MLIP
Machine Learning Interatomic Potentials

Frederik Schaub, Fritz Stange

Institute of Physics, University of Rostock

Report submitted: July 12, 2025
Supervisors: Mr. Mohammad Sadegh Talezadehlari,
                     Mr. Tolibjon Abdurakhmonov

Abstract: Machine learning interatomic potentials (MLIPs) aim to combine the accuracy of quantum methods with the speed of classical calculations. In this study, we trained a MACE-based MLIP on xTB-labeled molecular clusters and evaluated its accuracy, transferability, and stability. The model achieved low RMSEs on energies and forces and was able to generate stable molecular dynamics (MD) trajectories for both isolated molecules and a periodic liquid system. Structural and dynamical properties from MD agree well with expectations, demonstrating that MLIPs can generalize beyond their training domain and enable efficient simulations of complex systems. Additionally, we demonstrate that fine-tuning a foundation MACE model on a small subset of data can achieve comparable MD results at significantly reduced training cost.

1. Introduction

The development of accurate and efficient machine learning interatomic potentials (MLIPs) has opened up new possibilities for molecular simulations with quantum mechanical accuracy and classical efficiency. These potentials are particularly useful in systems where traditional ab initio methods such as density functional theory (DFT) or semi-empirical methods such as extended Tight Binding (xTB) are computationally intensive or incompatible.

In this work, we explore the neural network framework MACE (Multiplicative Atomic Cluster Expansion) [1], a MLIP approach that achieves high accuracy in modeling molecular systems while preserving physical symmetries such as rotational transformation invariance and translational invariance. Our goal is to evaluate the performance and applicability of MACE for simulating small organic molecules and molecular liquids.

To accomplish this, we train MACE models on clusters of carbonate molecules (Vinylene carbonate, Ethylene carbonate, Propylene carbonate, Dimethyl carbonate, Ethyl methyl carbonate, and Diethyl carbonate), systematically benchmark their performance in energy and force prediction, and investigate their ability to generate stable molecular dynamics (MD) trajectories. We compare our ML-based MD simulations with those using xTB, evaluate radial distribution functions (RDFs), and assess mean square displacements (MSDs).

In addition to training from scratch, we also explore the fine-tuning of a small foundation model trained on a reduced dataset. This strategy enables the rapid creation of usable MDs with minimal computational effort.

All models, scripts, and data are provided alongside this report [2].

2. Methods
2.1. Theoretical background

Classical force fields are based on fixed functional forms that can only capture the complexity of interatomic interactions to a limited extent. Neural network potentials, such as the Behler-Parrinello model [3], replace explicit functional forms with flexible functional approximators that learn the potential energy surface from data.

MACE extends this approach by combining the Atomic Cluster Expansion (ACE) framework with equivariant message-passing networks. Rather than only passing pairwise (2-body) messages, MACE uses many-body messages in conjunction with spherical harmonics and Clebsch–Gordan tensor products. The resulting architecture is linear in the number of interactions but supports higher-order correlations with low computational cost.

MACE is designed to take into account the symmetries of the potential energy surface: the predicted energies are invariant energies are invariant to translations, rotations, and permutations of identical atoms and the predicted forces transform equivalently under rotations. These properties are achieved by encoding local atomic environments in a rotationally equivalent basis and aggregating the contributions of neighbors

in a permutation-invariant manner. With this, the learned representation is physically consistent and transferable to different molecular orientations and geometries. MACE can efficiently capture both intramolecular and intermolecular interactions.

2.2. Data set and model training

The training data consists of molecular clusters with one to five solvent molecules, generated from representative liquid configurations and labeled with GFN2-xTB. The data set includes different intermolecular orientations and internal geometries. As required by the MACE architecture, the data was encoded in a local, rotationally equivalent frame.

First, a hyperparameter scan was performed on a small subset of the data to determine the optimal architecture. The best-performing configuration with the smallest root-mean-square errors (RMSEs) was then trained on increasingly larger subsets of the complete dataset to evaluate learning behavior. The model trained on 4000 structures was used for subsequent evaluations.

In addition, we fine-tuned a small foundation model provided by MACE on a reduced training set of 400 configurations. This fine-tuned model was only evaluated in molecular dynamics simulations to assess its utility in practice.

Every model was trained for 50 epochs and $10\,\%$ of each specified training set was used for validation. 1000 configurations were used for testing each model. Training and simulations were performed on a GeForce RTX™ 3070 graphics card.

2.3. Energy and force evaluation

The 4000-structure model was evaluated on both training and test configurations by comparing predicted total energies, atomization energies, and atomic forces to the xTB reference values. RMSEs were computed to quantify the model's accuracy. Evaluations were carried out using the Atomic Simulation Environment (ASE) [4].

2.4. Force decomposition

To better understand prediction errors, we decomposed the predicted forces into translational, rotational, and vibrational components. This allows for the separation of intra- and intermolecular contributions to the force error, which is relevant for molecular liquids. The analysis was performed using the aseMolec package [5].

2.5. Molecular dynamics simulations

To evaluate the stability and transferability of the trained MACE models, we conducted molecular dynamics simulations using the Langevin thermostat as implemented in ASE [4], with a time step of $1\,\mathrm{fs}$ over $5\,\mathrm{ps}$. Atomic velocities were initialized from a Maxwell-Boltzmann distribution at $300\,\mathrm{K}$ to ensure a physically meaningful starting point.

Two types of simulations were performed. First, to benchmark the 4000-configuration MACE model against the non-periodic xTB reference, we simulated a single, isolated molecule without periodic boundary conditions (PBC) at a high temperature of $T = 1200\,\mathrm{K}$. This comparison serves as a consistency check between MACE and xTB in a regime where both are applicable.

Second, to test the model in condensed phase liquid simulations, we performed MD of a periodic molecular liquid at $T = 500\,\mathrm{K}$. Since the used reference xTB function is not periodic and unsuitable for simulations of the condensed phase, a direct comparison was not possible. To evaluate whether the results of the 4000-configuration MACE model can also be achieved with minimal training effort, we repeated the liquid phase simulation using the fine-tuned model.

2.6. Dynamic analysis

The structural and dynamical properties of the generated MD trajectories were analyzed using radial distribution functions (RDFs) and mean square displacements (MSDs). RDFs characterize spatial correlations between atom types, while MSDs provide insight into atomic diffusion behavior. These analyses were performed using functions from aseMolec [5]. The Diffusion coefficients

$$D = \frac{\text{slope}}{6} \tag{1}$$

were estimated by performing a linear fit through the origin to the initial part of the MSD curve, where the dynamics are expected to be diffusive. We also show potential energy per atom and temperature over time, including the criteria for temperature fluctuation in thermal equilibrium [6]

$$\Delta T = T\sqrt{\frac{2}{3N}} \tag{2}$$
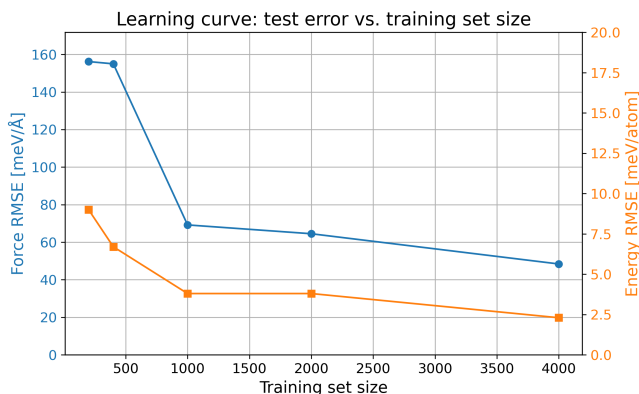
where N is the number of simulated molecules.

Figure 1 RMSE (test) for energies and forces as a function of training set size using configuration `mace_benchmark_best_guess_v3`.

3. Results and Discussion

3.1. Learning Curve and Model Accuracy

In the hyperparameter scan with the reduced training set, various configurations with different numbers of interaction layers, channels, correlation orders, and cutoff values for spherical harmonics were tested. Each model was evaluated in terms of energy and force RMSE. The results of this benchmark are summarized in Table 1 in the appendix. Deeper interaction networks and higher spherical harmonics improve accuracy at the expense of GPU memory and training time.

The configuration labeled `mace_benchmark_best_guess_v3` was selected for further training. This architecture was then trained on increasingly larger subsets of the complete training data (200, 400, 1000, 2000, 4000 structures) to evaluate the influence of the size of the training set on the prediction accuracy. The resulting learning curve is shown in Figure 1.

The final model trained on 4000 configurations achieved an energy RMSE below $3\,\mathrm{meV\,atom^{-1}}$ and a force RMSE under $50\,\mathrm{meV\,\mathring{A}^{-1}}$ with a training time of $86.8\,\mathrm{min}$. We also fine-tuned a foundation model on only 400 configurations. This required less than $3.3\,\mathrm{min}$ of training time. The fine-tuned model reached a validation RMSE of $100.2\,\mathrm{meV\,\mathring{A}^{-1}}$ for forces and $6.2\,\mathrm{meV\,atom^{-1}}$ for energies. It was not evaluated on static benchmarks, but later tested in liquid phase MD simulations.

3.2. Energy and Force Agreement with Reference Data

The 4000-structure model was applied to training and test sets. Figure 2 and Figure 3 show the correlation between MACE predictions and xTB reference values.
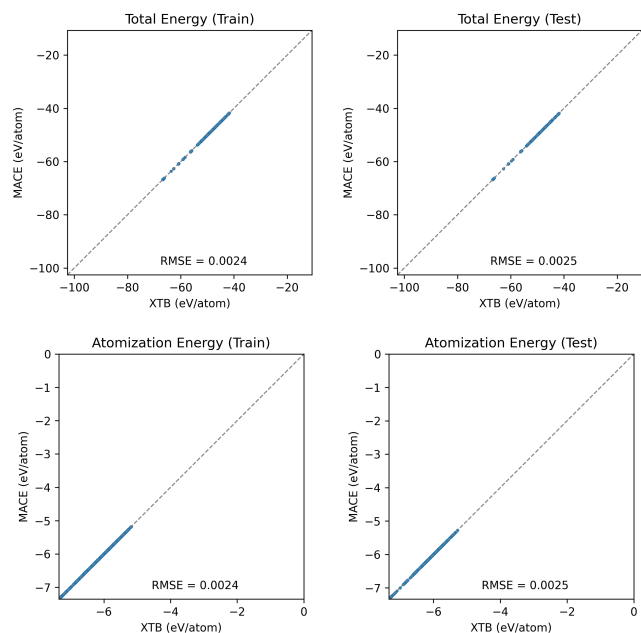

Figure 2 Correlation between predicted (MACE-4000) and reference (xTB) total and atomization energies for the training and test sets.
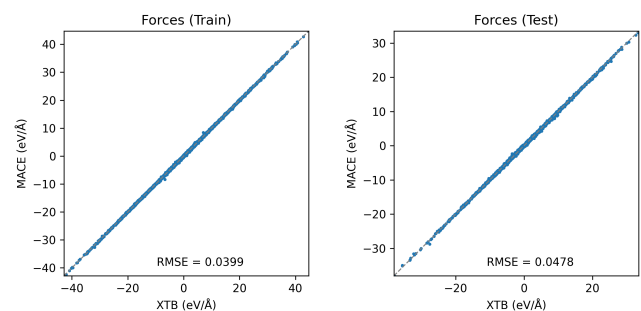

Figure 3 Correlation between predicted (MACE-4000) and reference (xTB) forces for the training and test sets.

The total energy and atomization energy predictions shows excellent agreement (RMSE < $3\,\mathrm{meV\,atom^{-1}}$), and the predicted forces follow the reference data with a force RMSE of $0.040\,\mathrm{eV\,\mathring{A}^{-1}}$ (train) and $0.048\,\mathrm{eV\,\mathring{A}^{-1}}$ (test).

3.3. Force Decomposition: Intra- vs Intermolecular Accuracy

To separate the accuracy of intra- and intermolecular forces, we decomposed each force into translational, rotational, and vibrational components. Figure 4 shows the results.

Vibrational components (intramolecular) are predicted with very low relative error (RRMSE $\sim 2\,\%$), while rotational and translational parts (intermolecular) have higher errors ($23\,\%$ and $32\,\%$, respectively). This is consistent with the physical expectation that
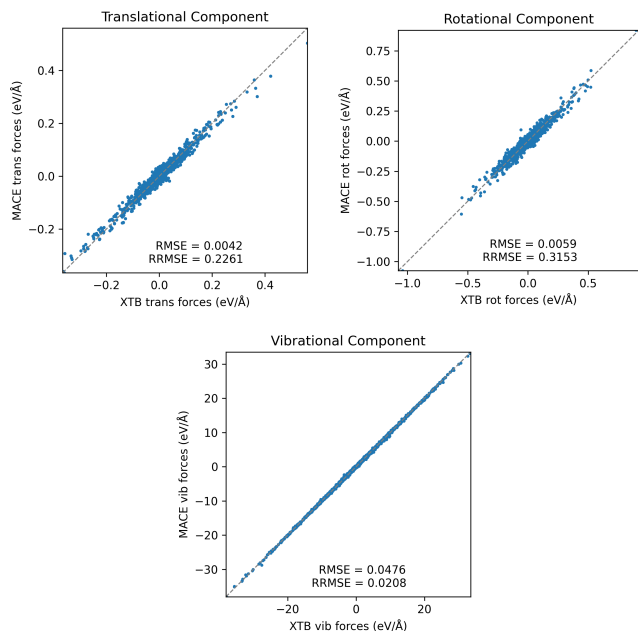
Figure 4 Force component comparison between predicted (MACE-4000) and reference (xTB) forces for the training and test sets: translational, rotational, and vibrational.

long-range, weak intermolecular forces are more difficult to learn.

3.4. Stability of Molecular Dynamics

Langevin dynamics simulations for a single Vinylene Carbonate molecule were performed using the 4000-structure model and xTB. Figure 5 compares the temperature and potential energy per atom.

The MACE simulation is slightly cooler on average, but shows similar fluctuations and no signs of instability. The simulation has not yet reached thermal equilibrium (2) in the chosen time frame. Both trajectories are included as video files in the repository [2] (`xtb_md_molecule.mp4` and `mace_md_molecule.mp4`).

3.5. Structural and Dynamical Properties of the Liquid

We performed a $5\,\text{ps}$ MD simulation at $500\,\text{K}$ with both the 4000-structure and fine-tuned model on the periodic carbonate liquid `input2.xyz`. The RDFs for intermolecular atom pairs (Figure 6) show a clear structural organization that is typical for a liquid. Figure 7 confirms the expected diffusive dynamics via the mean square displacement.

Videos of both simulations are included in the repository [2] (`mace_md_input2.mp4` and `mace_finetuned_md.mp4`) and show stable molecular motion and liquid-like structural dynamics.
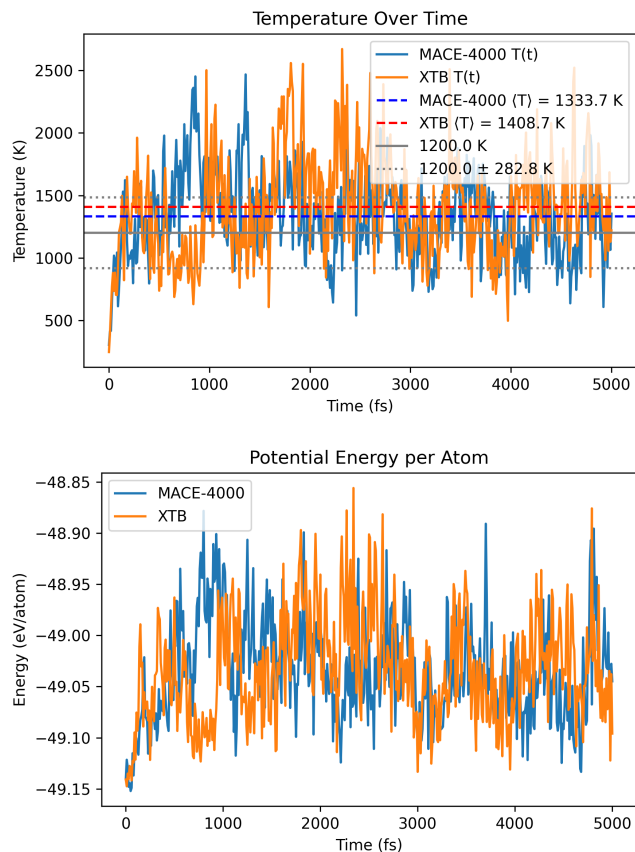

Figure 5 Temperature and potential energy over time for MACE-4000 and xTB single-molecule simulations.
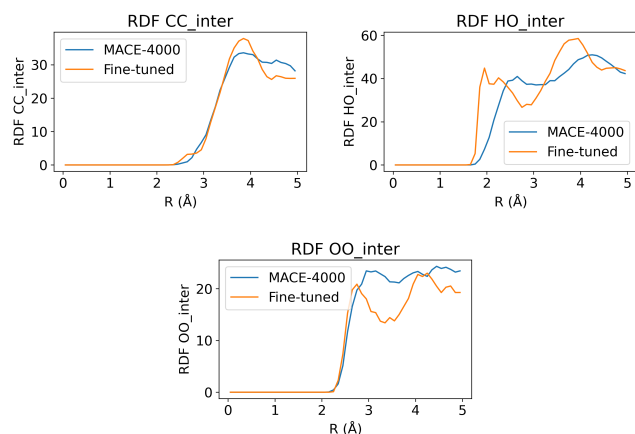

Figure 6 Comparison of intermolecular RDFs from the liquid MD simulation computed using the MACE-4000 model and the fine-tuned model.
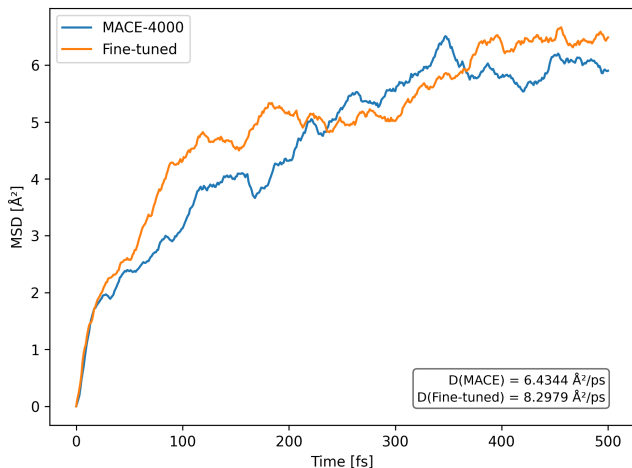
4

Figure 7 Comparison of the mean square displacement (MSD) over time for the liquid MD simulation computed using the MACE-4000 model and the fine-tuned model. The linear region between 0 and $1\,\mathrm{ps}$ was fitted to extract the diffusion coefficient.

This confirms that although the 4000-structure model is only fitted to isolated clusters, it can be generalized well to condensed phase systems and enables stable molecular dynamics. While the fine-tuned model exhibits deviations, particularly in peak intensities, the overall structural and dynamical behavior is consistent with that of the larger model. This demonstrates that useful MD models can be obtained with minimal data and training time. As expected for finite molecular systems, the MSD for both simulations becomes increasingly noisy. The shown diffusion coefficients (1) are based on the linear part up to $1\,\mathrm{ps}$.

4. Conclusion

This study demonstrates that MACE-based machine learning interatomic potentials can accurately reproduce the structural and dynamic properties of molecular systems while drastically reducing computational costs compared to quantum chemical methods.

The model trained on 4000 xTB-labeled cluster configurations achieved excellent accuracy with energy and force RMSEs below $3\,\mathrm{meV\,atom^{-1}}$ and $50\,\mathrm{meV\,\mathring{A}^{-1}}$, respectively. It generated stable MD trajectories for both isolated molecules and periodic liquids, even though it was only trained on non-periodic inputs. A detailed force decomposition confirmed that intramolecular forces are easier to learn, while weaker, long-range intermolecular interactions remain subject to larger errors.

Beyond this comprehensive training, we showed that fine-tuning a small base model to only 400 config-

urations within just three minutes of training yields a potential capable of generating stable and structurally plausible liquid MD. Although deviations from the larger model were observed, the core dynamic features were preserved. This underscores the practical potential of fine-tuning for rapid deployment in exploratory simulations.

Compared to traditional force fields or ab initio methods, MLIPs such as MACE offer a compelling trade-off: high prediction accuracy combined with massive speedups in simulation.

Future work should focus on improving transferability through active learning and hybrid schemes that combine ML with physical priors.

Appendix

Model Parameter Scan

Table 1: MACE hyperparameter scan on 400 training configurations.

| Config	BS	Ch	Int	L	Corr	Time [s]	GPU [%]	VRAM [MB]	RMSE$_E$ [meV/atom]	RMSE$_F$ [meV/Å]
mace_benchmark_best_guess_v1	10	64	3	0	3	117.4	21.3	569.7	34.8	176.6
mace_benchmark_best_guess_v2	10	64	3	1	3	171.8	39.0	1021.9	9.4	173.1
mace_benchmark_best_guess_v3	10	64	3	2	3	282.0	43.5	1574.2	7.5	157.3
mace_benchmark_baseline	10	32	2	0	2	92.6	10.7	365.5	11.2	210.3
mace_benchmark_batch32	32	32	2	0	2	47.2	13.3	449.7	24.7	291.3
mace_benchmark_batch64	64	32	2	0	2	34.6	13.9	564.1	2460.6	1822.2
mace_benchmark_channels64	10	64	2	0	2	95.8	14.9	432.7	8.4	195.9
mace_benchmark_channels128	10	128	2	0	2	108.2	26.3	891.2	11.7	188.6
mace_benchmark_interactions3	10	32	3	0	2	109.5	12.9	416.2	15.7	204.7
mace_benchmark_interactions4	10	32	4	0	2	121.3	14.6	466.4	27.5	217.8
mace_benchmark_maxL1	10	32	2	1	2	114.0	15.2	417.7	9.1	236.3
mace_benchmark_maxL2	10	32	2	2	2	122.7	22.1	508.4	9.8	186.4
mace_benchmark_correlation3	10	32	2	0	3	100.2	12.6	388.5	7.6	173.8
mace_benchmark_correlation4	10	32	2	0	4	111.9	24.3	937.9	7.3	174.8
mace_benchmark_chan64_corr3	10	64	2	0	3	101.0	17.8	478.4	13.7	176.7
mace_benchmark_chan64_inter3	10	64	3	0	2	109.6	17.6	503.3	10.0	185.5
mace_benchmark_inter3_corr3	10	32	3	0	3	114.0	14.9	446.8	12.2	197.4
mace_benchmark_maxL1_corr3	10	32	2	1	3	121.3	18.4	485.6	6.5	175.1

Single Molecule MD
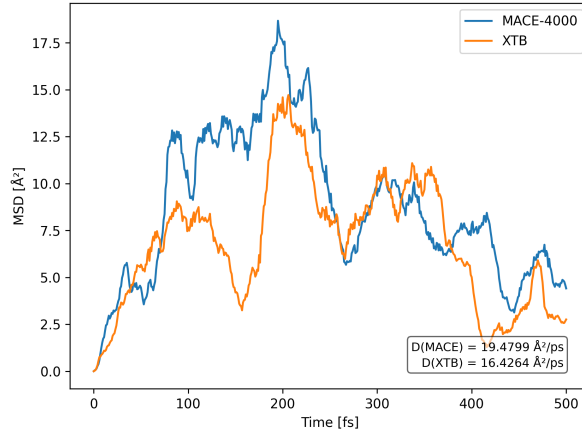
Figure 8 Comparison of the mean square displacement (MSD) over time for the single molecule MD simulation computed using the MACE-4000 model and the xTB. The linear region between 0 and $1\,\mathrm{ps}$ was fitted to extract the diffusion coefficient.
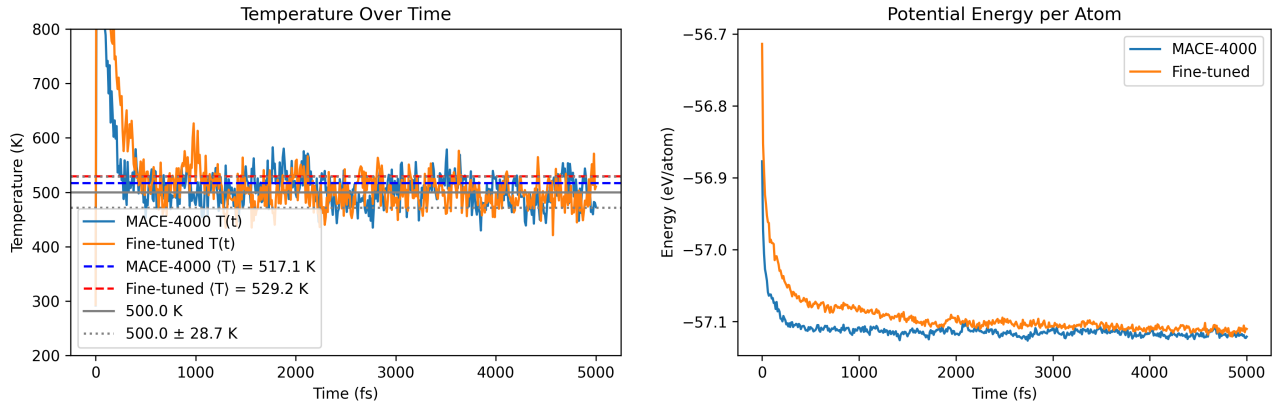
Liquid MD

Figure 9 Temperature and potential energy over time for the MACE-4000 and fine-tuned models liquid MD simulations.

References

[1] I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner and G. Csányi, *MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields*, Advances in neural information processing systems **35**, 11423–11436 (2022).

[2] F. Schaub and F. Stange, *Supplemental repository*, GitHub , (2025).

[3] J. Behler and M. Parrinello, *Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces*, Phys. Rev. Lett. **98**, 146401 (2007).

[4] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *The atomic simulation environment—a Python library for working with atoms*, Journal of Physics: Condensed Matter **29**, 273002 (2017).

[5] T. Abdurakhmonov, *aseMolec – MD analysis and plotting tools*, GitHub , (2025).

[6] Q. Zheng, *Molecular Dynamics at Constant Temperature*, University of Science and Technology of China , (2021).