Dyna-Mat: End-to-end benchmarking of foundation machine learning interatomic potentials in finite-temperature ensembles

Mikołaj J. Gawkowski [1,2], Nongnuch Artrith[3], Silvia Bonfanti[4], Abhijeet Sadashiv Gangan[5], Hendrik H. Heenen[6], Joseph Kioseoglou[7,8], Ivor Lončarić[9], Hemanadhan Myneni[10], Janosh Riebesell[5], Mariana Rossi[11,12], Matthias Rupp[13], Jonathan Schmidt[14], Shubham Sharma[11], Benjamin X. Shi[15], Antoni Wadowski[4,16], Lukas Hörmann[17,18], Venkat Kapil [1,2, †]

[1]Department of Physics and Astronomy, University College London, 7-19 Gordon St, London WC1H 0AH, UK
[2]Thomas Young Centre and London Centre for Nanotechnology, 9 Gordon St, London WC1H 0AH, UK
[3]Debye Institute for Nanomaterials Science, Utrecht University, 3508 TA Utrecht, the Netherlands
[4]NOMATEN Centre of Excellence, National Center for Nuclear Research, ul. A. Sołtana 7, 05-400 Swierk/Otwock, Poland
[5]Periodic Labs, Menlo Park, CA, USA
[6]Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany
[7]Physics Department, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece
[8]Center for Interdisciplinary Research and Innovation, Aristotle University of Thessaloniki, GR-57001 Thessaloniki, Greece
[9]Ruđer Boškovi'c Institute, Bijenička cesta 54, 10000 Zagreb, Croatia
[10]The Faculty of Industrial Engineering, Mechanical Engineering, and Computer Science, University of Iceland, Reykjavík, Iceland
[11]MPI for the Structure and Dynamics of Matter, Luruper Chaussee 148, 22761 Hamburg, Germany
[12]Yusuf Hamied Department of Chemistry, Lensfield Road, CB2 1EW Cambridge, UK
[13]Luxembourg Institute of Science and Technology (LIST), 5 Avenue des Hauts-Fourneaux, L-4362 Esch-sur-Alzette, Luxembourg
[14]Laboratory of Computational Science and Modeling, Institut des Mat'eriaux, 'Ecole Polytechnique F'ed'erale de Lausanne, 1015 Lausanne, Switzerland
[15]Initiative for Computational Catalysis, Flatiron Institute, New York, NY 10010, USA
[16]Faculty of Materials Science and Engineering, Warsaw University of Technology, Wołoska 141, 02-507 Warsaw, Poland
[17]Faculty of Physics, University of Vienna, Vienna, 1090, Austria
[18]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom
†corresponding author's email: v.kapil@ucl.ac.uk

Foundation machine learning interatomic potentials (MLIPs) are increasingly being used as drop-in replacements for first-principles calculations – enabling simulations of materials at length and time scales that were previously inaccessible. However, due to lack of ground truth data, their accuracy on structural and dynamical observables in finite thermodynamic ensembles is yet to be established. Here, we introduce Dyna-Mat-v1.0[1], a benchmark dataset of condensed-phase first-principles molecular dynamics trajectories designed to test foundation MLIPs at realistic finite-temperature conditions. Using this dataset, we evaluate 15 foundation MLIPs across four model tiers by comparing both single-point energy and force errors on first-principles configurations and observables generated from MLIP-driven trajectories. We find that "on average" models with lower single-point force errors also yield lower errors for structural and dynamical observables. However, there are individual systems for which low force errors lead to qualitative failures in the predicted structure. Pressure remains poorly described across most models, pointing to limitations in the density functional theory stress labels available in current large-scale training datasets. Finally, we construct an accuracy-cost Pareto frontier to identify the best trade-offs for molecular dynamics with foundation MLIPs, finding that the latest generation of cross-trained models is (close to) Pareto-optimal according to the accuracy metrics considered here. Overall, Dyna-Mat-v1.0 shows that end-to-end finite-temperature validation is essential for quantifying the predictive behaviour of foundation MLIPs, and provides a simple, scalable route for assessing them beyond static and harmonic benchmarks relevant to materials design.

References

[1] Gawkowski, Mikołaj J., et al. "Dyna-Mat: End-to-end benchmarking of foundation machine learning interatomic potentials in finite-temperature ensembles." arXiv preprint arXiv:2607.03433 (2026).