys.* **27**, 1208–1209 (1957).
4. Alder, B. J. & Wainwright, T. E. Studies in molecular dynamics. I. general method. *J. Chem. Phys.* **31**, 459–466 (1959).
5. Hollingsworth, S. A. & Dror, R. O. Molecular dynamics simulation for all. *Neuron* **99**, 1129–1143 (2018).
6. Richards, W., Tsujimura, T. & Miara, L. et al. Design and synthesis of the superionic conductor na10snp2s12. *Nat. Commun.* **7**, 11009 (2016).
7. Boero, M., Parrinello, M. & Terakura, K. First principles molecular dynamics study of Ziegler Natta heterogeneous catalysis. *J. Am. Chem. Soc.* **120**, 2746–2752 (1998).
8. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science* **334**, 517–520 (2011).
9. Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
10. Jones, J. E. & Chapman, S. On the determination of molecular fields. -ii. from the equation of state of a gas. *Proc. R. Soc. Lond. Ser. A Contain. Pap. A Math. Phys. Character* **106**, 463–477 (1924).
11. Cornell, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
12. MacKerell, A. D. J. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
13. Scott, W. R. P. et al. The gromos biomolecular simulation program package. *J. Phys. Chem. A* **103**, 3596–3607 (1999).

14. van Duin, A. C. T., Dasgupta, S., Lorant, F. & Goddard, W. A. Reaxff: A reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).
15. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* **9**, 646–652 (2002).
16. Tersoff, J. Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Phys. Rev. B* **39**, 5566–5568 (1989).
17. Iftimie, R., Minary, P. & Tuckerman, M. E. Ab initio molecular dynamics: concepts, recent developments, and future trends. *Proc. Natl. Acad. Sci. USA* **102**, 6654–6659 (2005).
18. Zhao, Q. et al. Harnessing machine learning to enhance transition state search with interatomic potentials and generative models. *Adv. Sci* **12**, e06240 (2025).
19. Batatia, I., Kovacs, D. P., Simm, G. N. C., Ortner, C. & Csanyi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. In Oh, A. H., Agarwal, A., Belgrave, D. & Cho, K. (eds.) *Advances in Neural Information Processing Systems* https://openreview.net/forum?id=YPpSngE-ZU (2022).
20. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
21. Zhang, D., Liu, X. & Zhang, X. et al. Dpa-2: a large atomic model as a multi-task learner. *npj Comput. Mater.* **10**, 293 (2024).
22. Behler, J. Perspective: machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
23. Bartók, A. P. et al. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **3**, e1701816 (2017).
24. Smith, J. S., Nebgen, B. T., Zubatyuk, R. & Isayev, O. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **10**, 2903 (2019).
25. Keith, J. A. et al. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* **121**, 9816–9872 (2021).
26. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
27. Zhang, D., Bi, H. & Dai, F. Z. et al. Pretraining of attention-based deep learning potential model for molecular simulation. *npj Comput. Mater.* **10**, 94 (2024).
28. Wang, H., Zhang, L., Han, J. & Weinan, E. Deepmd-kit: a deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **228**, 178–184 (2018).
29. Shoghi, N. et al. From molecules to materials: Pre-training large generalizable models for atomic property prediction https://arxiv.org/abs/2310.16802 (2024).
30. Qu, E. & Krishnapriyan, A. S. The importance of being scalable: Improving the speed and accuracy of neural network interatomic potentials across chemical domains. In *The Thirty-Eighth Annual Conference on Neural Information Processing Systems* https://openreview.net/forum?id=Y4mBaZu4vy (2024).
31. Liao, Y.-L., Wood, B. M., Das, A. & Smidt, T. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. In *The Twelfth International Conference on Learning Representations* (2024).
32. Gasteiger, J., Becker, F. & Günnemann, S. Gemnet: Universal directional graph neural networks for molecules. *Adv. Neural Inf. Process. Syst.* **34**, 6790–6802 (2021).
33. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet - a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
34. Smith, J. S., Isayev, O. & Roitberg, A. E. Ani-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
35. Deng, B., Zhong, P. & Jun, K. et al. Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
36. Choudhary, K. & DeCost, B. Atomistic line graph neural network for improved materials property predictions. *npj Comput. Mater.* **7**, 185 (2021).
37. Yang, H. et al. Mattersim: a deep learning atomistic model across elements, temperatures and pressures https://arxiv.org/abs/2405.04967. 2405.04967 (2024).
38. Satorras, V. G., Hoogeboom, E. & Welling, M. E(n) equivariant graph neural networks. https://arxiv.org/abs/2102.09844 (2021).
39. Thomas, N. et al. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. http://arxiv.org/abs/1802.08219 (2018).
40. Geiger, M. & Smidt, T. e3nn: Euclidean neural networks https://arxiv.org/abs/2207.09453 (2022).
41. Jumper, J., Evans, R. & Pritzel, A. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
42. Du, W. et al. Se (3) equivariant graph neural networks with complete local frames. In *International Conference on Machine Learning*, 5583–5608 (PMLR, 2022).
43. Du, W. et al. A new perspective on building efficient and expressive 3d equivariant graph neural networks. In *Thirty-seventh Conference on Neural Information Processing Systems.* https://openreview.net/forum?id=hWPNYWkYPN (2023).
44. Duval, A. A. et al. FAENet: Frame averaging equivariant GNN for materials modeling. In Krause, A. et al. (eds.) *Proceedings of the 40th International Conference on Machine Learning*, vol. 202 of *Proceedings of Machine Learning Research*, 9013–9033 (PMLR, Honolulu, Hawaii, USA, 2023). https://proceedings.mlr.press/v202/duval23a.html.
45. Chanussot, L. et al. Open catalyst 2020 (oc20) dataset and community challenges. *ACS Catal.* **11**, 6059–6072 (2021).
46. Passaro, S. & Zitnick, C. L. Reducing so(3) convolutions to so(2) for efficient equivariant gnns. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23 (JMLR.org, Honolulu, Hawaii, USA, 2023).
47. Gasteiger, J. et al. Gemnet-OC: Developing graph neural networks for large and diverse molecular simulation datasets. *Transactions on Machine Learning Research* https://openreview.net/forum?id=u8tvSxm4Bs (2022).
48. Gasteiger, J., Groß, J. & Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations (ICLR)* (2020).
49. Fu, X. et al. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *Transactions on Machine Learning Research* https://openreview.net/forum?id=A8pqQipwkt. Survey Certification. (2023).
50. Wines, D. & Choudhary, K. Chips-ff: evaluating universal machine learning force fields for material properties. *ACS Mater. Lett.* **7**, 2105–2114 (2025).
51. Riebesell, J. et al. Matbench discovery—a framework to evaluate machine learning crystal stability predictions https://arxiv.org/abs/2308.14920 (2024).
52. Park, Y., Kim, J., Hwang, S. & Han, S. Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations. *J. Chem. Theory Comput.* **20**, 4857–4868 (2024).
53. Du, Y. et al. A new perspective on building efficient and expressive 3d equivariant graph neural networks. *Adv. Neural Infor. Process. Syst.* **36**, 66647–66674 (2024).
54. Shaw, P., Uszkoreit, J. & Vaswani, A. 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).
55. Su, J. et al. Roformer: enhanced transformer with rotary position embedding. *Neurocomputing* **568**, 127063 (2024).
56. Schollwöck, U. The density-matrix renormalization group in the age of matrix product states. *Ann. Phys.* **326**, 96–192 (2011).

57. Chen, R. T., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. Neural ordinary differential equations. *Adv. Neural Inform. Process. Syst.* **31**, 6572–6583 (2018).

Acknowledgements

This work was supported by National Key Research and Development Project (2022YFA1503000), National Natural Science Foundation of China (No. 92261111), and the NSFC Center for Single-Atom Catalysis (No. 22388102). We are also grateful to the Center of High-Performance Computing at Tsinghua University for providing computational resources. We also acknowledge the Welch Foundation (F-1841) for support. We thank Prof. Aditi Krishnapriyan for her insightful suggestions and valuable discussions.

Author contributions

B.Y. conducted all experimental tests, developed the software, and wrote the original manuscript. P.W. assisted with speed testing and provided one of the experimental datasets. P.Y. and Z.Z. contributed additional datasets for the experiments. J.W. helped conceptualize the study from a chemistry perspective and contributed to manuscript drafting. W.D., Y.D., H.J., and C.D. provided guidance on machine learning techniques and refined the manuscript. C.G., G.H., and H.X. conceived the research idea and designed the study. All authors participated in discussions and provided critical feedback.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-025-01817-w.

Correspondence and requests for materials should be addressed to Jiaao Wang, Yuanqi Du, Chenru Duan, Graeme Henkelman or Hai Xiao.