

Text S1 Architecture of MACE models

MACE adopts the pipeline as follows. Structure \rightarrow Atomic Environment \rightarrow Embedding \rightarrow (Interaction–Product–Update)^K \rightarrow learnable features f_i \rightarrow Readout \rightarrow Atomic Energy E_i , as shown in Fig. 1. This architecture combines high-body-order ACE expansion with E(3)-equivariant message passing in an end-to-end framework¹⁻⁷. Relative to conventional pairwise-message MPNNs, MACE improves expressivity through explicit higher-order interactions while maintaining a favorable accuracy-efficiency balance at shallow depth^{2,5-7}.

MACE adopts atomic-energy decomposition:

$$E = \sum_i E_i \quad (1)$$

Forces are obtained from energy gradients with respect to atomic coordinates:

$$\mathbf{F}_i = -\frac{\partial E}{\partial \mathbf{R}_i} \quad (2)$$

This single-energy formulation with auto-differentiated forces enforces consistency between energy and force predictions and supports stable convergence in joint training.

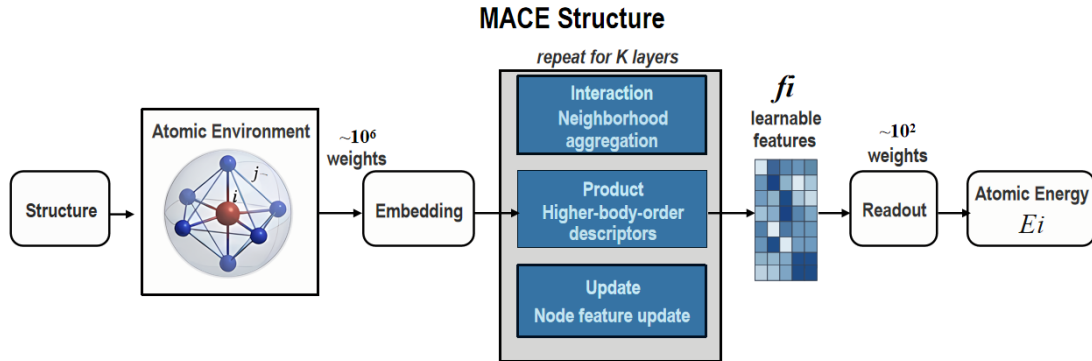


Fig.S1 Schematic illustration of the MACE architecture. The pipeline follows Structure \rightarrow Atomic Environment \rightarrow Embedding \rightarrow (Interaction–Product–Update, repeated for K layers) \rightarrow learnable atomic features f_i \rightarrow Readout \rightarrow atomic energy E_i .

The core characteristics in MACE are summarized as follows.

(1) Atomic Environment and Embedding: The structure is mapped to a local neighborhood graph. Nodes encode element identities. Edge features are encoded with radial basis functions and spherical harmonics. Initial node states $h_i^{(0)}$ and edge representations are constructed under rotation/reflection equivariance c

constraints, providing inputs for subsequent equivariant message construction. The corresponding node and edge representations can be written as:

$$\sigma_i^{(t)} = (r_i, z_i, h_i^{(t)}) \quad (3)$$

$$h_{i,k00}^{(0)} = W_{kz_i} \quad (4)$$

$$A_{i,klm}^{(1)} = \sum_{j \in \mathcal{N}(i)} R_{kl}^{(1)}(r_{ji}) Y_l^m(\hat{r}_{ji}) W_{kz_j} \quad (5)$$

where $(\sigma_i^{(t)})$ denotes the node state tuple at layer (t) , $(h_{i,k00}^{(0)})$ denotes the (k) -th invariant scalar channel of the initial node embedding, and $(A_{i,klm}^{(1)})$ denotes the first-layer equivariant atomic-environment feature centered at atom (i) , (i) is the center-atom index, (j) is a neighbor index, $(\mathcal{N}(i))$ is the neighbor set within cutoff, (r_i) is the atomic coordinate, (z_i) is the chemical species label, $(h_i^{(t)})$ is the node feature at layer (t) , (W_{kz}) is the learnable species-embedding weight, $(r_{ji} = |r_j - r_i|)$, $(\hat{r}_{ji} = (r_j - r_i)/r_{ji})$, $(R_{kl}^{(1)}(\cdot))$ is the radial basis embedding, and $(Y_l^m(\cdot))$ is the spherical-harmonic angular basis.

(2) Feature Construction (Interaction–Product–Update): This stage is repeated for K layers. Interaction aggregates neighborhood information. Product builds higher-order features through tensor products and symmetrization. Update writes messages back to nodes with residual updates, yielding $h_i^{(k+1)}$. The correlation order ν controls message body order (approximately $\nu + 1$) and governs the expressivity-cost trade-off.

(3) Learnable Atomic Features f_i : After K interaction layers, the model yields atom-wise representations f_i for property prediction. Readout mainly operates on invariant scalar channels (0e components), ensuring rotational invariance of predicted atomic energies.

(4) Readout: The readout module maps f_i to atomic energy E_i . A hierarchical aggregation scheme can combine information from different layers. Intermediate-layer readouts are typically linear; the final layer may use a lightweight MLP to improve nonlinear fitting capacity. This pipeline supports direct comparison of zero-shot, scratch, and fine-tune paradigms in terms of energy/force errors and transfer stability for doped systems.

73 **Table S1** Overview of the ML models and their settings including training paradigms,
74 data splitting, losses, and metrics in this work.

Module	Item	Setting
Training paradigm	Zero-shot	Direct inference (no parameter update)
Training paradigm	Scratch	Training from scratch
Training paradigm	Fine-tuning	Continued optimization from pretrained weights (full- parameter update)
Base model	Pretrained model	MACE-MP-0
Base model	Parameter scale	Small/Medium/Large models
Data setup	FT-Eq (equilibrium)	420 samples; 10-fold IID
Data setup	FT-NonEq (non-equilibrium)	6970 samples; 5-fold IID
Evaluation protocol	LODO-Strict (11 elements)	11 folds; per-fold test size 68
Evaluation protocol	LODO-Relaxed (11 elements)	11 folds; per-fold test size 8
Optimization target	Joint total loss	$L = \lambda_E L_E + \lambda_F L_F$, $\lambda_E = \lambda_F$
Metrics	Energy/Force	MAE (meV/atom)/RMSE (meV/Å)
Hyperparameters	Energy key	REF_energy
	Forces key	REF_forces
	Learning rate	3e-4
	Weight decay	1e-6
	Batch size	8
	Validation batch size	8
	Max epochs	200
	Energy loss weight	30
	Forces loss weight	30
	EMA	Enabled (ema_decay=0.99)
	Seed	42
	Default dtype	float32

References

- (1) Drautz, R. Atomic Cluster Expansion for Accurate and Transferable Interatomic Potentials. *Phys. Rev. B* 2019, 99 (1), 014104. <https://doi.org/10.1103/PhysRevB.99.014104>.
- (2) Batatia, I.; Kovacs, D. P.; Simm, G.; Ortner, C.; Csányi, G. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *Advances in neural information processing systems* 2022, 35, 11423–11436.
- (3) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. In *International conference on machine learning*; Pmlr, 2017; pp 1263–1272.
- (4) Satorras, V. G.; Hoogeboom, E.; Welling, M. E (n) Equivariant Graph Neural Networks. In *International conference on machine learning*; PMLR, 2021; pp 9323–9332.
- (5) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor Field Networks: Rotation- and Translation-Equivariant Neural Networks for 3D Point Clouds. *arXiv* May 18, 2018. <https://doi.org/10.48550/arXiv.1802.08219>.
- (6) Fuchs, F.; Worrall, D.; Fischer, V.; Welling, M. Se (3)-Transformers: 3d Roto-Translation Equivariant Attention Networks. *Advances in neural information processing systems* 2020, 33, 1970–1981.
- (7) Geiger, M.; Smidt, T. E3nn: Euclidean Neural Networks. *arXiv* July 18, 2022. <https://doi.org/10.48550/arXiv.2207.09453>.