The Journal of Chemical Physics

RESEARCH ARTICLE | NOVEMBER 13 2025

# A foundation model for atomistic materials chemistry

Ilyes Batatia ; Philipp Benner ; Yuan Chiang ; Alin M. Elena ; Dávid P. Kovács ;
Janosh Riebesell ; Xavier R. Advincula ; Mark Asta ; Matthew Avaylon; William J. Baldwin ;
Fabian Berger ; Noam Bernstein ; Arghya Bhowmik ; Filippo Bigi ; Samuel M. Blau ;
Vlad Cărare ; Michele Ceriotti ; Sanggyu Chong ; James P. Darby ; Sandip De ;
Flaviano Della Pia ; Volker L. Deringer ; Rokas Elijošius ; Zakariya El-Machachi ; Edvin Fako ;
Fabio Falcioni ; Andrea C. Ferrari ; John L. A. Gardner ; Mikołaj J. Gawkowski ;
Annalena Genreith-Schriever ; Janine George ; Rhys E. A. Goodall ; Jonas Grandel ;
Clare P. Grey ; Petr Grigorev ; Shuang Han ; Will Handley ; Hendrik H. Heenen ;
Kersti Hermansson ; Cheuk Hin Ho ; Stephan Hofmann ; Christian Holm ; Jad Jaafar ;
Konstantin S. Jakob ; Hyunwook Jung ; Venkat Kapil ; Aaron D. Kaplan ; Nima Karimitari ;
James R. Kermode ; Panagiotis Kourtis ; Namu Kroupa ; Jolla Kullgren ; Matthew C. Kuner ;
Domantas Kuryla ; Guoda Liepuoniute ; Chen Lin; Johannes T. Margraf ; Ioan-Bogdan Magdău ;
Angelos Michaelides ; J. Harry Moore ; Aakash A. Naik ; Samuel P. Niblett ; Sam Walton Norwood;
Niamh O'Neill ; Christoph Ortner ; Kristin A. Persson ; Karsten Reuter ; Andrew S. Rosen ;
Louise A. M. Rosset ; Lars L. Schaaf ; Christoph Schran ; Benjamin X. Shi ; Eric Sivonxay ;
Tamás K. Stenczel ; Christopher Sutton ; Viktor Svahn ; Thomas D. Swinburne ; Jules Tilly ;
Cas van der Oord ; Santiago Vargas ; Eszter Varga-Umbrich ; Tejs Vegge ; Martin Vondrák ;
Yangshuai Wang ; William C. Witt ; Thomas Wolf; Fabian Zills ; Gábor Csányi

Check for updates

*J. Chem. Phys.* 163, 184110 (2025)
https://doi.org/10.1063/5.0297006

CHORUS

View Online

Export Citation

## Articles You May Be Interested In

Δ-model correction of foundation model based on the model's own understanding

*J. Chem. Phys.* (May 2025)

Multi-head committees enable direct uncertainty prediction for atomistic foundation models

*J. Chem. Phys.* (December 2025)

How accurate are DFT forces? Unexpectedly large uncertainties in molecular datasets

*J. Chem. Phys.* (December 2025)

AIP Publishing

# A foundation model for atomistic materials chemistry

Cite as: J. Chem. Phys. 163, 184110 (2025); doi: 10.1063/5.0297006
Submitted: 16 August 2025 • Accepted: 25 August 2025 •
Published Online: 13 November 2025

Ilyes Batatia,[1] Philipp Benner,[2] Yuan Chiang,[3,4] Alin M. Elena,[5] Dávid P. Kovács,[1] Janosh Riebesell,[4,6] Xavier R. Advincula,[6,7] Mark Asta,[3,4] Matthew Avaylon,[8] William J. Baldwin,[1] Fabian Berger,[7] Noam Bernstein,[9] Arghya Bhowmik,[10] Filippo Bigi,[11] Samuel M. Blau,[12] Vlad Cărare,[1,6] Michele Ceriotti,[11] Sanggyu Chong,[11] James P. Darby,[1] Sandip De,[13] Flaviano Della Pia,[7] Volker L. Deringer,[14] Rokas Elijošius,[1] Zakariya El-Machachi,[14] Edvin Fako,[13] Fabio Falcioni,[15] Andrea C. Ferrari,[16] John L. A. Gardner,[14] Mikołaj J. Gawkowski,[17] Annalena Genreith-Schriever,[7] Janine George,[2,18] Rhys E. A. Goodall,[19] Jonas Grandel,[2,18] Clare P. Grey,[7] Petr Grigorev,[20,21] Shuang Han,[13] Will Handley,[6,22] Hendrik H. Heenen,[23] Kersti Hermansson,[24] Cheuk Hin Ho,[25] Stephan Hofmann,[1] Christian Holm,[26] Jad Jaafar,[1] Konstantin S. Jakob,[23] Hyunwook Jung,[23] Venkat Kapil,[7,17] Aaron D. Kaplan,[4] Nima Karimitari,[27] James R. Kermode,[28] Panagiotis Kourtis,[29] Namu Kroupa,[1,6,22] Jolla Kullgren,[24] Matthew C. Kuner,[3,4] Domantas Kuryla,[7] Guoda Liepuoniute,[1,16] Chen Lin,[1,30] Johannes T. Margraf,[31] Ioan-Bogdan Magdău,[29] Angelos Michaelides,[7] J. Harry Moore,[1] Aakash A. Naik,[2,18] Samuel P. Niblett,[7] Sam Walton Norwood,[10] Niamh O'Neill,[6,7] Christoph Ortner,[25] Kristin A. Persson,[3,4,32] Karsten Reuter,[23] Andrew S. Rosen,[33] Louise A. M. Rosset,[14] Lars L. Schaaf,[1] Christoph Schran,[6] Benjamin X. Shi,[7] Eric Sivonxay,[12] Tamás K. Stenczel,[1] Christopher Sutton,[27] Viktor Svahn,[24] Thomas D. Swinburne,[20,34] Jules Tilly,[15] Cas van der Oord,[1] Santiago Vargas,[35] Eszter Varga-Umbrich,[1] Tejs Vegge,[10] Martin Vondrák,[23,31] Yangshuai Wang,[25] William C. Witt,[36] Thomas Wolf,[37] Fabian Zills,[26] and Gábor Csányi[1,a)]

## AFFILIATIONS

[1] Engineering Laboratory, University of Cambridge, Trumpington St. and JJ Thomson Ave., Cambridge, United Kingdom
[2] Federal Institute of Materials Research and Testing (BAM), Berlin, Germany
[3] Department of Materials Science and Engineering, University of California, Berkeley, California 94720, USA
[4] Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
[5] Scientific Computing Department, Science and Technology Facilities Council, Daresbury Laboratory, Keckwick Lane, Daresbury WA4 4AD, United Kingdom
[6] Cavendish Laboratory, University of Cambridge, J. J. Thomson Ave., Cambridge, United Kingdom
[7] Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, United Kingdom
[8] Computing Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
[9] U.S. Naval Research Laboratory, Washington, District of Columbia 20375, USA
[10] Department of Energy Conversion and Storage, Technical University of Denmark, Anker Engelunds Vej 301, 2800 Kgs. Lyngby, Denmark
[11] Laboratory of Computational Science and Modeling, Institute of Materials, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[12] Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
[13] BASF SE, Carl-Bosch-Straße 38, 67056 Ludwigshafen, Germany
[14] Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford OX1 3QR, United Kingdom
[15] InstaDeep, London W2 1AY, United Kingdom

[16]Cambridge Graphene Centre, University of Cambridge, Cambridge CB3 0FA, United Kingdom
[17]Department of Physics and Astronomy, University College London, London WC1E 6BT, United Kingdom
[18]Institute of Condensed Matter Theory and Solid State Optics, Friedrich Schiller University Jena, Jena, Germany
[19]Chemix, Inc., Sunnyvale, California 94085, USA
[20]Aix-Marseille Université, CNRS, CINaM UMR 7325, Campus de Luminy, 13288 Marseille, France
[21]CNRS, INSA Lyon, Universite Claude Bernard Lyon 1, MATEIS, UMR5510, 69621 Villeurbanne, France
[22]Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom
[23]Fritz-Haber-Institute of the Max-Planck-Society, Berlin, Germany
[24]Department of Chemistry–Ångström, Uppsala University, Box 538, S-751 21 Uppsala, Sweden
[25]Mathematics Department, University of British Columbia, 1984 Mathematics Rd., Vancouver, British Columbia V6T 1Z2, Canada
[26]Institute for Computational Physics, University of Stuttgart, 70569 Stuttgart, Germany
[27]Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina 29208, USA
[28]Warwick Centre for Predictive Modelling, School of Engineering, University of Warwick, Coventry CV4 7AL, United Kingdom
[29]School of Natural and Environmental Science, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom
[30]Information Engineering, University of Oxford, Oxford OX1 3PA, United Kingdom
[31]University of Bayreuth, Bavarian Center for Battery Technology (BayBatt), Bayreuth, Germany
[32]Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
[33]Department of Chemical and Biological Engineering, Princeton University, Princeton, New Jersey 08544, USA
[34]Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA
[35]Department of Chemistry and Biochemistry, University of California – Los Angeles, 607 Charles E. Young Drive East, Los Angeles, California 90095, USA
[36]Department of Materials Science and Metallurgy, University of Cambridge, 27 Charles Babbage Road, CB3 0FS Cambridge, United Kingdom
[37]Hugging Face Inc., Brooklyn, New York 11201, USA

[a)]**Author to whom correspondence should be addressed:** gc121@cam.ac.uk

## ABSTRACT

Atomistic simulations of matter, especially those that leverage first-principles (*ab initio*) electronic structure theory, provide a microscopic view of the world, underpinning much of our understanding of chemistry and materials science. Over the last decade or so, machine-learned force fields have transformed atomistic modeling by enabling simulations of *ab initio* quality over unprecedented time and length scales. However, early machine-learning (ML) force fields have largely been limited by (i) the substantial computational and human effort required to develop and validate potentials for each particular system of interest and (ii) a general lack of transferability from one chemical system to the next. Here, we show that it is possible to create a general-purpose atomistic ML model, trained on a public dataset of moderate size, that is capable of running stable molecular dynamics for a wide range of molecules and materials. We demonstrate the power of the MACE-MP-0 model—and its qualitative and at times quantitative accuracy—on a diverse set of problems in the physical sciences, including properties of solids, liquids, gases, chemical reactions, interfaces, and even the dynamics of a small protein. The model can be applied out of the box as a starting or "foundation" model for any atomistic system of interest and, when desired, can be fine-tuned on just a handful of application-specific data points to reach *ab initio* accuracy. Establishing that a stable force-field model can cover almost all materials changes atomistic modeling in a fundamental way: experienced users obtain reliable results much faster, and beginners face a lower barrier to entry. Foundation models thus represent a step toward democratizing the revolution in atomic-scale modeling that has been brought about by ML force fields.

https://doi.org/10.1063/5.0297006

## I. INTRODUCTION

An overarching goal in the field of atomistic modeling is to develop an interatomic potential (alternatively also called a "force field") that quickly and accurately predicts the total energy and atomic forces for an arbitrary chemical structure. Existing methods are not capable of this feat: while *ab initio* methods such as density functional theory (DFT)[1–7] are widely applicable and accurate, their high computational cost prohibits their use in many important cases, including high-throughput workflows and those in which large (≫1000 atom) systems need to be simulated over long timescales. Conversely, empirical force-field models that use simple functional forms are extremely cheap and quick to use but fail to capture accurately the important subtleties of the many-body interactions between collections of atoms induced by quantum mechanics.[8] Finally, modern machine-learning based interatomic potentials (MLIPs) can faithfully approximate *ab initio* methods for orders of magnitude less cost but typically require significant upfront investment and human effort when generating and labeling the training dataset.[9–17] Furthermore, these datasets and models typically need to be redeveloped from scratch for each new system of interest.[18] As a remedy to these issues, and as a step toward a truly universal MLIP, we present MACE-MP-0, a foundation model for materials chemistry that displays an impressive out-of-the-box ability to model a wide variety of chemical systems. Crucially, we also demonstrate that fine-tuning MACE-MP-0 using just a handful of new configurations leads to quantitatively accurate models, dramatically reducing the cost and barrier to entry for the modeling of novel chemical systems.

MACE-MP-0 uses the MACE architecture,[19] which unifies the atomic cluster expansion (ACE)[14,20–22] and equivariant graph neural networks.[16,23] MACE was designed to keep only what appear to be essential components of the latter:[23] the element embedding[12,24] and the equivariant messages constructed through the symmetric tensor product operation. MACE's unique innovations are that (i) it uses high body-order equivariant features in each layer (four-body in the present case), and consequently only two layers of message passing are sufficient;[25] (ii) it is only mildly nonlinear, as the only nonlinear activations are in the radial basis and the final readout layer, hence its classification as a graph tensor network; and (iii) it uses tensor decomposition[24] for the efficient parameterization of high body-order features. The MACE architecture allows MACE-MP-0 to accurately model its training data while remaining competitively performant with other graph neural networks, presently allowing simulations of around a thousand atoms for nanoseconds per day on a graphics processing unit (GPU).

Despite training MACE-MP-0 on a dataset with a specific materials focus (MPtrj), the most striking finding is that our model shows remarkable out-of-distribution performance and leads to stable molecular dynamics (MD) simulations for arbitrary systems over long timescales, showing chemically sensible structures, reactions, and transformations. In the main body of this paper, we showcase the generality of MACE-MP-0 by considering three disparate classes of chemical systems: solid and liquid water, heterogeneous catalysis, and metal-organic frameworks. In Appendix A, we further demonstrate MACE-MP-0's capabilities on an unprecedented range of qualitative and quantitative examples drawn from computational chemistry and materials science, including running molecular dynamics simulations for a wide variety of chemistries, predicting phonon spectra, calculating activation energies for point defect and dislocation motion, simulating solvent mixtures, combusting hydrogen gas, modeling a complete rechargeable battery cell, and many more.

There are several versions of the MACE-MP-0 model, all trained on the same dataset, with minor variations in the model architecture. Unless otherwise stated, all results in this paper were obtained with the MACE-MP-0b3 version, and this is emphasized in figures and captions, while we retain the simpler "MACE-MP-0" name in the text for readability. All model versions are publicly available.

## II. APPLICATIONS

### A. Water and aqueous systems

Water is ubiquitous in nature and technology and has long been a major focus of computational study. Driven by the delicate balance between directional hydrogen bonding and primarily non-directional van der Waals interactions, aqueous systems remain a challenge for simulations.[31] For example, the study of proton transfer in water, a fundamental process characterized by the continuous breaking and forming of covalent bonds, has long required the use of *ab initio* molecular dynamics for detailed atomistic insight.[32–34] We demonstrate in this section how MACE-MP-0 describes various aqueous systems.

We start by examining the structure of liquid water and hexagonal ice (ice Ih). The oxygen–oxygen radial distribution function, depicted in Fig. 1(a), shows reasonable agreement with reference simulations. The infrared vibrational spectra of both phases, shown in panel Fig. 1(b), align well with experimental observations, albeit with a notable red shift in the stretching vibrations indicating a softer description of the O–H bond, as is well-known for PBE-D3.[29,31] In Fig. 1(d), the relative stabilities of 12 ice polymorphs with respect to ice Ih, used in a recent benchmark,[35] show excellent agreement with respect to PBE-D3 with a MAE of around 5 meV. Proton defects ($OH^-$ and $H_3O^+$) in ice Ih and liquid water were simulated, revealing robust descriptions of proton transfer, as shown in Fig. 1(c). The proton transfer barrier for hydroxide is higher than for hydronium in liquid water, consistent with experimental diffusion trends.

Next, we evaluate MACE-MP-0 for describing solid–liquid interfaces. First, we focus on NaCl in water in two cases: an NaCl(001) interface in contact with water and a small nanocrystal surrounded by water. Simulations were performed at 400 K to promote dissolution and compared to simulations with a custom-trained ML potential based on revPBE-D3 from Ref. 30. As expected, for the flat surface, the model predicts no dissolution events on the timescale of the simulation (0.5 ns). Meanwhile, for the nanocrystal surrounded by water, MACE-MP-0 captures a dissolution mechanism resembling that in Ref. 30, as shown in Fig. 1(e). The dissolution proceeds via a crumbling mechanism, where an initial steady loss of ions is followed by the rapid disintegration of the crystal. As ions dissolve from the crystal, they are hydrated by water. The dissolution process is stochastic, leading to an intrinsic variation between independent simulations, as demonstrated by three examples. The final structure of the dissolved ions in water also displays the expected orientation of the water molecules with respect to the ions.

We then model the $SiO_2$/water interface, Fig. 1(f), revealing the expected density modulations in the first few contact layers. As before, the liquid phase is found to be overstructured, a common characteristic of the Perdew–Burke–Ernzerhof (PBE) functional[31] used by MPTraj and, therefore, by MACE-MP-0. $SiO_2$ is known for its dissociative water adsorption, which we observe in our simulations. Deprotonation of water is evidenced by the shoulder in the water density plot and can also be seen in the inset of a snapshot of this system in Fig. 1(f).

Finally, we investigate nanoconfined water in graphene-like nanocapillaries,[36,37] which exhibit dramatically different properties from bulk water. MACE-MP-0 proved robust in simulating nanoconfined water. Stable simulations were conducted at 4 GPa

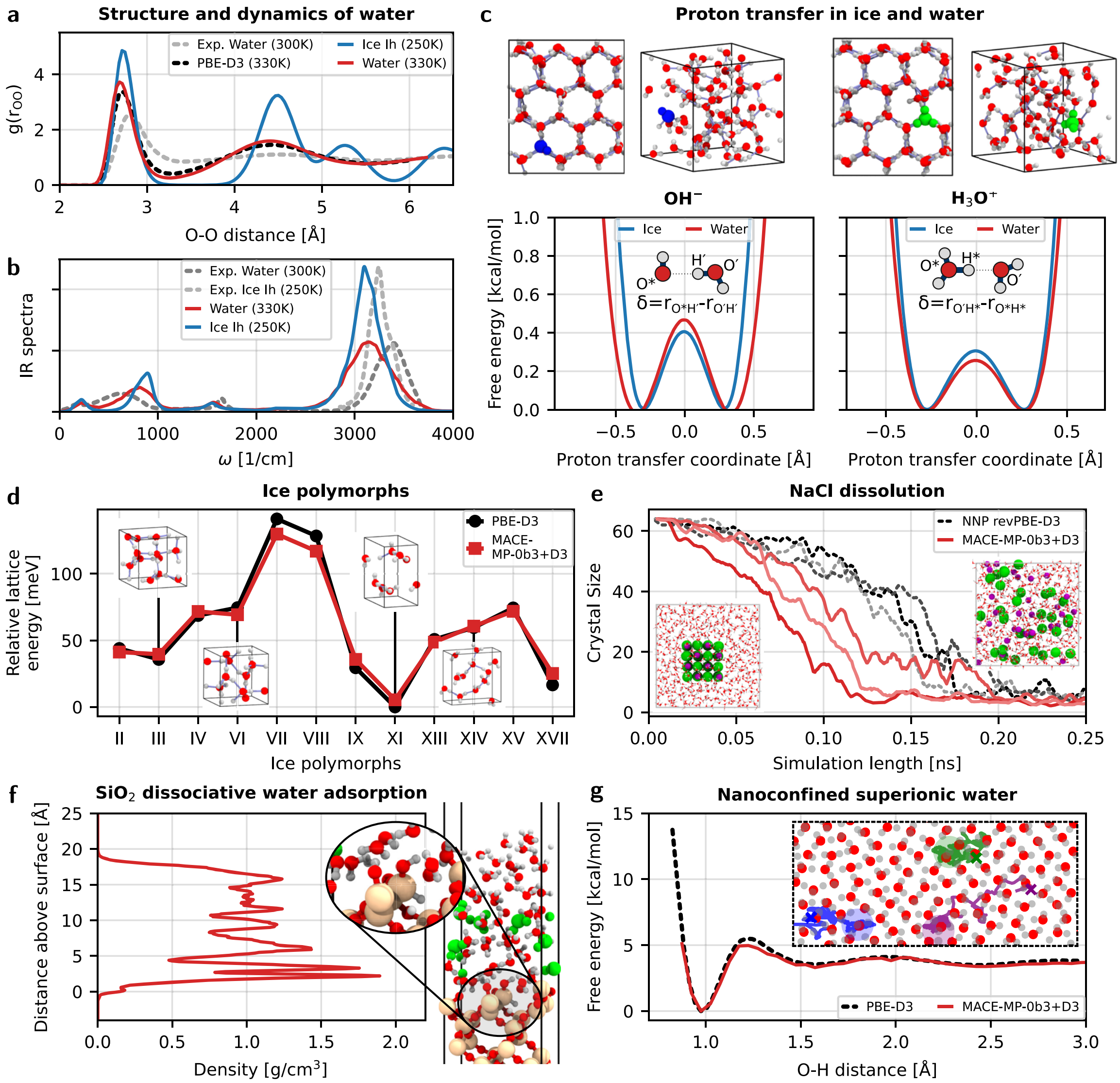


**FIG. 1.** Aqueous systems. (a) Oxygen–oxygen radial distribution function for bulk water (experimental result from Ref. 26) and ice Ih. (b) Experimental (Refs. 27 and 28) and computed infrared spectra of bulk water and ice Ih. (c) Free energy profiles as a function of the proton transfer barrier for a hydroxide ion and excess proton in ice Ih at 250 K and bulk water at 330 K. Snapshots at the top show the simulation cells. (d) Performance of MACE-MP-0b3 (red squares) on the relative lattice energies of the DMC-ICE13 dataset, compared to the reference method, PBE-D3[29] (black circles). (e) Dissolution of a 4 × 4 × 4 unit-cell NaCl nanocrystal in water at 400 K, monitoring the extent of dissolution over the simulation time via the crystal size. The performance of the MACE-MP-0b3 (red lines) is compared to a neural network potential[30] trained explicitly to capture NaCl dissolution (black dashed lines). (f) $SiO_2$/water interface simulation showing density modulations and dissociative water adsorption, with an inset highlighting the deprotonation of water as indicated by a shoulder in the water density plot. $H_3O^+$ defects in the liquid are highlighted in green. (g) The free energy profile of the O–H distance in the superionic phase of monolayer water in a confining potential. The inset shows a snapshot of the monolayer superionic phase with lines indicating the 50 ps-long trajectory of randomly chosen hydrogen atoms, with "×" indicating their initial positions.

and 600 K, conditions under which a superionic phase with high ionic conductivity was previously predicted[38] using a custom-trained ML potential. The MACE-MP-0 model accurately captured the dynamical characteristics of this phase, including extensive proton transfer on the 10 ps timescale, as illustrated in the inset of Fig. 1(g). Comparing the free energy profile associated with the O–H distance [Fig. 1(g)] against the PBE-D3 reference, MACE-MP-0 shows near quantitative agreement and an overall very good description of nanoconfined water.

## B. Catalysis

The study of heterogeneous[44–46] and electrocatalysis[47–49] is another major area where DFT excels. It provides atomistic insight into the underlying reaction mechanisms and enables the prediction of the properties of new catalytic materials,[50] including reaction barriers and rates, which are, in turn, used to predict turnover frequencies.[51] The latter is essential for the computational discovery of new solid catalysts to overcome dependence on rare and toxic elements, improving the efficiency of critical processes for energy conversion. However, the computational cost of DFT is a serious impediment. Empirical interatomic potentials are typically inadequate for catalysis applications, as they rarely describe chemical reactions accurately. Machine learning has already had a strong impact on computational catalysis,[41,52,53] e.g., enabling fast screening of material spaces,[54–56] and free energy calculations beyond the

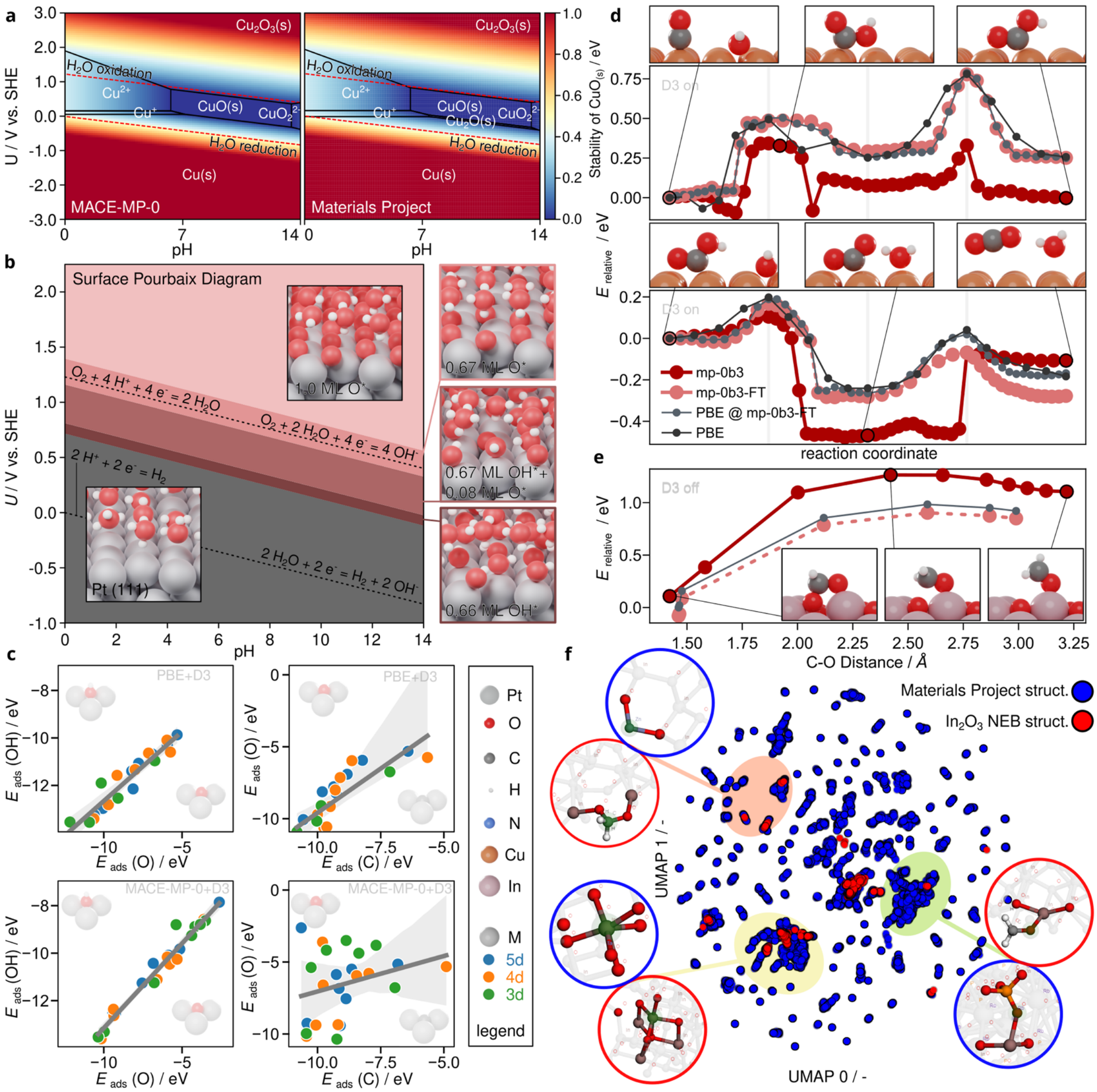


**FIG. 2.** Heterogeneous catalysis. (a) Pourbaix diagrams of CuO bulk systems constructed with MACE-MP-0b3 (left) and materials project reference data (right). (b) MACE-MP-0b3+D3-calculated Pt(111) surface Pourbaix diagram, in overall good agreement with the literature.[39] (c) The relative adsorption energy scaling relation between O and OH on transition metal surfaces is captured correctly by MACE-MP-0b3+D3, as is the lack of linear scaling between C and O.[40] Metals are colored according to rows in the Periodic Table as 3d, 4d, and 5d. (d) Reaction profile of multistep electrochemical CO oxidation on Cu. CO–OH coupling and dehydrogenation reactions are characterized in the upper and lower panels, respectively. Energy profiles from MACE-MP-0b3+D3 and PBE+D3 nudged elastic band (NEB) calculations show significant deviations, although the qualitative features agree. Fine-tuning (FT) yields a model that is in excellent agreement with the reference. (e) MACE-MP-0b3 reaction profile for a key reaction step ($CH_2O_2 \rightarrow CH_2 + O$) in the $CO_2$-to-methanol conversion on $In_2O_3$[41] and the profile of an FT model. (f) Comparison of the atomic environments in the training data (blue) and in the $In_2O_3$ NEB images (red) in the form of a UMAP plot.[42,43] Insets show local environments with similar MACE features (inset frames in blue for training data and in red for NEB configurations), exemplifying which bulk training environments influence predictions for the out-of-domain catalytic test case.

harmonic approximation.[41,57,58] However, developing such accurate potentials from scratch still requires significant human and computational effort. We now test the performance of MACE-MP-0 for different catalysis applications and summarize the results in Fig. 2.

Potential–pH Pourbaix diagrams are central to understanding the aqueous stability of solid materials in an electrochemical environment[59,60] and, therefore, allow predicting the active phase of an electrocatalyst under given conditions. Within the computational hydrogen electrode (CHE) framework,[61] these diagrams can be computed without an explicit electrostatic model. Figures 2(a) and 2(b) show the Pourbaix diagrams for bulk CuO and a Pt(111) surface calculated with MACE-MP-0 using the D3 correction. The Pourbaix diagrams are constructed via the formalism described in Refs. 62 and 63, where only the energies of the relevant solids are calculated while experimentally derived energies are used for the aqueous ions. In both cases, the MACE-MP-0 results show remarkably good agreement with DFT,[39] predicting the correct sequence of stable phases (with the exception of a very narrow region of $Cu_2O$ stability) and corresponding pH and potential ranges. While this accuracy may be expected for the bulk CuO system represented in the training set, the electrosorption at the Pt(111) surface is also well described despite being out of the domain.

In Fig. 2(c), adsorption energy scaling relations between atomic and hydrogenated adsorbates on transition-metal surfaces are shown for MACE-MP-0 and PBE (see Appendix A for more examples). Such scaling relations are central to understanding the activity of heterogeneous catalysts.[64,65] MACE-MP-0 captures these trends well, and the slopes of the linear fits are in reasonable agreement with DFT (e.g., 0.71 for O vs OH, compared to 0.64 for PBE). Importantly, the lack of correlation between O and C adsorption energies is also captured, indicating that the model is not merely sorting metals according to their general reactivity.[40,66] Figures 2(d) and 2(e) show reaction energy profiles for CO oxidation on Cu[67] and a key step in $CO_2$ conversion to methanol on $In_2O_3$,[41,68] respectively. While these are not quantitatively accurate when compared to DFT, MACE-MP-0b3 nevertheless captures the location and magnitude of the barriers surprisingly well. To obtain quantitative agreement, MACE-MP-0b3 is fine-tuned with five single-point DFT calculations from each energy profile. NEB calculations with the fine-tuned (FT) model then yield excellent agreement with the DFT reference in almost all cases, with the exception of the energy of the final state of CO oxidation, which is slightly overestimated by the FT model. Here, describing the subtle non-covalent interactions between the surface and molecular $CO_2$ and $H_2O$ would require additional training. Nonetheless, this shows that fine-tuning with very small datasets is sufficient to obtain quantitatively accurate potentials for heterogeneous catalysis.

Figure 2(f) illustrates how MACE-MP-0b3 generalizes to out-of-domain catalysis tasks from bulk training configurations. To this end, the high-dimensional MACE features are projected to 2D using a Uniform Manifold Approximation Projection (UMAP),[42] with local atomic environments (LAE) in the training set shown in blue and those found in the $In_2O_3$ transition path shown in red. Representative environments with similar features are highlighted, indicating that the internal representation of the atomic environments in the NEB configurations is similar to the representation of under-coordinated environments and metal–organic systems in the training set.

While MACE-MP-0b3 is not always quantitatively accurate for the most challenging catalysis applications, its stability in MD and exploration of reactive pathways are remarkable and provide a starting point for further optimizations. Relevant configurations or phase space regions thus identified may subsequently be validated either by first-principles calculations or serve to initiate active-learning for refining the model, as demonstrated for the NEB calculations. Even at its current foundation level, MACE-MP-0b3 already allows for statistical sampling far beyond the present DFT-based state of the art, which is still largely thermochemistry-centered, whereas the foundational MACE model will pave the way for true kinetic modeling through explicit evaluations of reaction profiles and the reactive flux along them.

## C. Metal–organic frameworks

Metal–organic frameworks (MOFs) are a class of nanoporous materials comprised of metal cations or clusters connected by organic linkers arranged in a periodic lattice.[75] Due to their large surface areas, tunable building blocks, and permanent porosity, MOFs hold substantial promise for various applications, including but not limited to catalysis, energy storage, gas adsorption and separations, and optoelectronic devices.[75] We tested our pre-trained model directly against version 14 of the Quantum MOF (QMOF) database, which contains DFT-computed properties at several levels of theory for 13 912 MOFs and structurally related coordination polymers.[69,70] MACE-MP-0b3 was not trained on any data from the QMOF database, making this a challenging test of its transferability to largely unseen chemistries.

As shown in Fig. 3(a), MACE-MP-0b3 performs well out-of-box in predicting the PBE energies of MOFs, achieving an MAE of 0.040 eV/atom (with the full range of energies spanning nearly 4 eV/atom, about 100 times larger), despite the pronounced difference between the inorganic crystals of the MPtrj training set and the MOF structures that make up the QMOF database. This accuracy spans most of the periodic table, after exclusion of elements with incompatible pseudopotentials and calculation parameters (see Appendix A 28 and Fig. 45).

To validate the use of MACE-MP-0b3 for capturing dynamic processes, we investigate $CO_2$ adsorption in a prototypical MOF known as Mg-MOF-74. The MOF-74 family, including the Mg-containing version, has been extensively studied for the selective adsorption of $CO_2$.[76–78] Of particular note, the coordinatively unsaturated metal sites[79] of Mg-MOF-74 enable chemical bonding interactions between the metal and $CO_2$ adsorbate[76] that cannot be captured from classical force fields alone. We directly compare the adsorption dynamics against the results presented in Ref. 80, which considered the same system using a custom-trained ML force field generated using DeePMD-Kit[71] and PBE-D3 calculations in CP2K.[3]

MACE-MP-0b3 accurately and efficiently captures the $CO_2$ adsorption process in Mg-MOF-74. As shown in Fig. 3(c), the $CO_2$ adsorbate favorably binds to the Mg center in a tilted configuration that is in agreement with both experimental neutron diffraction data[77,81] and the previous custom-trained ML model.[80] The mean bond distance between the Mg center and $CO_2$ adsorbate is predicted to be 2.27 Å from MACE-MP-0b3 [Fig. 45(a)], in agreement

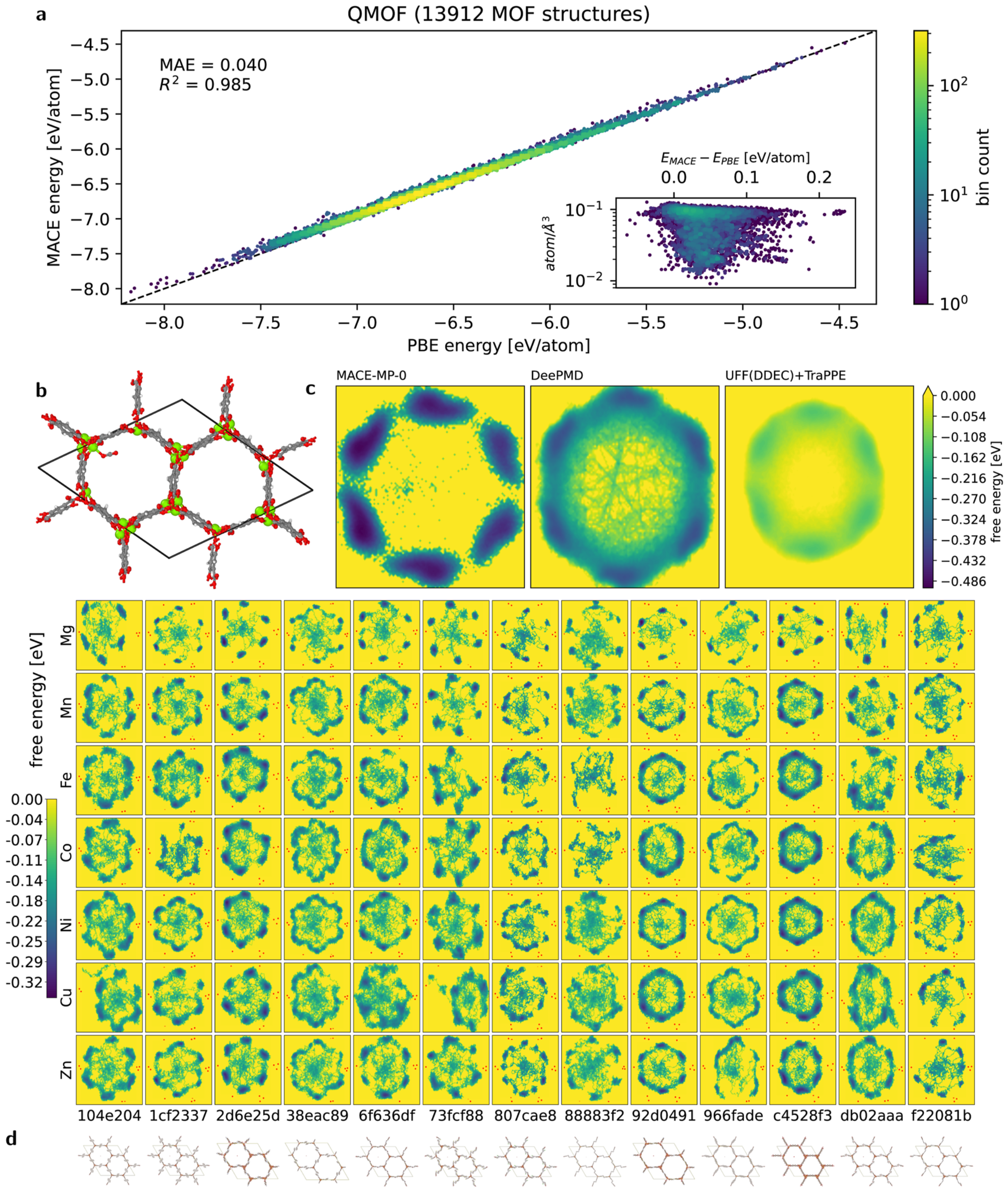


**FIG. 3.** Metal–organic frameworks. (a) Comparison between MACE-MP-0b3 and DFT (PBE) energies on 13 912 relaxed structures with compatible GGA calculations (i.e., without the Hubbard-U correction) and pseudopotentials in the QMOF database.[69,70] The inset presents the energy error distribution in relation to the atomic density (number of atoms per volume). The protocol for filtering incompatible calculations is provided in Appendix A 28. (b) Mg-MOF-74 structure with chemisorbed $CO_2$ optimized with MACE-MP-0b3. Color key: Mg (orange), O (red), C (brown), and H (white). (c) Left: Free energy landscape of $CO_2$ in Mg-MOF-74. Middle: Free energy landscape from Ref. 71 using a custom-trained DeePMD ML force field. Right: Free energy landscape using the UFF classical force field[72] with DDEC6 charges[73] for the framework and TraPPE for $CO_2$.[74] (d) Free energy maps of 91 hypothetical MOF-74 analogues, with the QMOF ID of the parent Mg-containing frameworks indicated at the bottom of each column and the transition metal to the left of each row.

with the experimental value of 2.27 Å[77] and the value of 2.23 Å from the custom ML model in Ref. 80. The mean Mg–O–C bond angle is predicted to be 137.3° from MACE-MP-0b3 [Fig. 45(a)], substantially closer to the experimentally determined bond angle of 131°[77] than the 118.6° value from the ML model in Ref. 80. The projected density map for the $CO_2$ adsorption site [Fig. 3(b)] is, again, in excellent agreement with prior studies[80,81] and shows how the adsorbed $CO_2$ molecules are mobile but largely confined to the vicinity of the Mg binding site due to chemisorption.

To showcase an example of how one might use the foundation model in a high-throughput setting, we considered 91 hypothetical MOF-74 analogues derived from those in Ref. 82 based on 13 (out of 58) different frameworks and seven different metal cations (M) that have been used to synthesize M-MOF-74.[77] Figure 3(e) shows the resulting free energy maps, comprising over 357 ns of simulation altogether, displaying diverse and dynamic behavior of the $CO_2$ adsorbate across the range of hypothetical MOF-74 analogues.

Given the nature of our foundation model, we anticipate many additional application areas where MACE-MP-0 (or one of its future variants) could be valuable in the MOF field. Based on the $CO_2$ adsorption example, we envision applications in capturing dynamic processes, particularly those that cannot be accurately modeled using classical force fields and are prohibitively expensive to carry out with *ab initio* MD, given the large unit-cell size required to describe most MOFs. Foundation models are promising for modeling competitive multi-component physisorption and chemisorption processes, especially across many families of compositionally different MOFs and combinations of gas mixtures, for which training a system-specific, on-the-fly active learning model would be expensive or even prohibitive. In addition to the compositional diversity relevant to high-throughput screening, not all MOFs can be described via a static picture based on an ideal crystalline structure: in fact, there has been recent interest in liquid and amorphous MOFs,[83,84] and the dynamic behavior of crystalline frameworks[85]—such as in the so-called "flexible" and "breathing" MOFs—has been leveraged for highly selective separation processes.[86] This dynamic behavior cannot be completely captured from static DFT calculations alone, and accurate and easily accessible interatomic potentials are expected to accelerate the modeling of spatiotemporal processes in future studies.[87]

## D. A wide range of applications and benchmarks

In Appendix A, in 32 subsections, we provide a rather wide ranging set of application examples to support the claim that the MACE-MP-0 is a robust modeling tool and, when fine-tuned, can reach *ab initio* accuracy. We also present the results of a comprehensive set of benchmarks, including the performance in calculating phonon dispersions, bulk and shear moduli of crystals, atomization energies, and lattice constants of elemental solids, the cohesive energies of the S66 set[88] of molecular dimers and the X23 set[89] of molecular crystals, the CRBH20 set[90] of reaction barrier heights, and the homonuclear diatomic binding curves. The full set of heteronuclear diatomic curves is provided in Appendix A.

We also provide more details on the training protocol, a graphical exploration of the data, including histograms of energies, forces, stresses, magnetic moments, and element and composition counts, and a discussion of the quantification of uncertainty in the model predictions.

## III. FINE-TUNING

Although the multitude of applications demonstrates that MACE-MP-0 is a robust model, it is also clear that, in many cases, it is not accurate enough out of the box to rival or replace *ab initio* calculations. For a selection of examples, we performed fine-tuning on configurations generated using MACE-MP-0, typically via molecular dynamics or other downstream tasks appropriate for the application. We used ~100 new configurations for each application during fine-tuning. To prevent catastrophic forgetting[91] and retain the robustness of the foundation model, we introduce a new fine-tuning protocol: *multi-head replay fine-tuning*. This approach includes a subset of the foundation model training data in the loss function while fine-tuning on the new data (see Appendix C 2 for details). We train a separate model for each case using this multi-head fine-tuning protocol with replay. Figure 4 shows the resulting force errors, which decrease significantly in every case. For comparison, we also present the force errors of a MACE model trained just on the small fine-tuning dataset. In almost all cases, the force errors of the model trained from scratch are significantly worse than those of the fine-tuned model. In each corresponding subsection of Appendix A, we demonstrate the performance of the fine-tuned model on application-relevant observables, showing considerable improvement over the original model in every case.

## IV. RELATED WORK: GENERAL PURPOSE MLIPs

The development of the MACE-MP-0 models as foundation models for atomistic materials simulation follows more than a decade of intense activity and progress in making MLIPs for specific materials. General purpose MLIPs—i.e., models that aim to target a wide range of chemical systems spanning many possible combinations of elements—are much more recent. Here, we summarize the brief history of such general-purpose models as well as the culmination of this trend into the creation of true "foundation models." Within this commentary, we seek to highlight the particular merits of MACE-MP-0 in comparison to existing alternatives. It is worth noting that the reason we call MACE-MP-0 "foundational" is because of how it can be used, as already mentioned in the introduction: the model is suitable for many different applications as a tool for initial exploration, but it likely requires fine-tuning for specific simulation tasks to achieve quantitatively accurate predictions.

A key advance toward making general-purpose MLIP models was made by MEGNet, introduced in 2019.[92] This model, which provides property predictions for inorganic crystals, was trained on minimum energy configurations in the Materials Project (MP)[93] that includes most elements of the periodic table.[89] Subsequently, models that predict atomic forces were also trained on MP-based datasets, including M3GNet[94] and CHGNet,[95] which were trained on snapshots of DFT relaxations of MP structures, with CHGNet using the MPtrj dataset introduced at the same time.[95] The ALIGNN-FF model[96] was trained on a database of inorganic crystals, JARVIS-DFT,[97] which covers 89 elements and uses

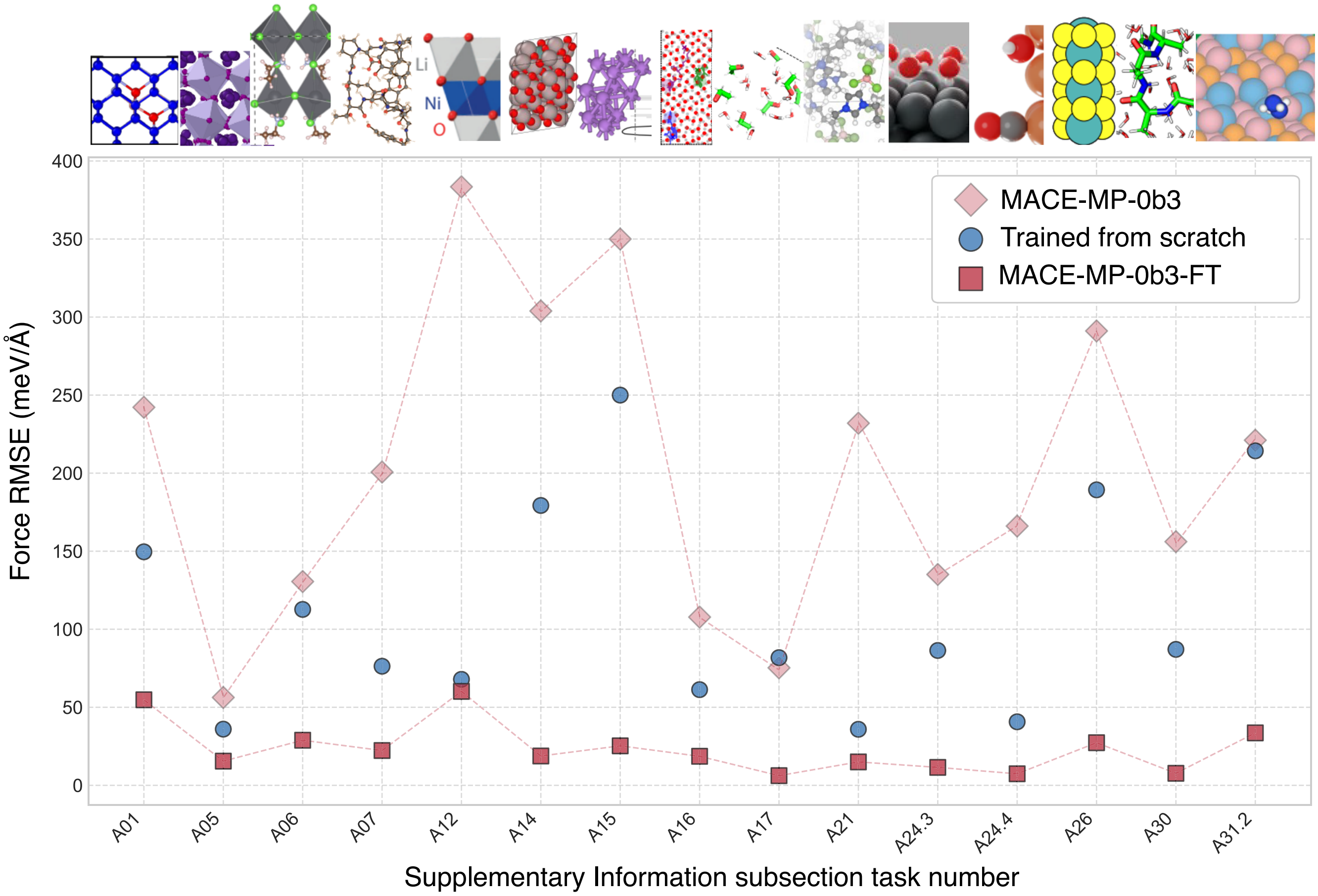


**FIG. 4.** Fine-tuning. A comparison of force rms error on selected applications in Appendix A for which fine-tuning was performed. The MACE-MP-0b3 model is shown with pink diamonds, and the fine-tuned model (MACE-MP-0b3-FT) for each application is shown with red squares. For comparison, in each case, we also show the results corresponding to a model trained "from scratch" on the small amount of fine-tuning data (blue circles).

the optB88vdW exchange-correlation functional.[98] The proprietary GNoME[99] (based on the NequIP architecture[16]) model also starts from MP but uses a complex active learning workflow to generate and train on a dataset of inorganic crystals nearly two orders of magnitude larger than MPtrj. The above-mentioned models were created primarily for the purpose of "materials discovery," i.e., predicting the thermodynamic stability of hypothetical inorganic crystals. In addition, they were capable of molecular dynamics for such crystals, and indeed, both CHGNet and GNoME were used to study alkali metal ion diffusion in battery materials. More recently, the DPA models (DPA-1[100] and DPA-2[101]) were trained on a wide variety of datasets (with 56 and 73 elements, respectively), a combination of some previously available and some released with the models (altogether 4M configurations). The second paper reports MD results for versions of the baseline model fine-tuned separately to specific systems (e.g., water, solid-state electrolytes, and ferroelectric oxides). To date, the most general and transferable force field for molecular dynamics is the PFP model[102] (TeaNet architecture[103]), which is also proprietary (including its training set that originally covered 45 elements, was recently updated to 72 elements,[104] and is significantly larger than MP and also covers molecules and surfaces). PFP was demonstrated for running simulations on solid-state ionic conductors and a molecular adsorption and a heterogeneous catalysis example—systems that formed part of its training dataset. There are also ML force fields specialized for organic molecules (with a much more limited number of elements), such as the ANI (and later AimNET) series of models[105–107] and the MACE-OFF models,[108] as well as for metal alloys.[109] However, there has yet to be a comprehensive demonstration that a single ML potential can describe solid, liquid, and gaseous systems of materials and molecules across the periodic table and well beyond the distribution of the underlying training set.

Since the first preprint version of this paper, a number of models have been fitted to the same MPtrj dataset and also to larger extended datasets, including the Alexandria[110] and OMat24.[111] Notable models (reported in preprint form) that showed high in-domain accuracy include SevenNet[112] (based on the NEquIP architecture[16]), GRACE,[113] Orb,[114,115] EquiformerV2,[111] MatterSim,[116] and eSEN.[117] Of these, the MatterSim models have been tested using molecular dynamics for some materials, including polymers and surfaces. The GRACE theoretical framework formally generalizes MACE, but the actual released GRACE models remain in close correspondence with the MACE design choices. To compare with these recent models, we also include results for a new model termed MACE-MPA-0 in Appendix A, which has been trained on an extended dataset including MPtrj and Alexandria.

## V. OUTLOOK

The stable MD propagation for a wide range of materials across the periodic table and the DFT-quality simulation (in some cases after fine-tuning) that we have shown here are landmark achievements for a single machine-learned interatomic potential. In this sense, we expect that the present study will have implications for the wider development of the field, beyond any specific model parameterization. Yet there are a number of limitations of the current ("b3" and "MPA-0") versions of the MACE-MP-0 foundation model. The exchange–correlation functional used in the MPtrj dataset is PBE,[118] which must be augmented with Hubbard *U* terms to improve electronic correlations for particular element combinations [introducing inconsistencies in the potential energy surface (PES) that must be compensated[6]] and dispersion corrections, such as D3.[29] Recent developments in DFT are beginning to supersede conventional generalized gradient approximation (GGA) functionals by achieving improved accuracy at comparable computational cost,[119,120] and methods beyond DFT, such as hybrid functionals[121] and the random phase approximation,[122] improve upon this even further, but at a much larger computational cost. Refitting or fine-tuning the model to a more modern function is expected to increase its predictive power and will reduce the need for system-dependent corrections such as the use of Hubbard *U* terms and dispersion. (Note that the above-mentioned inconsistency is not present in the more recent MATPES dataset,[123] which removes the Hubbard U correction altogether.)

The MACE architecture that we used to fit the data presently does not contain explicit long-range interactions (beyond the 12 Å receptive field afforded by two steps of message passing), nor does it take into account magnetic or spin degrees of freedom. Despite the success in describing many different chemistries demonstrated herein, there will be observables, particularly in the context of dilute solutions and at interfaces, that cannot be calculated with a short-range model. There are several approaches to incorporating explicit electrostatic interactions into atomistic ML models in the literature,[124–127] as well as spin degrees of freedom.[95,128,129] In the future, foundation models could undoubtedly benefit from such an extension.

Considering the results for the diverse systems shown in Appendix A, a particular area where the model clearly needs improvement is describing intermolecular interactions. While the overarching goal of MD stability is achieved, for many systems there is room for improvement in a quantitative sense, for example, in obtaining more accurate densities of molecular liquids, such as ethanol–water mixtures (Appendix A 17). The present version of the potential includes a repulsive pair potential[130] that helps describe the repulsive interaction of atoms at close range; the accuracy of the model (e.g., in predicting the equation of state) at high pressures is limited due to the absence of data in this regime. This can easily be remedied either by active learning[116] or a more systematic approach, e.g., replicating part of the MP dataset at lower and higher densities.

Although we have described an example of a model with wide generalization, we expect that there will be considerable improvements possible both in the model architecture and in optimizing the way in which data are assembled and the model is fine-tuned.[101,131,132] It is an open question whether the biggest gains will be obtained by improving the underlying data (both the amount and the consistency) or by scaling the size and expressivity of the model. There is good evidence that reaching higher levels of electronic structure theory (such as improved XC functionals) from a DFT baseline and beyond requires significantly less data than fitting to DFT itself,[106,133,134] and we show an example of this in Appendix A, where we fine-tune the model to data computed with the r2SCAN functional.[119]

Finally, there is the tantalizing possibility that, with some improvements, it will be possible to create an ML force field model that achieves quantitative agreement with explicit electronic structure theory across the full range of chemistry and structure. If this turns out to be true, future foundation models may truly provide a universal model for carrying out atomistic simulations at scale.

## VI. METHODS

### A. MACE

All models trained in the paper use the MACE[19] architecture implemented in PyTorch[135] and employ the *e3nn* library.[136] The MACE training and evaluation codes are distributed via GitHub under the MIT license, available at https://github.com/ACEsuit/mace/. The models used in this paper (and all previous and subsequent versions) are available at https://github.com/ACEsuit/mace-mp/. MACE is an equivariant message-passing graph tensor network where each layer encodes many-body information of atomic geometry. At each layer, many-body messages are formed using a linear combination of a tensor product basis.[23,24] This is constructed by taking tensor products of a sum of two-body permutation-invariant polynomials, expanded in a spherical basis. The final output is the energy contribution of each atom to the total potential energy. For a more detailed description of the architecture, see Refs. 19 and 137.

### B. Model versions

Different model versions have been released based on this study, including the models used in the first version of the paper, now named MACE-MP-0a, and the model used in the present version, named MACE-MP-0b3. Unless otherwise stated in the text, all models used in this paper correspond to MACE-MP-0b3. We use the label "MACE-MP-0" to refer to this model series generally.

### C. Hyper-parameters

The model referred to in this study uses two MACE layers, a spherical expansion of up to $l_{\max} = 3$, and 4-body messages in each layer (correlation order 3). The model uses a 128-channel dimension for tensor decomposition. We use a radial cutoff of 6 Å and expand the interatomic distances into ten Bessel functions multiplied by a smooth polynomial cutoff function to construct radial features, which are in turn fed into a fully connected feed-forward neural network with three hidden layers of 64 hidden units and SiLU nonlinearities. We fit an $L = 1$ model, corresponding to a "medium sized" model, as it represents a good compromise. More efficient models that only pass invariants during the message passing step ($L = 0$) or those that pass higher order tensors ($L \geq 2$) are straightforward to train and can form part of the accuracy/efficiency tradeoff in selecting the optimal model in the future. The irreducible representations of the messages have alternating parity (in *e3nn* notation, `128x0e + 128x1o`).

### D. Distance transforms and pair repulsion

Smooth behavior of the potential at close approach is essential for a broadly applicable model. We use a combination of Ziegler–Biersack–Littmark (ZBL)[138] core potential for the short-range repulsive forces and distance transformation to smoothly connect this behavior to equilibrium interactions. The ZBL energy is given by

$$E_{\mathrm{ZBL}} = \sum_{j} \frac{14.3996 \cdot Z_i \cdot Z_j}{r_{ij}} \cdot \phi(r_{ij}/a) \cdot \mathrm{Envelope}(r_{ij}, r_{\max}, p), \quad (1)$$

$$\phi(r/a) = c_0 e^{-3.2(r/a)} + c_1 e^{-0.9423(r/a)} + c_2 e^{-0.4028(r/a)} + c_3 e^{-0.2016(r/a)}, \quad (2)$$

where $Z_u$ and $Z_v$ are the atomic numbers of the interacting atoms, and $r_{ij}$ is the interatomic distance between atoms $i$ and $j$. The screening length $a$ is given by $a = 0.529 \cdot a_{\mathrm{prefactor}}/(Z_i^{a_{\mathrm{exp}}} + Z_j^{a_{\mathrm{exp}}})$. The coefficients are $c = \{0.1818, 0.5099, 0.2802, 0.028\,17\}$. The maximum cutoff radius is defined as $r_{\max} = R_{\mathrm{cov}}(Z_u) + R_{\mathrm{cov}}(Z_v)$. The envelope function $\mathrm{Envelope}(r, r_{\max}, p)$ is a polynomial cutoff function applied to smooth the potential. We use the same envelope as the radial basis. To smoothly transition from the ZBL to the MACE energy, we use the Agnesi distance transform,[22]

$$y_{ij} = \left(1 + \frac{a \cdot (r_{ij}/r_0)^q}{1 + (r_{ij}/r_0)^{q-p}}\right)^{-1}, \quad (3)$$

where $y_{ij}$ is the transformed distance and the parameters $a$, $q$, and $p$ control the shape of the transformation, $r_0 = \frac{1}{2}(R_{\mathrm{cov}}(Z_u) + R_{\mathrm{cov}}(Z_v))$. We then evaluate the radial basis in this transformed space instead of directly on the distances.

### E. Normalization

To ensure internal normalization of the weights and smooth extrapolation to high pressure systems, we divide the atomic basis in each layer by a learnable quantity called density normalization $e_i$,

$$e_i = 1 + \sum_{j} \tanh\left(\mathrm{SiLU}\left(\left[\sum_{k} W_k B_k(r_{ij})\right]\right)^2\right), \quad (4)$$

where $B$ denotes a set of Bessel basis and $W$ are learnable weights. The predicted density normalization varies between 1 and the number of neighbors of atom $i$, depending on the local environment. This normalization corresponds to a smooth version of the node degree normalization in graph neural networks.[139] The node energy $\epsilon_a$ of atom $a$ is shifted by the isolated atoms' energies. Therefore, the prediction of the energy for the whole structure is constructed as

$$\hat{E} = \sum_{a=1}^{N} \left[\sigma\left(\sum_{k=1}^{K} \epsilon_a^{(k)}\right) + \mu_{Z_a}\right],$$

where $K$ denotes the total number of message passing layers and $\epsilon_a^{(k)}$ is the energy of atom $a$ at layer $k$. $\mu$ and $\sigma$ are the isolated atomic energies and the mean square of the atomic forces computed on the training set. The predicted forces and stresses are computed as derivatives of the total energy with respect to the atomic positions and the strain tensor, respectively.

### F. Training loss

The models were trained using a weighted sum of Huber losses of energy, forces, and stress,

$$\begin{aligned}\mathcal{L} = {} & \frac{\lambda_E}{N_b} \sum_{b=1}^{N_b} \mathcal{L}_{\mathrm{Huber}}\left(\frac{\hat{E}_b}{N_a}, \frac{E_b}{N_a}, \delta_E\right) \\ & + \frac{\lambda_F}{3\sum_{b=1}^{N_b} N_a} \sum_{b=1}^{N_b} \sum_{a=1}^{N_a} \sum_{i=1}^{3} \mathcal{L}^{\star}_{\mathrm{Huber}}\left(-\frac{\partial \hat{E}_b}{\partial r_{b,a,i}}, F_{b,a,i}, \delta_F\right) \\ & + \frac{\lambda_\sigma}{9N_b} \sum_{b=1}^{N_b} \sum_{i=1}^{3} \sum_{j=1}^{3} \mathcal{L}_{\mathrm{Huber}}\left(\frac{1}{V_b} \frac{\partial \hat{E}_b}{\partial \varepsilon_{b,ij}}, \sigma_{b,ij}, \delta_\sigma\right), \end{aligned} \quad (5)$$

where $\lambda_E, \lambda_F, \lambda_\sigma$ are predetermined weights of energy ($E$), forces ($F$), and stress ($\sigma$) losses; the symbols under a hat correspond to predicted values; and $N_b$ and $N_a$ are the batch size and the number of atoms in each structure. In the last term involving the stress, $\varepsilon_b$ and $\sigma_b$ correspond to the strain and stress tensors, respectively. We used $(\lambda_E, \lambda_F, \lambda_\sigma) = (1, 10, 10)$ and Huber deltas of $\delta_E = 0.01$, $\delta_F = 0.01, \delta_\sigma = 0.01$. We use a conditional Huber loss $\mathcal{L}^{\star}_{\mathrm{Huber}}$ for forces, where the Huber delta $\delta_F$ is adaptive to the force magnitude on each atom. The Huber delta $\delta_F$ decreases stepwise by a factor from 1.0 to 0.1 as the atomic force increases from 0 to 300 eV/Å. For more details, see Appendix C 1.

### G. Optimization

The models are trained with the AMSGrad[140] variant of Adam[141] with default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. We use a learning rate of 0.001 and an exponential moving average (EMA) learning scheduler with a decaying factor of 0.999 99. We employ a gradient clipping of 100. The training curves for the medium model are presented in Fig. 67 in Appendix A. The model is trained for 100 epochs on 40–80 NVIDIA H100 GPUs across 10–20 nodes. Training the medium-sized model took ~2600 GPU hours. We find that MACE-MP-0 achieves an energy MAE of 18 meV/atom and a force MAE of 39 meV Å$^{-1}$ for the medium model. After fine-tuning with higher weights for energies for an additional 50 epochs, the small model is able to achieve an energy MAE of 13 meV/atom (see Appendix C 1).

### H. Performance

The speed of evaluation of the MACE-MP-0 model depends on the atomic density, hardware, floating point precision, size of model, etc. (see Appendix A 32 for details), but a rough guide is that on a single NVIDIA A100 GPU with 80 GB of RAM, it can do several nanoseconds per day for 1000 atoms. When run in parallel using domain decomposition, weak scaling at around 1 ns/day is perfect up to $1 \times 10^6$ atoms and 256 GPUs (4000 atoms/GPU) for a dense metallic alloy and can exceed 4 ns/day using 500 atoms/GPU.

## I. Training data

All data pertaining to this paper are available at https://zenodo.org/communities/mace-mp. The MACE-MP-0b3 model was trained on the `MPtrj` dataset, which was compiled originally for CHGNet.[95] This dataset consists of a large number of static calculations and structural optimization trajectories from the Materials Project (MP).[93] These include ~1.5M configurations (roughly ten times the ~150k unique MP structures), mainly small periodic unit cells (90% under 70 atoms) describing inorganic crystals with some molecular components. The DFT calculations use the PBE exchange-correlation functional with Hubbard $U$ terms applied to some transition metal oxide systems, but no additional dispersion correction.[142]

Since the potential we fit calculates the energy based only on structural information, ideally, we would like to use consistent electronic calculation parameters and the lowest energy electronic state for each configuration. One significant source of inconsistency is the application of Hubbard $U$, which is used in MP calculations only when O or F are present together with any of eight transition metals (Co, Cr, Fe, Mn, Mo, Ni, V, and W).[143] The application of $U$ leads to a shift in energy correlated with the value of $U$, i.e., a few eV, not explicitly accounted for in our fit. Therefore, energies from calculations using those eight elements with and without O or F are inconsistent (in the sense that the energy along a continuous deformation path that removes the O or F atoms from around these metals would be discontinuous). The pre-trained CHGNet fit to MPtrj used energies corrected to account for the presence or absence of $U$.[144] In our fit, this shift only occurs between structures with different compositions, and for any given composition, the energies should be consistent. As a result, we expect configurations that include local regions of these metals with very different O or F content, e.g., an interface between a metal and an oxide, to be poorly described.

In addition, the current fitting database includes a variety of magnetic orders generated as part of a systematic search for the magnetic ground state,[145] chosen from the full database only based on calculation type ("GGA Static" and "GGA Structure Optimization") and energy-difference criteria.[95] To quantify the effect of this additional and unaccounted-for degree of freedom, we classify the magnetic order associated with each calculation task into one of four categories: (1) no atomic magnetic moment listed, (2) moment converged to zero on all atoms, (3) converged to ferromagnetic order, and (4) converged to another magnetic order. Of the ~150k MP-IDs present, about 48k have more than one magnetic order present in the fitting database. In the vast majority of cases, this includes a calculation where the moments are *unknown* (i.e., not recorded) and a single other magnetic order, and we can hope that they are actually consistent. However, for 5186 MP-IDs, we find multiple non-trivial magnetic orders. To quantify the effect on the fitting quantities, we calculate the minimum energies of each magnetic order for each material and analyze the range of minimum values seen for each material (the distribution is plotted in Appendix A, Fig. 73). While the vast majority of materials have negligible variation, there are hundreds with variation >100 meV/atom (i.e., an order of magnitude larger than the energy error on the validation set) and a few that vary by <0.5 eV/atom.

## J. Long-range dispersion corrections

Dispersion interactions, sometimes called van der Waals interactions, are crucial for describing the weak, long-range interactions between electrons. Common approximations in DFT, such as PBE,[118] cannot capture such long-ranged interactions, motivating the use of additive non-local corrections, such as DFT-D3[29] or rVV10.[146] Inclusion of a dispersion correction to DFT is necessary to describe the dynamics of liquid water,[147] the geometries and binding energies of layered solids,[148] and the stability of metal–organic frameworks,[149] among many other examples.

Additive dispersion corrections typically employ a physical model for dispersion interactions with empirical parameters optimized to cut off the correction at interatomic distances where approximate DFT is reliable. DFT-D3 is an interatomic potential that uses tabulated values of atomic polarizabilities to describe two-body and, optionally, three-body Axilrod–Teller[150] dispersion interactions. As MACE-MP-0b3 is trained to PBE energies, forces, and stresses, it inherits PBE's lack of long-range dispersion interactions. An optional, additive DFT-D3 dispersion correction can be applied to MACE-MP-0b3. The PyTorch implementation of DFT-D3 used in this study is described in Ref. 102. The same parameters used in PBE-D3(BJ), i.e., DFT-D3 with a Becke–Johnson damping function,[151] are used in the D3 correction to MACE-MP-0b3.

## ACKNOWLEDGMENTS

Model training made use of resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 using Award Nos. BES-ERCAP0023528 and BES-ERCAP0022838. Part of this study was performed using the Cambridge Service for Data-Driven Discovery (CSD3), part of which is operated by the University of Cambridge Research Computing on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The DiRAC component of CSD3 was funded by BEIS capital funding via STFC capital Grant Nos. ST/P002307/1 and ST/R002452/1 and STFC operations Grant No. ST/R00689X/1. DiRAC is part of the National e-Infrastructure. The study of Y.C., J.R., A.D.K., M.C.K., M.A., and K.A.P. was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract No. DE-AC02-05-C.H.11231 (Materials Project Program No. KC23MP). We could not have performed this study without the DFT relaxation trajectories freely provided by the Materials Project and carefully curated into the MPtrj training set by Bowen Deng.[95] J.R. acknowledges support from the German Academic Scholarship Foundation (Studienstiftung). Y.C. acknowledges financial support from UC Berkeley and the Taiwan-UC Berkeley Fellowship from the Ministry of Education in Taiwan. M.C.K. acknowledges support by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2146752. Any opinions, findings, and conclusions or recommendations expressed in this study are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. N.B. was supported by a fundamental-research base-program funding from the U.S. Naval Research Laboratory. A.S.R. acknowledges support via a Miller Research Fellowship from the Miller Institute for Basic Research in Science, University of California, Berkeley. L.L.S. acknowledges support from the EPSRC Syntech CDT with grant Reference No. EP/S024220/1. A.M. and X.R.A. acknowledge support from the European Union under the "n-AQUA" European Research Council project (Grant No. 101071937). S.W.N., A.B., and T.V. acknowledge support from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Actions (Grant Agreement No. 945357) as part of the DESTINY Ph.D. program. G.C., C.P.G., T.V., A.B., and S.W.N. acknowledge support from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 957189 (BIG-MAP). V.K. acknowledges support from the Ernest Oppenheimer Early Career Fellowship and the Sydney Harvey Junior Research Fellowship, Churchill College, University of Cambridge. V.K. acknowledges computational support from the Swiss National Supercomputing Center under Project No. s1209. Z.E.-M. acknowledges support from the EPSRC Center for Doctoral Training in Theory and Modeling in Chemical Sciences (TMCS), under Grant No. EP/L015722/1. V.L.D. acknowledges support from UK Research and Innovation (Grant No. EP/X016188/1) and the John Fell OUP Research Fund. C.H. and F.Z. acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the framework of the priority Program No. SPP 2363, "Utilization and Development of Machine Learning for Molecular Applications—Molecular Machine Learning" Project No. 497249646, as well as further funding through the DFG under Germany's Excellence Strategy—EXC 2075—Grant No. 390740016 and the Stuttgart Center for Simulation Science (SimTech). A.M.E.'s work used the DiRAC Extreme Scaling service (Tursa) at the University of Edinburgh, which is part of the STFC DiRAC HPC Facility (www.dirac.ac.uk) and the scarf cluster (www.scarf.rl.ac.uk/) maintained by the Scientific Computing Department, STFC. A.M.E.'s access to DiRAC resources was granted through a Director's Discretionary Time allocation in 2023/24, under the auspices of the UKRI-funded DiRAC Federation Project. A.M.E.'s work was also supported by the Ada Lovelace Centre at STFC (https://adalovelacecentre.ac.uk/), Physical Sciences Databases Infrastructure (https://psdi.ac.uk), and EPSRC under Grant Nos. EP/W026775/1 and EP/V028537/1. I.B., R.E., and N.K. were supported by the Harding Distinguished Postgraduate Scholarship. H.J. gratefully acknowledges support from the Alexander von Humboldt (AvH) Foundation. H.H.H., J.T.M., and K.R. acknowledge support from the German Research Foundation (DFG) through DFG CoE e-conversion Grant No. EXC 2089/1. F.B. acknowledges the Alexander von Humboldt Foundation for a Feodor Lynen Research Fellowship and the Isaac Newton Trust for an Early Career Fellowship. B.X.S. acknowledges support from the EPSRC Doctoral Training Partnership (Grant No. EP/T517847/1). I.B., D.P.K., X.R.A., W.J.B., F.D.P., R.E., V.K., D.K., G.L., N.O., L.L.S., Christoph Schran, T.K.S., C.v.d.O., E.V.-U., and W.C.W. acknowledge access to CSD3 GPU resources through a University of Cambridge EPSRC Core Equipment Award (No. EP/X034712/1). We acknowledge project/application support by the Max Planck Computing and Data Facility. K.H., J.K., and V.S. acknowledge the Swedish Research Council (Vetenskapsrådet, Project No. 2021-06757) and the National Strategic e-Science program eSSENCE for funding, as well as the Swedish National Infrastructure for Computing (SNIC/NAISS) for providing computer resources used in this project. S.W.N., A.B., and T.V. acknowledge the Pioneer Center for Accelerating P2X Materials Discovery (CAPeX), DNRF Grant No. P3. Y.W. acknowledges support from the Shanghai Jiao Tong University. W.C.W. acknowledges support from the EPSRC (Grant EP/V062654/1). CO and C.H.H. acknowledge support from NSERC (Discovery Grant No. GR019381) and NFRF (Exploration Grant No. GR022937). Janine George, Jonas Grandel, and A.A.N. would like to acknowledge the Gauss Center for Supercomputing e.V. (https://www.gauss-centre.eu) for funding workflow-related developments by providing generous computing time on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Center (www.lrz.de) (Project No. pn73da). Janine George was supported by ERC Grant MultiBonds (Grant Agreement No. 101161771), funded by the European Union. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.) S.H. and J.J. acknowledge funding from EPSRC (Grant Nos. EP/T001038/1 and EP/S022953/1). A.D.K. acknowledges the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer). A.C.F. acknowledges funding from the EU Graphene Flagship, ERC grants Hetero2D and GIPT, EU grants Graph-X and CHARM, and EPSRC Grant Nos. EP/K01711X/1, EP/K017144/1, EP/N010345/1, EP/L016087/1, EP/V000055/1, and EP/X015742/1. W.J.B., Christopher Sutton, and C.P.G. acknowledge the US AFRL for partial funding of this project through Grant No. FA8655-21-1-7010. J.P.D., J.R.K., and G.C. acknowledge funding from the NOMAD Center of Excellence (European Commission Grant Agreement No. 951786). P.G. and T.D.S. acknowledge the support from the Cross-Disciplinary Program on Numerical Simulation of CEA, the French Alternative Energies and Atomic Energy Commission. P.G. and T.D.S. used access to the HPC resources of IDRIS under the Allocation No. A0120913455 attributed by GENCI. S.C. and M.C. acknowledge the support by the Swiss National Science Foundation (Project No. 200020_214879). F.B. and M.C. acknowledge support from NCCR–MARVEL, funded by the Swiss National Science Foundation (Grant No. 182892). M.C. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant No. 101001890–FIAMMA). We acknowledge the Jean Zay cluster of access to compute as part of the Grand Challenge No. GC010815458 (Grand Challenge Jean Zay H100). The work of I.-B.M. and P.K. made use of the facilities of the N8 Centre of Excellence in Computationally Intensive Research (N8 CIR) provided and funded by the N8 research partnership and EPSRC (Grant No. EP/T022167/1). The Centre is coordinated by the Universities of Durham, Manchester, and York. P.K. was supported by the Engineering and Physical Sciences Research Council (Grant No. EP/W524700/1). G.C. is grateful to Ágnes Borszéki for help with graphics.

## AUTHOR DECLARATIONS

### Conflict of Interest

G.C., J.P.D., C.v.d.O., and C.O. are partners in Symmetric Group LLP, which licenses force fields commercially, and G.C. and J.H.M. also have equity interests in Ångström AI.

### Author Contributions

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, Dávid P. Kovács, and Janosh Riebesell contributed equally to this study.

**Model training:** Y.C., P.B., I.B., and C.L.; **Data/Model analysis:** P.B., Y.C., J.R., N.B., R.E., M.C.K., E.S., and I.B.; **MACE code:** I.B., Y.C., S.W.N., D.P.K., P.B., W.C.W., M.A., S.V., E.S., and C.L.; **Application examples:** W.J.B. ($CsPbI_3$, Appendix A 5); L.L.S. (catalysis: $In_2O_3$, Sec. II B and Appendix A 24 d); I.B. (a-C quenches, Appendix A 3 a); Z.E.-M. (a-C graphitization, Appendix A 3 b); N.K. (Si interstitials, Appendix A 1); E.V.-U., X.R.A., and N.O. (aqueous interfaces, Sec. II A and Appendix A 19); Y.C. (molten salts, Appendix A 20); Christoph Schran, V.K., F.D.P., and X.R.A. (water and ice, Sec. II A and Appendix A 16); S.P.N. and A.G.-S. ($LiNiO_2$, Appendix A 12); S.W.N. (lithiated graphite, Appendix A 11); A.M.E. (zeolites Appendix A 10); J.J. (transition metal dichalcogenides, Appendix A 26); J.H.M. (ethanol/water, Appendix A 17, trialanine, Appendix A 30); G.L. and L.A.M.R. (a-Si Appendix A 2); D.K. (carborane, Appendix A 25, ammonia-borane, Appendix A 23); V.C. (S polymerization, Appendix A 9); J.R., Janine George, Jonas Grandel, and A.A.N. (phonons, Appendix B 1); J.R. and R.E.A.G. (materials discovery: formation energy, Appendix A 29 a); K.S.J. (materials discovery: stoichiometric substitutions, Appendix A 29 b); A.D.K. (materials discovery: highly coordinated structures, Appendix A 29 c); A.S.R., Y.C., and A.M.E. (MOFs, Sec. II C and Appendix A 28); M.V. (solvent mixtures, Appendix A 18); D.P.K., E.S., and S.M.B. (hydrogen combustion, Appendix A 8); Nima Karimitari, Christopher Sutton (HOIPs, Appendix A 6); F.F. and J.T. (protein dynamics and stability Appendix A 7); P.G., Y.W., T.D.S., J.R.K., and C.O. (point and extended defects in BCC metals, Appendix A 13); B.X.S. and F.B. (molecule–surface interactions, Appendix A 31); W.C.W. (HEA, Appendix A 32); E.F. and S.D. (catalysis: linear scaling relationships, Sec. II B and Appendix A 24 b); H.J. and H.H.H. (catalysis: CO oxidation on Cu, Sec. II B and Appendix A 24 c); S.H. and S.D. (catalysis: Pourbaix diagrams, Sec. II B and Appendix A 24 a); M.C.K. (benchmarks: bulk and shear moduli, Appendix B 2); F.D.P. (benchmarks: cohesive energies and lattice constants of solids, Appendix B 3, atomization energies Appendix B 4, and reaction barrier heights, Appendix B 5); T.K.S. ($Al_2O_3$, Appendix A 14, diatomics, Appendix B 6); J.P.D. (Arsenic random structure search, Appendix A 15); I.-B.M. and P.K. (high-pressure hydrogen, Appendix A 22); I.-B.M., P.K., and C.v.d.O. (electrode–electrolyte interface/battery system, Appendix A 27); J.K., V.S., and K.H. ($CeO_2$, Appendix A 4); F.Z. (ionic liquids, Appendix A 21); **Fine-tuning**: I.B., N.B., and T.W. **Uncertainty quantification:** Filippo Bigi, M.C., S.C., C.H.H., C.O., and Y.W. (Appendix D); **Supervision of research:** A.B., A.C.F., A.M., A.S.R., C.H., C.O., C.P.G., Christopher Sutton, G.C., H.H.H., Janine George, J.K., J.T.M., K.A.P., K.H., K.R., M.A., M.C., S.D., S.M.B., T.V., V.L.D., and W.H.; **Drafted manuscript:** I.B., N.B., Y.C., G.C., S.D., H.H.H., M.C.K., J.R., A.S.R., Christoph Schran, and J.T.M.; **Edited manuscript:** I.B., N.B., Y.C., G.C., V.L.D., Janine George, R.E.A.G., J.R., M.C.K., K.A.P., A.S.R., L.L.S., J.T.M., A.M., C.O., A.M.E., W.C.W., and J.L.A.G.; **Supervised manuscript writing:** N.B., G.C., V.L.D., V.K., J.T.M., and Christoph Schran.

**Ilyes Batatia**: Conceptualization (lead); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (lead); Software (lead); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Philipp Benner**: Conceptualization (equal); Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Resources (equal); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Yuan Chiang**: Data curation (lead); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (supporting); Resources (equal); Software (supporting); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Alin M. Elena**: Data curation (supporting); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Dávid P. Kovács**: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (supporting); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Janosh Riebesell**: Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (supporting); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Xavier R. Advincula**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Mark Asta**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **Matthew Avaylon**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – review & editing (supporting). **William J. Baldwin**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Fabian Berger**: Formal analysis (supporting); Investigation (supporting); Supervision (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Noam Bernstein**: Data curation (equal); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Project administration (supporting); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Arghya Bhowmik**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Project administration (supporting); Supervision (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **Filippo Bigi**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Project administration (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Samuel M. Blau**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Vlad Cărare**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Michele Ceriotti**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Sanggyu Chong**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **James P. Darby**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Sandip De**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Flaviano Della Pia**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Volker L. Deringer**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Rokas Elijošius**: Formal analysis (supporting); Investigation (supporting); Visualization (equal); Writing – review & editing (supporting). **Zakariya El-Machachi**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Edvin Fako**: Formal analysis (supporting); Investigation (supporting); Visualization (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **Fabio Falcioni**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Andrea C. Ferrari**: Supervision (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **John L. A. Gardner**: Writing – review & editing (supporting). **Mikołaj J. Gawkowski**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Annalena Genreith-Schriever**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Janine George**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Rhys E. A. Goodall**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Jonas Grandel**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Clare P. Grey**: Formal analysis (supporting); Supervision (equal); Writing – review & editing (supporting). **Petr Grigorev**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Shuang Han**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Will Handley**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Writing – review & editing (supporting). **Hendrik H. Heenen**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (supporting). **Kersti Hermansson**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Cheuk Hin Ho**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Stephan Hofmann**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Writing – review & editing (supporting). **Christian Holm**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Writing – review & editing (supporting). **Jad Jaafar**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Konstantin S. Jakob**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Hyunwook Jung**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Venkat Kapil**: Formal analysis (supporting); Investigation (supporting); Project administration (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Aaron D. Kaplan**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Nima Karimitari**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **James R. Kermode**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Panagiotis Kourtis**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Namu Kroupa**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Jolla Kullgren**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Matthew C. Kuner**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (supporting). **Domantas Kuryla**: Formal analysis (supporting); Investigation

(supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Guoda Liepuoniute**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting). **Chen Lin**: Data curation (supporting); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Writing – review & editing (supporting). **Johannes T. Margraf**: Formal analysis (supporting); Investigation (supporting); Project administration (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Ioan-Bogdan Magdău**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Angelos Michaelides**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **J. Harry Moore**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Aakash A. Naik**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Samuel P. Niblett**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Sam Walton Norwood**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Niamh O'Neill**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Christoph Ortner**: Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Supervision (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **Kristin A. Persson**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Writing – review & editing (supporting). **Karsten Reuter**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Writing – review & editing (supporting). **Andrew S. Rosen**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Louise A. M. Rosset**: Formal analysis (supporting); Investigation (supporting); Writing – review & editing (supporting). **Lars L. Schaaf**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Christoph Schran**: Formal analysis (supporting); Investigation (supporting); Project administration (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Benjamin X. Shi**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Eric Sivonxay**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Tamás K. Stenczel**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Christopher Sutton**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Viktor Svahn**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Thomas D. Swinburne**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Jules Tilly**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Cas van der Oord**: Conceptualization (supporting); Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Santiago Vargas**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Eszter Varga-Umbrich**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Tejs Vegge**: Formal analysis (supporting); Investigation (supporting); Supervision (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **Martin Vondrák**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Yangshuai Wang**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **William C. Witt**: Formal analysis (supporting); Investigation (supporting); Software (equal); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Thomas Wolf**: Formal analysis (supporting); Investigation (supporting); Resources (equal); Writing – review & editing (equal). **Fabian Zills**: Formal analysis (supporting); Investigation (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting). **Gábor Csányi**: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing – original draft (lead); Writing – review & editing (lead).

## DATA AVAILABILITY

The data that support the findings of this study are openly available at http://github.com/acesuit/mace-mp.

## APPENDIX A: FURTHER APPLICATIONS

The following sections contain a diverse set of examples where the MACE-MP-0b3 foundation model is applied to a variety of material and chemical systems, with each subsection containing one application with one or more related examples.

**Similarity statement**

Each subsection also contains a statement (both qualitative and quantitative) about the extent to which the training data contain configurations similar to those relevant to the application in that section. This should inform the reader about the degree of extrapolation inherent in the particular example. In order to facilitate further scrutiny, we provide a data file for most applications that can be used in conjunction with the `chemiscope` tool (at https://chemiscope.org/) to explore the chemical environments in the training data and the application example, as well as their relation to one another.

**Performance summary**

Each subsection also contains a concise statement summarizing the performance of the MACE-MP-0b3 model in the application.

### 1. Self-interstitials in silicon

Di-interstitial silicon ($I_2$) constitutes a test case for the transferability of MACE-MP-0 to point defects in a periodic lattice. In the following, self-diffusion coefficients $D$ and interstitial migration energies $E_\mathrm{m}$ are obtained from MD simulations without a D3 dispersion correction. Consistency tests with a $(64 + 2)$-atom silicon structure from Ref. 152 relaxed with the PW91 functional are performed. Relaxing with MACE-MP-0 showed no change in energy at a force tolerance of 0.05 eV/Å.

The $I_2$ structure was generated by relaxing a $(216 + 2)$-atom diamond structure with a lattice constant of 5.4 Å. After running NVT MD, the distance to the closest lattice site for each atom was plotted. The interstitials propagate consistently with the accepted mechanism.[153] Isolating the trajectories of the point defects shows a characteristic jump length corresponding to expected jumps between stable sites.

At temperatures below 700 K, the interstitials almost exclusively remain bound in a state corresponding to the ground state with $C_{1h}$ symmetry[154] and transition between symmetry-equivalent ground states. At higher temperatures, higher-energy states are sampled in which the interstitials separate into tetrahedral interstitial states, lying at a higher energy of 0.3 eV (Fig. 5). We calculate $D$ from the mean-square-displacement (MSD) of all atoms in the unit cell.[155] The fit to the Arrhenius law (Fig. 5) gives $E_\mathrm{m} = (0.36 \pm 0.07)$ eV, in agreement with Refs. 156 and 157.

The above-mentioned calculations were repeated for a single interstitial ($I_1$) in a 64-atom cell. The interstitial predominantly occupies the tetrahedral state and transitions between symmetry-equivalent states via the split $\langle 110 \rangle$ state. While these states are expected, the occupancy of each state shows larger deviations from those reported in Ref. 155. Figure 5 shows a nudged elastic band (NEB) between two tetrahedral interstitial sites via a $\langle 110 \rangle$ split interstitial site, from which an energy barrier of 0.6 eV is calculated. Repeating the MD simulations at several temperatures (Fig. 5) gives $E_\mathrm{m} = (0.53 \pm 0.06)$ eV, in agreement with the Local density approximation (LDA).[155] While the prefactor is strongly system-dependent,[158] the migration energy is an intrinsic property of the energy landscape and may be compared with other calculations. However, the finite system size and the interaction between interstitials result in the difference between the single- and di-interstitial migration energies.

MACE-MP-0b3 displays the correct qualitative dependence of the phonon instability temperature $T_\mathrm{crit}$ on interstitial density (Fig. 5). Specifically, $T_\mathrm{crit} = 1400$ K of the $(216 + 2)$-system, which has an interstitial density of ~0.9%, is lower than the melting temperature of $(1449 \pm 10)$ K calculated for cubic diamond silicon with PBE.[159] A further decrease to $T_\mathrm{crit} = 1200$ K is observed for the $(64 + 1)$-system, which has a larger interstitial density of 1.5%. However, the phonon instability temperature for the $(64 + 1)$-system, calculated with LDA, is 1473 K,[155] which MACE-MP-0b3 consequently underestimates. The melting points and phonon instability temperatures are expected to have a dependence on the exchange-correlation functional.

**Similarity statement**

The MP dataset contains 41 pure silicon structures, including the diamond structure but no self-interstitial defects.

**Performance summary**

Silicon interstitials display the correct set of local minima, while their relative occupancy at finite temperature is not correct. The predicted activation energies for self-diffusion of the single- and di-interstitials are consistent with previous force field and DFT calculations. The dependence on interstitial density is at least qualitatively correct, whereas the phonon instability temperature of the single interstitial system is too low by about 20%.

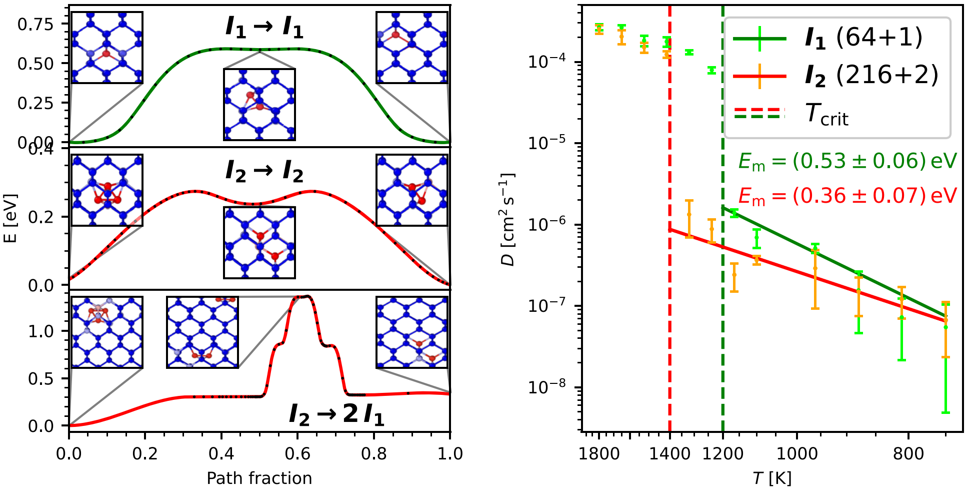


**FIG. 5.** Single-interstitial ($I_1$) and di-interstitial ($I_2$) defects in silicon. Left: Nudged elastic band paths between two metastable sites with a subset of images shown (black dots). Right: Diffusion coefficient against inverse temperature and Arrhenius laws with migration energy $E_\mathrm{m}$ and phonon instability temperature $T_\mathrm{crit}$.

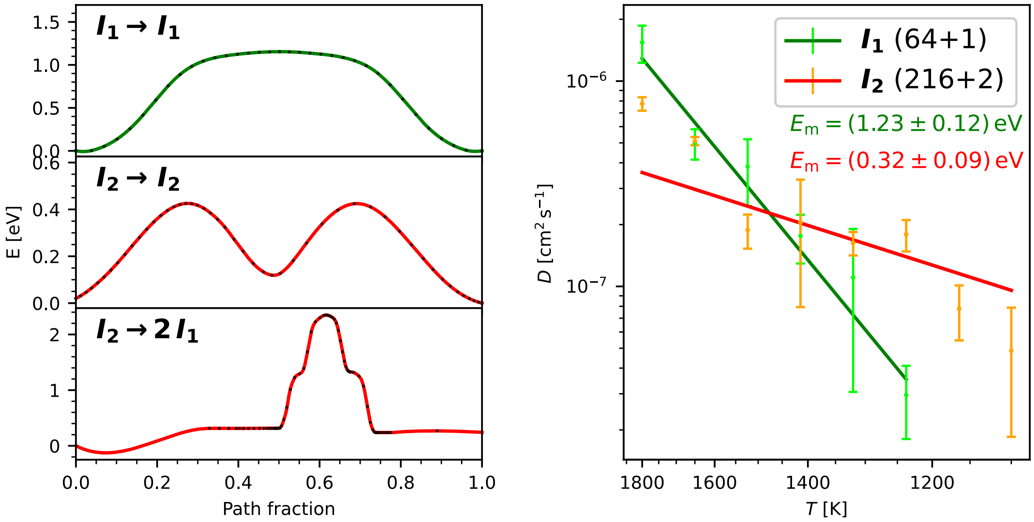


**FIG. 6.** Results for the fine-tuned model. Plots are the same as in Fig. 5.

**Fine-tuning**

Fine-tuning was performed on 100 configurations sampled from an MD trajectory at 1800 K from the $I_2$ system. Figure 6 shows the NEBs and diffusion coefficients, recomputed with the fine-tuned model. The barrier heights of all NEBs have increased. The migration energy for $I_1$ has increased to $E_m = (1.23 \pm 0.12)$ eV, in agreement with the increase in barrier height. However, this is significantly larger than the LDA value in Ref. 155. For $I_2$, the migration energy has not changed, within the error, and remains consistent with the barrier height. Finally, we note that the phonon instability temperatures for $I_1$ and $I_2$ are now above the melting temperature of cubic diamond silicon calculated with PBE.[159]

## 2. Amorphous silicon from melt-quench simulations

Amorphous silicon (a-Si) is a prototypical disordered material and has served as an example of both the physical and chemical insight afforded by machine-learned potentials[160,161] and as a challenging benchmark for the development of new methodology and potential models.[21,152,162]

Here, we assess the performance of the MACE-MP-0b3 potential on standard melt-quench simulations and compare it to the well-established Si-GAP-18 potential, a general purpose GAP model for Si[152] that has been extensively validated in previous literature.[160,161,163] We perform melt-quench simulations of 512-atom systems quenching from the liquid state at 1400 K to the amorphous state at 300 K at a rate of $1 \times 10^{12}$ K s$^{-1}$, for a total of 1100 ps. The simulations were run in Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS),[164] in the NpT ensemble using a Nosé–Hoover thermostat and barostat with a time step of 1 fs. A well-equilibrated liquid starting structure was generated by annealing a random hard-sphere model at 1400 K for 10 ps using Si-GAP-18. The protocol was repeated three times for each model to assess the variability in predictions.

Figure 7(a) presents the temperature, density, and average coordination within a cutoff of 2.85 Å as a function of the progress of the quench simulation. The density of the liquid structure stays constant for the Si-GAP-18 model but increases slowly to around 2.65 g cm$^{-3}$ for the MACE-MP-0b3 model, above the experimental density of liquid Si at 2.57 g cm$^{-3}$.[165] When quenched from the liquid state, Si undergoes a vitrification transition that is accompanied by a sudden decrease in density and concomitantly by a decrease in the average atomic coordination number, as the highly coordinated metallic liquid transforms into the mainly fourfold coordinated semiconducting a-Si. The MACE-MP-0b3 model predicts a glass transition temperature around 600 K, compared to around 1050 K for Si-GAP-18, and experimental observations of around 1000 K.[166,167] Hence, the MACE-MP-0b3 model appears to overstabilize the liquid phase below the experimental glass transition temperature. After the transition, the resulting bulk structures are mostly fourfold-connected, with average densities of 2.266 g cm$^{-3}$ for Si-GAP-18 and 2.233 g cm$^{-3}$ for MACE-MP-0b3, close to the experimental density of 2.285 g cm$^{-3}$.[168]

The final structure from one of the three quench simulations driven by MACE-MP-0b3 is presented in Fig. 7(b), where the Si atoms are color-coded by coordination numbers. This confirms the insight from Fig. 7(a), which is that the majority of atoms are 4-fold coordinated, with some 3- and 5-fold coordination defects.

We further compare the excess energies ($\Delta E$), calculated relative to crystalline diamond-type silicon (**dia**), as shown in Fig. 7(c). These values were computed by taking structural snapshots throughout the quench and relaxing them either with Si-GAP-18 (left) or MACE-MP-0b3 (right), similar to Ref. 163. Both Si-GAP-18 and MACE-MP-0b3 show small energy fluctuations between individual runs for the Si-GAP-18 simulations, while MACE-MP-0b3 had some minor variability in each run. MACE-MP-0b3 predicts the same overall trend in energies as Si-GAP-18, but the energy predictions have an offset of around +0.05 eV at$^{-1}$ for the liquid structures, while the predictions from both models on the amorphous structures are very close. The final structures quenched by MACE-MP-0b3 have a higher energy (0.18 eV at$^{-1}$) than those quenched by Si-GAP-18 (0.16 eV at$^{-1}$). This compares to a previous result of 0.14 eV

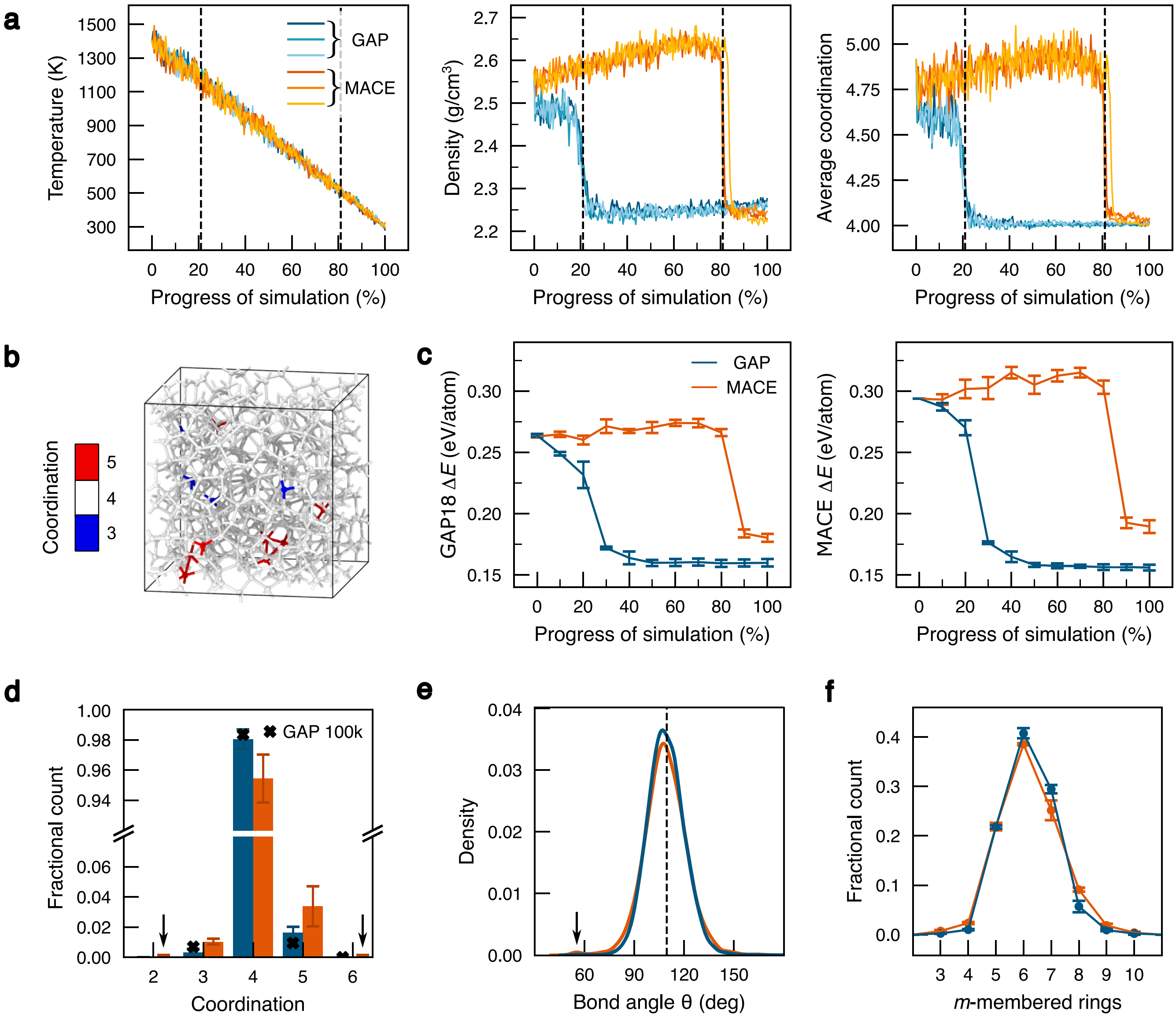


**FIG. 7.** Amorphous silicon. We characterize structural models of amorphous silicon simulated by melt-quenching using MACE-MP-0b3 and benchmark against the well-established Si-GAP-18 model.[152] (a) Evolution of selected properties during the quench simulations: temperature (left), density (center), and average coordination number, defined using a bond-length cutoff of 2.85 Å (right). (b) Amorphous silicon structure generated using the MACE-MP-0b3 potential, where atoms are color-coded by coordination number. (c) Excess enthalpies (Δ*E*) relative to diamond-type silicon, computed using Si-GAP-18 (left) and MACE-MP-0b3 (right), for relaxed snapshots from simulations with either potential. (d) Coordination count, where the black crosses represent a study of a 100 000-atom system driven by Si-GAP-18.[160] (e) Bond-angle distribution, where the dashed line is the equilibrium 109.5° angle for tetrahedral environments. (f) Ring-size distribution.

at$^{-1}$ for a 4096-atom system quenched at a rate of $1.0 \times 10^{11}$ K s$^{-1}$ driven by Si-GAP-18[163] and the experimental excess enthalpy of a-Si after deposition and annealing of 0.14 eV at$^{-1}$.[169] The present quench rate is one order of magnitude faster; hence, higher average enthalpies and variability can be expected. Nonetheless, the excess enthalpies predicted by MACE-MP-0b3 are qualitatively in the correct ballpark, viz., slightly above the corresponding crystalline phase.

Probing the short range order of the structure, Fig. 7(d) presents the distribution of Si coordination, averaged over all three repeats for each model. While the Si-GAP-18 structures present fewer than 2% of 3- and 5-fold defects, the MACE-MP-0 structures feature around 4% of coordination defects, mostly 5-fold coordinated atoms, but also 2-, 3-, and 6-fold defects in smaller concentrations. The 100k-atom simulations in Ref. 160 predicted that a well-relaxed a-Si would contain on the order of 1.5%–2%

of defects, consistent with the 4096-atom simulations of Ref. 163. Faster quenching leads to higher defect counts, as seen from the Si-GAP-18 results in Fig. 7, whereas the overall defect count predicted by MACE-MP-0 is still notably higher than that of Si-GAP-18 at the same quench rate. Another measure of the quality of the short-range structure is the bond angle distribution, shown in Fig. 7(e), where both distributions appear very similar, but for a small feature at 60° for the MACE-MP-0 structures [arrow in Fig. 7(e)].

Finally, we assess the quality of the medium-range order of the network using shortest-path ring statistics with the `matscipy` package,[170] depicted in Fig. 7(f). The distribution of $m$-membered rings is very similar for both models, but the MACE-MP-0 structures have an increased count of small ($m < 5$) and large ($m > 7$) rings, with the presence of $m = 3$ rings that support the feature in Fig. 7(e) at 60° and the cluster of 5-fold coordinated atoms in Fig. 7(b).

**Similarity statement**

The MP dataset includes 41 different silicon-only structures; however, many of them are very high-density (high coordination number) or crystalline (all 4-fold). There were no cases of wide coordination number distribution, as seen in liquid Si; however, we found five unique a-Si structures (with 100 atoms each) with a mix of slightly higher and lower coordination numbers, providing information about a-Si. Based on a UMAP analysis, the closest structures in the training set are mp-1244971, mp-1245242, and mp-1245041. To help with visualization, we provide `amorphous-silicon.json` on https://chemiscope.org/.

**Performance summary**

The model performs reasonably well for the description of the melt–quench process, leading to good-quality a-Si structures, albeit markedly underestimating the vitrification temperature and the excess enthalpy and overestimating the number of 5-fold coordinated atoms compared to a domain-specific ML potential.

### 3. Amorphous carbon

#### a. Melt-quench simulations

Carbon forms many different crystalline and amorphous modifications. The structural diversity of amorphous carbon (a-C),

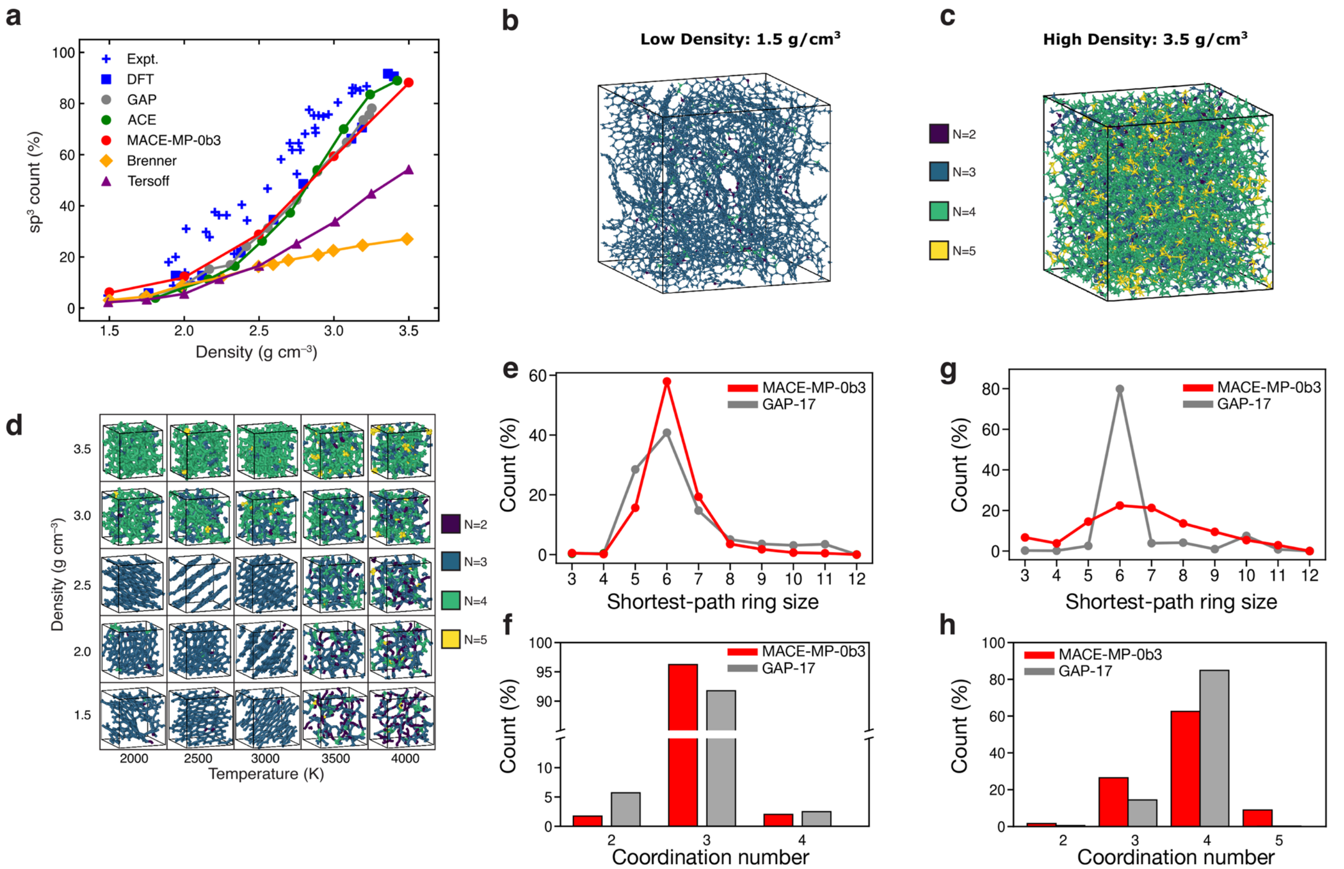


**FIG. 8.** Amorphous carbon. (a) Count of sp$^3$ (fourfold coordinated) carbon atoms in melt-quenched carbon structures as a function of density. The results obtained with the MACE-MP-0b3 model are compared to computational and experimental data compiled in Ref. 171 and references therein, as well as Refs. 172 and 173. (b) and (c) 4096-atom structures generated using the MACE-MP-0b3 potential. (d) Results from 25 × 200 atom graphitization simulations spanning relevant temperature and density ranges, similar to Ref. 174. The structures in panels (b)–(d) are color-coded according to coordination numbers as indicated in the legends. (e) Shortest-path ring size count for 4096-atom low-density structures as determined using `matscipy`.[170] (f) Coordination number count for 4096-atom low-density structures. (g) and (h) Same as panels (e) and (f) but for high-density structures. The results in panels (e)–(h) are shown for a C-GAP-17-driven simulation (gray) and for a comparable simulation driven by the MACE-MP-0b3 potential (red).

characterized by the simultaneous presence of threefold coordinated ($sp^2$) and fourfold coordinated carbon atoms ($sp^3$), makes it a challenging system for both classical and ML force fields.[175,176] The correct description of its growth mechanism has been among the early successes of ML-driven materials modeling.[177]

We assess the accuracy of the MACE-MP-0 model in reproducing the structural complexity of amorphous phases by plotting the concentration of fourfold coordinated atoms ($sp^3$) as a function of density in Fig. 8. To generate amorphous structures with a given density with the MACE-MP-0 model, we perform melt-quench simulations. We start by melting diamond structures at a given density by running NVT simulations at 8000 K for 3 ps. We then perform a fast quench, reducing the temperature from 8000 to 300 K at a cooling rate of 1000 K $ps^{-1}$. Finally, we optimize the geometry with Limited-memory Broyden–Fletcher–Goldfarb–Shanno (LFBGS) to obtain the final structure and determine the count of $sp^3$ atoms using a bond-length cutoff of 1.85 Å. We observe in Fig. 8 that the MACE-MP-0 model predictions reproduce the trend observed in both the DFT[172] and the experimental data extracted from.[171] We also see good agreement with results of quenches using the carbon ACE reported by Qamar *et al.* in Ref. 173 and using C-GAP-17 reported in Ref. 172, noting that both potentials had been specifically trained on large carbon datasets.

### *b. Graphitization*

MACE-MP-0 was used for two annealing runs for simulation cells containing 4096 atoms at low density (1.5 g $cm^{-3}$) and high density (3.5 g $cm^{-3}$), respectively. The low-density simulation was run at 2000 K, and the high-density simulation was run at 4000 K. In addition, 25 × 200 atom annealing runs spanning from 2000 to 4000 K and 1.5–3.5 g $cm^{-3}$ were conducted to explore a finer grid of parameters. Both 4096-atom structures were compared to structures generated using C-GAP-17 using the same protocol, which was also used recently for an "on-the-fly" generated GAP potential.[174] The protocol has two stages. The goal of stage I is to prepare the starting configuration for the annealing in stage II, and it begins with a random structure with a hard-sphere constraint of $r_{\min} \geq 1$ Å and equilibrating it at 9000 K for 40 ps, followed by cooling to 5500 K over 40 ps and subsequent quenching to 300 K over 10 ps. The structures are then held at 300 K over 50 ps before being rapidly heated up to the annealing temperature over 10 ps. This concludes stage I. In stage II, the structure is annealed at 2000 or 4000 K for 350 ps using a time step of 1 fs. We used C-GAP-17 to perform stage I because MACE-MP-0 was found to be unstable at 9000 K.

For the low-density case, the structures generated by MACE-MP-0 in annealing simulations agree with the predictions of the established C-GAP-17 model in terms of overall trends [Figs. 8(b) and 8(d)]. For more detailed insight, Fig. 8(e) shows the shortest-path ring distribution for the low-density structure, indicating that MACE-MP-0 predicts a greater number of six-membered rings and fewer large rings for the low-density structure compared to C-GAP-17 (thus suggesting a higher degree of crystallinity in the MACE-MP-0 prediction)—this might be correlated with the higher relative count of graphite vs amorphous-like structures in the training dataset, although further analysis is required.

Figure 8(g) shows poor agreement between MACE-MP-0 and C-GAP-17 in terms of ring count for the high-density structure. Whereas annealing with C-GAP-17 led to a partly crystallized structure under the chosen conditions (indicated by the large abundance of six-membered rings, as are found in the diamond structure), the MACE-MP-0b3 model gave rise to a highly disordered structure—this is evident from a visual inspection of Fig. 8(c) and reflected in the ring-count plot in Fig. 8(g), showing a notable number of (presumably strained) three-membered rings as well as larger ring sizes of >7. In other words, the MACE-MP-0b3 appears to fail to describe the crystallization in this case [we note that a partially ordered structure was obtained in a 200-atom simulation at 3000 K but not at 4000 K at the same density; Fig. 8(d)]. Finally, Figs. 8(f) and 8(h) show the coordination number for both 4096 atom structures. MACE-MP-0b3 leads to more $sp^2$ environments and fewer sp and $sp^3$ environments compared to C-GAP-17 [again likely consistent with a higher degree of crystallinity, as suggested by the ring count in Fig. 8(e)]. In the high-density 4096-atom structure obtained with MACE-MP-0b3, a notable amount of 5-fold coordinated atoms is observed, on the order of about 10%—a behavior that would not be expected in a simulation of carbon at diamond-like density.

We note that these issues are not expected to be a fundamental shortcoming of the MACE architecture: we show that training on a wider-ranging dataset with more crystal structures like the Alexandria dataset (which notably has no amorphous structures) results in a model, MACE-MPA-0, that describes the structure of amorphous carbon in better agreement with C-GAP-17 at high and low density, as shown in the section dedicated to the MACE-MPA-0 model (A.33).

**Similarity statement**

The MP dataset contains 89 different all-carbon structures, most of which correspond to diamond and related stacking polytypes (lonsdaleite-like and more complex ones, all purely $sp^3$-bonded), as well as graphite in various forms. These structures include a range of mixed configurations with $sp^2/sp^3$ coexistence—the latter is expected to be critical for a correct description of a-C. The dataset also contains a number of compressed and defective fullerene configurations, with one of those cells containing sp-, $sp^2$-, and $sp^3$-like environments and a few hypothetical allotropes (notably "T-carbon" and a cubane-motif-based form, representing three- and four-membered shortest-path rings, respectively). In essence, the dataset does contain relevant carbon environments but does not contain a significant share of highly disordered carbon configurations. Based on UMAP analysis, we find that the closest structures in the training set are mp-568028 and mp-568806. We provide `amorphous_carbon.json` to help visualize the interactive UMAP on https://chemiscope.org/.

**Performance summary**

The MACE-MP-0 model correctly captures the $sp^3$ content as a function of density in melt–quench simulations. Detailed analysis of long-annealing simulations shows qualitative agreement with a purpose-trained ML force field for low-density a-C graphitization and poor agreement for high-density annealing.

### 4. Ceria nanoparticles

Cerium oxide (ceria, $CeO_2$) is a reducible metal oxide with intriguing chemical and physical properties and important technological applications, especially for nanostructured ceria. Experiments in the literature have, for example, shown that the oxygen storage capacity (OSC) of ceria at the nanoscale is strongly shape- and size-dependent.

Behind the versatile usage of ceria lies one overriding feature, namely, its exceptional reduction–oxidation (redox) properties enabled by the duality of the cerium ion ($Ce^{4+} \leftrightarrow Ce^{3+}$). It is generally a formidable task to try to mimic interactions, structure, and energetics simultaneously for a compound like ceria without having access to explicit electrons. In an earlier study,[178] we constructed a reactive interaction model using the ReaxFF modeling framework[179] with the aim of handling stoichiometric and partially reduced ceria bulk, surfaces, and nanoparticles (NPs). The model was based on a training set of DFT calculations for a large number of ceria systems in various forms and configurations (bulk, clusters, and surface systems; stoichiometric as well as reduced systems). With some exceptions, that model performed very well. In the present study, instead, our force field is MACE-MP-0, where the content of ceria in the training set comprises only bulk structures, namely exactly 18 bulk polymorphs (stoichiometric or partially reduced).

Here, we assess the ability of the MACE-MP-0 model to describe small ceria nanoparticles of different shapes, sizes, and reduction degrees without the training ever including any ceria NPs or surfaces. The optimized NP structures/shapes, as well as their formation energies (with respect to stoichiometric bulk $CeO_2$ and gas-phase $O_2$ as needed), will be assessed.

The left panel in Fig. 9 shows the MACE-MP-0 results for optimized particles up to 140 formula units: stoichiometric tetrahedra, stoichiometric truncated octahedra, and perfect octahedra (which are partially reduced by virtue of their shapes). The bottommost curve pertains to "supercharged" NPs, i.e., perfect octahedra that are decorated with oxygen molecules. The right panel shows the corresponding results from independent reference calculations at the DFT (PBE+U) level, taken from Refs. 178 and 180.

The agreement between the energetics in the two panels in Fig. 9 is good overall, which is satisfying. However, we note that as far as structures are concerned, the MACE-MP-0 model (leftmost panel of Fig. 10) is unable to distinguish between $Ce^{3+}$ and $Ce^{4+}$ ions, both of which should in fact be present in a partially reduced, perfectly octahedral ceria nanoparticle. This deficiency of the MACE-MP-0 structure is evident from a comparison with the independent electronic PBE+U calculations in the rightmost panel of Fig. 9, labeled "DFT+U." Such calculations involve a Hubbard correction, which enforces a stronger and more adequate localization of electrons at $Ce^{3+}$ sites than what is achieved by standard PBE without U, which is the DFT method used in the Materials Project for ceria. The presence of both $Ce^{3+}$ and $Ce^{4+}$ ions in the PBE+U results is seen to lead to local relaxations of the nearest-neighbor oxygen ions around the Ce ions, resulting in symmetry breaking of the NP; see, for example, the lack of symmetry with respect to the NP edges in the rightmost panel.

Neither the proper local relaxation nor the symmetry breaking, both seen in the PBE+U results, is captured by the MACE-MP-0 model. On the other hand, the middle panel of Fig. 10 shows our PBE-optimized results for the same NP. The structural similarity between the MACE-MP-0 result and the PBE-optimized nanoparticle is evident.

**Similarity statement:**

There are altogether 18 $CeO_2$ and $CeO_{2-x}$ bulk structures present in the MP dataset. No examples of stoichiometric or reduced ceria surface structures or ceria nanoparticles are present in MP.

**Performance summary**

Broadly correct prediction of the energy of nanoparticles as a function of size, including overoxidized particles, with respect to reference DFT results.

### 5. Inorganic halide perovskite

Halide perovskites have been shown to exhibit subtle phase transitions and long-range structural correlations. The MACE-MP-0 model has been applied to these systems by predicting phase transitions in the inorganic perovskite $CsPbI_3$. This material shows two solid–solid phase transitions between room temperature and 600 K, both of which involve small rotations of the octahedral units and accompanying changes in pseudo-cubic lattice parameters.[181]

To analyze these transitions, we ran constant pressure simulations of a 14 000 atom supercell with a slowly varying temperature. Figure 11 shows the variation in pseudo-cubic lattice parameters

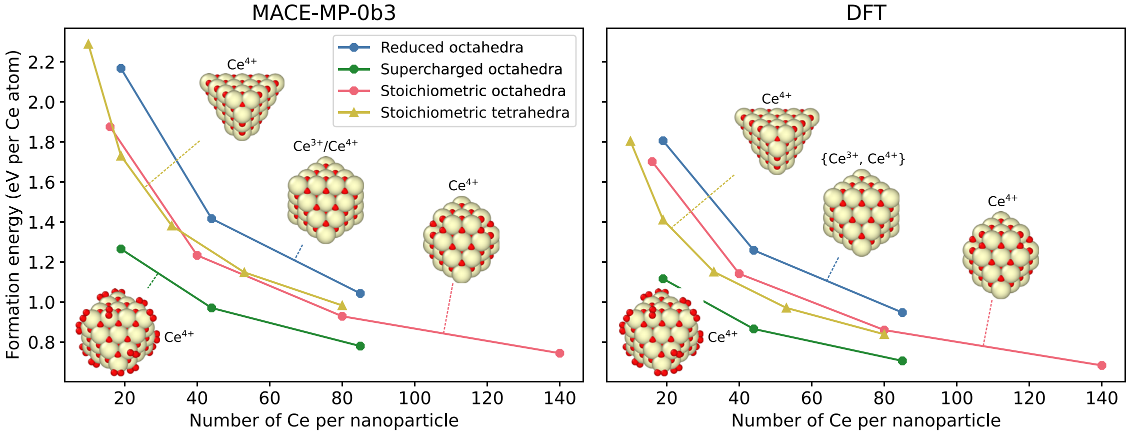


**FIG. 9.** Size-dependent formation energies for different shapes and stoichiometries of ceria nanoparticles (NPs). Left panel: The results calculated with the MACE-MP-0 model, which was not trained on any data from ceria surfaces or NPs. Right panel: Independent validation data from PBE+U calculations. The NP images are just illustrations; the optimized structures from MACE-MP and DFT are compared in more detail in Fig. 10. The formation energy is calculated with respect to stoichiometric bulk $CeO_2$ (and gas-phase $O_2$ molecules as needed).

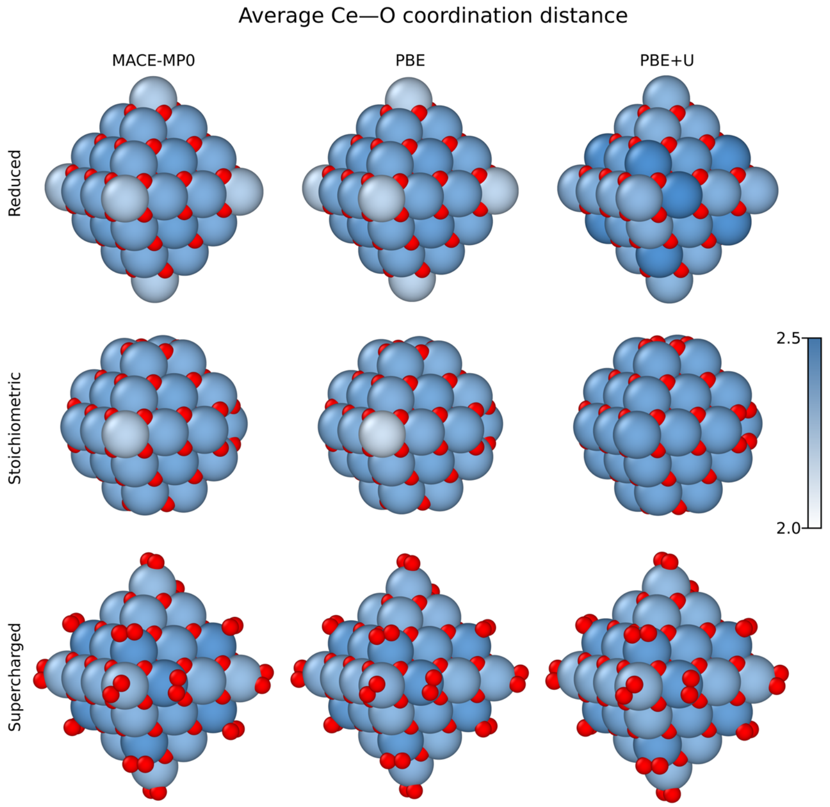


**FIG. 10.** Optimized structures of the perfect $Ce_{44}O_{80}$ octahedron with three different methods: MACE-MP-0, PBE, and PBE+U. The purpose of the figure is to highlight the "pattern of distances" rather than quantitative values. The large spheres represent the Ce ions (regardless of charge), and the small red spheres represent oxygens. The color scheme indicates the optimized interatomic distances in the following way: for each Ce atom, the distances to its O neighbors in the coordination figure of nearest-neighbors are measured, and the average value is reflected in the color of the sphere. Light blue indicates a short average Ce–O distance, and dark blue indicates a long average Ce–O distance. The distance scale is shown in the color bar to the right, covering the range from 2.0 to 2.5 Å.

with temperature, compared to experimental data. These data were obtained from a 3 ns simulation during which the temperature was raised from 350 to 550 K. The MACE-MP-0 model correctly predicts the qualitative nature of both phase transitions. There is a shift in both the transition temperatures and the average lattice constant, which has also been observed in other studies of these materials with DFT.[182] It is also known that the choice of exchange-correlation functional has a large effect on transition temperatures for these materials.[183,184]

**Similarity statement**

There are 57 structures in the MP dataset containing some combination of Cs, Pb, and I, without other elements. Of these, five structures contain all three of these elements in different compositions spanning several phases of this material. Based on UMAP analysis, these five structures are close to the training dataset. In particular, the cubic and orthorhombic phases, which are studied in this example, are present. Several similar structures with Br replacing I are also in the training set.

**Performance summary**

Both structural phase transitions and their transition temperatures are well captured, and the 10% discrepancy in the latter with respect to experimental values is likely due to the PBE functional.

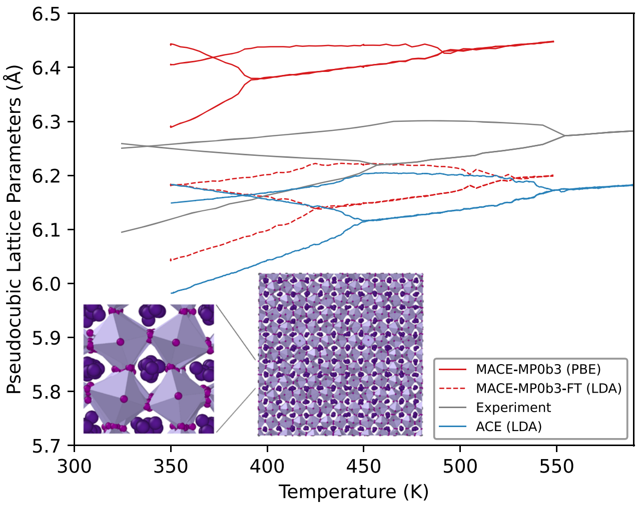


**FIG. 11.** Variation of pseudo-cubic lattice parameters with temperature for $CsPbI_3$. The MACE-MP-0 model is compared to experimental data reported by Even and co-workers[181] and an atomic cluster expansion (ACE) model trained for this material.[182] The inset shows an illustration of the system used to calculate the lattice constant dependence on temperature.

**Fine-tuning**

Fine-tuning was performed using 100 configurations sampled randomly from the database used in Ref. 182, wherein an ACE model was constructed specifically for studying phase transitions and nanostructural features of $CsPbI_3$ and the LDA exchange correlation functional was used. This database was obtained by running NPT and NVT MD simulations of the material at a range of temperatures. Figure 11 compares the predictions of the lattice constant variation from the MACE-MP-0 foundation model to those of the fine-tuned model and the ACE model trained in Ref. 182. One can see that the model prediction is shifted substantially toward that of the ACE model.

## 6. Hybrid organic–inorganic perovskites (HOIPs)

Hybrid organic–inorganic perovskites (HOIPs)[189] are a promising class of perovskites that have been studied extensively due to their strong chiral response, optical absorption, and high carrier mobility. However, the most-studied three-dimensional $MAPbI_3$ suffers considerably from low stability.[190] Two-dimensional HOIPs have the advantages of enhanced stability and structural tunability, which make them candidates for promising applications in photoluminescence (PL), solar cells, and light emitting diodes (LEDs).[191] However, 2D HOIPs are computationally expensive to examine using DFT because of their complicated unit cells formed when the organic cations separate the inorganic layers in the Ref. 100 direction, giving the modified general formula $A'_mA_{n-1}B_nX_{3n+1}$, corresponding to $n$ layers of the 3D-parent $ABX_3$ structure, separated by a layer of $A'$ organic cations that carry either a single charge ($m = 2$) or two charges ($m = 1$); the diversity of these systems can be seen in Fig. 12(a).

Using MACE-MP-0, we investigated a set of 159 experimentally synthesized 2D HOIPs from the Cambridge Structural Database (none are in the MP) with B = Pb and X = Cl, I, and Br. The organic cations ($A'$ and A) comprise only the elements C, H, N, and O, with either a +1 or +2 charge. MD simulations were performed within the NPT ensemble using MACE-MP-0 with the D3 correction at a temperature of 300 K and a pressure of 1 atm for 10 ps. For all 159 MD, the trajectories remained stable and reached equilibrium in about 3 ps [Fig. 12(a)]. No bond-breaking or surface cleavage between the organic/inorganic layers occurred. To test the accuracy of the model, we applied it to a set of 3007 configurations drawn from a recent study of HOIPs,[192] also sampled from multiple MD trajectories, with the errors in forces [root mean square errors (RMSEs) = 137 meV/Å] and energies (RMSE = 35.9 meV/atom) [Figs. 12(b) and 12(c)].

To fine-tune the model, a total of 159 configurations (one per composition) were randomly sampled from the MD trajectories. The errors in forces and energies are significantly reduced to 51 meV/Å and 6.7 meV/atom and are similar to the errors reported in Ref. 192. However, in that study, the training procedure required 2457 configurations from MD trajectories, about 15× more than what is used in the fine-tuning in this study.

**Similarity statement**

In the MP training set, there are in total 398 structures with PbX (X: Cl, Br, I) and 1627 structures with organics made of the elements C, H, N, and O. Based on UMAP analysis, we observe that some have similarities in environments to the MP training set, but

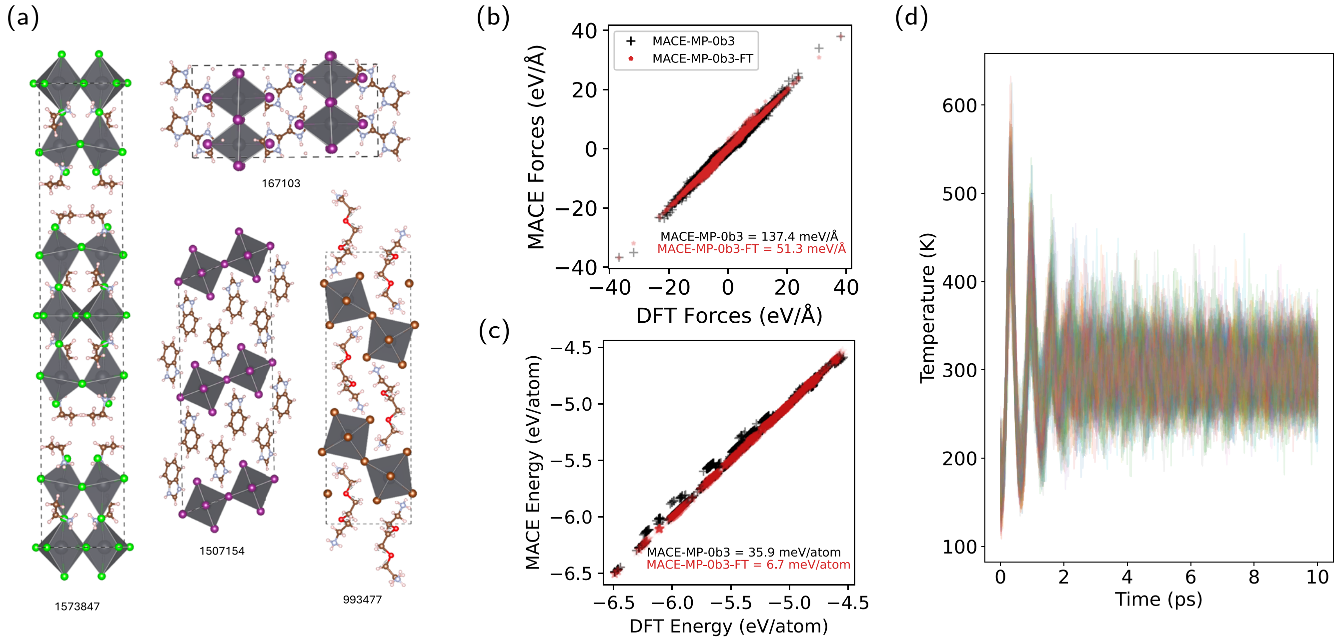


**FIG. 12.** (a) Four selected HOIP structures[185–188] with CCDC deposition numbers provided below each structure. (b) Force parity and (c) energy parity plots, with samples taken from MD trajectories using both the MACE-MP-0 and a version fine-tuned on 159 configurations, compared directly with the corresponding DFT values. (d) The MD trajectories for 159 compositions for 10 ps in the NPT ensemble at $T = 300$ K.

only 22 structures have a mixture of PbX and CHNO. Among these, there are 14 3D HOIPs, of which 12 have methylammonium (MA) as the organic cation, 3 0D HOIPs, and 4 cases of non-perovskite systems. In the MP, there is one 2D HOIP, which is the most similar to our dataset (mp-1194995), but this structure was not included in our dataset.

**Performance summary**

Stable NPT MD at ambient conditions for all 159 2D hybrid perovskite materials.

## 7. Protein dynamics and stability

Understanding the dynamics of proteins is crucial for deciphering their biological function and remains a core challenge of computational chemistry. Machine learning potentials have the capability of modeling non-covalent interactions, which are key components of secondary and tertiary structures of proteins, due to the quantum mechanical (QM) data on which they are trained. In this section, we perform simulations on the well-known engineered Chignolin peptide (PDB: 1UAO), an artificial prototype for the protein folding phenomenon. This study diverges significantly from the majority of studies discussed in this paper, as we are employing a machine-learning potential trained on inorganic crystals for a biological purpose.

We perform three separate simulations for each trained or fine-tuned model as follows:

- A simulation starting from a partially folded structure of Chignolin.
- A simulation starting from an unfolded structure of Chignolin.
- A simulation of a mutant of Chignolin (PDB: 5AWL) starting from its PDB structure, which contains 12 crystallographic waters.

Representative starting structures for the folded and unfolded Chignolin were obtained from Ref. 193, specifically the first (folded) and last (unfolded) geometries from the 9543 conformations sampled by replica exchange molecular dynamics. We protonate negatively charged residues (1UAO: ASP3, GLU5, GLY10 | 5AWL: ASP3, GLU5, TYR10) and remove the proton on GLY1 for 1UAO and on TYR1 for 5AWL. We are aware that the real stable state of Chignolin is non-neutral and solvated, but this test is performed in a neutral environment, as MACE-MP-0 was not trained with a charge-based loss, and testing the stability of the simulation is our primary goal. Simulations were performed at 300 K for 1 ns at a 1 fs time step and a sampling frequency of 1 ps in the NVT ensemble using the Atomic Simulation Environment (ASE) package. The Langevin thermostat was employed with a friction coefficient of 10 $\text{ps}^{-1}$. MACE-MP-0b3 with the D3 dispersion correction was used for all simulations. Most simulations were completed successfully in less than a day (elapsed time), while one simulation ("Mutant") with the pre-trained MACE-MP-0b3 model took about three days.

We first analyze the results obtained from the pre-trained-only version of MACE-MP-0. From Fig. 13, we can see that from its unfolded state (green line), Chignolin is quickly compacting at the start of the simulation and then maintains a stable radius of gyration, while the folding state (blue line) is maintained when starting from a partially folded state. Both simulations of Chignolin try to converge around the radius of gyration computed for the crystal structure 1UAO (red line), with the "Folded" simulation undershooting and the "Unfolded" simulation slightly over-shooting it. On the other hand, the Chignolin mutant (yellow line) shows a perturbed radius of gyration due to some of the crystallographic waters losing contact with the protein early on in the simulation. Note that the radius of gyration was computed considering all the system's atoms. This was performed to highlight any lack of stability in the simulation.

The simulations pursued with the fine-tuned (FT) model show improvements in terms of stability for the Chignolin mutant (red line) as the crystallographic waters interact with the protein at any given time during the simulation. The unfolded Chignolin

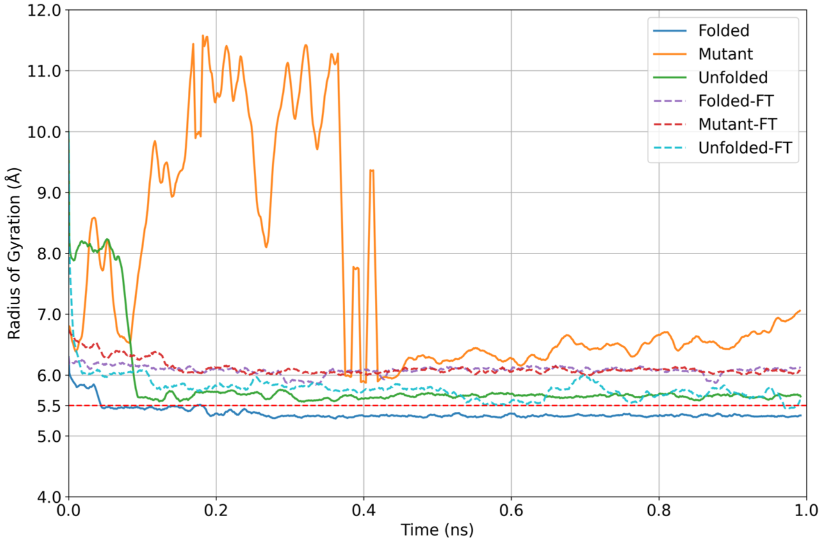


**FIG. 13.** Plot of the radius of gyration (Å) vs the time of simulation (ns) for the three simulations performed on Chignolin (PDB: 1UAO) and Chignolin mutant (PDB: 5AWL) with both MACE-MP-0b3 and the fine-tuned (FT) version. The radius of gyration curves is smoothed with a running mean over ten steps to better visualize their overlap. The "Unfolded" simulation trajectories start from an unfolded structure of Chignolin, the "Mutant" lines correspond to the simulations for the Chignolin mutant, and "Folded" corresponds to the simulations that start from a partially folded structure of Chignolin. The red line corresponds to the computed radius of gyration for the PDB 1UAO structure.

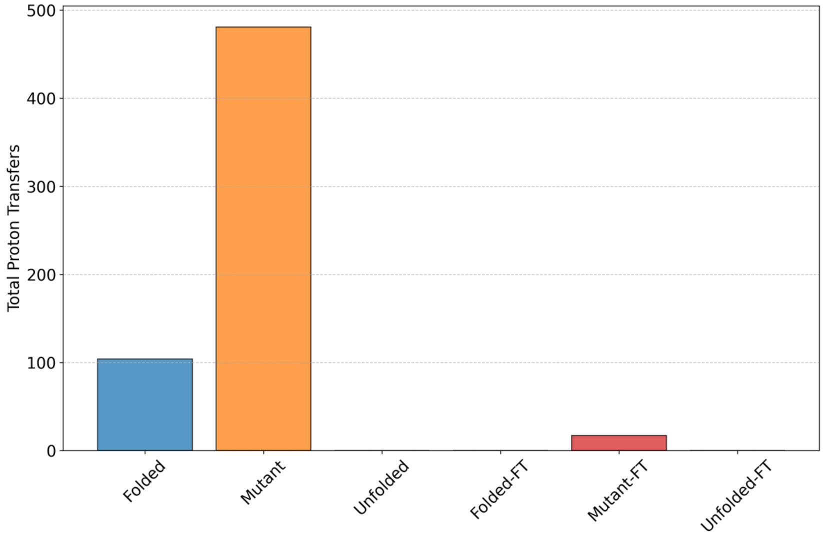


**FIG. 14.** Histogram of proton transfers occurring during the 1 ns simulations. Total count of proton transfers throughout the simulation for each system.

simulation (light-blue line) has an even faster formation of a compact protein structure compared to the pre-trained-only model, while the folded Chignolin simulation (purple line) is stable but shows a shift toward a higher-average radius of gyration. To investigate this further, we performed a proton transfer analysis, shown in Fig. 14.

On the left, we estimate the amount of proton transfers happening across time steps of each simulation. We use a simple distance-based detection to define the transfer of a proton. At every time step, we keep track of the closest heavy-atom neighbor of each proton. If in a time step the proton has a new heavy atom neighbor with respect to the previous time step and its distance to it is less than or equal to 1.2 Å (resembling the fact that a bond has likely formed), then we increase the total transfer count by one. We observe a large number of proton transfers for the folded Chignolin and Chignolin mutant in the pre-trained MACE-MP-0, while none in the unfolded. The number of proton transfers drops drastically in the fine-tuned version of the model. In theory, we expect the transfers to be low as the system is neutralized and there are little to no water solvation shells with which the system could interact. Therefore, it is clear that the fine-tuned model is closer to our expectations. This, however, might also mean that the fine-tuned model is biased toward weaker covalent and non-covalent interactions, which could also explain the higher radius of gyration observed in Fig. 13 for all of the fine-tuned simulations.

We want to point out that some proton transfers are still observed in the Chignolin mutant even after fine-tuning. The proton transfer encountered happens between the neutralized COOH and $NH_2$ functional groups of GLY1 and GLY10 of Chignolin. This is a typical case of acid-base chemistry where we would expect the proton to transfer from the more acidic carboxylic acid group to the basic amine group, forming a zwitterion in physiological conditions (i.e., neutral pH). Overall, this small case study shows the capability of MACE in modeling reactivity in molecular dynamics simulations. We point out that these observed processes of proton transfer cannot be observed with classical force fields, as they are generally not parameterized for such effects. Moreover, we conclude that fine-tuning the pre-trained foundational model is overall necessary to find a better agreement with the experiment and realistic physicochemical behavior. However, more real-world tests, which are outside the scope of this section, are needed to confirm that the model behaves correctly. Such tests would include a fully solvated system in physiological pH conditions and the use of a better-tailored fine-tuning dataset.

**Similarity statement**

The MP dataset encompasses only 99 structures exclusively composed of the elements hydrogen (H), carbon (C), oxygen (O), and nitrogen (N). Based on the UMAP analysis, the atomic environments of Chignolin and the Chignolin mutant are clustered similarly to those of the filtered 99 structures (see Fig. 15), which could mean that the environments of the training set are fairly similar to those of the test structures. However, after manual inspection, we observe that the filtered structures only resemble the protein under study by certain functional groups such as carboxylic acid (–COOH), amino groups ($–NH_2$), and aromatic and amide groups. Some of the most similar structures are, e.g., mp-998880, mp-1203308, mp-556151, mp-707289, and mp-1203544.

**Fine-tuning**

Fine-tuning of the MACE-MP-0b3 model was performed with 100 configurations obtained from the simulations conducted with the model itself. Specifically, there were 50 samples from the "Unfolded" Chignolin simulation, with 25 in the first third of the simulation, where the system is less compact, and 25 in the rest of the simulation; 25 from the "Folded" Chignolin simulation; and 25 from the "Mutant" Chignolin. For the latter, we only considered the water molecules that were interacting with the system and removed those that moved toward very large distances from the protein center of mass (>20 Å).

**Performance summary**

Most simulations performed showed no specific unphysical phenomena (e.g., no "explosions" were observed); however, many proton transfers that can be deemed unchemical were observed for

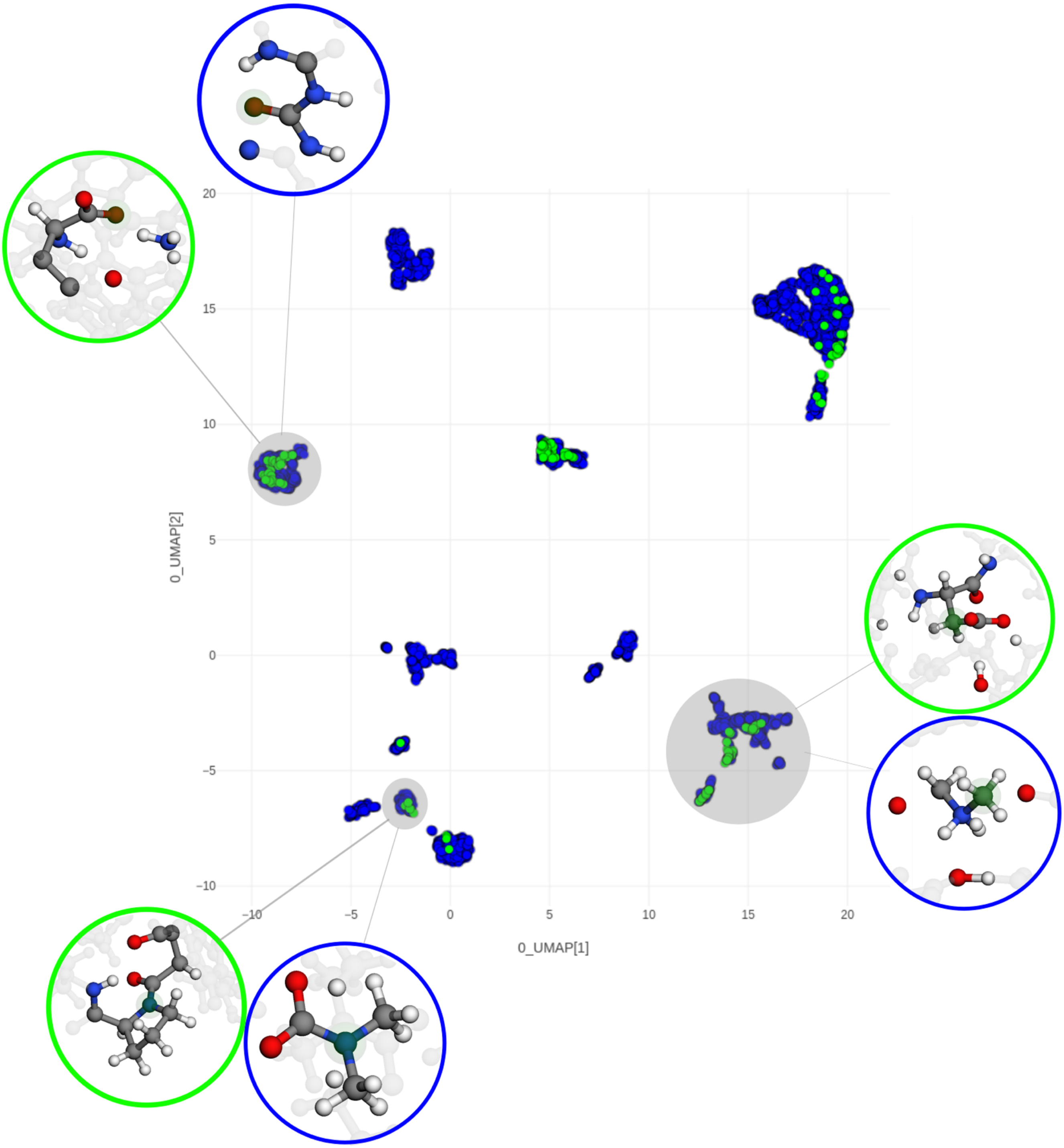


**FIG. 15.** Comparison of the atomic environments in the 99 filtered structures of the training data (blue dots) and in the Chignolin mutant structure (green dots) in the form of a UMAP plot. Insets of some local environments are shown for a few key functional groups of the protein (green circles) with corresponding similar clustered environments from the training set (blue circles).

the pre-trained model. For a few systems, the radius of gyration is maintained or converges to the value computed from the PDB structures. However, we note that residues that were charged in the PDB structures were protonated; therefore, the experimental folding (i.e., formation of secondary structure) cannot be fully recovered in this context.

### 8. Hydrogen combustion

Describing the complex reactivity in hydrogen combustion systems is a challenging task, often approached through thermodynamics via quantum mechanical (QM) calculations. However, accurately capturing kinetics in molecular dynamics simulations is hindered by the lack of a suitable transferable empirical force field, a difficulty compounded by the challenge of collecting experimental data in high-pressure, extreme explosion conditions. In this context, we compare our machine learning force field, MACE-MP-0, against the reactive classical empirical force field ReaxFF.[194] The ReaxFF model was created by fitting to an extensive QM dataset. The creation involved the identification of key reactions and components and the collection of data on formation heats, reaction heats, and energy barriers, as well as computing bond stretching energies and valence angle distortion energies for all combinations of hydrogen and oxygen. In our study, the performance of MACE-MP-0 in describing these reactions is evaluated and compared to both ReaxFF and experimental values. It is important to note that MACE-MP-0 has not been specifically trained for hydrogen combustion reactions, as the training data primarily comprise periodic representations of strongly bonded inorganic materials and a smaller fraction of molecular crystals (see Appendix A 8).

As seen in Fig. 16, the ReaxFF model agrees well with experimental values[195] for heats of reaction, with an RMSE of 4.82 kcal/mol. The MACE-MP-0b3 model (green in the figure) that is used in almost all other examples, while qualitatively showing the right trend, is significantly less accurate (with an RMSE of 14.7 kcal/mol). It is interesting to note that this example is one of the very few cases where the original version of the foundation model, MACE-MP-0a (orange in the figure), does much better, with an RMSE of 6.6 kcal/mol (and its "large" version, with more free parameters, reaches the accuracy of ReaxFF). The latest version of the foundation model is significantly more stable in general than the original versions, due to better architectural choices. Nevertheless, this example shows that there is still more understanding to be gained, and in the future, it should be possible to improve the out-of-the-box accuracy of the foundation model without sacrificing its stability. Furthermore, a detailed investigation of the accuracy of different foundation models would necessarily include a study of the accuracy of DFT and the exchange-correlation functional that is used to generate the training data, rather than assuming that it is close to the experimental values. For example, the PBE functional has a mean absolute deviation (MAD) of about 9 kcal/mol for a set of reactions relevant to hydrocarbon combustion.[196]

We used Packmol[197] to randomly arrange a 1:1 fuel mixture comprising 128 $H_2$ and 64 $O_2$ molecules within cubic cells (side length, a = 25–42 Å), yielding densities ranging from 0.05 to 0.25 kg $dm^{-3}$. Employing NVT simulations for 100 ps with the MACE-MP-0b3 potential, we tracked the evolution of $H_2/O_2$ mixtures. Reactivity analysis focused on water formation, identified via pairwise cutoffs derived from the first minima of the radial distribution function. The relationship between temperature/density variations in the fuel mixture and the water formation rate is depicted in Figs. 17(a) and 17(c). We find qualitative agreement with the water formation curves of Ref. 194, with a max conversion of ~80%.

**Similarity statement**

We analyze the MPtrj training dataset for the key species in hydrogen combustion ($O_2$, $H_2$, $H_2O$, $H_2O_2$, $HO_2$, and OH). These species are present as minority units in other structures, appearing in 2277, 1310, 1342, 232, 21, and 0 structures, respectively. We find 21 molecular crystals composed exclusively of $O_2$, 17 for $H_2$,

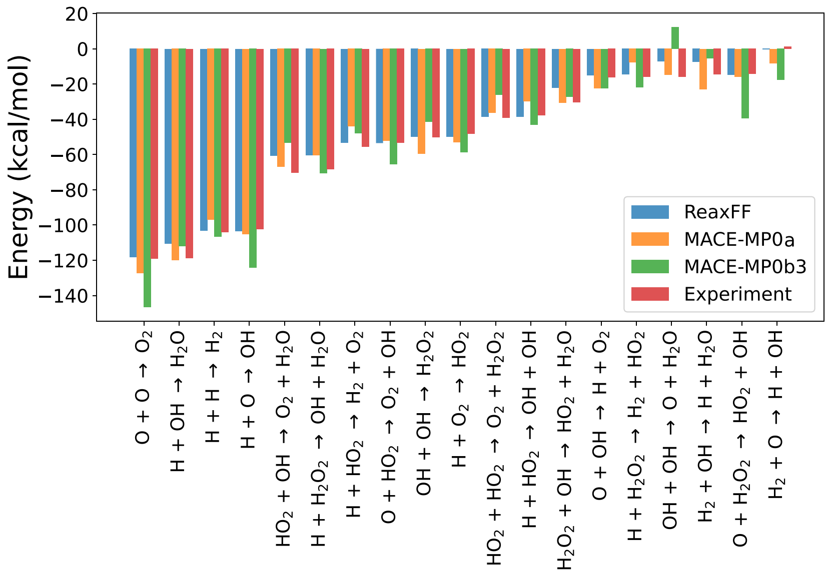


**FIG. 16.** Comparison of heats of reaction of key hydrogen combustion reactions with ReaxFF and the literature.

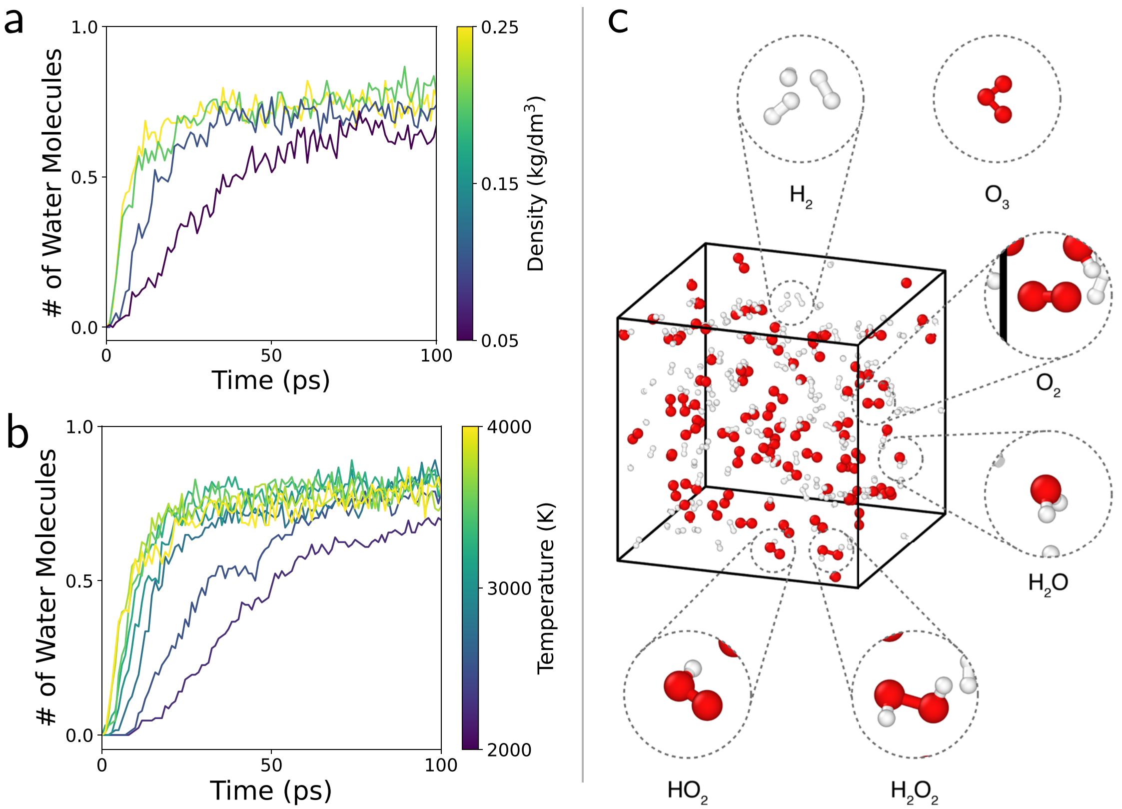


**FIG. 17.** Analysis and visualization of hydrogen combustion in MD simulations using the MACE-MP-0b3 potential. Water formation as a function of elapsed time for a range of (a) densities and (b) temperatures. (c) Representative snapshot during MD simulations, with key species highlighted. Note, although $O_3$ is not present in the particular simulation frame shown, it is occasionally observed in other simulations.

11 for $H_2O$, and 2 for $H_2O_2$. There are only eight structures made up of multiple key reaction species. UMAP analysis reveals only 2 MPtrj structures (mp-684678 and mp-1181087) with high similarity to frames within MD simulations.

**Performance summary**

Heats of reaction show the right trend for 19 reactions. Chemically correct species were produced during combustion, with the final yield also consistent with reference methods.

## 9. Sulfur polymerization

We ran a 20 ps NPT heat-up of a 128-atom structural model of elemental sulfur from 300 to 1400 K at 1 bar. At ambient pressure, experiments show that the molecular crystal formed of $S_8$ rings melts at 392 K and starts polymerizing at 432 K,[199] forming large chains that result in a 4-fold increase in viscosity of the liquid. We can qualitatively reproduce this melting and chain formation with MACE-MP-0b3+D3 correction (Fig. 18). The simulated melting temperature does not exactly match the experiment, as expected for a very fast run with only 128 atoms; however, the simulation does qualitatively reproduce the onset of depolymerization with increasing temperature, where large chains break down into smaller ones. For these reasons, we can say that MACE-MP-0b3+D3 is at least qualitatively applicable to simulate the polymerization of elemental sulfur starting from $\alpha$-rhombohedral crystalline $S_8$ (with further study required to test the quantitative agreement).

**Similarity statement**

The MP dataset contains 31 structures that only contain the element sulfur. Based on UMAP analysis, we see that a large part of the atomic environments in the example system are similar to environments in the training data. The database contains geometry optimizations of sulfur crystals formed of rings with various sizes: 6 (mp-7), 7 (mp-557559), 8 (mp-77), 9 (mp-556269), 10 (mp-557031), 11 (mp-561370), 12 (mp-558014), 13 (mp-583072), 14 (mp-561513), 10 × 6 (there exists a sulfur crystal form comprised of $S_{10}$ and $S_6$ rings, 557031), 18 (mp-555915), and 20 (mp-558964). It also contains crystals with planar strands of sulfur (mp-1179643), trigonal polymeric sulfur (mp-555760), and so-called fibrous sulfur (quenched polymeric liquid sulfur, mp-1196831), as well as isolated dimers (mp-1179639), trimers (mp-655141), and single atoms (mp-1063988). It does not contain melt or polymeric liquid structures. Based on the UMAP analysis, the closest (most relevant) structures in the training set are mp-556269, mp-555915, mp-83, mp-557031, mp-557559, and mp-666931. We provide `sulfur.json` to help visualize the interactive UMAP on https://chemiscope.org/.

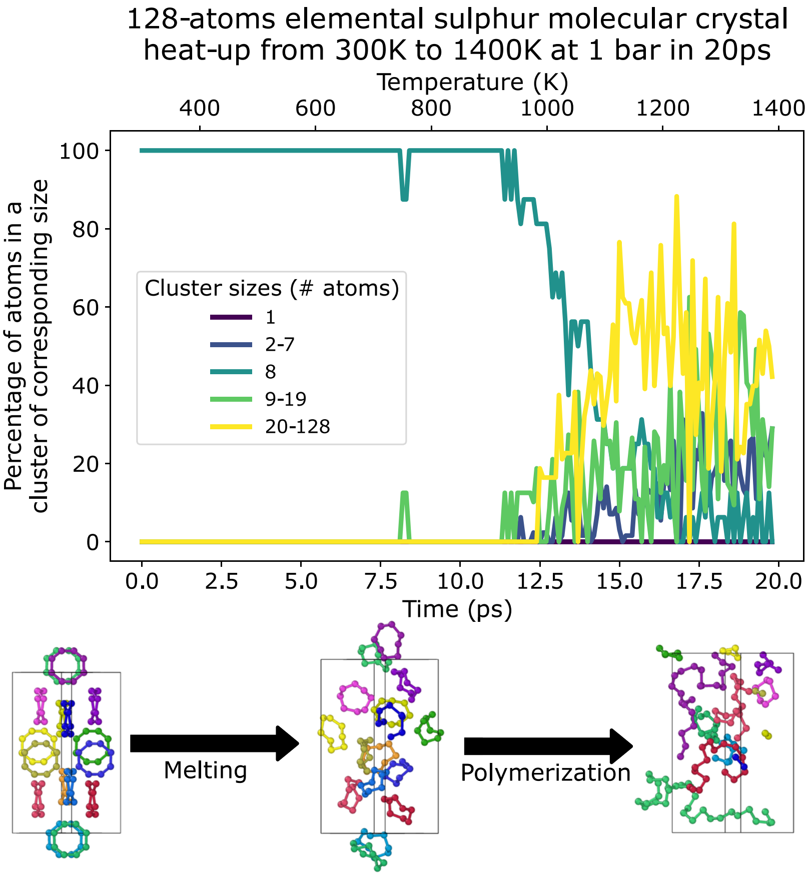


**FIG. 18.** Elemental sulfur. Evolution of the cluster size as a function of temperature in a 20 ps heat-up simulation with a linear increase in temperature applied over time. The pictures below are representative snapshots from the simulation, visualized in OVITO.[198]

**Performance summary**

Qualitatively correct polymerization, with large clusters forming and then subsequently breaking down to smaller sizes with an increase in temperature.

### 10. Zeolites

Zeolites are mesoporous materials with an important role as heterogeneous catalysts in several industrial processes. In this section, we assess the suitability of MACE-MP-0 to model these materials. We chose two zeolites, Modernite and Zeolite Socony Mobil–5 (ZSM-5), or MOR and MFI by their International Zeolite Association names. We have investigated the dynamic stability of the zeolite frameworks themselves, MOR with $NH_3$, acetone, and formamide inside one channel, and MFI with water, cyclohexane, and a mixture of $N_2$ in one channel and $CO_2$ in the other. Each zeolite was modified by adding a Brønstead acid site (BAS), Al, and the compensating H on the adjacent oxygen (see Fig. 19). Another set of simulations was carried out on MOR-Al, wherein, in addition to acetone, we introduced 20 water molecules and 32 water molecules, and similarly for MFI-Al with cyclohexene instead of acetone. MACE-MP-0 correctly identified the adsorption sites for ammonia, acetone, and formamide and the structural motifs in agreement with DFT calculations from Ref. 202, in NPT ensemble[203,204] simulations carried out at 300 and 400 K for 125 ps each using ASE. Furthermore, in the case of ammonia, MACE-MP-0 correctly predicts the formation of the $NH_4^+$ and its stabilization around the BAS by the creation of hydrogen bonds with adjacent oxygen atoms [see Fig. 19(c)]. MACE-MP-0 also correctly reproduces the DFT findings that acetone does not deprotonate the BAS but forms hydrogen bonds, while formamide predominantly forms hydrogen bonds but deprotonates the BAS occasionally. In addition, for the system MOR-Al with acetone and 20 water molecules, we have computed the barrier of the keto-acetone to enol-acetone conversion. MACE-MP-0 gave a barrier of 2.11 eV, and with D3 correction, 2.20 eV; the numbers are in good agreement with PBE calculations reported in Ref. 205. The code used to generate the trajectories is available in the repo.[206]

**Similarity statement**

For the system MOR-Al, the training set contains 145 structures that have Si, O, Al, and H elements on their own or along with other elements. Based on UMAP analysis, we see that almost all atomic environments in the example system are similar to environments in the training data. Similar findings hold true for MFI-Al with 145 structure matches. The structures with adsorbates inside have very low structural matches; for example, acetone in MOR-Al matches only three structures, and formamide only one, and none are very close to the studied zeolites. Adsorbents on their own match 1029 structures for formamide, 1892 for acetone, and 3139 for $NH_3$.

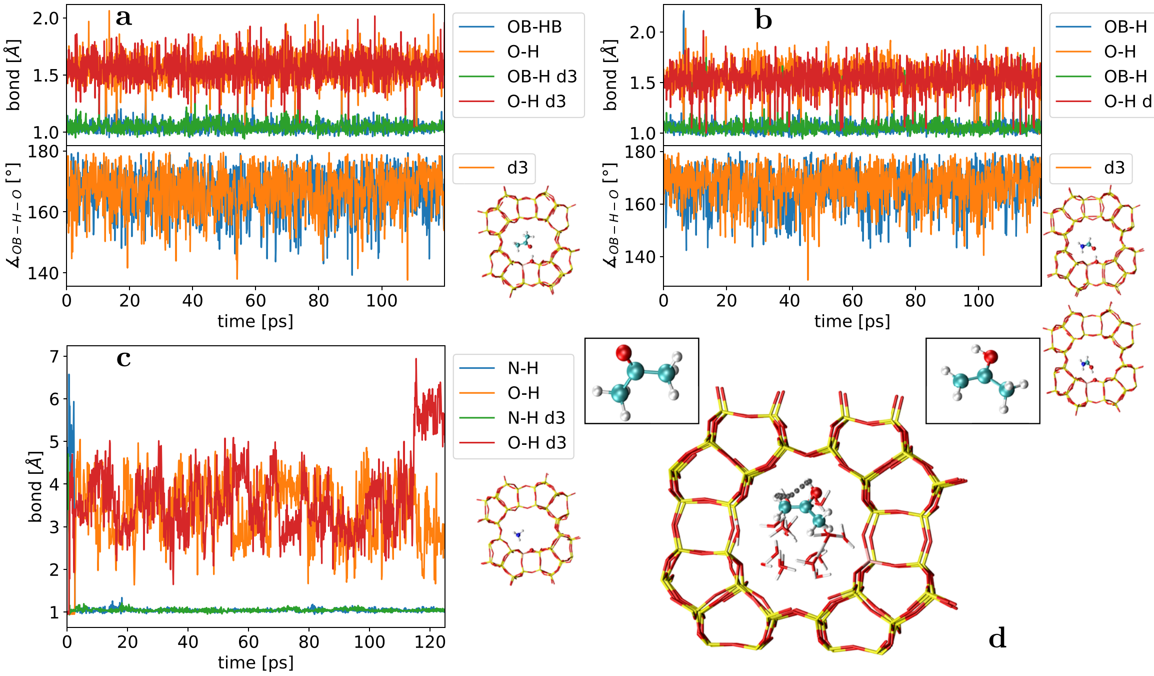


**FIG. 19.** (a) MOR-Al with acetone, instantaneous bond lengths between O@BAS, H@BAS, O@acetone, and H@BAS upper panel and angles lower panel—for both MACE-MP-0b3 and D3, a hydrogen bond is formed. (b) Same as (a), but for formamide, one can see deprotonation events when the O–H bond value approaches 1 Å. (c) $NH_3$ case with bonds same as (a) $NH_3$ quickly deprotonates the BAS to form $NH_4^+$ (d). The NEB[200,201] path for keto–enol tautomerism of acetone is shown; the path is indicated by the H atom traces, and only the final position of the rest of the atoms is shown.

If we consider only Si, O, Al, and H elements, we have only 11 similar structures for both zeolites considered, and none is an exact match, but they offer good representability of the local environments.

The closest (most relevant) structures in the training set are $CO_2$ (mp-556034, mp-20066, mp-995224, mp-11725, mp-644607, mp-1102227, mp-1190685, mp-995198, mp-1190699, mp-1077906, mp-1077316, and mp-729728). $CO_2$ alone matches 4896 structures with C, O and alongside other elements.

We provide

- `MOR-Al_FilterType.exclusive_SiOAlH_chemiscope_input.json`
- `MFI-Al_FilterType.exclusive_SiOAlH_chemiscope_input.json`
- `MFI-Al-H2O_FilterType.exclusive_SiAlOH_chemiscope_input.json`
- `MFI-Al-H2O-cyclohexene_FilterType.exclusive_SiAlOCH_chemiscope_input.json`
- `MFI-Al-H2O_FilterType.exclusive_SiAlOH_chemiscope_input.json`
- `MFI-Al-cyclohexene_FilterType.exclusive_SiAlOCH_chemiscope_input.json`
- `MOR-Al_FilterType.inclusive_SiOAlH_chemiscope_input.json`
- `MFI-Al_FilterType.inclusive_SiOAlH_chemiscope_input.json`

to help visualize the interactive UMAP on https://chemiscope.org/.

**Performance summary**

Correct prediction of binding sites and qualitatively correct reaction behavior for a range of structures and ligands, including good agreement of predicted reaction barrier with DFT.

## 11. Open-circuit voltage of lithiated graphite

The open-circuit voltage profile of an electrode material is an example of a technologically relevant macroscopic observable that can be accessed through atomistic simulation. We apply MACE-MP-0b3 to model the thermodynamics of lithium-ion intercalation in graphite using hybrid grand canonical Monte Carlo/molecular dynamics (GCMC/MD).

Beginning from a 720-atom cell of pristine graphite containing 10 graphene layers, we generate 40 parallel simulation trajectories of 30 000 steps each at a system temperature of 300 K. In our GCMC/MD protocol, at every simulation step, we update the ionic positions according to Verlet dynamics. Every five steps, we generate a Monte Carlo proposal on the system volume, followed by a proposal on the system's composition. For the volume proposal, we sample a perturbation of the unit cell: this is a set of three Euclidean vectors sampled component-wise from a normal distribution with a mean of zero and variance of 0.01 Å. We add these random vectors to the existing lattice vectors and rescale the atomic positions to generate the proposed unit cell update. For the composition proposal, we randomly make one of three modifications to the population of lithium ions: insertion, deletion, or swapping. We choose one of these types of modifications at random and then generate five candidate structures, each with either a single new lithium atom placed in a void in the host lattice (insertion), an existing lithium atom displaced into a void (swapping), or an existing lithium atom removed (deletion). As the lattice in our simulations is *not* fixed but evolves under molecular dynamics, we use Voronoi triangulation to identify void sites in the atomic lattice, excluding all sites subject to steric overlap according to the atomic radii. Once the set of composition candidates is generated, their energy is evaluated, and the lowest-energy candidate is used as the composition proposal. If a composition proposal is accepted, before proceeding to the next simulation step, we relax the ionic positions and the unit cell for up to ten ionic steps with a force tolerance of 0.05 eV/Å using the FIRE algorithm.

After sampling configurations with this protocol, we compute the open-circuit voltage as a function of lithium concentration over the sampled ensemble. Following a previous study,[208] the open-circuit voltage is estimated as the negative of the free energy of formation per atom of the phase with composition $\mathrm{Li}_x\mathrm{C}_6$ from reference states of graphite ($\mathrm{C}_6$) and metallic BCC lithium, divided by the lithium concentration: $V(x) = -\Delta G_{f,\mathrm{Li}_x\mathrm{C}_6}/x$. The free energy per atom of the metallic lithium reference state is taken as the potential energy predicted by MACE-MP-0 of BCC lithium after structural optimization, neglecting entropy. To determine the free energies per atom of the $\mathrm{C}_6$ and $\mathrm{Li}_x\mathrm{C}_6$ phases, we compute the internal energy $U$ and Gibbs entropy $S$ as Boltzmann averages over the sample distribution at concentration $x$: $G(x) = U(x) - TS(x)$, $U(x) = \sum_j (E_j \cdot p_j)$, $S(x) = -k_B \cdot \sum_j (p_j \cdot \ln(p_j))$, with probabilities $p_j = e^{-\frac{E_j}{k_B T}} / \sum_j e^{-\frac{E_j}{k_B T}}$, where $E_j$ is the potential energy per atom predicted by MACE-MP-0 and $j$ indexes the set of all simulation frames with composition $\mathrm{Li}_x\mathrm{C}_6$.

MACE-MP-0 reproduces the experimentally known voltage profile of Li/graphite with good quantitative accuracy [Fig. 20(a)]. In the regime of $x > 0.04$, the error is $<0.1$ V, which reflects the combined error of the model as well as the limitations of the GCMC/MD protocol. This may be compared favorably with a recently purpose-developed model for lithium–graphite energetics,[209] which reported $<0.1$ V error for $x > 0.0833$ in an open-circuit voltage profile produced through GCMC; that model was trained on more than 8000 system-specific DFT calculations, while MACE-MP-0 obtains comparable accuracy zero-shot. We note that at very low concentrations, our predicted voltage is higher than the experimental voltage by as much as a factor of 3, indicating over-stabilization of dilute lithium. Since the lithium fraction appears in the denominator of the expression for the open-circuit voltage,

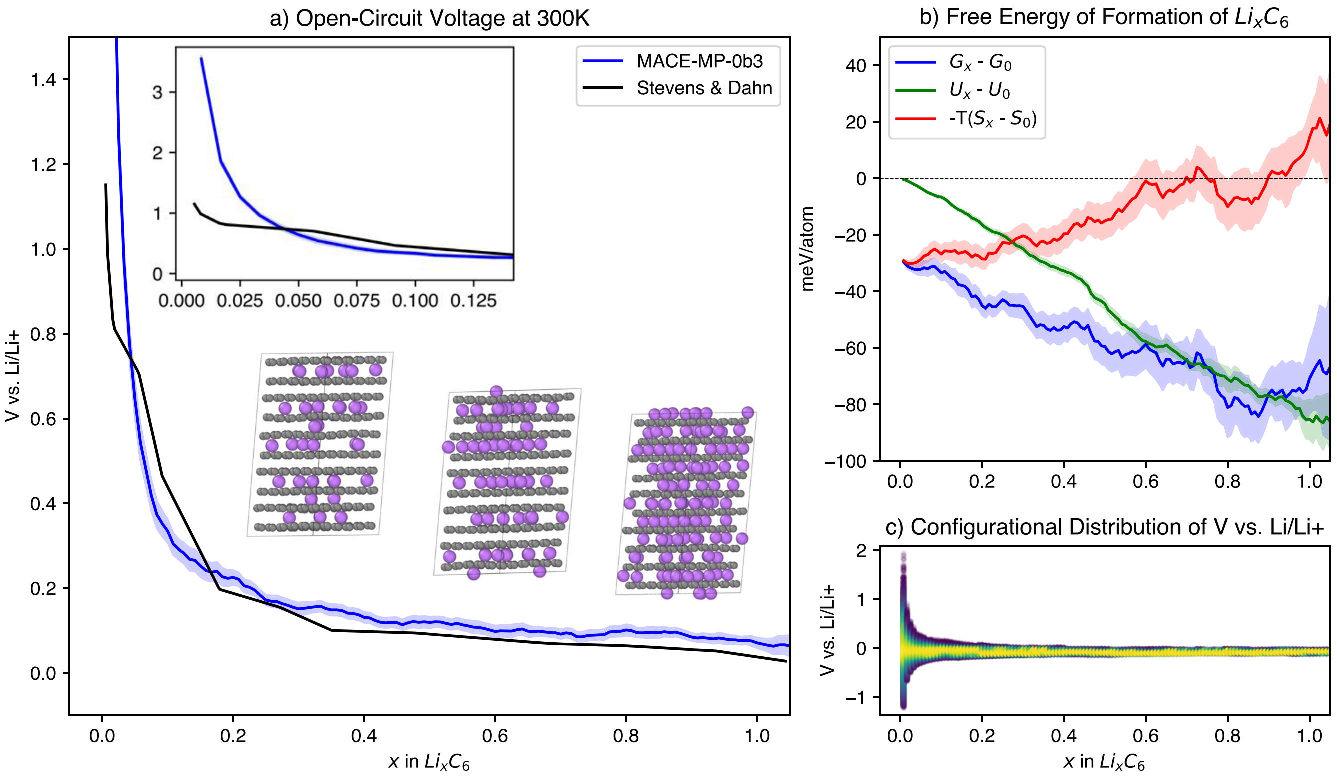


**FIG. 20.** (a) The open-circuit voltage profile of lithium in graphite vs lithium metal, computed with MACE-MP-0b3 (blue) using a hybrid Grand Canonical Monte Carlo (GCMC)/molecular dynamics protocol, contrasted with an experimental reference (black).[207] The variance of the simulated voltage is estimated over 100 bootstrapped samples of GCMC/MD trajectories. Representative lithium–graphite configurations are shown at $x = 0.3, 0.5, 1.0$. (b) The contributions to the free energy of formation of sampled lithium–graphite phases. (c) The density plot of individual Li/graphite structures sampled during a GCMC/MD simulation, showing the distribution of potential values over the configurational ensemble.

very slight energetic deviations are magnified in this range; moreover, the free energy at low concentrations is dominated by the entropic contribution [Fig. 20(b)], for which the limited sample size introduces uncertainty. Beyond this lowest-concentration regime, MACE-MP-0 provides good agreement with experiment.

**Similarity statement**

There is a skew toward battery materials in MP. Given this, there are several Li–C structures that are relevant to this application: mp-1210743 ($Li_2C$), mp-976060 ($Li_3C$), mp-1223102 ($Li_7C_{120}$), mp-1378 (LiC), mp-1021323 ($LiC_{12}$), mp-1232339 ($LiC_{12}$), and mp-1001581 ($LiC_6$). There also exist 62 pure carbon structures, including graphite (mp-48).

**Performance summary**

Correct prediction of voltage as a function of Li concentration with reference to the experimental curve.

### 12. Jahn–Teller distortions in $LiNiO_2$

$LiNiO_2$ (LNO) is an important material in lithium-ion battery research, serving as a model for future high-performance cathode materials with reduced or no cobalt content. Structural and chemical degradation of these materials is a key challenge, so understanding their thermodynamic and dynamic properties is critical. In particular, the presence of ordered regions of Jahn–Teller distorted $NiO_6$ octahedra may influence electron and Li-ion transport. Recent *ab initio* molecular dynamics (AIMD) and variable-temperature X-ray diffraction (VT-XRD) measurements have suggested a phase transition from a low-temperature phase with Jahn–Teller distortions to a dynamic phase without permanent Jahn–Teller distortions at high temperatures.[210] Above this displacive phase transition, the material does not exhibit Jahn–Teller distortions either in the time-averaged bulk structure or in instantaneous snapshots of the local structure. A similar displacive transition was also observed in the related sodium-ion battery cathode material $NaNiO_2$.[211]

We have investigated the ability of MACE-MP-0 to reproduce the temperature dependence of the structural distortions in LNO. We used a 256-atom supercell, initialized from the established DFT ground state structure, which has distorted $NiO_6$ octahedra with zigzag long-range ordering of their long axes [see Fig. 21(a)]. We performed NPT-ensemble dynamics using MACE-MP-0 without a D3 correction, in line with the earlier AIMD simulations at the PBE+$U$ level of theory.

Trajectories were propagated for ~250 ps at 13 temperature values from 25 to 1000 K. These trajectories were used to analyze the phase transition behavior [Fig. 21(c)], and a sample of 104 configurations was used to fine-tune a new model following the protocol outlined in Appendix C 2. Specifically, we extracted configurations from each trajectory at 20 ps intervals, after discarding the first 100 ps for equilibration. The performance of the fine-tuned model is compared with that of the original MACE-MP-0 in the next section.

**Performance**

The MACE-MP-0 model demonstrates stable dynamics during a heating trajectory from 0 to 1000 K with temperature increments

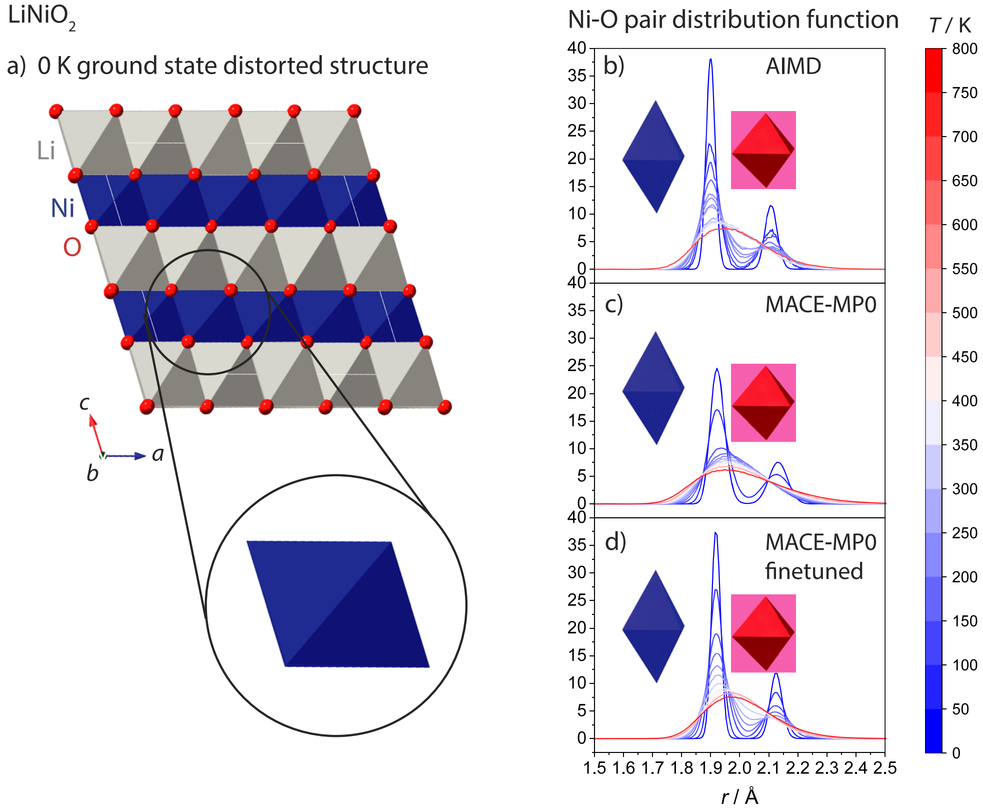


**FIG. 21.** (a) Ground state structure of LNO. The magnified insert shows a $NiO_6$ octahedron with Jahn–Teller distortion producing long and short axes. (b) Ni–O pair distribution functions (PDFs) as a function of ion separation computed by AIMD.[210] Line color corresponds to trajectory temperature as shown in the color bar. (c) PDFs computed by MACE-MP-0b3. Inserts indicate shapes of $NiO_6$ octahedra (blue: distorted, red: undistorted). (d) PDFs computed by the fine-tuned MACE model as described in the text. The same temperature values are used in panels (c) and (d); slightly fewer values are used in AIMD but span a similar range of temperature.

of 10 K $ps^{-1}$ and a time step of 1 fs. Long simulations (at least 500 ps) could be run at 1000 K without significant energy drift or noticeably unphysical behavior.

The Ni–O pair distribution functions predicted by both AIMD and MACE-MP-0 [Figs. 21(b) and 21(c)] show two peaks at low temperature corresponding to the short and long Ni–O bonds in Jahn–Teller distorted octahedra. At higher temperatures, one broad asymmetric peak appears, which is characteristic of octahedra in a displacive high-temperature phase.[210] MACE-MP-0 thus correctly predicts a phase transition from Jahn–Teller-distorted to undistorted octahedra on heating. However, it significantly underestimates the temperature at which the transition occurs. We diagnose the onset of the transition at the lowest temperature where the PDF has only one maximum and the second peak (at longer bond lengths) becomes a shoulder on the short-bond peak. Above this temperature the trajectories show dynamic Jahn–Teller distortions, i.e., reorientations of the long O–Ni–O axes over time. This onset occurs between 50 and 100 K in the MACE-MP-0 simulations, compared to ~250 K with AIMD.[210] The transition is complete when the Ni–O PDF exhibits only one broad asymmetric peak without a shoulder. MACE-MP-0 predicts that this completion temperature is below 150 K, compared with 350 K in AIMD.

By contrast, the fine-tuned model predicts an onset temperature between 250 and 300 K and the transition to complete by 350 K, both in excellent agreement with the AIMD results. This improvement suggests that the deficiency of the original model results from undersampling of important configurations in the training data.

**Similarity statement**

Battery materials are well represented in the MP database. 1393 structures in the training set contain Li, Ni, or O atoms; 143 contain all three elements, and 23 have the exact formula $LiNiO_2$ representing different crystal structures. However, many of these structures are obtained from geometry optimization and hence neglect the temperature dependence of the equilibrium geometry. The substantial improvement of MACE-MP-0 performance for this system after fine-tuning on variable-temperature trajectories suggests that incorporating these temperature-dependent configurations is necessary to describe structural properties of the material.

**Performance summary**

MACE-MP-0 correctly captures the loss of Jahn–Teller distortions on heating, with an underestimation of the transition temperature compared with AIMD that is largely corrected with a small amount of fine-tuning.

**TABLE I.** Lattice constants (in units of Å) for W, Mo, and Nb computed by variable-cell minimization. DFT reference data are from Ref. 214 for W, Ref. 215 for Mo, and Ref. 216 for Nb.

| | DFT | MACE-MP-0b3 | MACE-MP-0b3-FT |
|---|---|---|---|
| W | 3.185 | 3.203 | 3.191 |
| Mo | 3.163 | 3.182 | ⋯ |
| Nb | 3.322 | 3.321 | ⋯ |

### 13. Point and extended defects in BCC metals

This test explores bulk and extended defect properties of three prototypical BCC metals: W, Mo, and Nb. An accurate description of these properties is essential to enable predictive modeling of mechanical responses to applied loads such as dislocation glide,[212] dislocation climb through interaction with point defects, grain boundary motion, and the competition between cleavage and dislocation emission that underpins the brittle-to-ductile transition in fracture. Across-the-board accuracy for bulk and defect properties in these systems is challenging, even for bespoke machine learning potentials fit to carefully curated datasets.[213]

Lattice and elastic constants are shown in Table I and Fig. 22, respectively. Lattice constants are generally well reproduced, but there is a general softening of the elastic response. In Fig. 23, we compare the energy–volume (E–V) curves predicted by the MACE-MP-0b3 model for BCC and FCC phases of the three metals with reference DFT data from Ref. 217. The BCC cases show generally good agreement. There is room for improvement in the curvature of the E–V curves (critical for the elastic properties) for W and Mo, while Nb is well described. FCC energies are underestimated, while the curvature is approximately correct, giving reasonable predictions of the elastic response.

We next investigated point defect formation energies, including vacancies and self-interstitial atoms (SIAs). Calculations were performed in a 5 × 5 × 5 supercell and were relaxed to a force tolerance of $1.0 \times 10^{-5}$ eV/Å. For the SIAs, a short MD run was performed to escape the initial metastable configuration. The results, illustrated in Fig. 24, show good agreement with reference DFT data from Ref. 218: vacancy energies are predicted within ~20% of the DFT value and SIA energies within 50%. For all three elements, the MACE-MP-0b3 predicts that the ⟨111⟩ dumbbell is the most stable SIA configuration, in agreement with DFT.

We next looked at generalized stacking fault energy profiles for the (112) and (110) Γ-surfaces along the [111] direction as shown in

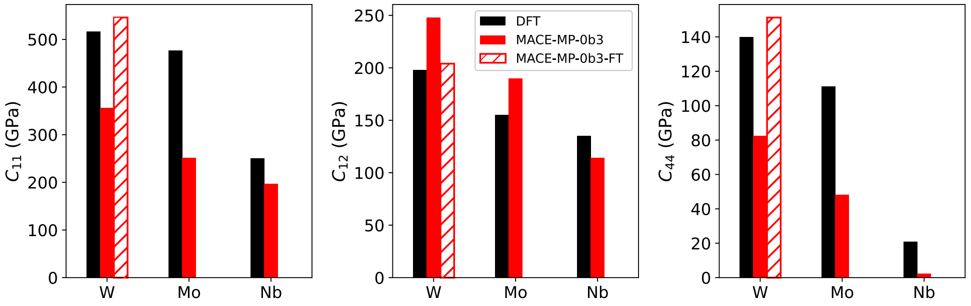


**FIG. 22.** Cubic elastic constants for W, Mo, and Nb computed with linear stress–strain fits for DFT (black; from Ref. 214 for W, Ref. 215 for Mo, and Ref. 216 for Nb), the MACE-MP-0b3 (solid red). The MACE-MP-0b3-FT model fine-tuned on W data is shown with hatched red bars.

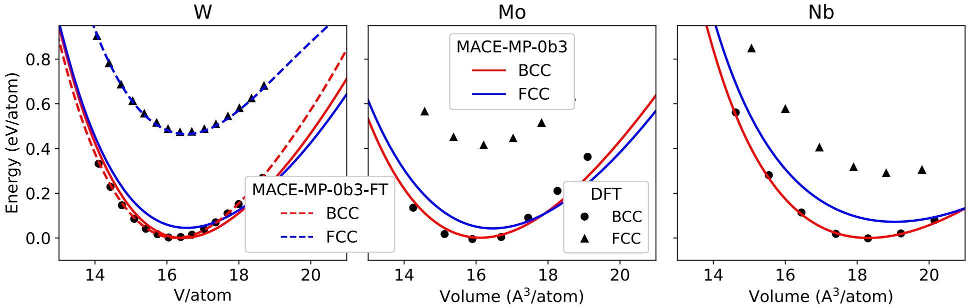


**FIG. 23.** Energy vs volume curves for W, Mo, and Nb computed with the medium MACE-MP-0b3 model (solid lines) for FCC (red) and BCC (blue) crystals. The MACE-MP-0b3-FT model fine-tuned on W data is shown with dashed lines. The DFT reference data shown with the black circular (BCC) and triangular (FCC) points are from Ref. 217.

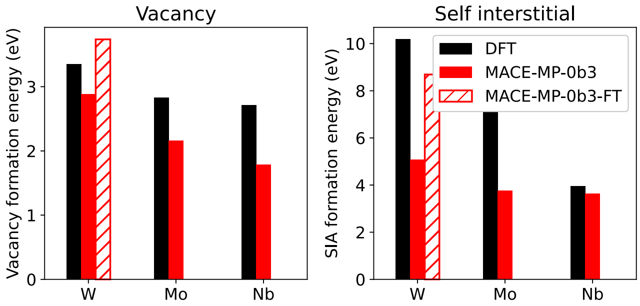


**FIG. 24.** Vacancy and SIA formation energies for W, Mo, and Nb computed with the MACE-MP-0b3 model (solid red), the W fine-tuned MACE-MP-0b3-FT model (hatched red), and DFT reference data from Ref. 218 (black).

Fig. 25. These results were obtained with constrained minimization, where atoms were allowed to move only in the direction perpendicular to the cut surface and with a force tolerance of 1.0 × $10^{-3}$ eV/Å. The details of the method are explained in Ref. 220. Nb is very well described by the MACE-MP-0b3 model, while there is an underestimate in the stacking fault energies in W and Mo by around a factor of two, which explains the underestimates in the screw dislocation glide barriers discussed below.

Dislocations in BCC materials lie predominantly in the $\langle 111\rangle\{110\}$ and $\langle 100\rangle\{010\}$ slip systems. We investigate the characteristics of $\langle 111\rangle$ screw and $\langle 100\rangle$ edge dislocations by calculating the transition pathways and Peierls barriers using the nudged elastic band (NEB) method and the MACE-MP-0b3 potential, comparing against DFT[221] and hybrid QM/MM calculations[222,223] where the data are available.

The cells contained ≈1400 atoms for the [111] screw dislocation and ≈2200 for the [100] edge dislocation. Geometry optimizations to obtain starting configurations used the FIRE algorithm with a force tolerance of 1 × $10^{-6}$ eV/Å. To create and analyze atomistic dislocation configurations, we employed the `matscipy.dislocation` module.[170] The transition path calculation is performed with an adaptive ordinary differential equation (ODE) solver[201] following the approach of Refs. 224 and 222, using 15 intermediate images with a stopping force tolerance of 0.025 eV/Å. Starting positions for the NEB relaxation were obtained by linear interpolation between initial and final configurations.

Figures 26 and 27 illustrate the NEB minimum energy path depicting the Peierls barriers for $\langle 111\rangle$ screw and $\langle 100\rangle$ edge dislocations in W, Mo, and Nb. We compare to DFT results where they are available. Insets within Fig. 26 illustrate the dislocation core structures at the initial, intermediate, and final positions along the MACE-MP-0b3 minimum energy path. Screw dislocations are known to be a sensitive probe of potentials, since the accuracy required is on the meV/atom level. DFT predicts dislocations to move between highly symmetric easy core configurations via split core configurations at the saddle point.[221] For all three metals, the degenerate core is incorrectly predicted to be the most stable configuration by the MACE-MP-0b3 model. At the same time, the

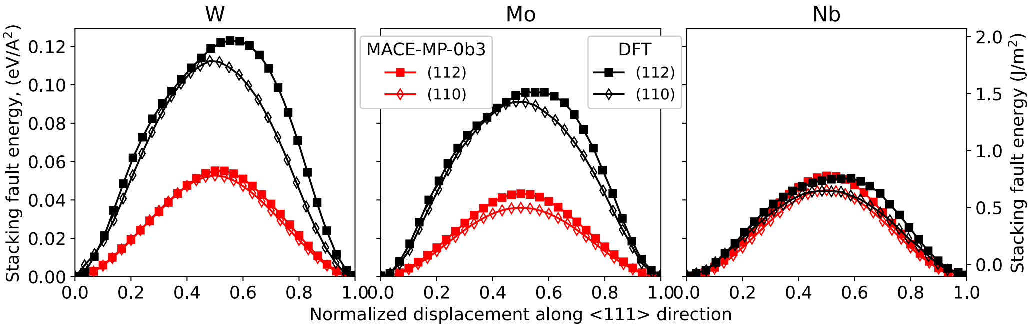


**FIG. 25.** Generalized stacking fault profiles for (112)[111] and (110)[111] Γ-surfaces predicted by the MACE-MP-0b3 model, shown in red. The DFT reference data, shown in black, are from Ref. 219.

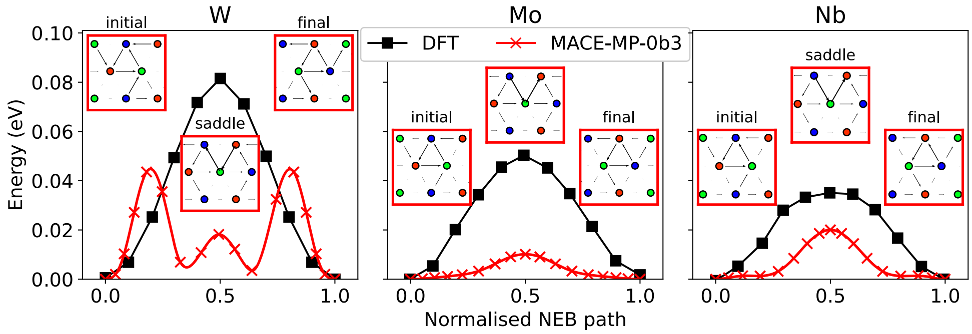


**FIG. 26.** Screw dislocation glide barriers for W, Mo, and Nb. DFT data from Ref. 221.

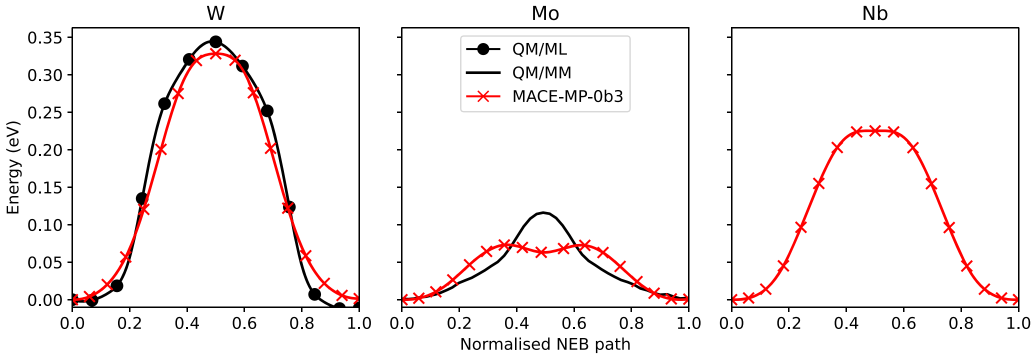


**FIG. 27.** Edge ⟨100⟩ dislocation glide barriers for W, Mo, and Nb. QM/Ml data for W from Ref. 222 and QM/MM data for Mo from Ref. 223.

split core configuration at the saddle point is correctly predicted for Mo and Nb, while for W, the single hump shape of the glide barrier is not reproduced, with some intermediate configurations having energy close to the degenerate core. The barrier height is underestimated for all three metals compared to reference DFT results.[221] For the edge dislocation, where energy differences are larger, we find that the barrier height aligns well with QM/Ml results in W[222] and Mo.[223] However, the presence of a minimum along the transition path for Mo results in a spurious stable dislocation configuration.

We anticipate that the performance of MACE-MP-0b3 for all properties considered here would be substantially improved by enhanced accuracy in stress and a more precise agreement on the elastic constants, followed by fine-tuning on defect configurations where necessary.

**Fine-tuning**

Fine-tuning was performed on 142 W configurations containing an isolated atom, deformed perfect BCC and FCC structures, a vacancy in BBC, and molecular dynamics snapshots for BCC. This is sufficient to correct softening of BCC and the overstability of FCC as seen in the updated elastic constants and equation of state plots in Figs. 22 and 23, and also improves the agreement of the lattice parameter and the vacancy and SIA formation energies, as shown in Table I and Fig. 24. The improvement in the SIA formation energy demonstrates transferability, as there are no interstitial configurations included in the fine-tuning dataset.

**Similarity statement**

The MP dataset includes seven elemental tungsten, seven elemental molybdenum, and four elemental niobium structures. They are all crystalline without any defects. Based on UMAP analysis, we find that the closest structures in the training set are mp-8641 for tungsten, mp-8637 for molybdenum, and mp-8636 for niobium. We provide `W_input.json`, `Mo_input.json`, and `Nb_input.json` to help visualize the interactive UMAP on https://chemiscope.org/.

**Performance summary**

Energy–volume curves for BCC are well reproduced, while for FCC structures the MACE-MP-0b3 model shows a ~0.5 eV/atom shift in energy; this is corrected by fine-tuning. Stacking fault profile energies for Nb are well reproduced, while for W and Mo, they are underestimated by a factor of around two with respect to DFT. Relaxed point defect structures are reasonable in all cases, with formation energies within 50% of reference DFT values (again, improved by fine-tuning). Peierls energy barrier profiles for dislocation glide are qualitatively correct for the edge dislocations and underestimated for the screw dislocation, together with incorrect screw dislocation core stability. There is a small spurious local minimum near the top of the barrier for the edge dislocation in Mo.

### 14. Alumina defects and bulk diffusion

#### a. Bulk diffusion

One Al and one O vacancy were introduced into a 270-atom alumina supercell, and over 2 ns (at 2.5 fs) MD was used at temperatures between 2600 and 3200 K to measure diffusivities of the two elements. Diffusivities agree within one order of magnitude for Al, and activation energies are underestimated for both compared to extrapolated experimental values.[225] This demonstrates the long-timescale stability of the model, even at high temperatures and for long MD trajectories, but highlights a shortcoming of the model for quantitative prediction of macroscopic observables.

### *b. Elemental defects*

Elemental defects in $Al_2O_3$ were investigated by substituting Si, S, Ti, V, Cr, Fe, Co, Ni, Cu, Y, Ag, and Pt into lattice sites in a $2 \times 2 \times 1$ supercell (120 atom), and minimal energy paths to neighboring sites were obtained using NEB[201,226] starting from a linear interpolation. Paths were converged (max 100 steps, 0.3 eV/Å tolerance on projected forces), and the lowest energy one was tested with PBE single point evaluations using CASTEP.[227] Comparing MACE-MP-0b3 and PBE on Figs. 28(c)–28(e), there are large discrepancies, with a total force component RMSE of 0.35 eV/Å across 152 structures evaluated.

**Fine-tuning**

Fine-tuning was performed on 72 configurations, which included substitutions of each element and snapshots from paths found with MACE-MP-0b3. The barrier energies computed with the fine-tuned model are shown in Fig. 28(c), showing improved agreement with the PBE reference in barrier energies and a lower 0.21 eV/Å force component RMSE on 149 PBE evaluations. Notably, the MACE-MP-0b3 model found a lower energy path for S, Ni, Cu, and Y than the fine-tuned model.

**Similarity statement**

There are 109 structures in the MP dataset containing exclusively Al and O; pure $Al_2O_3$ appears as mp-1143 (used to generate supercells). There is a total of 243 structures in the training set with Al and O and exactly one of Si, S, Ti, V, Cr, Fe, Co, Ni, Cu, Y, Ag, or Pt.

**Performance summary**

Activation energies for self-diffusivity are underestimated compared to experimental values (extrapolated from lower temperatures). Dopant atom migration minimum energy paths are all stable and are sometimes accurate (e.g., Y) and sometimes only qualitative (e.g., Co) with respect to DFT single-point reevaluations. Fine-tuning substantially improves agreement with DFT.

## 15. Random structure search: Arsenic

*Ab initio* Random Structure Searching (AIRSS)[228] is a simple, yet highly successful, approach for discovering new materials computationally. Multiple candidate structures are generated randomly, subject to physically motivated constraints, and then relaxed to local enthalpy minima using *ab initio* methods such as DFT. There is great interest in accelerating structure prediction by using surrogate models,[99,229–231] such as ML potentials, to perform the initial structural relaxations. Here, we test the suitability of MACE-MP-0 (without the D3 correction) for this task by searching for structures of arsenic at 0.1, 10, and 50 GPa. The exceptional structural variety encountered during RSS probes the robustness of the model in an extremely extrapolative regime; there are only six As structures and no high-pressure data in the training set.

At each pressure, 2000 ($\times 100n$) random structures were generated using $n = 2$–$6$ atoms per primitive unit cell, 2–4 randomly chosen symmetry operations, minimum distance constraints of 2 Å, and a volume per atom of 15–40 Å$^3$. The structures were then relaxed with MACE-MP-0 (ASE, force tolerance of $1 \times 10^{-3}$ eV/Å) and CASTEP;[227] PBE exchange-correlation functional,[118] 400 eV cutoff energy, k-point spacing of $2\pi\times$ 0.05 Å$^{-1}$, and Vanderbilt ultrasoft pseudopotentials[232] with a force tolerance of 0.05 eV/Å and stress tolerance of 0.1 GPa. With these settings, 96% and 99% (at 0.1 and 10 GPa) of structural relaxations were successful with DFT and MACE-MP-0, respectively. The distributions of relaxed structures are depicted in Fig. 29, and the known structures listed in Table II are highlighted with colored symbols. The energy and

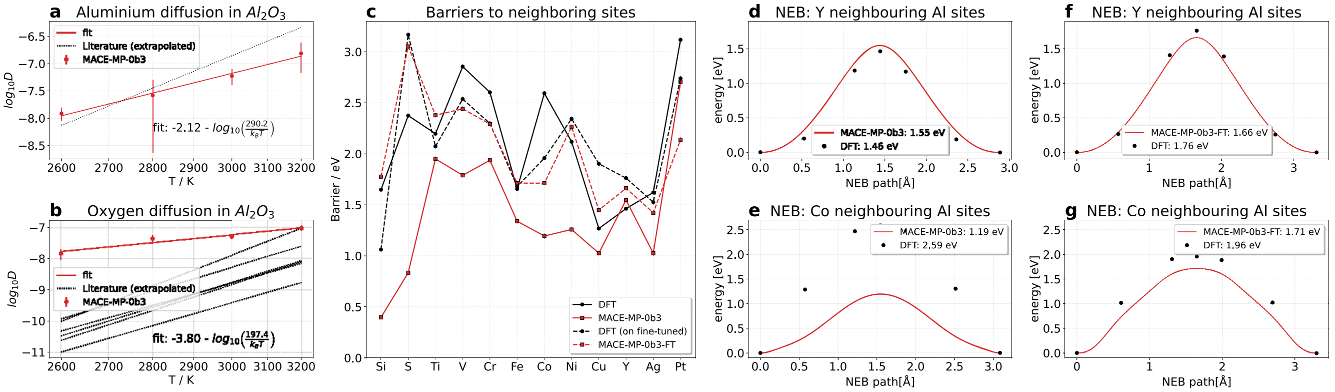


**FIG. 28.** (a) and (b) Arrhenius plot of elemental diffusion in $Al_2O_3$ compared with experimental results from Ref. 225. (c) Comparison of MACE-MP-0b3 NEB barrier paths (red) and PBE single point evaluation of the MACE-MP-0b3 transition state (black) for elements in Al or O sites in $Al_2O_3$ moving to neighboring sites (connecting lines are guides to the eye); the same protocol repeated with MACE-MP-0b3-FT as well is shown with dashed lines. (d) and (e) Lowest energy NEB path for Y (where MACE-MP-0b3 is accurate) and Co (where there is a substantial discrepancy), with single-point PBE evaluations on selected images of the obtained NEB path, indicated by black points. (f) and (g) Y and Co NEBs with the fine-tuned MACE-MP-0b3-FT model and the corresponding single point PBE evaluations.

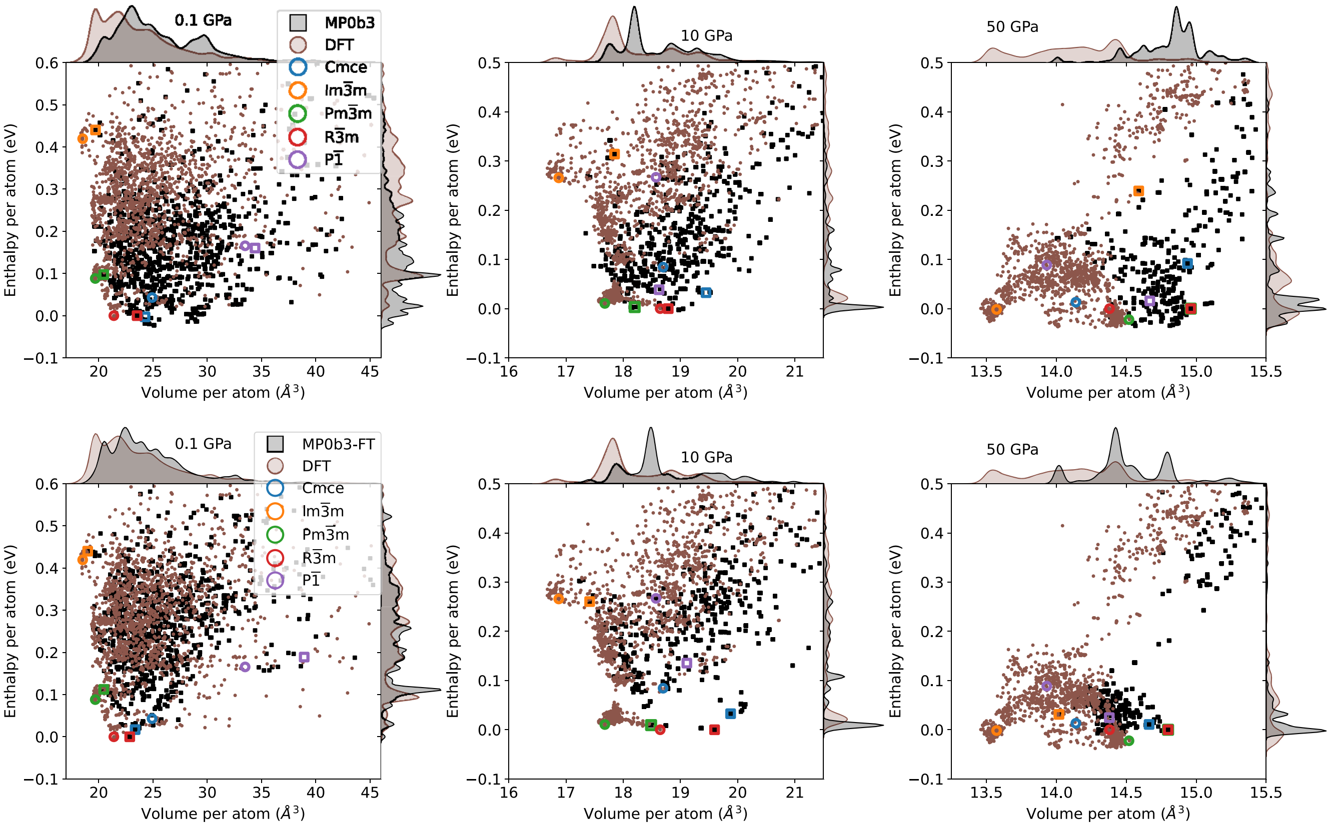


**FIG. 29.** Densities of states of random structure search (RSS) minima for As at 0.1, 10, and 50 GPa, obtained using DFT (brown circles) and MACE-MP-0b3 (black squares) shown before (top row) and after fine-tuning (bottom row). Known As structures are highlighted in various colors (see legend).

**TABLE II.** Summary of known arsenic structures. The white P structure type is used as a proxy for yellow As, as the structure is unknown.[233]

| Structure | Pressure | In training set? | Space group | Z | Found with 0.1 GPa MACE-MP-0b3? |
|---|---|---|---|---|---|
| A7, gray As[234] | Ambient | Yes | $R\bar{3}m$ | 2 | Yes |
| Black P[235] | Ambient | Yes | $Cmce$ | 4 | Yes |
| White P | Ambient | No | $P\bar{1}$ | 24 | $As_4$ tetrahedra found |
| Simple cubic[236] | 27–57 GPa | Yes | $Pm\bar{3}m$ | 1 | Yes |
| bcc[236] | ≥110 GPa | No | $Im\bar{3}m$ | 1 | Yes |

volume distributions are visually similar, with the relative energy differences between the highlighted structures, particularly the low-energy ones, generally being small compared to the overall range. Fine-tuning improves the agreement further and reduces the systematic shift in volume distribution. Inspection of the structures at 0.1 GPa reveals that similar 3-fold coordinated 3D, layered, and 1D structures are found with both MACE-MP-0 and DFT. Furthermore, all known structures listed in Table II were found using MACE-MP-0, including a simplified packing of the $As_4$ tetrahedra found in yellow As.

The original MACE-MP-0a foundation model suffers from "holes" in the potential energy surface, where exceptionally dense, highly coordinated structures are predicted to be overly stable, as shown in Fig. 30. The repulsion seen in the As–As dimer curve indicates that these holes are caused by higher body-order terms in an extrapolative regime—holes occur at shorter As–As distances than those found in the radial distribution function of the 0.1 GPa RSS results. These holes are typically not an issue during ambient pressure MD, due to the large energy barriers seen in Fig. 30. The updated model does not exhibit these holes due to the use of the

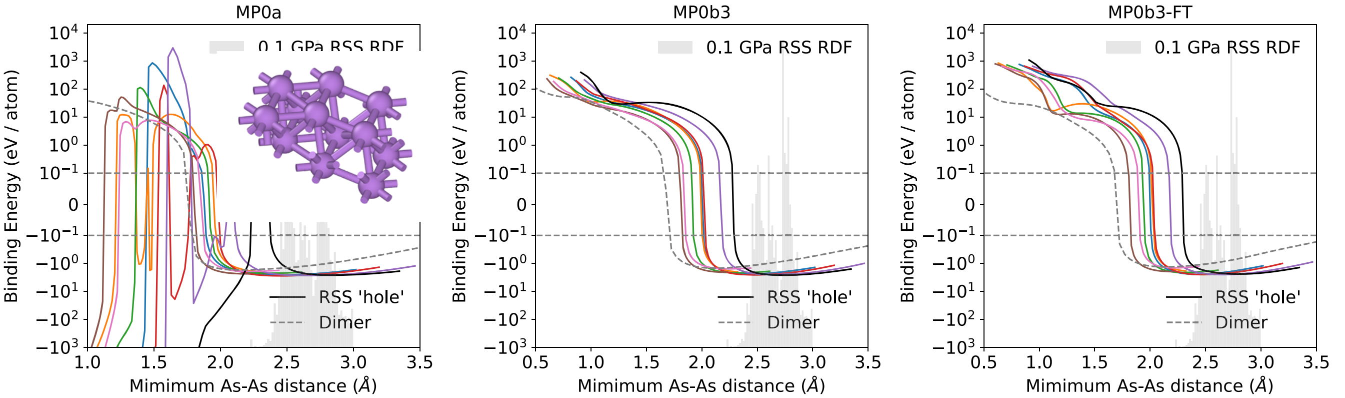


**FIG. 30.** Binding energy of seven (unrelaxed) randomly generated structures (colored lines) and one pathological structure found during RSS with MACE-MP-0a, the original version of our model (black line and inset), are shown as the structures are compressed uniformly with fixed fractional atomic coordinates. The As–As dimer energy and the RDF for the 0.1 GPa As RSS structures are shown for comparison. The left panel corresponds to the original MACE-MP-0a model, the middle panel to MACE-MP-0b3, and the right panel to the fine-tuned MACE-MP-0b3-FT model.

ZBL repulsive pair potential and density normalization (see Sec. VI). Bar a systematic shift in the volume distribution, there is reasonable agreement with DFT even at 50 GPa.

**Fine-tuning**

Fine-tuning was performed on a total of 200 configurations. The five known structures were relaxed at each pressure, contributing 15 configurations to the fine-tuning. A further 185 configurations were selected from the relaxed structures from the original RSS using furthest-point sampling on averaged (across atomic sites) MACE descriptors. This procedure yielded a roughly even split between configurations at each pressure, with 77, 51, and 72 configurations selected at 0.1, 10, and 50 GPa, respectively.

**Similarity statement**

The MP dataset contains six pure As structures. Gray arsenic ($R\bar{3}m$)[234] and the orthorhombic allotrope (*Cmce*, isostructural with black phosphorus)[235] have been observed at ambient conditions, while the simple cubic structure ($Pm\bar{3}m$)[236] is stable at moderate pressure between 27 and 57 GPa. The remaining three structures are > 0.4 eV/atom above gray arsenic. There are an additional 3857 unique structures that contain As and other elements. Within these structures, there are a total of 22 047 As environments, of which 1606, 1537, 534, and 12 are 1-, 2-, 3-, and 4-fold coordinated by neighboring As atoms, respectively (2.7 Å cutoff). Many of the 3-fold As environments are found in AsX compounds, where X is a group I or II element and the As atoms are arranged in local clusters. There is one As atom, which is bonded to four neighboring As atoms [$Cs_7(InAs_2)_3$, mp-1203378], one structure containing isolated $As_4$ tetrahedra ($AsO_3$, mp-1215144), and two structures containing connected $As_4$ tetrahedra ($Re_4As_5S_4$ mp-1209063 and $Re_4As_6S_3$ mp-1219545).

**Performance summary**

All expected low enthalpy stable structures were found.

### 16. Properties of bulk and nanoconfined water

See main text Sec. II A for results and discussion.

**Similarity statement**

The MP dataset contains 21 structures composed of O and H elements and 7769 structures that have O and H elements alone or together with other elements. Based on UMAP analysis, we see that some atomic environments in the example system are similar to environments in the training data. For instance, bulk water and ice comprise typical molecular environments (e.g., the environment of atom 20 in structure 13 of water_exclusive_OH_chemiscope_input.csv) but also environments of hydrogen peroxide (e.g., the environment of atom 1 in structure 14 of ice_exclusive_OH_chemiscope_input.csv). Despite being two-dimensional, the superionic phase also comprises distinct environments mimicking those of water molecules (e.g., the environment of atom 3 in structure 20 of

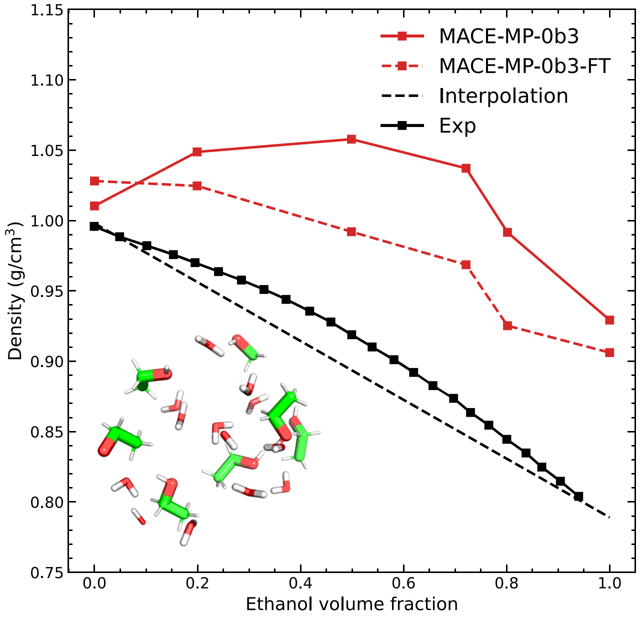


**FIG. 31.** Ethanol–water density curves obtained by NPT MD using MACE-MP-0b3 in single precision, compared to experimental data taken from Ref. 237.

superionic_exclusive_OH_chemiscope_input.csv) and dissociated environments mimicking those of hydrogen peroxide (the environment of atom 12 in structure 12 of superionic_exclusive_OH_chemiscope_input.csv). The environments farthest from the MP dataset are the monolayer oxygen environments surrounded by a (flat) hexagon of 6 other oxygen atoms.

**Performance summary**

The MACE-MP-0 model demonstrates stability and reliable performance in conducting simulations across diverse conditions for both bulk and confined water. It maintains stability in NVT simulations at experimental densities and temperatures for bulk water, ice Ih, and reactive proton defects ($OH^-$ and $H_3O^+$). The model describes extensive proton transfer in nanoconfined water at 4 GPa and 600 K, in good agreement with reference methods.

## 17. Ethanol–water density–composition curves

In this section, we investigate the ability of the MACE-MP-0b3 model to describe mixtures of molecular liquids. In particular, we study the density–composition curve for a range of volume fractions of ethanol in water. Initial configurations were generated with Packmol,[197] with 120 molecules per box, and the initial box vectors were set to be slightly below the experimental density for each composition. Initial structures were minimized to a tolerance of 0.01 eV/Å with the L-BFGS algorithm. Trajectories were generated in the NPT ensemble using ASE, including a D3 dispersion correction with the Becke–Johnson damping function. Final densities were computed as the average of the final 1000 snapshots from the simulation, once the density had converged.

**Similarity statement**

The MP dataset contains 37 structures that contain only the elements C, H, and O. Based on UMAP analysis, we observe that almost all atomic environments in the example system are similar to environments in the training set. On closer inspection, we find that the most similar environments to the majority of the example configurations are clusters primarily containing water, hydroxide, and atomic hydrogen and oxygen, with a few examples containing small hydrocarbon-type fragments.

**Performance summary**

For low volume fractions of ethanol, MACE-MP-0b3 predicts a spurious density maximum at 50% volume fraction, while correctly predicting the lower density for high ethanol volume fractions. MACE-MP-0b3-FT densities more closely follow the shape of the experimental curve, successfully capturing the deflection from linear behavior. MACE-MP-0b3-FT overpredicts the absolute densities by ~4% with respect to experiments, which is not uncommon for a GGA DFT functional on molecular liquids.

## 18. Solvent mixtures

Modeling solvent mixtures requires an accurate description of intermolecular forces within highly disordered systems. To investigate the performance of MACE-MP-0 in this setting, MD simulations were performed for four mixtures of solvents of varying polarity. The investigated systems are water–benzene, water–heptane, water–ethanol, and benzene–ethanol. Simulations were performed at 300 K in the NVT ensemble via the ASE interface. A time step of 1 fs and a friction constant of 0.001 $fs^{-1}$ were used. In the case of immiscible solvents, a mixture of equal volumes of both solvents with their corresponding densities was assumed. In the case of miscible solvents, the experimental density of the mixtures was used. All systems were initialized with a uniform random mixture of both solvents using the `packmol` code.[197]

Figure 32 shows the states of all systems after 1 ns. Notably, mixtures of water with apolar solvents (heptane and benzene) quickly form separate phases, whereas the ethanol–water and ethanol–benzene systems remain mixed on the timescale of the simulation. This is in good agreement with the experiment.

**Similarity statement**

The MP dataset contains 37 structures composed exclusively of C, H, and O, and 1902 structures that contain C, H, and O along with other elements. Regarding the specific molecules, several ice structures but none of the other molecules are included as pure compounds. The closest to benzene (with the ratio of C:H 1:1) is mp-995197, containing chains of dimethylbenzenes with methyl–methyl bridges. The UMAP analysis shows that many atomic environments from our structures have similar environments in the training data. However, no liquid configurations are included in the MP. We provide two files to visualize the interactive UMAP at https://chemiscope.org/. `solvents_mixtures_CHO.json` contains structures exclusively containing C, H, and O. `solvents_mixtures_CHOplus.json` includes structures containing C, H, and O along with other elements.

**Performance summary**

The miscibility of all four mixtures was correctly predicted, at least on the limited timescale of the MD simulations.

## 19. Aqueous interfaces

Simulating complex systems, such as solid–liquid interfaces, is a difficult endeavor, as the potential must simultaneously describe the two materials and their interface. We tested the effectiveness of MACE-MP-0 on a wide range of aqueous interfaces, from oxides and metals to confinement.

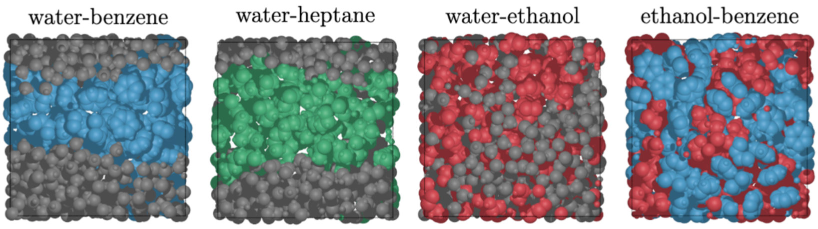


**FIG. 32.** Snapshots of solvent mixtures after 1 ns of NVT MD with MACE-MP-0b3. The solvents shown are benzene (blue), heptane (green), ethanol (red), and water (gray). Axis orientations in the figures were chosen to highlight the phase separation in these systems.

NVT MD simulations were performed on a variety of surfaces at a temperature of 330 K. The average density of water above the surface is shown in Fig. 33.

$SiO_2$ and $TiO_2$ were two notable oxide systems in which dissociative and molecular adsorption were observed, respectively. Deprotonation of water was expected on the surface of silicon dioxide, which is evidenced by the shoulder in the water density plot. These figures show that the interfacial water property is accurately reproduced; however, the liquid phase is overstructured, which is a common characteristic of the PBE functional[31] used in the Materials Project.

Water in confinement was also investigated within $MoS_2$ slit pores. The simulation captures the pronounced stratification characteristic of the aqueous phase perpendicular to the two-dimensional layers. This was also observed for water confined between graphene sheets and boron nitride nanotubes. In particular, in Fig. 33, we show sharply defined interfacial water layers between the $MoS_2$ sheets. Upon the addition of interlayer spacing (not shown), we also capture additionally smoother intermediate layers, noting that with more layers, we lose the sharp peaks at the surfaces.

Finally, the NaCl (001) surface in contact with water was simulated. The system comprised a (3 × 4) NaCl(001) supercell containing 24 atoms, with 3 NaCl layers and a unit cell lattice constant of 5.72 Å, on top of which were 89 water molecules. A subsequent 25 Å of vacuum was added between the adsorbed water layer and the lower layer of the surface. The layered structure of the water, as previously observed in *ab initio* PBE simulations in Ref. 238, is captured by the MACE-MP-0 model, with the positions of the density minima and maxima qualitatively agreeing with the PBE simulations.

Simulation of dissolution processes is another challenge for the MACE-MP-0 model. It must be able to describe the very different chemical environments of the bulk crystal surrounded by water going through the stages of ions detaching from the crystal to fully solvated ions in solution. In Fig. 34, we compare the MACE-MP-0 model in NVT simulations of the pristine NaCl (001) interface in contact with water and a NaCl nanocrystal surrounded by water at 400 K. The nanocrystal system simulated comprised a 4 × 4 × 4 NaCl nanocrystal comprising 32 ions, with a lattice constant of 5.72 Å surrounded by 625 water molecules, giving a final concentration when dissolved of 2.84 mol/kg.

As expected, for a pristine NaCl surface, the model predicts no dissolution events on the time scale of the simulation. Meanwhile, for the nanocrystal surrounded by water, the model captures a dissolution mechanism similar to that reported by Ref. 30 with an ML model trained specifically to capture NaCl dissolution at the revPBE-D3 level of theory. The dissolution proceeds via a crumbling mechanism, where an initial steady loss of ions precedes the rapid disintegration of the crystal. Moreover, the dissolution process is stochastic, leading to an intrinsic variation between independent simulations, as shown by three examples. The resulting solution of ions in water also displays the correct expected orientation of the water molecules with respect to the ions.

**Similarity statement**

The MP dataset contains 460, 100, 112, 13, and 29 structures composed exclusively of [H, O, Si], [H, O, Ti], [H, O, Cu], [H, O, Mo, S], and [H, O, Na, Cl], respectively. The corresponding number of structures inclusive of the given atoms along with other elements is 477, 215, 435, 260, and 190. Based on UMAP analysis, the closest atomic environments for each of these systems are mp-626085, mp-626550, mp-697660, mp-990086, and mp-504600. Two files are provided for each of the systems for visualization using Chemiscope, one for inclusive and one for exclusive matches in the training set.

**Performance summary**

All interface structures correctly predicted, including dissociative adsorption on $SiO_2$ and molecular adsorption on $TiO_2$. At the salt/water interface, it correctly predicted dissolution from nanocrystals and no dissolution from flat surfaces on a nanosecond time scale.

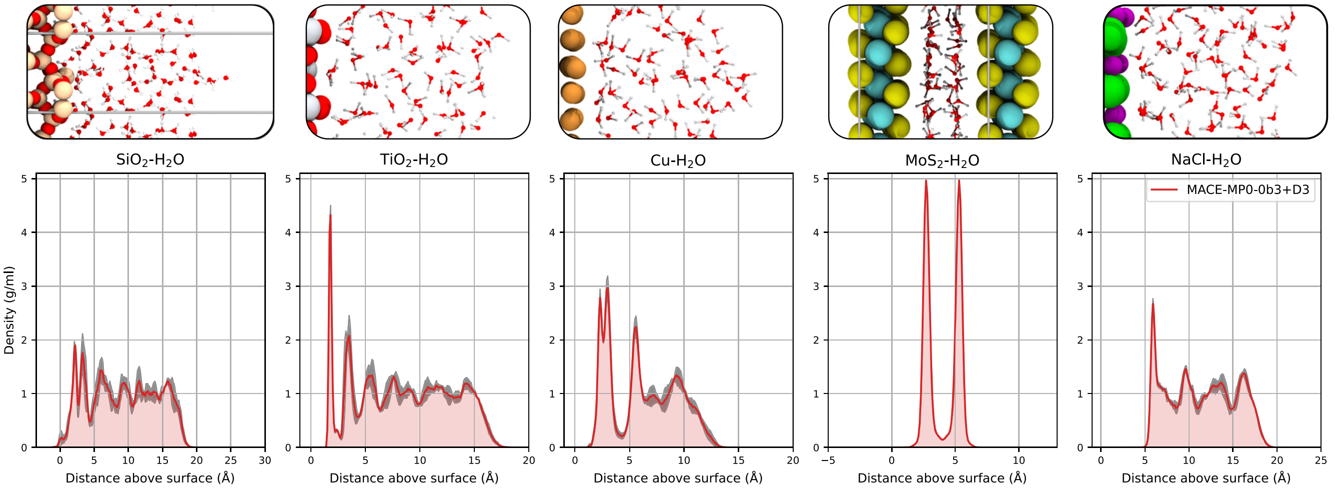


FIG. 33. Water structure and density at various interfaces. Representative snapshots of investigated systems are shown with corresponding water density profiles. The red curve represents the average water density profile obtained from three independent MD simulations; the gray shading indicates the standard deviation. The systems depicted are water on silicon dioxide ($SiO_2$), titanium dioxide ($TiO_2$), copper (Cu), between two layers of molybdenum disulfide ($MoS_2$), and on sodium chloride (NaCl).

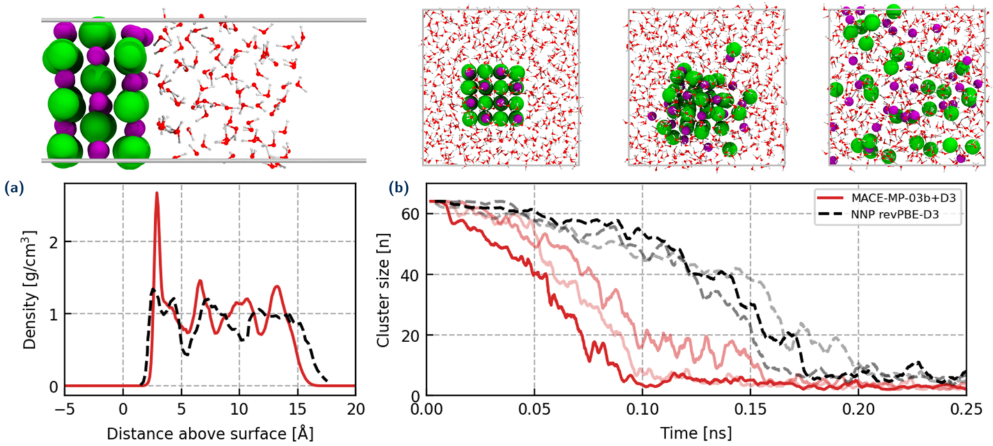
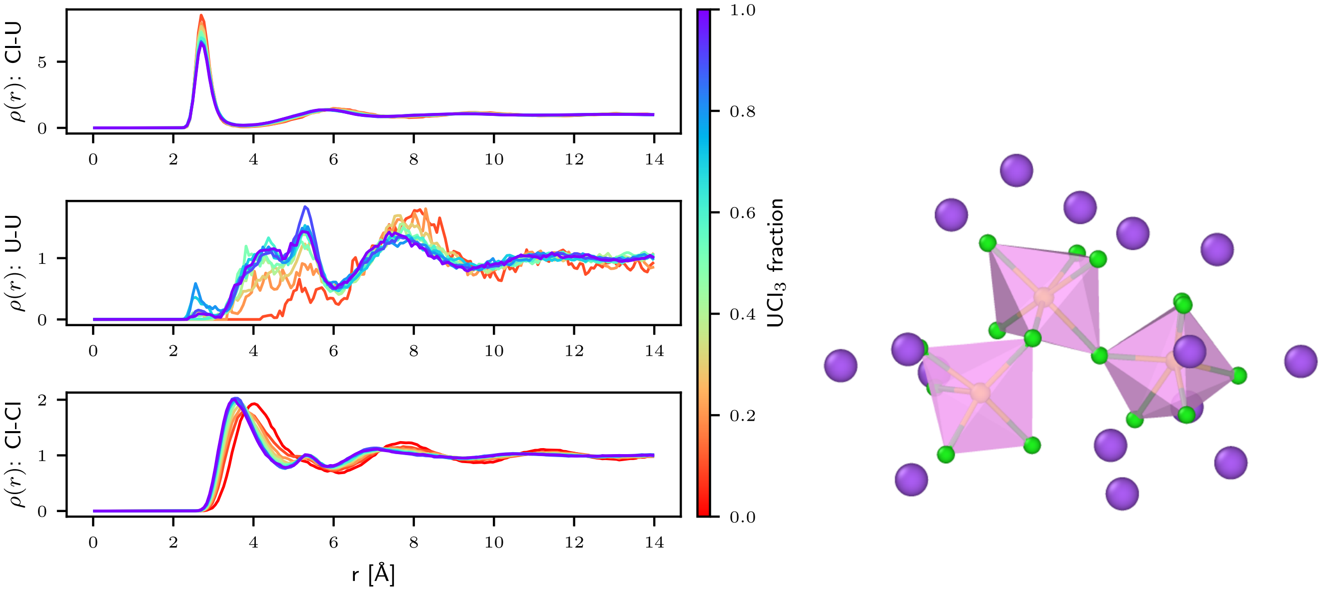


**FIG. 34.** Dissolution of NaCl in water. (a) Density profile of water in contact with the NaCl(001) surface, with a representative snapshot from the simulation showing no dissolution events from the pristine surface. (b) Evolution of crystal size of 4 × 4 × 4 NaCl nanocrystal in water over time, comparing the MACE-MP-0 (red line) with an ML model explicitly trained to capture NaCl dissolution (black dashes).[30] Representative snapshots showing the dissolution progress of the crystal are shown above the plot.

## 20. Molten salts

With increasing interest in molten-salt energy technologies, we have simulated binary NaCl–UCl$_3$ salt mixtures $(\mathrm{NaCl})_{(1-x)}(\mathrm{UCl}_3)_x$ at different compositions using MACE-MP-0. The initial structures were randomized using Packmol in a cubic cell at the density estimated by the linear interpolation of the densities of constituent solid-state salts at 0 K.[197,239] We then implemented geometric optimization using Lennard-Jones potential and further relaxed structures using MACE-MP-0+D3 with a two-step process: NVT relaxation through annealing at 1.2× target temperature for 5 ps and NPT relaxation at 1100 K and zero pressure for 10 ps. Figure 35(a) presents the pair correlation functions between Cl–U, U–U, and Cl–Cl in salt mixtures. The characteristic peaks and transitions



(a) MACE-MP-0b3+D3 (b)

**FIG. 35.** NaCl–UCl$_3$ molten salt mixtures at 1100 K. (a) Pair correlation function of NaCl–UCl$_3$ mixtures at 1100 K. (b) Example U–Cl oligomers forming vertex sharing coordination polyhedra (U: yellow, Cl: green, and Na: purple).

are consistent with previous polarizable ion models[240] and AIMD simulations,[241] except for a noticeable shift of the U–U peak from 4.5 to 4 Å and a U–U peak formation at 5.5 Å at high $UCl_3$ concentration. The shift could be explained by the lack of Hubbard *U* correction for rare earth elements in MP, leading to unrealistic ionic radii and solvation shells in the mixture. We also note that there is a small U–U peak around 2.5 Å. This peak is absent in previous molten salt studies at high temperature,[241] but as it is close to the equilibrium distance between U as demonstrated by the homonuclear diatomic curve (Fig. 66), its appearance indicates the formation of a few U–U bonds at a high fraction of molten $UCl_3$ salt.

**Similarity statement**

The MP dataset contains 573 structures composed of at least one Na, U, or Cl atom, 14 elemental Na crystals, 14 elemental U crystals, and 3 $Cl_2$ molecular crystal structures. Based on UMAP analysis, we see that all atomic environments in the example system are similar to environments in the training data. We found that $Cl_2$ molecular crystals are close to the molten salts, but most of the pure U metals are found separated from the molten salt in terms of MACE descriptors. We provide

- `T_1100-P_0-seed_3-npt-5_chemiscope_input.json`

to help visualize the interactive UMAP of molten $Cl_{64}Na_{28}U_{12}$ on https://chemiscope.org/.

**Performance summary**

Correct pair distribution peaks and variation of peak positions as a function of concentration, with a notable shift in the first U–U peak position due to the absence of the Hubbard-*U* correction.

## 21. Room temperature ionic liquids

Room temperature ionic liquids provide a class of organic solvents with desirable properties such as low melting and high boiling points, chemical inertness, and good ionic conductivity, making them applicable to different chemical and physical applications. Furthermore, these properties can be tuned by changing substituents on the anion or cations. The vast availability of substituents makes simulations at quantum mechanical accuracy to optimize these *in silico* a very interesting approach.

For the given example, the class of imidazolium-based ionic liquids was chosen. Simulations using the $BF_4^-$ anion with the 1-butyl-3-methylimidazolium (BMIM) cation were conducted. A single MD simulation was performed starting at the experimental density[243] of BMIM $BF_4$ at 273 K. The temperature was stepwise increased from 273 to 323 K. Between each increase, the cell was adjusted to the new density and equilibrated over 500 fs. All simulations are conducted using the MACE-MP-0b3 with additional D3(BJ) corrections.

At each temperature, an NVT simulation using a Langevin thermostat was conducted for 50 ps with a time step of 0.5 fs. From the final trajectory spanning 250 ps over 5 different temperatures, data points were uniformly selected, and energies and forces were compared to DFT.[244] In addition, the radial distribution function was compared to an AIMD simulation, indicating a shift in the hydrogen positions compared to DFT. Furthermore, interatomic interactions are probed using a volume scan, demonstrating the importance of the additional D3(BJ) correction to stabilize the correct volume (see Fig. 36). Finally, a MD simulation in an NPT ensemble at 1 atm and 300 K showed that the model reproduces the experimental density within 5%.

All models struggle to run simulations using a $Cl^-$ anion instead of $BF_4^-$, as this results in Cl atoms bonding to the aromatic ring. To address this, fine-tuning was performed on 100 configurations of 16 ion pairs of BMIM Cl. A workflow using IPSuite[245] was used to set up the simulation box from SMILES,[246] generate GROMACS[247] input files, and sample random configurations. The CL & P force field[248] was used in the 50 ns NVT sampling simulation. After fine-tuning, the MACE-MP-0b3-FT model was able to run stable simulations for BMIM Cl.

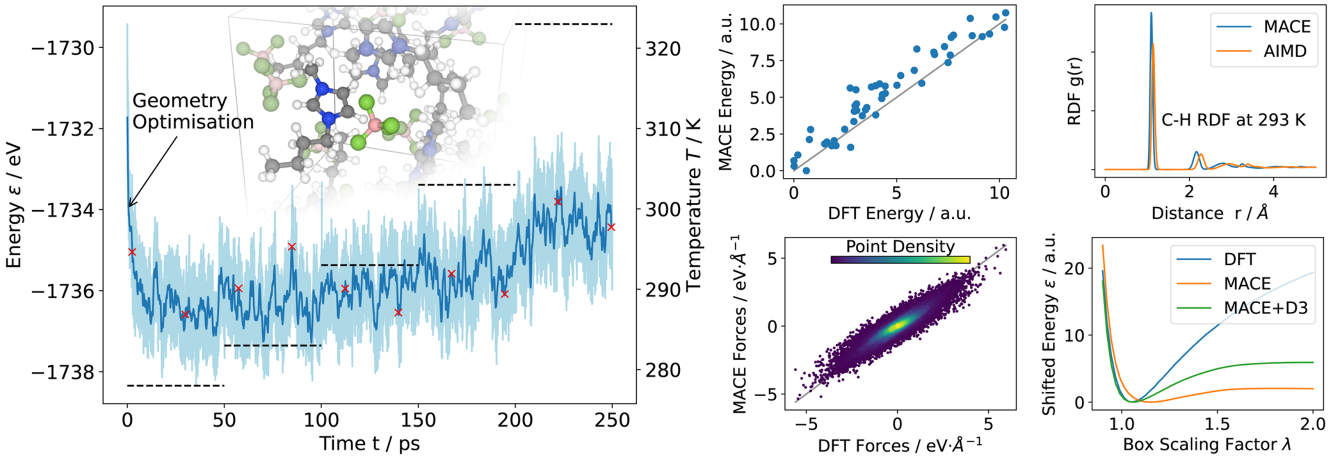


**FIG. 36.** MD simulation of the BMIM $BF_4$ room temperature ionic liquid. The left panel shows energy as a function of the trajectory, starting from an energy minimization, followed by MD simulation in an NVT ensemble with a stepwise increase in temperature (dashed lines, right axis). The middle column of panels shows energy and force parity plots for configurations from the trajectory (red markers). The right top panel compares RDF to AIMD, and the right bottom panel shows a rigid-molecule volume-scan test.[242]

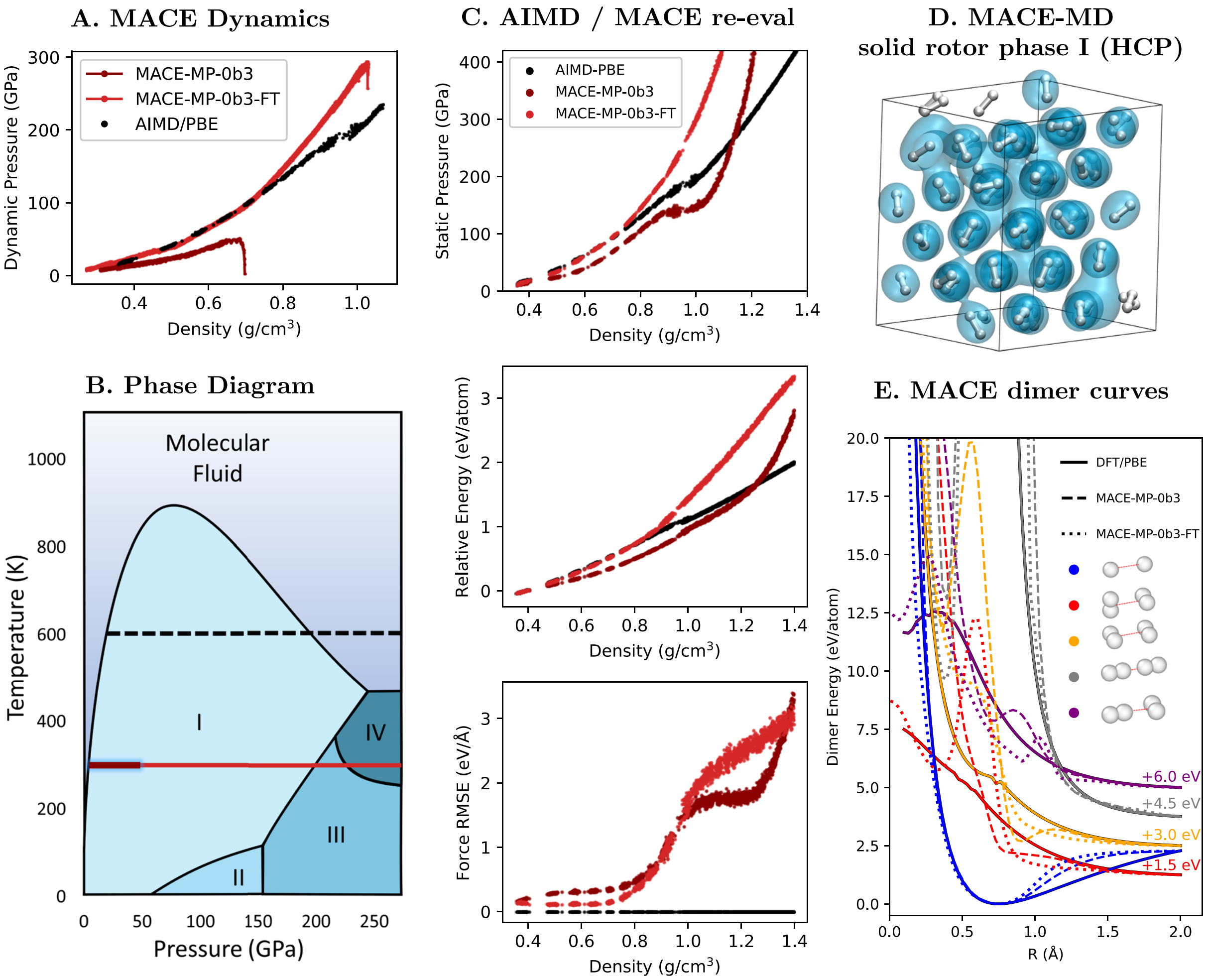


FIG. 37. (a) Pressure/density dependence in MACE-MD compared to AIMD. (b) Illustrative phase diagram of high-pressure hydrogen.[249,250] Horizontal lines show the NPT-MD simulations: MACE-MP-0b3 (dark red), MACE-MP-0b3-FT (red) (see text), and AIMD/PBE (black).[251] (c) MACE properties (pressure, energy, forces) evaluated on the AIMD trajectory and compared to the original PBE result. (d) MACE reproduces solid hydrogen phase I as a hexagonal-close-packing (HCP) lattice of free rotors. (e) MACE energy curves computed for different orientations of the $H_2$–$H_2$ dimers compared to DFT/PBE results. The distance is measured between the centers of mass of the two molecules. The curves were offset along the y axis to improve readability.

**Similarity statement**

There are 52 structures in the MP dataset that explicitly include the $BF_4$ anion. Although there are organic nitrogen-containing molecules as well as heterocyclic systems, there are no alkyl-substituted imidazolium derivatives like BMIM in the training dataset. We provide `BMIM_BF4.json` for a comparison of snapshots from the MD trajectory to the training dataset at https://chemiscope.org/.

**Performance summary**

Stable MD for $BF_4$ anion, but Cl anion bonded to imidazolium. Bond formation is prevented by applying fine-tuned models. Intermolecular distance distribution shows a small peak shift, and intermolecular attraction is underestimated.

## 22. High-pressure hydrogen

Condensed-phase hydrogen is an exotic state of matter that forms at extreme conditions in the core of larger planets and in specially designed laboratory diamond anvil cells. Despite the simplicity of the hydrogen atom and molecule, the condensed phase exhibits fascinating phenomena such as entropy-driven phase transitions,[252,253] phonon localization,[254,255] quantum rotor solid phases,[256] and an insulator-to-metal transition.[257,258] AIMD has been used extensively to study the molecular mechanisms underlying these phenomena; however, simulations are often affected by finite-size effects. Bespoke ML potentials,[259] fitted to reproduce the interaction of hydrogen molecules with *ab initio* accuracy, have demonstrated simulations at an unprecedented level of detail, unlocking new scientific observations.[257] Being so different from other materials in the MP database, solid hydrogen is a uniquely challenging test for MACE-MP-0b3.

The stability of the potential on this system was tested by running MD simulations at high pressure and by investigating the H–H and $H_2$–$H_2$ dimer curves for molecules in different orientations. The MD simulations started from an AIMD thermalized Pc-48 crystal structure,[252,253,260] and pressure was slowly ramped up at constant temperature. The MACE-MP-0b3 potential is stable up to a density of 0.7 g $cm^{-3}$ and pressure of 50 GPa; however, it does not accurately reproduce the AIMD equation of state (EoS). At low pressures, the dynamics does reproduce the expected behavior of hydrogen in phase I—HCP crystal of freely rotating hydrogen molecules. A fine-tuned variant of the potential, MACE-MP-0b3-FT, was obtained by sampling 50 configurations with densities ranging from 0.5 to 0.9 g $cm^{-3}$ from a MD trajectory generated with an earlier MP potential. These configurations were evaluated with PBE, and the MACE-MP-0b3 model was fine-tuned on this new dataset. The resulting potential is more robust and remains stable up to a remarkable 300 GPa and a density of 1.0 g $cm^{-3}$. In addition, the MACE-MP-0b3-FT model accurately reproduces the AIMD EoS up to 130 GPa.

The accuracy of both potentials was also quantified by reevaluating the pressure, energy, and forces on an existing AIMD trajectory.[251] MACE-MP-0b3 follows the overall trend of the PBE EoS up to 0.9 g $cm^{-3}$, but shows a constant erroneous shift to lower pressures. MACE-MP-0b3-FT significantly improves the result and accurately captures the EoS up to 0.8 g $cm^{-3}$. A similar result is found for energy and forces, where MACE-MP-0b3-FT significantly improves the agreement with PBE in the density range 0.5–0.9 g $cm^{-3}$, where new data were added. Finally, the energy dissociation curves for $H_2$–$H_2$ dimers are smooth at separation distances above 1.4 Å, which corresponds to distances observed in equilibrium MD. Similar smooth behavior is observed for the H–H energy dissociation curve, where both models accurately reproduce PBE around the 0.75 Å equilibrium bond length of the $H_2$ molecule.

**Similarity statement**

The MP dataset contains 17 structures with a crystal structure composed exclusively of H. From these, only 2 structures are above the 0.2 g $cm^{-3}$ density value: mp-1096977 (hexagonal P4/mmm 0.24 g $cm^{-3}$) and mp-754417 (hexagonal P6/mmm 0.24 g $cm^{-3}$), yet MACE-MP-0b3 was found to be stable up to around 0.7 g $cm^{-3}$. We provide `hydrogen_exclusive.json` for visualization on https://chemiscope.org/.

**Performance summary**

The correct solid hydrogen structure is reproduced at low to moderate pressures by MACE-MP-0b3 with an inaccurate EoS. At pressures above 50 GPa, the potential failed and resulted in unphysical structures. A MACE-MP-0b3-FT potential fine-tuned on 50 configurations successfully reproduced the EoS up to 130 GPa and remained stable up to 300 GPa.

## 23. Ammonia and borane thermal decomposition

Ammonia and borane form an adduct $NH_3BH_3$. At high temperatures, these molecules lose hydrogen gas to produce increasingly heavier B- and N-containing molecules, ultimately resulting in the growth of hexagonal boron nitride (hBN).[261] We simulated this process with MACE-MP-0b3 by running a NVT molecular dynamics simulation of 32 ammonia and 32 borane molecules in a cubic box of length 16 Å at a temperature of 1600 K for 1 ns with a 0.5 fs time step. We analyzed the time evolution of the system in terms of the sizes of heavier atom clusters (excluding hydrogen), illustrated in Fig. 38. Initially, the formation of ammonia borane adduct and small borane clusters is observed, while in 100 ps timescales, B and N atoms are increasingly more clustered, with a preference for B–N bonds over homonuclear bonds. Ultimately, an hBN-like fragment is formed.

**Similarity statement**

The training set contains 67 structures composed of H, B, and N elements. The training set contains various structures encountered during the simulation, including ammonia, borane, and HBN compounds of various stoichiometries, for example, borazine ($B_3N_3H_6$) and $(BNH_2)_n$ chains. We performed UMAP analysis for 100 frames

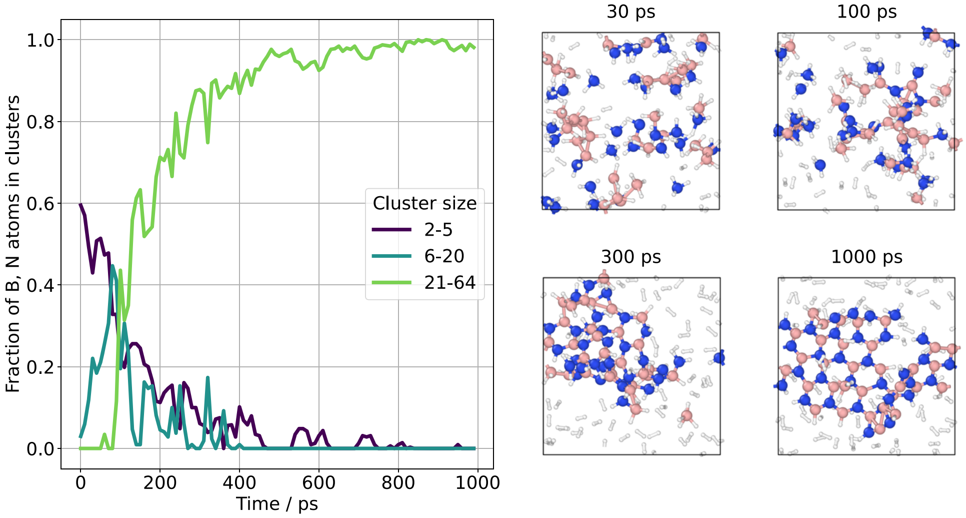


FIG. 38. Decomposition of ammonia and borane at 1600 K. The plot shows the time evolution of cluster size in terms of the fraction of heavier atoms (B and N) found in each cluster size group. Snapshots of the system at different times show the growth of BN clusters and the evolution of $H_2$ molecules.

taken from a 1 ns MACE-MP-0 MD simulation against training data containing at least one of the HBC elements and any other elements. Based on the UMAP values, most of the simulation atomic environments are clustered near the training data, with exceptions being species with unusual valency (e.g., $BH_2$). The closest structures in the training set are mp-1197795 and mp-1203334 (both containing B and N, among other elements) and mp-1214811 ($B_6N_6H_{10}$ bicyclic aromatic compound). We provide `ammonia-borane.json` for visualization on https://chemiscope.org/.

**Performance summary**

Model correctly predicts hydrogen production and BN cluster formation from thermal decomposition of $NH_3BH_3$.

## 24. Heterogeneous catalysis

Computational heterogeneous catalysis revolves around the exploration of *operando* catalyst stability and catalytic reaction mechanisms to provide information about the nature of the active site that defines a catalyst's performance. This information provides

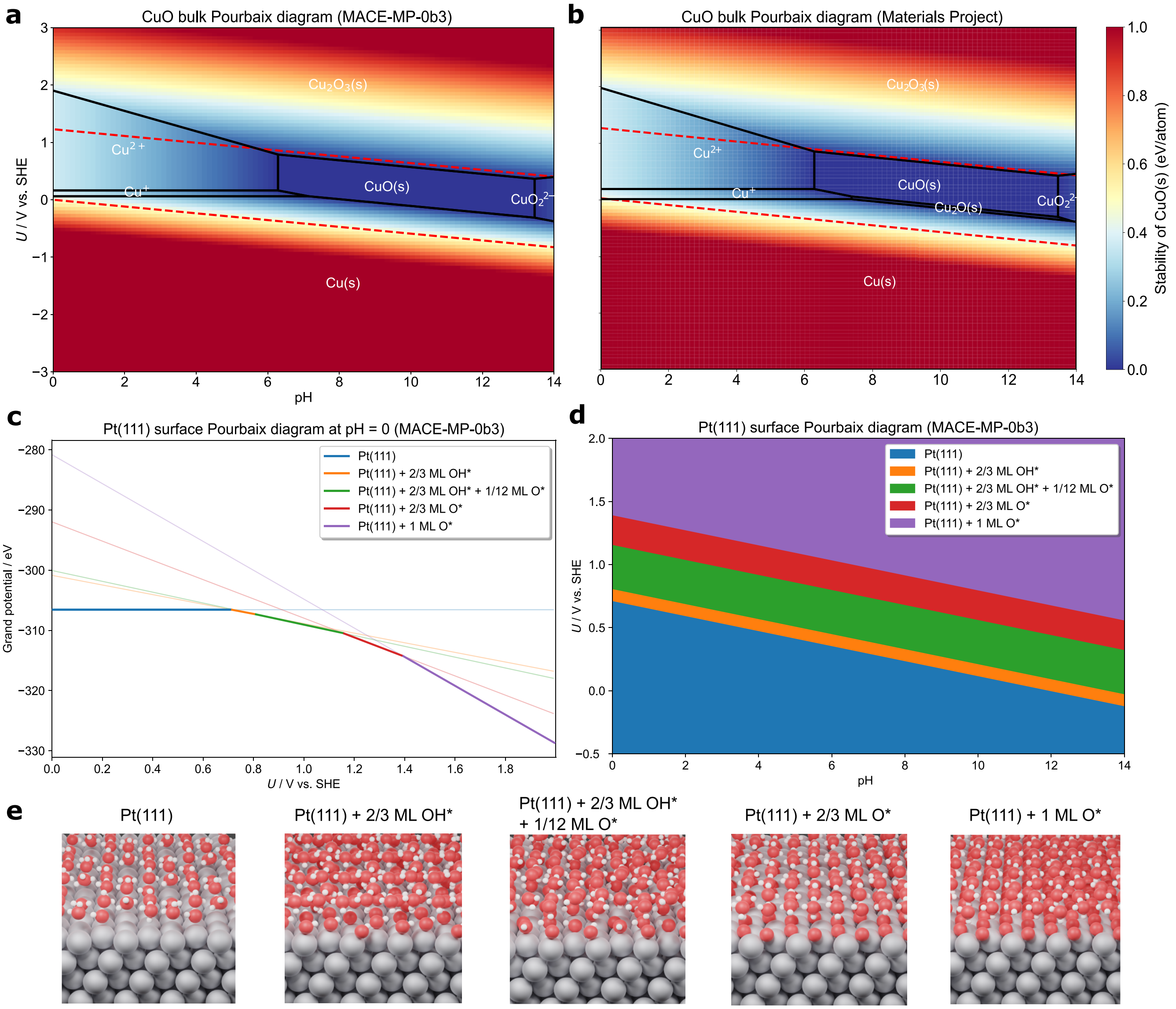


**FIG. 39.** Pourbaix diagrams of CuO bulk systems with energies of relevant solid compounds taken from (a) the MACE-MP-0 calculations and (b) the MP reference. (c) and (d) show the MACE-MP-0-calculated Pt(111) surface Pourbaix diagrams at pH = 0 and at various pH, respectively, which are in good agreement with Ref. 39. Different stable surface structures are represented in different colors. (e) shows the stable Pt(111) surface structures from low to high applied potentials. The red dashed lines indicate the stable window of water ranging from $U_{RHE} = 0$ V to $U_{RHE} = 1.23$ V ($U_{RHE} \approx U_{SHE} + 0.059 \cdot \text{pH}$).

a basis for screening applications to find efficient and, ideally, non-precious and non-toxic catalysts. To this end, a variety of atom-scale properties are investigated, including bulk and surface energies to evaluate catalyst stability in (surface) phase or Pourbaix diagrams, as well as adsorption energies, reaction thermodynamics, and reaction barriers that are key to elucidating mechanisms and catalytic activity.[51] Local geometry optimizations and transition state searches via, e.g., NEB calculations[201,226] that yield target properties are usually conducted on slab models that exemplify the catalyst surface. By expanding the usual surface science approaches via thermodynamic referencing of protons and electrons to pH and applied potential on the basis of the computational hydrogen electrode (CHE),[61] concepts in thermal catalysis can be extended to electrocatalysis. This approach provides fairly robust results even though the simulation of the electrolyte, charged species, or an applied potential and, therefore, the direct influence of the electrified solid/liquid interface is omitted. The computationally involved methodology is fully transferable to MACE-MP-0, and we apply it in full to the examples presented in the main text and below.

### *a. Pourbaix diagrams*

In Fig. 39, we show the Pourbaix diagrams, calculated by MACE-MP-0 with D3 corrections, which illustrate the aqueous stability for a CuO bulk and a Pt(111) surface in dependence of applied potential and pH as referenced by the CHE. Structures for bulk CuO and all other related oxide and peroxide compounds are taken from MP and are subsequently optimized (both atomic positions and cell

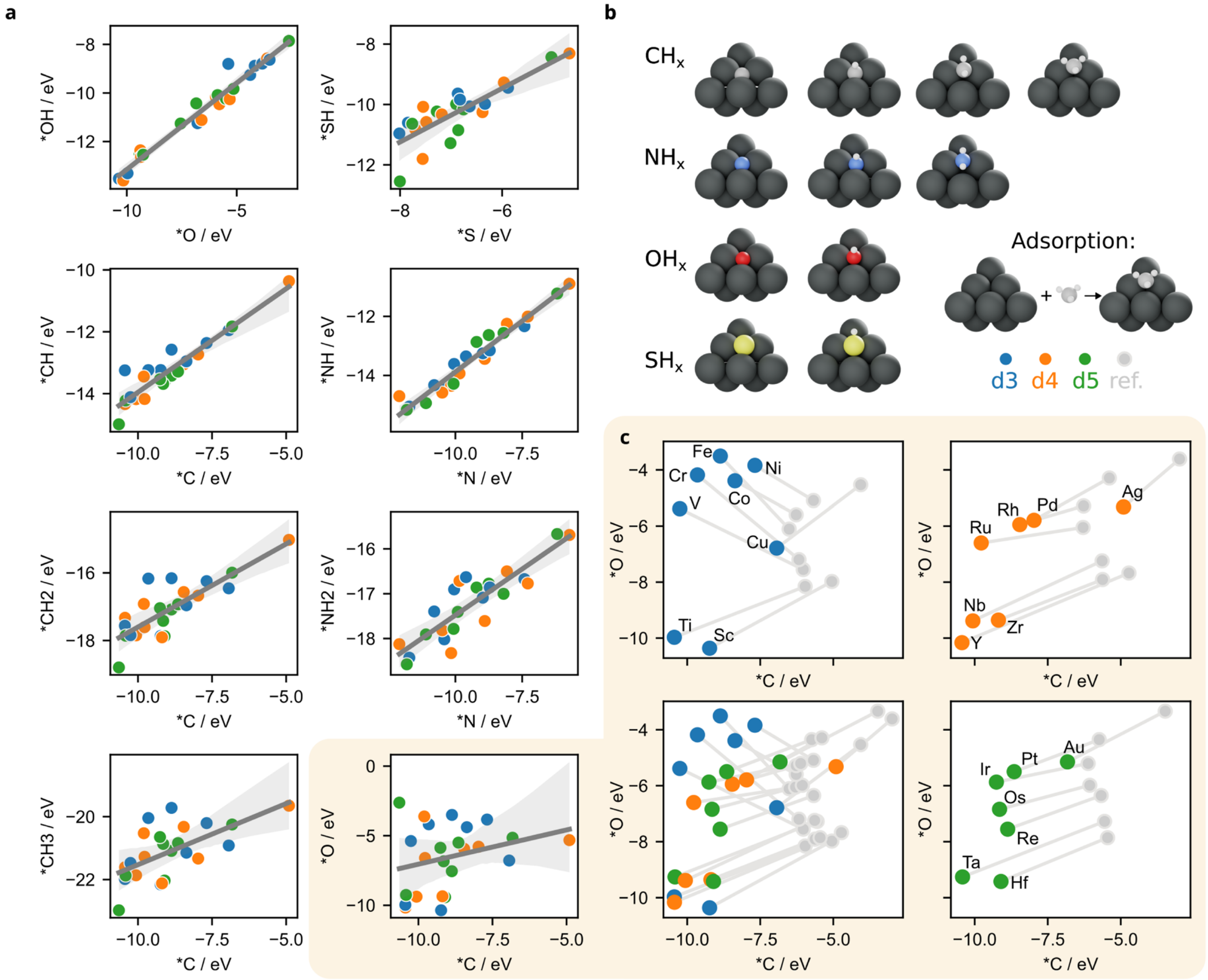


**FIG. 40.** Correlation plots (a) between the adsorption energies of two intermediates at the same (hollow site) of tightly packed metal surfaces (b). The correlation between *O and *C (c) is not linear (in agreement with the literature).

parameters) using MACE-MP-0. The energy corrections for oxides and peroxides, as well as the free energies for aqueous ions, are consistent with the values used in MP. As shown in Figs. 39(a) and 39(b), the overall trend of the CuO stability predicted by MACE-MP-0 is well-aligned with the result given by MP, except for the narrow region of the $Cu_2O$ phase that is not reproduced by MACE-MP-0. As depicted in Figs. 39(c)–39(e), MACE-MP-0 predicts that the Pt(111) surface starts to oxidize at $U_{\mathrm{SHE}} = 0.72$ V (pH = 0), followed by a stepwise increasing $OH^*/O^*$ surface coverage with more positive electrode potential. This is generally in good agreement with the Pt(111) surface Pourbaix diagram reported previously (58),[39] despite the predicted starting oxidation potential being 0.06 V too low.

### b. Linear scaling relationships (LSRs)

Adsorption energies of molecules and intermediates are indicative of catalyst reactivity and often used as descriptors in screening studies for catalyst materials. The adsorption energies are governed by electronic and geometric factors. Provided a consistent geometric environment (e.g., a hollow site of a tightly packed metallic lattice) across different metal surfaces, the trend in adsorbate binding energies resulting primarily due to electronic effects can be observed. A well-known property of catalytic surfaces (e.g., transition metals) is that the binding energies of individual intermediates are not independent of each other as a consequence of the varying degree of occupation of the metallic d-band.[40,64] Linear scaling relationships (LSR) were found for a range of metallic surfaces and molecules that bind to this surface through the same atom (e.g., E scales with $EH_x$, where E = C, O, N, S and x = 1,2,3); however, this scaling is not linear when comparing adsorbates that bind through different atoms (e.g., C vs O).

In Fig. 40, we show the correlation between the adsorption energies of $EH_x$, where E = C, O, N, S, and x = 0, 1, 2, 3. The structures [Fig. 40(b)] were relaxed with the MACE-MP-0 model with D3 correction (cutoff = 4 nm), and the adsorption energy was computed as $\Delta E_{\mathrm{ads}} = E(a*) - E(*) - E(a)$, with $a$ as the adsorbate and $^*$ as the empty surface site. The observed correlations are linear in all cases except for the correlation between O and C. In Fig. 40, the MACE-MP-0 computed adsorption energies (blue/green/orange circles) are compared to the corresponding DFT values (connected with faint gray lines to faint gray circles) as reported by Nørskov.[40] In this plot, although the absolute error of the obtained adsorption energies in comparison to the DFT values is high (which is not surprising as the model is extrapolating in this example), the trend of grouping metals into passive (noble, e.g., Au), catalytic (the so-called Pt group), and non-reducible (e.g., Zr) is correctly captured, and the essence of the LSR relationships was reproduced with the MACE-MP-0+D3 model.

### c. CO (electro-)oxidation on Cu

We test the ability of MACE-MP-0 to predict the catalytic reaction mechanisms for the oxidation of CO on different facets of Cu, as previously explored via DFT.[67] Specifically, we evaluate two complex potential-independent reaction steps of the CO oxidation, $OH^* + CO^* \rightleftharpoons COOH^*$ and $OH^* + COOH^* \rightleftharpoons CO_2(g) + H_2O(g)$, as shown in Fig. 41. Each reaction contains two reaction barriers, labeled i, ii and iii, iv, respectively. Other reaction

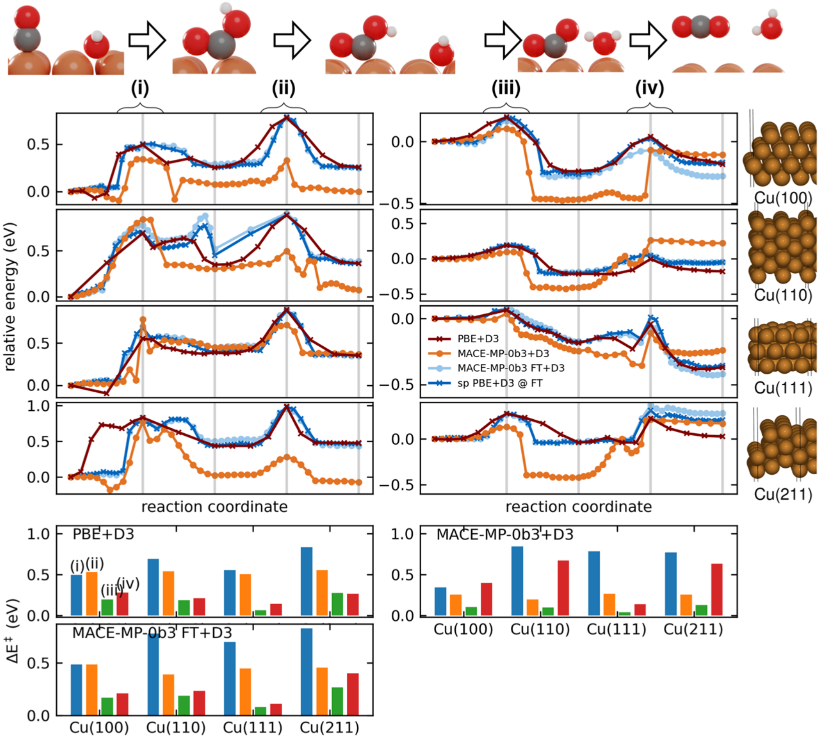


**FIG. 41.** NEB profiles (top) and extracted barriers (bottom) of reactions (i–iv) computed for the potential independent steps in the multistep reaction mechanism of CO electro-oxidation on Cu for the low index facets (100), (110), (111), and (211). The reaction mechanism without implicit solvent is adapted from Ref. 67. The NEB calculations were carried out with MACE-MP-0b3 (orange), a fine-tuned model (light blue), and DFT (dark red), each with the D3 dispersion correction. In addition, single point PBE+D3 calculations (deep blue) were performed for the fine-tuned NEB-path. Faint vertical lines indicate minima and saddle points along the reaction coordinate for reaction steps i–iv.

steps of the CO oxidation, which we do not include in our evaluation, are the adsorption of $CO^*$ and $OH^*$, where the latter is an electrochemical process that can be described via the CHE.[61] We recompute the reaction pathways for two Cu terraces (111) and (100) and two step-sites (110) and (211) via geometry optimizations of the initial and final states and subsequent NEB calculations (one for each reaction barrier i–iv) in lattice-parameter-adjusted simulation cells.

The reaction profiles and reaction barriers from the converged NEB calculations are shown in Fig. 41 for PBE+D3, MACE-MP-0+D3, and the fine-tuned model (+D3). Fine-tuning was performed via a training set of 70 structures, which was collected by drawing every fifth MACE-MP-0 NEB trajectory image and recomputing energies and forces with the computational settings of the `MPtrj` dataset. The foundation model MACE-MP-0+D3 shows in most cases qualitative agreement with the PBE+D3 reference, capturing the trends in barrier height differences for reaction steps i–iv. Quantitatively, the MACE-MP-0 barriers are underestimated, in line with systematic softening behavior noted in Ref. 262. Except for this are the cases of $CO_2$ and $H_2O$ desorption (vi) for the (110) and (211) facets. Here, the barriers are overestimated due to an underestimated final binding energy and (relatedly) a different final state geometry. In contrast, the fine-tuned model shows near-quantitative agreement in terms of the barrier heights, deviating on average by only 0.056 eV. This close agreement does not become immediately apparent when comparing the NEB profiles, which show some deviations. This can be attributed to a different number of images in the climbing image NEB calculations (43 for MACE and 19 for PBE+D3). To better demonstrate the accuracy of the fine-tuned MACE, we additionally performed PBE+D3 single point calculations for images of the fine-tuned NEB path, illustrating excellent agreement with the PBE-D3 energies.

### *d.* $In_2O_3$

As a final test system, we investigate a key step ($CH_2O_2 \rightarrow CH_2 + O$) in carbon dioxide hydrogenation to methanol over indium oxide via an NEB transition state search. This reaction has been extensively studied with *ab initio* methods due to indium oxide's promising selectivity compared to conventional modified copper catalysts.[41,68] First, we perform a global geometry optimization of the reactant near an oxygen vacancy. MACE-MP-0 correctly identifies the three-oxygen-coordinated indium as the active site.[41] Following a NEB calculation, MACE-MP-0 predicts the reaction barrier within 15% of that investigated with DFT (1.16 vs 0.98 eV), as visible in Fig. 2. Fine-tuning with just five single-point DFT calculations recovers the barrier with quantitative accuracy.

**Similarity statement**

With the exception of the CuO bulk Pourbaix diagram, which is based on structures from the Materials Project, all presented examples treat surface slab models (with and without reacting adsorbates) and represent a significant extrapolation. The dataset does not include any such slab models but only related bulk structures. These bulk structures include six bulk structures that contain Pt, O, and H and 15 bulk structures that contain Pt and O for the Pt Pourbaix diagram and the corresponding LSR example (similar number for other metals in the LSR); 8 different Cu-bulk phases and 111 structures composed of Cu, O, Cu, and H along with other elements for the example of CO oxidation on different Cu facets; and nine structures that contain In, O, H, and C and 825 bulk structures that contain indium and oxygen for the $In_2O_3$ example. The most similar configurations for the CO oxidation example are $Cu_2H_4C_4N_3O$ with the Materials Project ID mp-686268 and for the $In_2O_3$ example $(NH_4)In(OH)PO_4$ with the ID mp-764968. We provide `Pt_LSR.json` and `COoxCu_closest_training_points.csv` for the LSR and CO oxidation examples to help visualize the interactive UMAP on https://chemiscope.org/.

**Performance summary**

Solid energies were accurately predicted, leading to correct Pourbaix diagrams. Adsorption energies are overestimated, but linear scaling relationships between different surface/adsorbate pairs are preserved. Minimum energy paths for reactive steps are qualitatively correct, in some cases small (0.2 eV), and in others larger (0.5 eV) energy errors.

## 25. Carborane rearrangement

Carborane ($C_2H_{12}B_{10}$) is an organoboron compound with uses in drug discovery[263] and organometallic chemistry.[264] It adopts icosahedral cluster structures, with three isomers based on different relative positions of carbon atoms: *ortho* (*o*), *meta* (*m*), and *para* (*p*). The thermally activated rearrangements between these isomers have been thoroughly studied,[265] with several mechanisms proposed involving triangular face rotation (TFR). We used MACE-MP-0b3 to study two pathways from the *ortho* isomer to the *meta* isomer: one involving an anticuboctahedral transition state (by mutual rotation of two opposite faces), the other involving the rotation of a single triangular face.

We used MACE-MP-0b3 to obtain rearrangement pathways for each mechanism. The *o* and *m* isomer structures were relaxed with a force tolerance of 0.01 eV/Å using MACE-MP-0b3. Pathways were obtained using a nudged elastic band with 50 images, relaxed with a force tolerance of 0.05 eV/Å, first with spring constants of 0.5 and then 0.05 eV/Å2. We also evaluated energies of the images at the PBE/def2-TZVPPD level of theory using ORCA 5.0.3.[266]

**Fine-tuning**

We fine-tuned MACE-MP-0b3 using the *ortho*, *meta*, and *para* structures and 7 high-energy structures from various pathways connecting these isomers. MACE-MP-0b3 and MACE-MP-0b3-FT pathways, along with some reference DFT energies, are shown in Fig. 42. MACE-MP-0b3-FT gave larger energy barriers than MACE-MP-0b3: 2.56 vs 1.52 eV for the anticuboctahedral mechanism and 2.32 vs 1.45 eV for the single TFR mechanism. While fine-tuning improved energy errors in the high-energy regions, the reaction energy errors were not improved significantly. The reference DFT reaction energy is −0.647 eV, while MACE-MP-0b3 and MACE-MP-0b3-FT predicted −0.355 and −0.380 eV, respectively.

**Similarity statement**

The training set contains 5 structures composed of H, B, and C elements and 21 837 structures that have H or B or C along with any other elements. The dataset contains 2 structures containing icosahedral $C_2H_{11}B_{10}$ clusters linked by a C–C bond to form a dimer, 1 icosahedral borane $B_{12}H_{12}$ cluster, and 18 other borane clusters containing trigonal $B_3$ faces. Along with these, the dataset contains several hundred structures involving derivatives of borane and carborane clusters, such as salts, metal complexes,

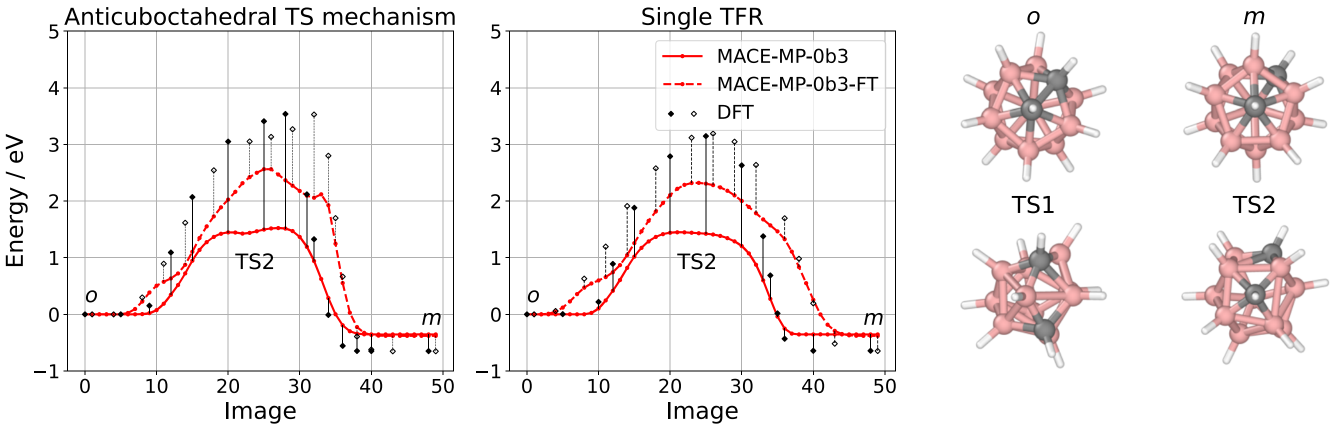


**FIG. 42.** MACE-MP-0b3 and MACE-MP-0b3-FT NEB minimum energy paths for the isomerization reaction of *ortho*-carborane to *meta*-carborane. DFT energies, evaluated on selected structures, are shown. All energies are presented relative to the *ortho* isomer energy. On the right, endpoint and transition state structures are illustrated.

and halogenated species. The closest structures in the training set are mp-1194548 ($C_4H_{22}B_{20}$ containing two icosahedral carborane clusters joined with a C–C bond) and metal complexes containing carborane ligands: mp-759303, 705569, 1199795, and 1198024.

**Performance summary**

Carborane rearrangement pathways found with MACE-MP-0b3 are qualitatively correct but do not lie on the reference DFT minimum energy path, as shown by the >1 eV energy errors relative to DFT in the high energy regions of the pathway.

## 26. Transition metal dichalcogenides

Edges in $MoS_2$ and transition metal dichalcogenides (TMDs) more broadly are known to be sites of high reactivity with much relevance to TMD-catalyzed reactions and material aging studies. Studies have recently addressed the structure of $MoS_2$ edges under various conditions using DFT.[267] Here, we examine stable Mo-edge and S-edge configurations with sulfur coverage $\theta_S = 0.5$ and $\theta_S = 1$ from Ref. 267. The edge models are multi-layer models, i.e., stacked

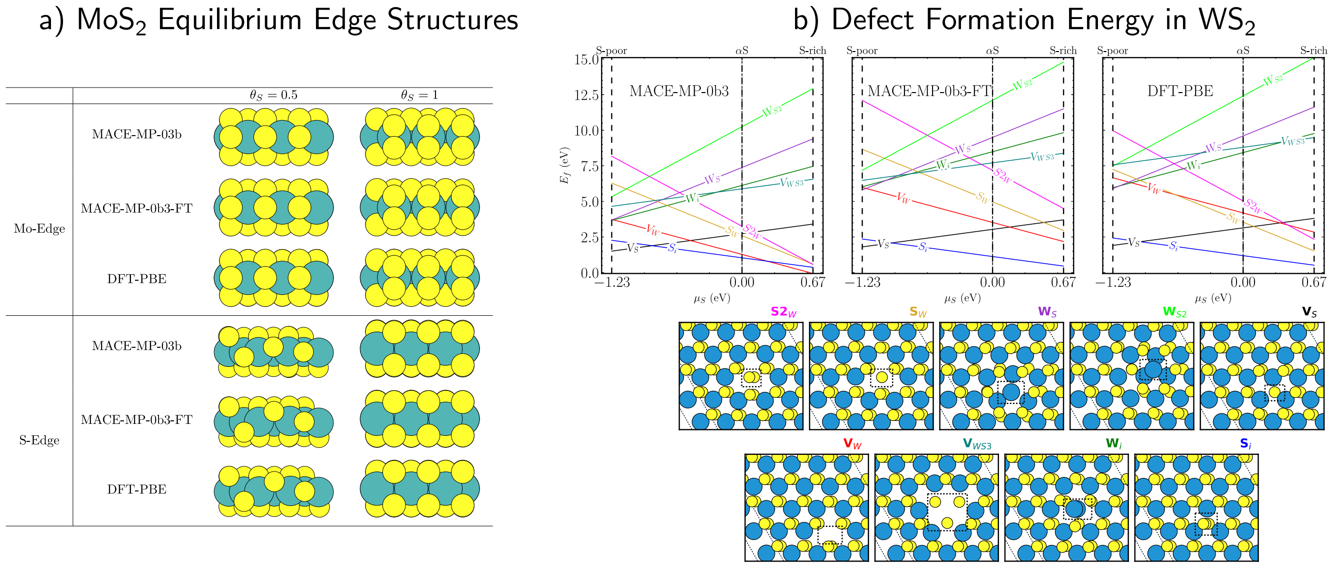


**FIG. 43.** Panel (a) shows four $MoS_2$ edge models of different types and sulfur coverage ($\theta_S$). DFT-predicted structures are obtained from Ref. 267. Geometry optimizations using MACE-MP-0b3 and MACE-MP-0b3-FT starting from DFT-optimized structures are shown. Panel (b) compares $WS_2$ defect formation energies predicted by MACE-MP-0b3, MACE-MP-0b3-FT, and DFT-PBE results from Ref. 268. Geometries of defects considered here as relaxed using MACE-MP-0b3 are also shown in panel (b).

infinite stripes. Figure 43(a) shows the $MoS_2$ edge structures considered. Geometry optimization performed on the DFT-optimized structures using MACE-MP-0b3 and MACE-MP-0b3-FT generally preserved the configurations. One notable deviation was observed for the MACE-MP-0b3 model prediction. The $\theta_S = 0.5$ S-edge configuration slightly deviated from the DFT-predicted zigzag structure, which was corrected following fine-tuning. To assess the extent of deviation, the $\theta_S = 0.5$ S-edge was further geometry-optimized using DFT, quickly converging in nine steps to the DFT minimum, which was 1.6 eV lower in energy. We further assessed the MD stability of MACE-MP-0b3 + D3 dispersion on these structures by running 100 ps of MD (NVT ensemble) at 300 K. This was also repeated using the fine-tuned model MACE-MP-0b3-FT. The MD simulations were found to be stable for the edge configurations above, using both models. Selected MD trajectories are available online.

Defects have a significant impact on the optical and electronic properties of 2D TMDs and come in various types, including vacancies. We assess the ability of MACE-MP-0 to describe defect formation energies $(E_f)$ of various defects in $WS_2$ as compared to PBE results from Ref. 268. The formation energy is calculated using the formula

$$E_{\mathrm{f}} = E_{\mathrm{defect}} - E_{\mathrm{pristine}} - \sum \Delta n_i \mu_i, \tag{A1}$$

where $E_{\mathrm{defect}}$ and $E_{\mathrm{pristine}}$ are energies of the $WS_2$ with and without defects and $n_i$ and $\mu_i$ are the number of atoms and chemical potential of element $i$. The chemical potential of S obeys the equilibrium condition $\mu_{\mathrm{WS}_2} = \mu_{\mathrm{W}} + 2\mu_{\mathrm{S}}$ and is bounded by predefined S-poor and S-rich conditions: $\mu_S^{\mathrm{bulk}_\mathrm{W}} \le \mu_S \le \frac{1}{2}\mu_S^{\mathrm{S}_2}$. It is calculated with respect to the $\alpha$-S as the reference state.[268] Figure 43(b) shows that $E_F$ values predicted by MACE-MP-0b3, MACE-MP-0b3-FT, and DFT are in qualitative agreement with regard to trend. MACE-MP-0b3-FT showed significant improvements in quantitative agreement with DFT-PBE, particularly for high-energy defects such as $V_{WS_3}$ and $V_W$.

**Fine-tuning**

We fine-tuned MACE-MP-0b3 on a collection of data comprising both $MoS_2$ and $WS_2$ structures. For $MoS_2$, 60 configurations were added in total. These comprise 15 snapshots from MACE-MP-0b3 MD runs of each of the four edge models. For $WS_2$, 45 configurations in total were added, which include three rattled copies of each of the defects and reference state configurations: pristine $WS_2$, $\alpha$-sulfur, tetragonal $WO_3$, body-centered cubic tungsten, and $S_2$ and $SO_2$.

**Similarity statement**

The majority of the Mo–S or W–S containing structures in the database include primitive units of layers of $MoS_2$ or $WS_2$. However, clusters of Mo–S and W–S were found as well. Edge models for $MoS_2$ are rare, with three structures identified (mp-990083, mp-989179, and mp-990086). These include a variant of the $\theta_S = 1$ S-edge and the $\theta_S = 0$ Mo-edge. No defect models of $WS_2$ were found in the dataset.

**Performance summary**

Geometric reconstruction of nanoribbon edges is mostly correct, apart from small deviations. The ordering of defect formation energies is qualitatively correct, with an overall tendency to underestimate precise values.

## 27. Electrode–electrolyte interface/battery system

Atomic-level interactions between the electrodes and electrolyte play a crucial role in determining the performance of electrochemical devices, including batteries, fuel cells, and electrocatalysts. Understanding these interactions is essential for optimizing the energy storage, conversion, and catalytic properties of these devices, and to this end, molecular modeling plays a crucial role. The remaining challenge is that processes underpinning transport and degradation in these devices take place on a long time scale, inaccessible to *ab initio* simulations. MLIPs are ideally suited to bridge this gap, bearing in mind that these complex materials and heterogeneous interfaces cover an extensive chemical space, which poses a big challenge to ML models. Here, we test the performance of MACE-MP-0b3 on two separate systems—pure (EC/EMC $LiPF_6$) electrolyte and the complete battery including the copper interface, anode, electrolyte, and NMC cathode (totaling nine chemical elements). In addition, we test a fine-tuned model, MACE-MP-0b3-FT, for the pure electrolyte. We performed MD simulations at 500 K using the models plus the D3 correction to stress-test the qualitative robustness of the potentials. Furthermore, we quantitatively assessed the potentials on a separate 200-config PBE test set in order to establish the accuracy of describing intra- and intermolecular interactions.

Previous studies[242,270,271] have shown that modeling even the neat solvent is a challenge to MLIPs owing to the weak, but crucially important, intermolecular interactions. Here, we find that MACE-MP-0b3 is stable in the NPT ensemble at 500 K; the density is preserved while the electrolyte (solvent + salt) remains liquid, and all molecules remain intact for the entire duration of the simulation (1 ns). A second NPT simulation was performed with MACE-MP-0b3-FT—a model fine-tuned on eight electrolyte and two pure solvent configurations with densities between 0.1 and 2.3 g $\mathrm{cm}^{-3}$ that were selected from the 200-config PBE dataset. The dynamics was found to be similarly robust with stable density and intact molecules; however, the densities and diffusivities found with the fine-tuned model are more physically realistic, meaning the electrolyte is less dense and significantly more diffusive, as expected at 500 K. The accuracy of both models was independently tested on the 200-config PBE test set of electrolytes spanning all physical compositions and densities (will be published elsewhere). The MACE-MP-0b3-FT model yields lower RMSEs on both intra- and intermolecular properties, demonstrating that fine-tuning can significantly specialize the potential with a minimal number of new training data points (ten configs). That MACE-MP-0b3-FT reproduces intermolecular interactions better than MACE-MP-0b3 is also clearly illustrated on the rigid-molecule volume scans[242] test. MACE-MP-0b3-FT accurately describes the PES as compared to PBE for both the neat solvent and the full electrolyte. Note that the potential correctly captures the $1/r$ Coulomb scaling of the energy up to large volume (low density), despite its short-range nature.

The full battery was simulated by MACE-MP-0b3 in the NVT ensemble, since the volume of the entire system was not stable in NPT simulations (possibly the result of large compressibility differences along the x-direction normal to the liquid electrolyte layer vs the y/z-directions along the solid electrode slabs). All NVT simulations were stable at 500 K for the entire simulation time (1 ns). The electrolyte Li-ions were found to deintercalate from both the graphite anode and the NMC cathode, and the electrolyte was mobile. The H-capped graphite was found to be inert, whereas the cathode–electrolyte interface exhibited pronounced reactivity. Evident from the start of the simulation was the extensive proton transfer from the carbonate solvent (EMC in particular) to the oxygen atoms in NMC. This, in turn, led to the continuous breakdown of the solvent molecule (which became a radical) and chemisorption onto the cathode surface, possibly demonstrating the initial steps of SEI formation. Notably, substantial amounts of $CO_2$ and $H_2O$ were generated in the process. Furthermore, oxygen atoms were easily extracted from the cathode, leaving behind binding sites for the oxygen-rich carbonate molecules. A separate simulation setting was tested where the exposed oxygen atoms of the cathode were hydrogenated before the simulation. Similar reactivity was observed, albeit with different outcomes; notably, more water molecules and less carbon dioxide were generated in the process. These early simulations demonstrate MACE-MP-0b3 is robust for battery interfaces and showcase the initial steps in modeling the SEI formation with *ab initio* accuracy—which has been a long-held dream of the scientific community.

**Similarity statement**

To perform the similarity analysis, 100 representative (decorrelated) structures were taken from the previously described simulations. While the MP dataset does not contain liquids, for all three simulations, the UMAP analysis showed that all atomic environments were well represented in the MP dataset. More specifically, the pure electrolyte was found to contain environments close to mp-995234 and mp-995218, which correspond to HCO and $H_4C_5O_2$, respectively. These configurations were also found to be similar for the snapshots obtained by running the electrolyte–anode interface and also notably included mp-707412 corresponding to $H_{22}C_{10}O_3$. For the entire battery system configurations, such as mp-1194779 and mp-698267, were found to be similar, which correspond to $CuH_3C_3O_4$ and $CoHCO_3$, respectively. We provide

- `interface_chemiscope_input.json`
- `battery_chemiscope_input.json`
- `electrolyte_chemiscope_input.json`

to help visualize the atomic environments against the MP dataset at https://chemiscope.org/.

**Performance summary**

Fine-tuning considerably improved solvent and electrolyte property (density, diffusivity) prediction. Full battery simulations were stable at fixed volume but unstable with variable volume and showed electrolyte reactions at the unpassivated electrode.

## 28. Metal–organic frameworks

### *a. QMOF*

Given that MACE-MP-0 is pretrained against PBE, whereas QMOF was constructed from PBE-D3(BJ) calculations to account for dispersion corrections,[29,151] in the comparison with QMOF, we

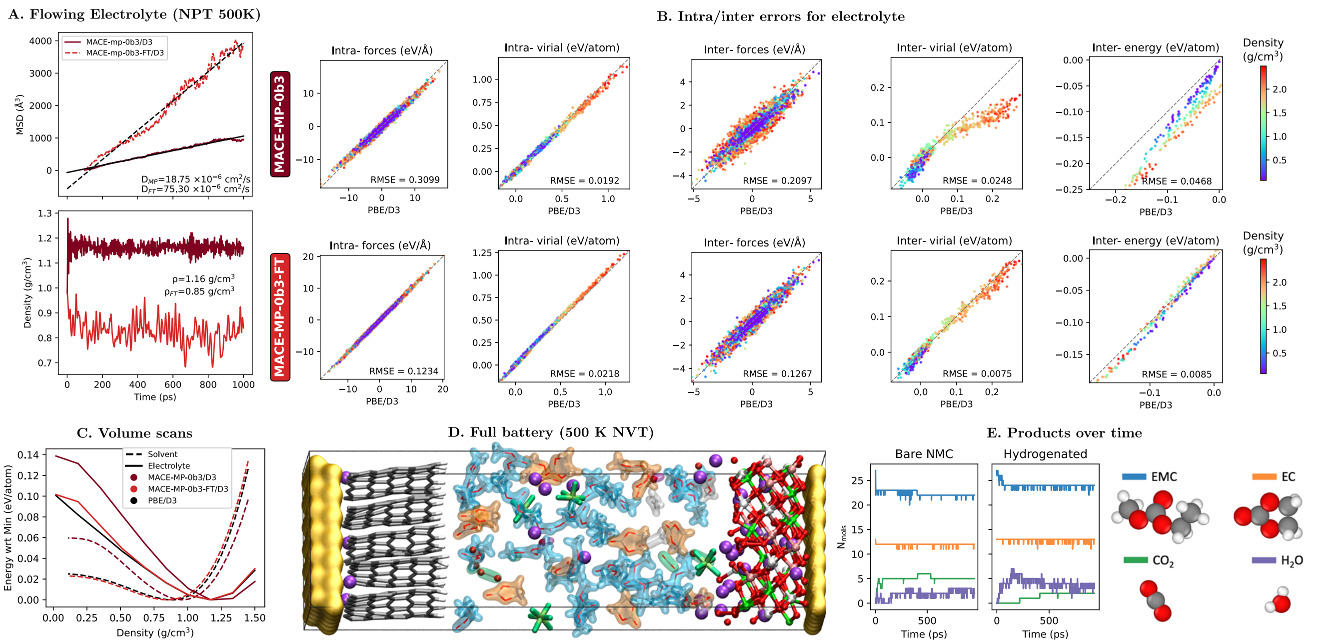
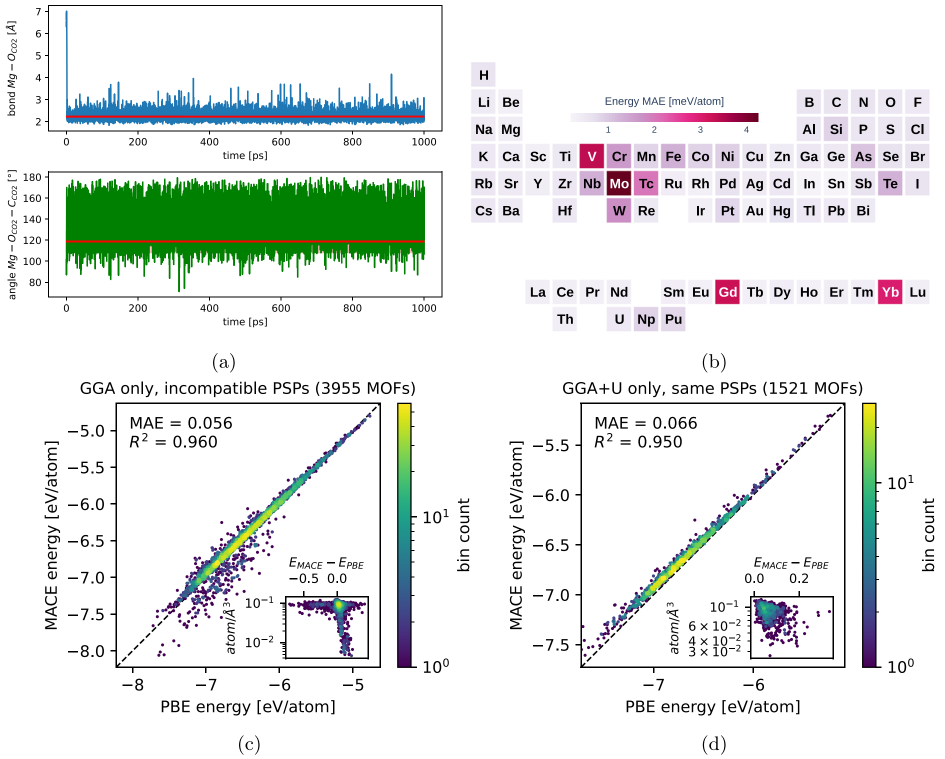


**FIG. 44.** (a) Flowing EC/EMC $LiPF_6$ liquid electrolyte with stable density and intact molecules at 500 K NPT MD. (b) MACE-MP-0b3 and MACE-MP-0b3-FT intra-/inter-molecular energy/forces/virial evaluated on an independent PBE test set spanning all possible compositions of EC/EMC $LiPF_6$ electrolyte at densities between 0.1 and 2.5 g $cm^{-3}$. Note that the 10 FT configurations were randomly selected from this set of 200 configurations. (c) Rigid-molecule volume-scan test[242] for both neat EC/EMC solvent and full EC/EMC $LiPF_6$ electrolyte, compared to PBE. (d) Full battery simulation (Cu | H-capped graphite+Li | EC/EMC+$LiPF_6$ | NMC+Li[269]), final snapshot of 500 K MACE-MP-0b3 NVT-MD, showing degraded solvent (gray iso-surface molecules), new $CO_2$ (green iso-surface) and $H_2O$ (purple iso-surface) molecules, and oxygen atoms originating from the cathode floating in the electrolyte (solid red spheres). (e) Time progression of the predominant molecular species in the electrolyte for the two NVT simulation settings (1. neat NMC, 2. H-capped NMC).



(a) (b)

(c) (d)

**FIG. 45.** (a) Top panel: Instantaneous distance between the Mg center and the closest oxygen from $CO_2$ in Mg-MOF-74; the red line shows the predicted average value from a previous study.[71] Bottom panel: Instantaneous angle between the Mg center, the closest oxygen, and the carbon in the $CO_2$, the red line indicates an average value from a previous study from Ref. 71. (b) Elemental dependence of energy error distribution over the periodic table on the complete QMOF dataset (20 375 relaxed structures),[69,70] correlating with incompatible pseudopotentials used between the MPtrj training and QMOF test sets (Table III). The absolute error of total energy was equally distributed to all atoms in the single MOF structure, and the average of each absolute error across different MOF structures was averaged for each element. (c) MACE-MP-0b3 energy predictions compared to QMOF GGA-PBE calculations using pseudopotentials (PSPs) incompatible with the MPtrj. Structures that would be calculated at the GGA-PBE+U level by MP standard, i.e., structures containing oxygen (or fluorine) and any element of V, Cr, Mn, Fe, Co, Ni, W, or Mo, are excluded. (d) MACE-MP-0b3 energy predictions compared to QMOF GGA-PBE calculations on *only* the structures that would be calculated at the GGA-PBE+U level by MP standard. QMOF structures calculated with incompatible PSPs are excluded.

compare MACE-MP-0 predicted energies with QMOF PBE energies by subtracting the dispersion correction from total QMOF energies.

We noted that most of the high-energy error MOF structures have high atomic density [Fig. 45(c) inset], and that the errors cannot be canceled by adding dispersion correction. In Fig. 45(b), we analyze the element-wise energy error per atom by distributing the absolute energy error per atom to the constituent elements by the corresponding composition in each MOF structure. As presented in Fig. 45(b), there is a strong elemental dependence of energy error per atom. Most of the high error elements can be attributed to the difference in the chosen pseudopotentials used by the MP and QMOF databases (in particular, the choices of which electrons are treated as valence states); see Table III.

Figures 45(c) and 45(d) present the incompatible QMOF entries excluded from the analysis in Fig. 3(a). The incompatibility stems from (i) different, incompatible pseudopotentials (PSPs), or (ii) mismatched calculation parameters (with or without Hubbard U correction) used in the QMOF and MP databases, or the combinations of both. Figure 45(c) compares MACE energies with QMOF PBE energies calculated with incompatible PSPs (case i). Figure 45(d) compares MACE energies with QMOF PBE calculations on structures that would be calculated at the PBE+U level by MP standard at the curation time of MPtrj.

**TABLE III.** Difference in VASP POTCARs used by MPtrj and QMOF.

| MPtrj | Be | Bi | Cu_pv | Eu | Fe_pv | Gd | Li_sv | Mg_pv | Mo_pv | Nb_pv | Ni_pv | Os_pv | Re_pv | Ti_pv | V_pv | W_pv | Yb_2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| QMOF | Be_v | Bi_d | Cu | Eu_3 | Fe | Gd_3 | Li | Mg | Mo_sv | Nb_sv | Ni | Os | Re | Ti_sv | V_sv | W_sv | Yb_3 |

#### *b. $CO_2$ adsorption*

All the calculations for the $CO_2$ dynamics with MOFs were performed with MACE-MP-0 by adding the D3 dispersion correction[29] to the MACE-MP-0 potential. The simulations were carried out with ASE[272] on a cell containing 165 atoms with one $CO_2$ molecule, initialized at the center of the pore, using NVT Langevin dynamics[273] with a friction factor of $5 \times 10^{-3}$ fs$^{-1}$. The temperature was set to 600 K, with a time step of 1 fs. Twenty-four 1 ns trajectories were generated using different initial velocities, and all quantities presented were averaged over all of them, discarding the first 2 ps from each to account for equilibration. All structures had their cells and positions optimized at the start using `FrechetCellFilter` from ASE. The code used to generate the trajectories is available in Refs. 206 and 274.

**Similarity statement**

The MP dataset does not contain MOFs. There are eight structures containing all MgOCH elements and 62 structures that have MgOHC elements on their own or along with other elements. Based on UMAP analysis, we see that most atomic environments, both MgO and linkers, in the example system, are similar to environments in the training data, but none is Mg-MOF-74 specific.

The closest (most relevant) structures in the training set are $CO_2$ (mp-556034, mp-20066, mp-995224, mp-11725, mp-644607, mp-1102227, mp-1190685, mp-995198, mp-1190699, mp-1077906, mp-1077316, and mp-729728). $CO_2$ alone matches 4896 structures with C, O and alongside other elements.

We provide

- `co2_FilterType.exclusive_OC_chemiscope_input.json`
- `mg-mof-74-co2_FilterType.exclusive_MgOCH_chemiscope_input.json`
- `mg-mof-74-co2_FilterType.exclusive_MgOCH_chemiscope_input.json`that contain exact matches of Mg, O, C, and H and the inclusive versions
- `mg-mof-74-co2_FilterType.exclusive_MgOCH_chemiscope_input.json`
- `mg-mof-74-co2_FilterType.inclusive_MgOCH_chemiscope_input.json`
- `mg-mof-74_FilterType.inclusive_MgOCH_chemiscope_input.json`

to help visualize the interactive UMAP on https://chemiscope.org/.

**Performance summary**

Excellent energy prediction for a large database of MOFs. Correct prediction of binding structure and free energy of $CO_2$ in Mg-MOF-74.

### 29. Combinatorial materials discovery

#### *a. Formation energy of hypothetical materials*

MACE-MP-0b3, trained on the MPtrj dataset, and MACE-MPA-0, trained on the combined MPtrj and sAlex[111] datasets, generalize to out-of-distribution (OOD) chemistries, as shown by their performance on the WBM dataset,[275] which was generated using elemental substitutions drawn according to a data-mined chemical similarity measure.[276] The initial set of 9524 structures for substitution was taken from the Materials Project convex hull. The substituted structures were relaxed using the `MPRelaxSet` PBE DFT workflow. After this, the convex hull was recalculated using the new structures, and then subsequent rounds of substitution were carried out on the new structures that ended up on the combined MP plus growing WBM convex hull. In total, five rounds of substitutions were carried out, yielding a dataset of 257 487 inorganic crystals that are OOD with respect to MP and, therefore, well suited to benchmarking. For this investigation, we use the cleaned version of the WBM dataset released in Matbench Discovery (MBD),[277] which first discards 524 crystals with unphysical or missing labels. Subsequently, all structures in WBM with composition + prototype matching a structure in MP are removed. Within the WBM composition + prototype, duplicates are dropped, leaving only the lowest energy structure. The final test set consists of 215 488 materials, and of these, we obtained MACE-MP-0b3 relaxed structures for 215 488 and 215 486 for MACE-MPA-0.

Following the MBD protocol, we use MACE-MP-0b3 to relax the initial substituted structures and compare these predictions against the ground truth formation energy calculated with DFT. The predictions on the MACE-MP-0b3 self-relaxed structures result in an MAE of 58 meV/atom, as shown in Fig. 46. MACE-MPA-0 cuts this error more than in half, achieving a formation energy MAE of 28 meV/atom. Similarly, MACE-MPA-0's $R^2$ improves upon MACE-MP-0b3's $R^2$, going from 0.987 to 0.993. However, both models' energy error exhibits a systematic tendency to underpredict the PBE formation energy, as seen in the least-squares fits of $y = 0.94x - 0.19$ for MACE-MP-0b3 and $y = 0.9x - 0.22$ for MACE-MPA-0. Interestingly, despite the lower error and higher $R^2$, MACE-MPA-0's error is more systematic since the slope and intercept are further from 1 to 0, respectively. Therefore, a simple affine fit to correct for the systematic component in the energy error would reduce MACE-MPA-0's already lower error more than MACE-MP-0b3's. While Alexandria is overall a highly valuable training set, this result is consistent with the observation that it is even more biased toward low-energy structures than MPtrj.

The OOD nature of the WBM dataset can be seen in the increase in MAE between WBM batches from 48 meV/atom for the first batch to 78 meV/atom for the final batch as the structures become increasingly dissimilar to MP due to accumulated substitutions. When these predictions are used to attempt to classify whether the structure lies above or below the MP convex hull training set, MACE-MP-0b3 achieves an F1 score of 0.66 and a discovery acceleration factor (DAF) of 3.95 compared to an F1 of 0.85 and a DAF of 5.58 for MACE-MPA-0. The DAF is the ratio of the precision (TP/PP) to the prevalence (P/N) of the test set. Here, TP = True Positives, PP = Predicted Positives, P = Total Positives, and N is the test set size. These results show that MACE-MP-0b3 and especially MACE-MPA-0 can extrapolate to novel chemistries and are well-suited to high-throughput materials discovery.

**Similarity statement**

By construction, there is no overlap in terms of both composition and prototype together between MP and the WBM test set studied here. However, there is an overlap for both compositions and prototypes separately. MP contains 105 583 unique reduced formulae, while the WBM test set contains 160 055. Of these 15 782 overlap MPs, all instances are examples of different prototype structures. MP contains 32 933 different isopointal prototypes, while the

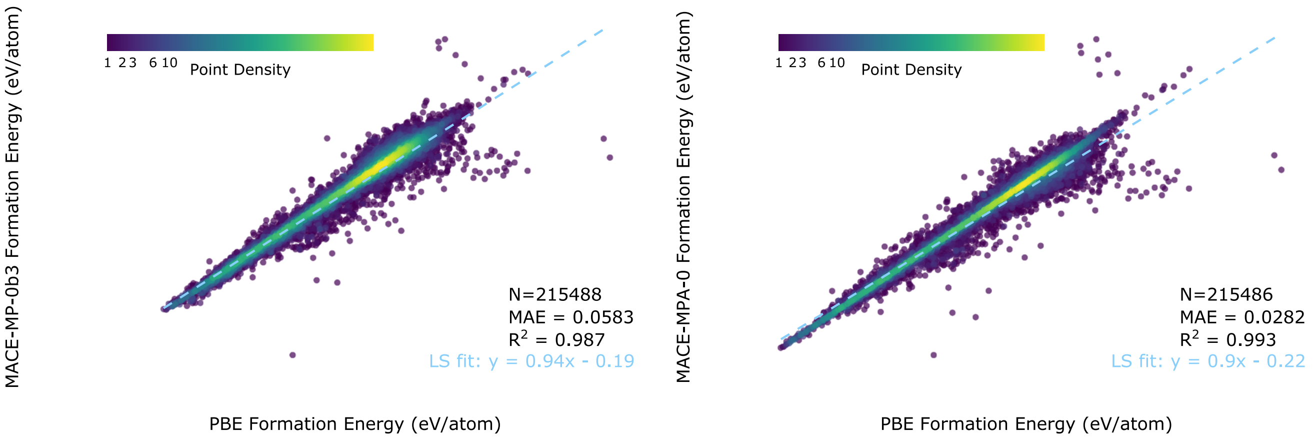


**FIG. 46.** Formation energy parity plot showing the difference between the DFT-relaxed energy and the MACE-MP-0b3-relaxed energy (left) and the MACE-MPA-0-relaxed energy (right) starting from WBM initial structures.

WBM test set contains just 2816. Of the prototypes found in WBM 1813, some are also found in MP. Of the isopointal prototypes not seen in MP, the five most common are (occurrences in parentheses): ABC2_oI8_71_a_b_f,[317] ABC2_hP12_181_c_d_i,[213] AB2_hP9_189_f_adg,[154] ABC2_oI8_44_a_b_c,[117] and AB4C6_mC22_8_a_2 ab_2a2b.[106] These arise due to changes in symmetry during the relaxation of the substituted structures.

### *b. Stoichiometric substitutions*

To test the interpolative and extrapolative capabilities of the model within MP chemistry, exhaustive element substitutions for 100 randomly selected compositions were performed. Specifically, MP was first filtered to remove all compositions with more than 16 atoms in the reduced formula to ensure that there were a sufficient number of possible substitutions. This set was randomly split into host and target compositions with a 60:40 ratio. 100 compositions were randomly drawn from the target split and substituted into each stoichiometrically matching host, yielding 154 685 substituted structures. These were optimized with MACE-MP-0 using full unit cell relaxations (without the D3 correction). In Fig. 47, the results are shown in terms of energies and volumes relative to the most stable structure of that composition within the MP. The distribution is sharply peaked at $\frac{V}{V_{\mathrm{ref}}} = 1$ and $E - E_{\mathrm{ref}} = 0$, indicating that the substituted cells often relax back to the known ground state structure from MP. This makes MACE-MP-0 suitable for predicting the crystal structures of unknown materials. For a random sample of 16 126 structures, the MACE-MP-0 relative energies were validated with MP-compatible PBE DFT calculations, yielding an RMSE of 0.097 eV/atom.

Beyond recovering the MP ground state, the wide range of substitutions tested also yields a large number of alternative structures

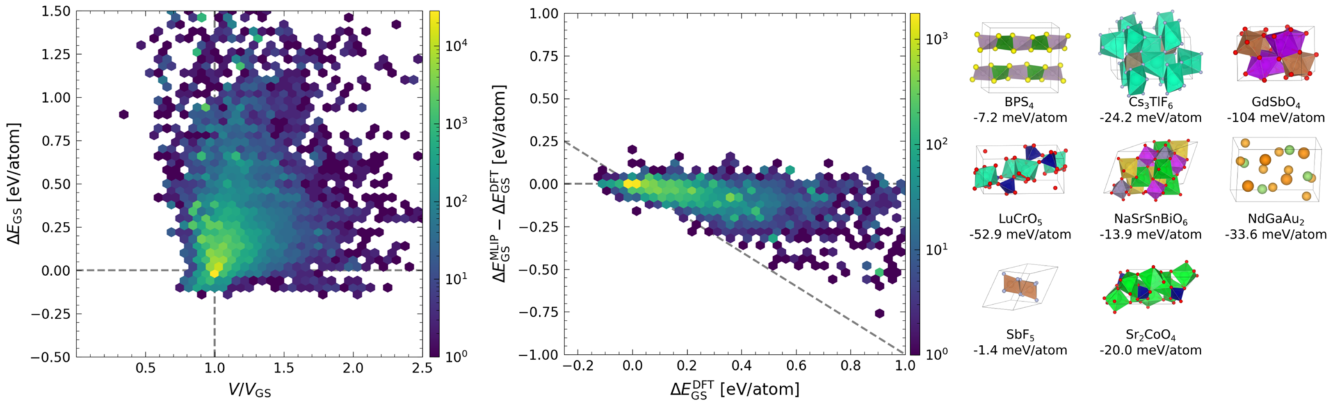


**FIG. 47.** Left: Relative energies vs volumes of ~150k relaxed structures generated via exhaustive element substitutions for 100 MP compositions. Center: Error of relative energies calculated with PBE and MACE-MP-0b3 for 6909 randomly drawn structures from the left plot, as a function of the PBE relative energy. Right: DFT validated examples of newly discovered stable phases, along with their energies relative to the MP convex hull.

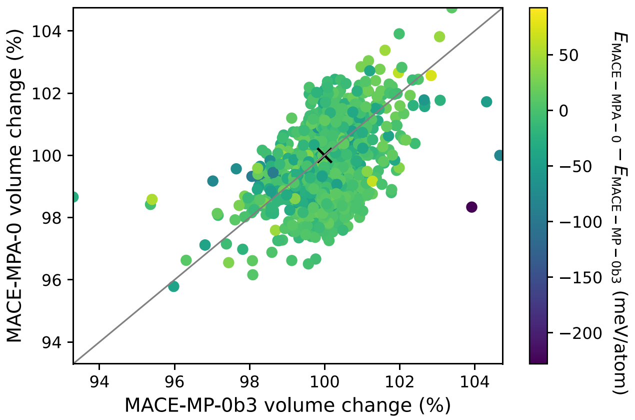


**FIG. 48.** Comparison of the percentage volume change under relaxation relative to the initial structure volume as taken from the GNoME dataset. MACE-MP-0b3 is plotted on the horizontal axis, and MACE-MPA-0 on the vertical. The line indicates perfect agreement between both MACE models, and the "x" indicates perfect agreement with the GNoME volume. We observe a slight tendency of MACE-MP-0b3 to predict larger volume changes and, therefore, larger volumes than MACE-MPA-0. The points are colored by the difference in energy between MACE-MPA-0 and MACE-MP-0b3; therefore, darker colors indicate MACE-MP-0b3 predicts a lower total energy than MACE-MPA-0, and brighter colors indicate that MACE-MP-0b3 predicts a higher total energy than MACE-MPA-0.

for each composition. 3522 of these have relative energies of zero or lower and are thus predicted to be more stable than the MP reference. To validate these predictions, DFT relaxations were performed for all structures with negative relative energies. The results confirm that 2724 of them are indeed more stable than the corresponding MP reference. Importantly, not each of these is a new stable phase, as some relaxations converged to the same minimum and multiple lower energy structures are found for some compositions. Nonetheless, for eight of the 100 compositions considered here, new structures with energies below the current MP hull were discovered (for $BPS_4$, $Cs_3TlF_6$, $GdSbO_4$, $LuCrO_5$, $NaSrSnBiO_6$, $NdGaAu_2$, $SbF_5$, and $Sr_2CoO_4$). These constitute genuine predictions of new, thermodynamically stable phases relative to the MP convex hull. These phases were also not reported in the recent GNoME effort (which focused on compositions not included in MP). However, the WBM dataset does report slightly more stable structures for $Cs_3TlF_6$ and $SbF_5$. It also includes structures for $NdGaAu_2$, which are less stable than the one reported herein.

**Similarity statement**

The MP contains at least one (e.g., $NaNiIO_6$ and $NaSrSnBiO_6$) and at most 423 ($Sr_2CoO_4$) structures that contain all elements in the discovered stable phases. These matches are similar to the discovered structures, as host and reference structures are present in the MP. Yet, the discovered phases are unique and by definition not part of the training set. For each discovery, the closest structures in the training set are mp-1106139 ($Bi_3Pb$), mp-559695 ($Cs_3BiF_6$), mp-561827 ($Cs_3TlF_6$), mp-1211553 ($LuCrO_5$), mp-545399 ($NaNiIO_6$), mp-1522253 ($NaSrSnBiO_6$), mp-1220399 ($NdGaAu_2$), mp-1100068 ($Sr_2CoO_4$), and mp-775149 ($TiS_2O_8$). We provide json files to visualize interactive UMAPs on https://chemiscope.org/.

### *c. Analysis of highly coordinated theoretical structures*

The recent GNoME database[99] published ~381 000 stable materials that could not be matched to any experimental structure in the Inorganic Crystal Structure Database (ICSD)[278] nor any public repository of theoretical structures such as the Open Quantum Materials Database (OQMD) or Materials Project. An analysis of the local coordination environment of these structures using the `CrystalNN` algorithm[279] reveals 21 300 structures with a maximum coordination number greater than 16.

To evaluate the performance of MACE-MP-0b3 and MACE-MPA-0 against the GNoME potential that generated these structures, we have re-relaxed 1199 structures with a predicted maximum coordination number greater than 20. In addition, to compute formation energies, we re-relaxed all elemental structures in the Materials Project identified to be the lowest energy configuration from PBE/ + $U$ calculations. The `FrechetCellFilter` class in the Atomic Simulation Environment (ASE)[272] was used to relax structures until the maximum (absolute) Cartesian component of any inter-atomic force was less than $2 \times 10^{-2}$ eV/Å. This force convergence setting is consistent with the most stringent force relaxation criterion used by the Materials Project. No dispersion correction was used to augment the MACE models, consistent with the lack of a dispersion correction in GNoME. All relaxations ran successfully on the central processing unit (CPU) resources.

The MACE-relaxed structures are highly similar to the GNoME ones. 48 plots the percentage change in volume

$$\delta V = (100\%) \times \frac{V_{\mathrm{MACE}}}{V_{\mathrm{GNoME}}}, \tag{A2}$$

for both MACE-MP-0b3 and MACE-MPA-0. MACE-MPA-0 tends to predict smaller volumes than MACE-MP-0b3, with a mean absolute deviation (MAD) from the GNoME structure volumes of 0.117 Å$^3$/atom, compared to the 0.085 Å$^3$/atom MAD of MACE-MP-0b3. However, their energies (indicated by the colors of the points) are nearly the same in all cases.

In virtually all cases, the MACE models relax structures to an indistinguishable configuration from their GNoME starting point. Only 14 (1) structures undergo an energy lowering of greater than 10 meV/atom during relaxation with MACE-MP-0b3 (MACE-MPA-0). 15 (14) structures underwent an increase of symmetry under relaxation with MACE-MP-0b3 (MACE-MPA-0), as measured by their space group number. Only one structure decreased in symmetry when relaxed with MACE-MP-0b3. The maximum coordination number decreased for 82[87] structures under relaxation with MACE-MP-0b3 (MACE-MPA-0) and stayed the same otherwise. Figure 49 presents these observations for both MACE models.

Finally, we consider the formation enthalpies, defined as

$$\delta H_f = \frac{1}{\sum_i c_i}\left[H_{\mathrm{solid}} - \sum_i c_i \mu_i\right], \tag{A3}$$

the total enthalpy of a solid $H_{\mathrm{solid}}$ minus the chemical potentials of its elemental constituents $\mu_i$ weighted by their stoichiometry $c_i$ in the structure. Note that $\sum_i c_i$ is the total number of atoms in the computational cell of the structure. After relaxation, the enthalpies and chemical potentials in Eq. (A3) should be well-approximated by

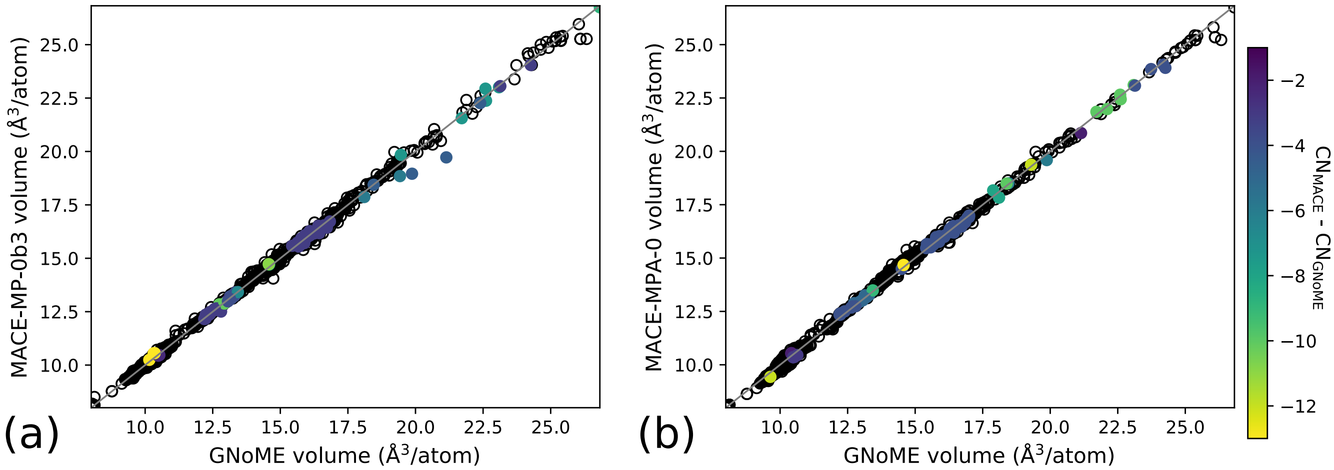


**FIG. 49.** Final volume of the highly coordinated GNoME materials before (horizontal) and after relaxation with a MACE model (vertical). Panel (a) plots the MACE-MP-0B3 model results, and (b) the MACE-MPA-0 results. The points are colored by their change in maximum observed coordination number as measured by `CrystalNN`:[279] a hollow circle indicates no change. Nonzero changes in the maximum coordination number, suggesting a decrease in dense bond patterns, are indicated by increasingly light colors for larger magnitude changes. In all cases, the maximum coordination number was observed to decrease or stay the same.

energies. Figure 50 presents the distribution of formation energies of the GNoME structures. Two distinct peaks can be observed, one for the intermetallics and covalently bound structures between −1 and 0 eV/atom, and one for the oxides below −1 eV/atom. In addition, note that no materials were found to be unstable as indicated by a positive formation energy.

Cumulatively, these observations suggest that both MACE models perform comparably to the GNoME UIP for these highly coordinated structures and yield equilibrated structures of equal or higher symmetry than the GNoME UIP.

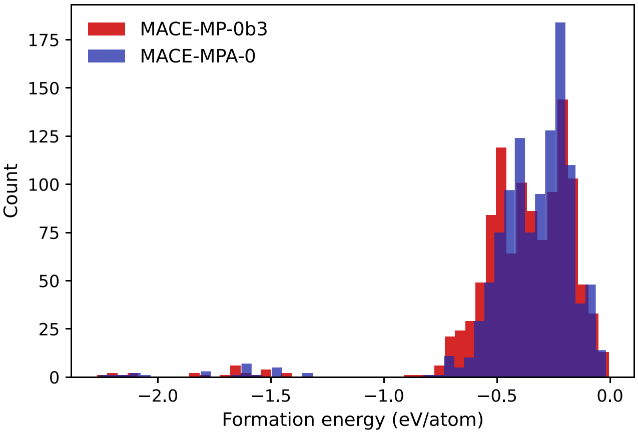


**FIG. 50.** Distribution of formation energies (zero-pressure limit of the formation enthalpies at zero temperature) of the highly coordinated GNoME structures after relaxation with MACE-MP-0b3 (red) and MACE-MPA-0 (transparent purple). Note the dual peaks in formation energies: between −1 and 0 eV/atom, most structures are intermetallics or covalently bound; below −1 eV/atom, most structures are oxides.

**Similarity statement**

By construction, the materials in the GNoME set do not exist in the MP dataset. The GNoME set was constructed in part by substituting elements on structures that originated in the Materials Project. However, some of the GNoME materials exist in the same chemical space as those in MP. For the 1199 GNoME materials considered here, 131 (10.9%) contain exactly the same elements as structures in MP. The chemical space of maximum overlap, Cu–La–Zn, contains four GNoME (7ee54b7a37, 3422b9acb0, f36439bb45, and 68a5e7535f) and two MPtrj (mp-1223296 and mp-1093834) materials. We provide two JSON-format dictionaries, `MPtraj_chem_env.json.gz` and `GNoME_chem_env_w_ovlp.json.gz`, which tabulate the chemical environments spanned by MP and the GNoME subset, respectively. The GNoME subset is further categorized by overlap with MP.

**Performance summary**

Excellent energy prediction of WBM and GNoME hypothetical materials, and able to newly discover stable materials as validated by DFT.

## 30. Alanine tripeptide free energy surface

In this section, we investigate the ability of MACE-MP-0b3 to construct the free energy surface of a simple peptide. We simulate a periodic box containing an alanine tripeptide solvated with explicit water for 1 ns without D3 correction. Sampling of the central backbone torsions was accelerated by well-tempered metadynamics, as implemented in OpenMM, using an initial Gaussian height of 1 kJ/mol and a temperature scaling factor of 10. MACE-MP-0b3 only approximately identifies the features of the free energy landscape, while MACE-MP-0b3-FT identifies the positions of the minima in better qualitative agreement with the classical force field. MACE-MP-0b3-FT also more accurately predicts the relative population of the minima. However, neither model fully resolves the $a\beta$

and PPII minima around ($\phi = -100^\circ$ and $\psi = 150^\circ$). MACE-MP-0b3 in particular overpopulates the $\alpha$-H(L) basin at $\phi = 80^\circ$ and $\psi = -100^\circ$.

**Similarity statement**

The MP dataset contains 99 structures containing exclusively the elements H, C, O, and N. Based on UMAP analysis, we observe that all atomic environments in the example system are similar to environments in the training set. On closer inspection, we find that the most similar environments to the majority of the example configurations are clusters of water, ammonia, and $NO_2$ molecules. Several examples contain oxygen-bearing molecules near water molecules, allowing for sampling of hydrogen bonding. We also find configurations containing clusters of large aromatic compounds containing carbon, nitrogen, oxygen, and hydrogen. Interestingly, we also find a large cluster containing diverse configurations of small C,H,N,O-containing fragments, including carboxyl and amide fragments, e.g., mp-997182.

**Performance summary**

Stable MD at ambient conditions, but incorrect free energy of conformers.

## 31. Molecule–surface interactions

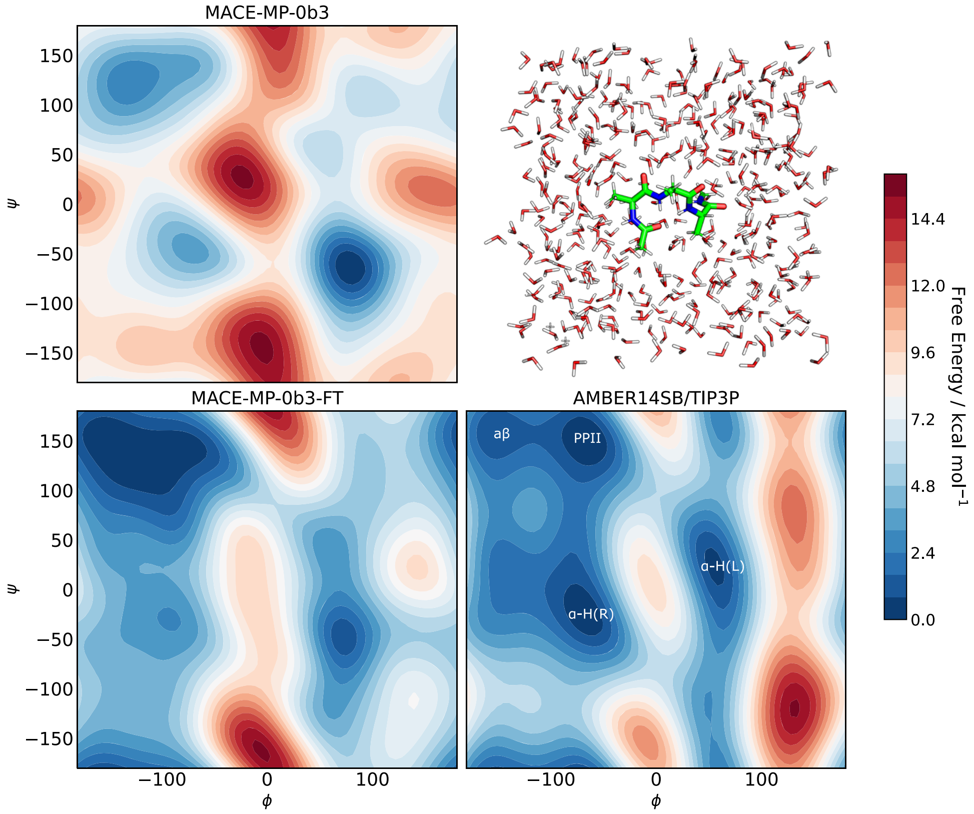
### *a. Adsorption energies—S24 dataset*

Describing the interaction between a molecule and a surface with first-principles accuracy is central to designing new and improved materials for heterogeneous catalysis, gas storage and separation, and many more.[44] Here, we test the accuracy of MACE-MP-0 augmented with the Becke–Johnson D3 dispersion correction for a set of prototypical systems found in surface chemistry, encompassing metallic, covalent, and ionic-bonded surfaces, together with porous metal-organic frameworks (MOFs) and zeolites. The structures used within this study—dubbed the S24 dataset—were taken from an amalgamation of published[280–290] and unpublished studies. Figure 52 summarizes the computed PBE-D3(BJ) adsorption energies using PYMATGEN to generate VASP[4,291,292] inputs with the `MPRelaxSet` settings.

We summarize the overall MAD as well as the MAD within the subcategories of surfaces for MACE-MP-0b3+D3 in Table IV. Performance is moderately good, with an overall MAD of 305 meV, in line with the errors observed from a typical DFT functional. Performance across the various sub-categories of surfaces is relatively uniform, with the best performance for zeolites with an MAD of 229 meV. It performs poorly for $H_2$ on Si(100), with errors exceeding 600 meV.

**Similarity statement**

The S24 dataset contains 72 structures comprising 24 adsorbates, surfaces, and adsorbate–surface combinations. Based on the UMAP analysis, the elemental compositions of the adsorbates occur 6–27 times, the surfaces 0–9043 times, and the adsorbate–surface combinations 0–4896 times on their own or along with other elements. Importantly, the training dataset does not contain any


MACE-MP-0b3

MACE-MP-0b3-FT

AMBER14SB/TIP3P

**FIG. 51.** Free energy surfaces constructed from 1 ns sampling with well-tempered metadynamics. Enhanced sampling was performed on the central $\psi$ and $\phi$ backbone angles.

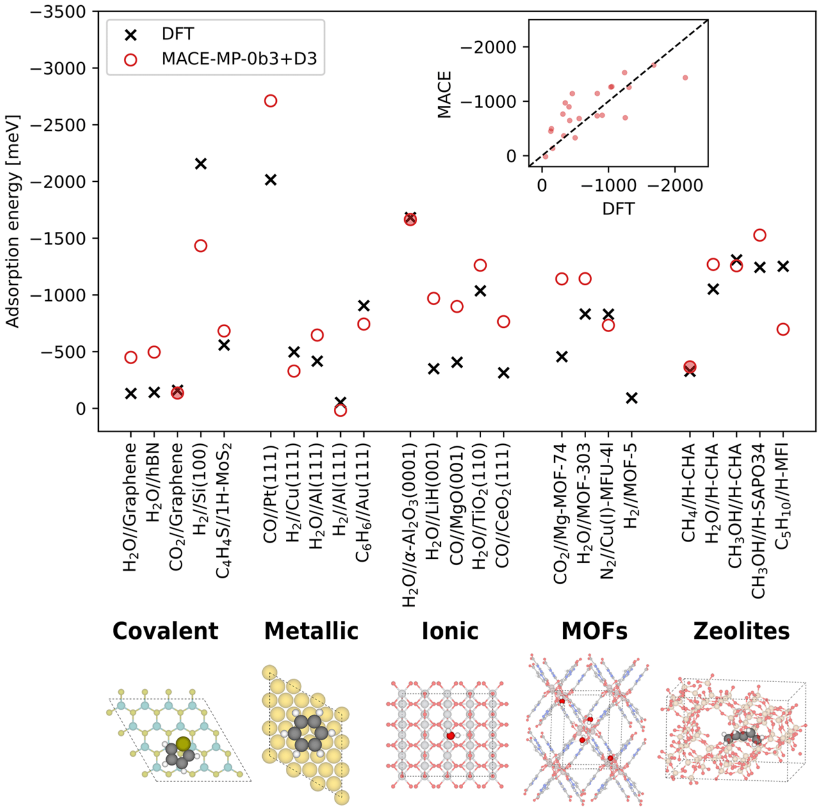


**FIG. 52.** Comparison between DFT (PBE+D3(BJ), red crosses) and MACE-MP-0b3+D3 (blue circles) adsorption energies calculated for a diverse set of surfaces. These surfaces consist of covalent, metallic, and ionic bonds as well as the porous material classes: MOFs and zeolites. Filled circles indicate surfaces where the MACE model reaches chemical accuracy (43 meV) agreement with DFT. The inset shows that there is a good correlation between MACE-MP-0b3+D3 and the DFT references.

gas-phase molecules, surface-truncated models, or MOFs. For bare surfaces, on average, similar element compositions occur in the training data as follows: 3997 covalent, 5805 metallic, 2859 ionic, 46 MOFs, and 88 zeolites. For adsorbate–surface combinations, on average, the number of similar element compositions occurring in the training data is lowered to 1492 covalent, 340 metallic, 499 ionic, 46 MOFs, and 59 zeolites. There is no clear correlation between the number of similar training data for a surface type and the accuracy of MACE-MP-0b3+D3.

**Performance summary**

Moderately good agreement is generally achieved for adsorption energies, with improved performance for adsorption onto zeolites for MACE-MP-0b3+D3. Performance appears to become poorer with dissociative chemisorption of $H_2$ on Si(100), albeit still reproducing the adsorption energy to the correct order of magnitude.

**TABLE IV.** Mean absolute deviation in meV for the S24 dataset and its subcategories of surfaces.

| | MACE-MP-0b3+D3 |
|---|---|
| Covalent | 308 |
| Ionic | 361 |
| MOF | 380 |
| Metal | 265 |
| Overall | 305 |
| Zeolite | 229 |

### *b. Relative energies—OC157 dataset*

Identifying the most stable structures for a molecule–surface system is pivotal toward predicting the activity and selectivity of a catalyst, facilitating rational design of new catalyst materials.[293] We compare MACE-MP-0b3+D3 and the model following fine-tuning against `MPRelaxSet` DFT at predicting the relative energies between three structures for 157 molecule–surface combinations. The fine-tuned (FT) models were trained on a dataset containing one high-energy structure (not from the test set) for each of the 157 molecule–surface combinations. These surfaces were taken from the Open Catalyst Challenge 2023,[58,294] using structures generated by the baseline EquiformerV2 model trained on the OC20–S2EF-2M dataset. While 200 molecule–surface combinations were originally provided, we have excluded those containing oxygen (O) in combination with several transition metals (Co, Cr, Fe, Mn, Mo, Ni, V, and W) as this leads to complications with the Hubbard U correction (see main text) with `MPRelaxSet` settings. Figure 53 shows the 471 relative energies ($\Delta_{12}$, $\Delta_{13}$, and $\Delta_{23}$) for each of the remaining 157 molecule–surface systems, where $\Delta_{XY}$ is the relative energy between structures $X$ and $Y$, predicted with the MACE models and calculated with DFT. Overall, we observe a strong correlation between MACE-MP-0b3+D3 and DFT on the relative energies, providing a Pearson correlation coefficient of 0.86 and an MAD of 0.38 eV. We find that

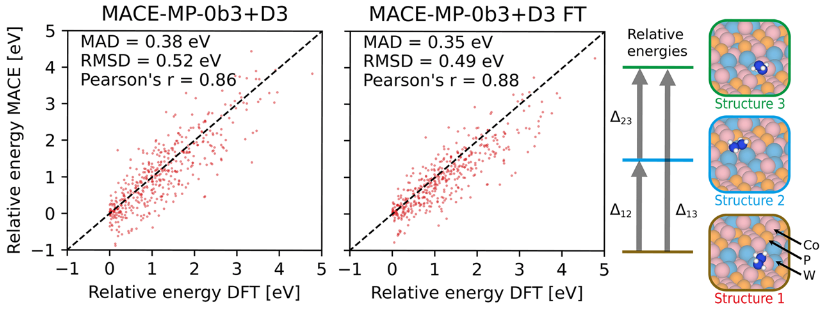


**FIG. 53.** Comparison of MACE-MP-0b3+D3 and its fine-tuned (FT) version (see text) against DFT [PBE-D3(BJ)] in predicting the relative energies between three structures across 157 molecule–surface combinations. This contains a highly diverse set of molecule–surface combinations that span 54 elements of the periodic table, involving up to three elements per surface.

fine-tuning produces a small improvement, decreasing the MADs by 0.03 meV. In particular, out of the 157 molecule–surface combinations, the lowest DFT energy configuration was correctly identified by MACE-MP-0b3+D3 for 126 of the surfaces. Fine-tuning further increases this number to 132 for MACE-MP-0b3+D3 FT.

**Similarity statement**

The molecule–surface combinations in the OC157 dataset cover all elements up to and including Bi (atomic number 83), except He, Li, Be, B, F, Ne, Mg, Cl, Ar, Br, Kr, I, Xe, Ba, and all lanthanoids. The UMAP analysis shows that 126 of the 157 molecule–surface combinations have no similarity to the training dataset, with a further 20 having less than 10, and only 3 having more than 20 similar compositions.

**Performance summary**

Good performance at predicting the absolute value of the relative energies between various molecule–surface configurations. The lowest energy (DFT) structure was correctly predicted in 80% of the cases for MACE-MP-0b3+D3, with fine-tuning further improving performance to 84%.

## 32. Computational efficiency (20-element alloy)

This section demonstrates the computational performance of the medium MACE-MP-0b3 model (without dispersion correction) while also illustrating stable dynamics for a large, diverse system. High entropy alloys are multicomponent mixtures with at least four or five distinct elements, where each component appears in non-trace proportions. Here, we consider the 20-component alloy investigated by Cantor,[295] which contains equal amounts of Mn, Cr, Fe, Co, Ni, Cu, Ag, W, Mo, Nb, Al, Cd, Sn, Pb, Bi, Zn, Ge, Si, Sb, and Mg (see also the recent computational study by Ceriotti and co-workers[109,296]). Our simulations were performed with LAMMPS,[297] utilizing its Kokkos extensions,[298] on AMD CPUs and NVIDIA A100 GPUs.

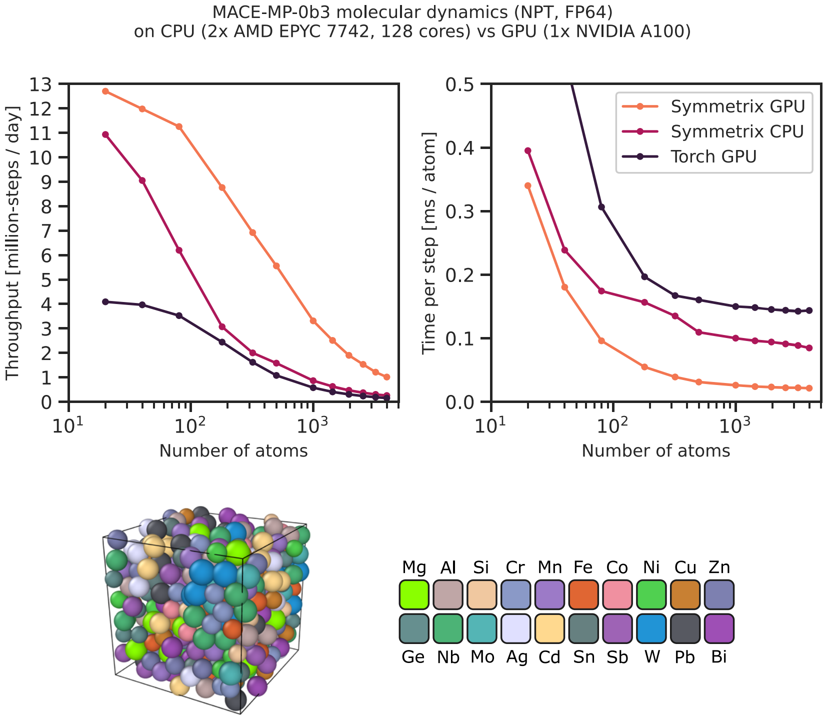


**FIG. 54.** Double-precision computational performance for NPT dynamics in LAMMPS, measured for the 20-element Cantor alloy at ambient pressure and 300 K. The two plots show the same data, highlighting throughput (left) and speed (right). The results are shown for a 128-core CPU node utilizing AMD EPYC 7742 CPUs, as well as a single NVIDIA A100 GPU. Each plot has three curves: one for the Torch baseline on GPU and two for an optimized MACE evaluator available from the `symmetrix` package. The `symmetrix` evaluator uses Kokkos, a vendor-agnostic library, to enable fast CPU and GPU performance.

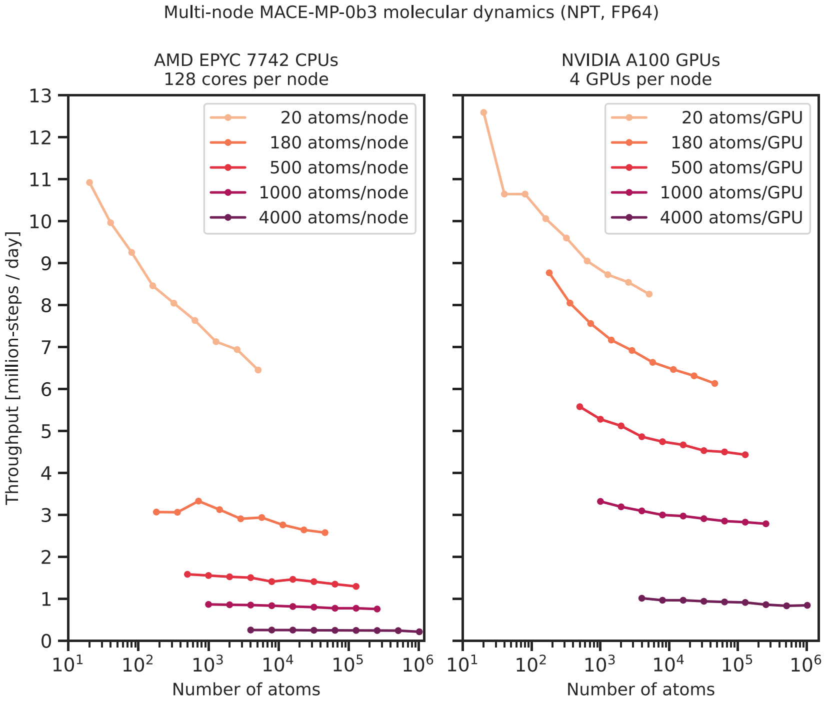


**FIG. 55.** Double-precision multi-node performance for NPT dynamics in LAMMPS, measured for the 20-element Cantor alloy at ambient pressure and 300 K. The left plot contains data for 1–256 CPU nodes, with 128 cores per node, and the right plot has data for 1–256 A100 GPUs, with four GPUs per node. Near-perfect weak scaling is achieved with as few as a few thousand atoms per CPU node or GPU.

Figure 54 shows single-node computational performance, and Fig. 55 demonstrates multi-node scaling. For each simulation, we began with a 20-atom FCC primitive cell and generated randomized supercells with sizes up to 1 024 000 atoms, measuring the performance over at least 5 ps of NPT dynamics at ambient pressure and 300 K. In all cases, the primary aim was to measure performance for a reasonably well-mixed system, and we did not attempt to reach full equilibration.

**Similarity statement**

Of the 150k MP structures, roughly 100k have at least one of the 20 elements used in this example, but only 3461 have compositions drawn exclusively from that set. Of this latter group, 193 are single-component structures, 2079 are binaries, 1155 are ternaries, and 34 are quaternaries. Moreover, no structure in the entire database has five or more of the 20 elements considered here. While it is encouraging that the dynamics appear stable for such a diverse, out-of-sample composition, we expect that incorporating dedicated datasets like those reported in Ref. 109 could enhance quantitative predictions.

## 33. MPA-0: Scaling up with the Alexandria dataset

To investigate the effect of training set size on model performance, we developed a MACE model, MPA-0, using identical hyperparameters to the MP-0b3 medium model but trained on an expanded dataset combining MPtraj and sAlex.[299] The sAlex dataset comprises 10.5M structures extracted from the original Alexandria dataset, curated for compatibility with the Matbench

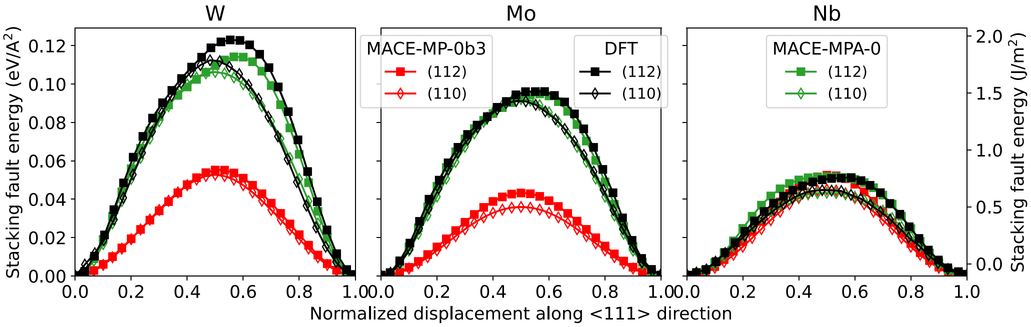


**FIG. 56.** Generalized stacking fault profiles for (112)[111] and (110)[111] Γ-surfaces predicted by MP-0b3 (red) and MPA-0 (green) compared to DFT reference data (black). MPA-0 demonstrates improved agreement across all three BCC metals, especially for Nb.

Discovery benchmark. The Alexandria dataset consists only of randomly generated and relaxed crystal structures. Here, we demonstrate that increased data volume on just crystal structures significantly enhances model robustness across multiple applications. Notably, we observe a strong enhancement in the accuracy of the MPA-0 model compared to MP-0b3 on applications for which coverage was not increased, namely, amorphous phases and small molecules on surfaces.

### *a. Metals*

Figure 56 compares generalized stacking fault profiles for body-centered cubic (BCC) metals (W, Mo, Nb) along (112)[111] and (110)[111] Γ-surfaces. MPA-0 shows markedly improved agreement with DFT reference data compared to MP-0b3 for W and Mo, where MP-0b3 substantially underestimates the fault energies. For W and Mo, MPA-0 correctly captures the peak heights and the overall energy landscape while maintaining proper symmetry across the normalized displacement. This improved accuracy in stacking fault energetics is critical for predicting mechanical properties and plastic deformation mechanisms in structural metals. This enhancement is likely due to an increase in chemical coverage for these metals in the Alexandria dataset.

### *b. Surfaces*

The prediction of surface energetics represents a critical test for materials models. Figure 57 shows parity plots comparing the performance of MP-0b3+D3 and MPA-0+D3 in predicting relative energies between three structures across 157 molecule–surface combinations. MPA-0+D3 achieves a significantly lower MAD (0.28 vs 0.38 eV) and RMSD (0.37 vs 0.52 eV) and a higher Pearson correlation (0.92 vs 0.86) compared to MP-0b3+D3. Fine-tuning further improves these metrics for MPA-0+D3 FT, reducing MAD to 0.25 eV and RMSD to 0.34 eV while increasing correlation to 0.94. This enhanced accuracy spans a diverse range of molecule–surface interactions involving 54 elements, indicating improved transferability for heterogeneous catalysis and materials interface applications. It is notable that Alexandria does not contain any surfaces or isolated small molecules, meaning that better coverage of crystal structures even improves indirectly related downstream tasks.

### *c. Amorphous carbon*

Accurate modeling of amorphous materials requires precise representation of local atomic environments and bonding. Figure 58 compares the structural properties of amorphous carbon predicted by MP-0b3, MPA-0, and the reference GAP-17 model. While MP-0b3 performs qualitatively well at low density [panel (a)–(c)], MPA-0 significantly improves the distribution of shortest-path ring sizes for the high density phases, showing a sharper peak at six-membered rings [panel (b)] that closely aligns with GAP-17, whereas MP-0b3 displays a broader, less defined distribution across ring sizes 6-8. For coordination numbers at high density [panel (d)], MPA-0 reproduces the correct $sp^3$:$sp^2$ ratio with ~85% four-coordinated and 15% three-coordinated carbon atoms, closely matching GAP-17. In contrast, MP-0b3 shows excess three-coordinated ($sp^2$) and five-coordinated carbon atoms, indicating structural deficiencies in the predicted amorphous networks. The Alexandria dataset does not contain any amorphous structures, demonstrating that an increase in broad chemical coverage of crystals helps in predicting amorphous phases.

### *d. Metastable cell*

A critical test for materials discovery applications is a model's ability to correctly rank metastable structures relative to the ground state. Figure 59 compares the error distributions of MP-0b3 and

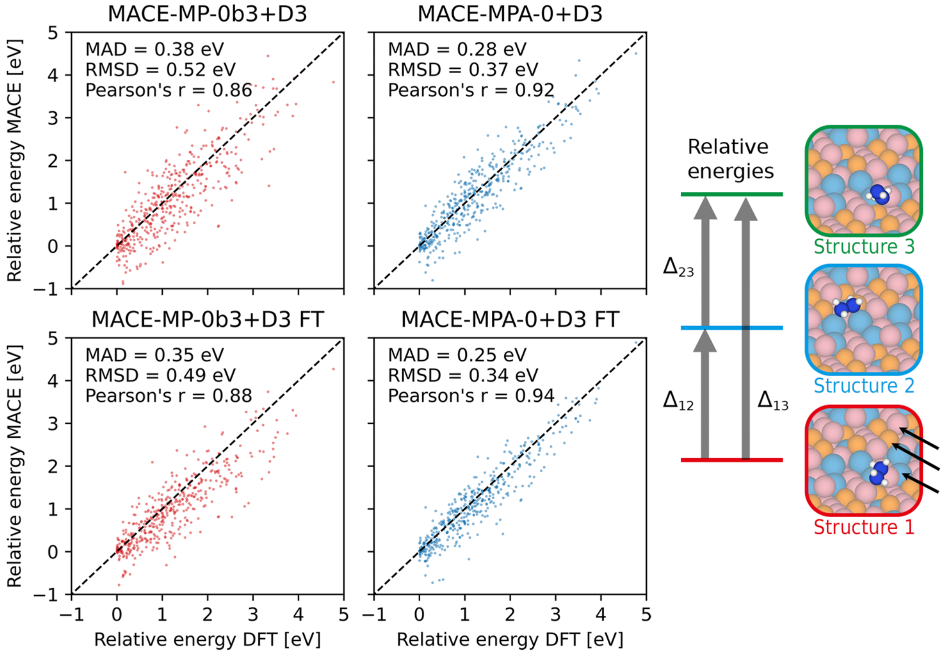


**FIG. 57.** Comparison of MP-0b3+D3, MPA-0+D3, and their fine-tuned (FT) versions against DFT (PBE-D3(BJ)) in predicting relative energies between three structures across 157 molecule–surface combinations. MPA-0 demonstrates consistently smaller errors and improved correlation with DFT values.

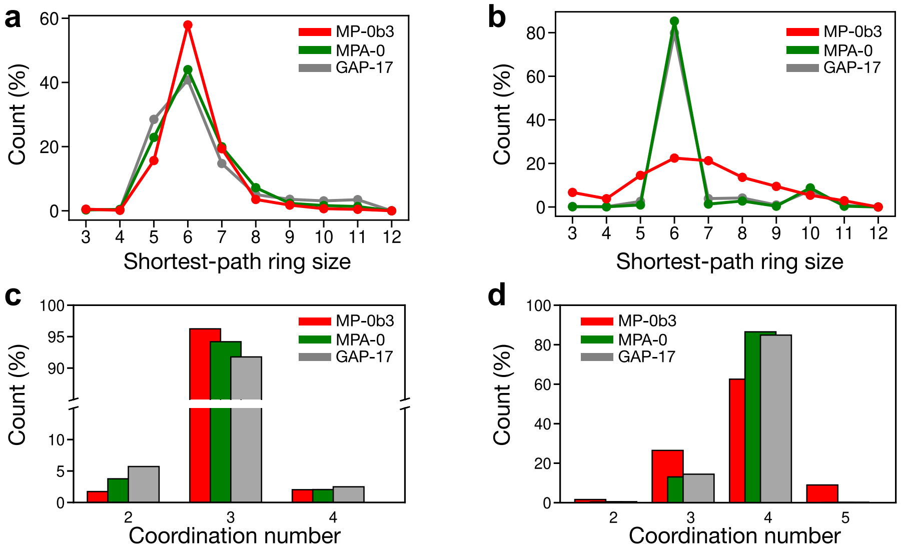


**FIG. 58.** Structural analysis of 4096-atom amorphous carbon configurations. (a) and (b) Shortest-path ring size distributions, and (c) and (d) coordination number distributions for MP-0b3 (red), MPA-0 (green), and GAP-17 reference model (gray). MPA-0 better captures the correct bonding environments and medium-range order.

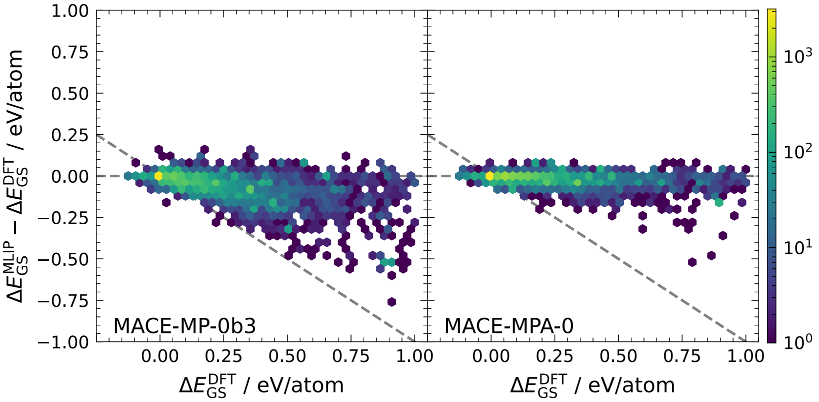


**FIG. 59.** Errors in prediction of energies above the ground state for MP-0b3 (left) and MPA-0 (right) across 16 126 relaxed structures from 100 different compositions. The color scale indicates point density. Dashed lines represent zero error (horizontal) and the boundary where structures would be erroneously predicted to have lower energy than the true ground state (diagonal). MPA-0 shows reduced error magnitude and spread, particularly for higher-energy structures.

MPA-0 in predicting energy differences above the ground state ($\Delta E_{\text{MLIP}} - \Delta E_{\text{DFT}}$) across 16 126 relaxed structures from 100 different compositions. Both models exhibit a systematic tendency to overstabilize local minima, as indicated by negative errors that increase with distance from the ground state. However, MPA-0 shows a considerably narrower error distribution and significantly reduced overstabilization, particularly for structures with $\Delta E_{\text{DFT}} >$ 0.5 eV/atom. This improved energy ranking enhances MPA-0's reliability for polymorph prediction, phase stability assessment, and materials discovery applications.

## APPENDIX B: BENCHMARKS

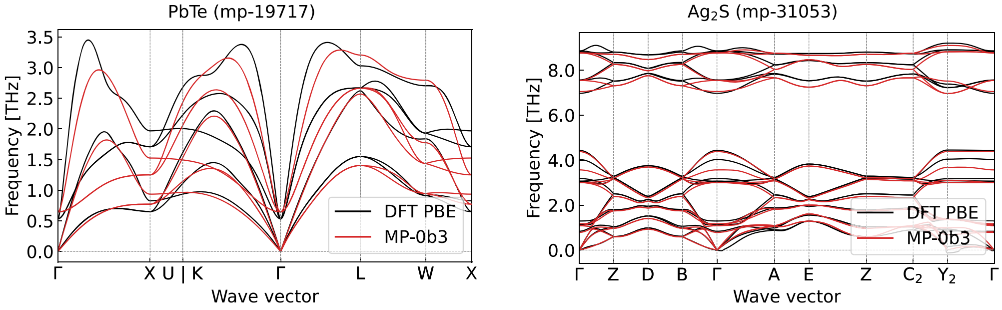
### 1. Phonons

Accurate modeling of phonons is essential for determining the dynamic stability of crystals, as well as thermoelectric properties and entropic contributions to the free energy.[300,301] These are important factors in the discovery of new materials. Harmonic phonons are typically calculated from the restoring force on atomic displacements and require highly accurate force predictions to be physical. An ML potential trained on PBE forces should be able to reproduce PBE lattice vibrations.[302,303] To assess the accuracy of MACE-MP-0b3 restoring forces, we compare harmonic phonon modes predicted via the finite-displacement method as implemented in Phonopy[304,305] and atomate2.[306,307] For this purpose, we use a benchmark set of 70 thermoelectric, phase-change, and chalcogenide-based materials. Furthermore, the reference set only contains materials in which magnetism does not play a role and U-correction is not applied in the MP.

For all materials in the benchmark set, we used the MPRelaxMaker[306] for structure optimization and chose a k-point density of 300 k-points per reciprocal volume and a cutoff energy of 520 eV (1.3*ENMAX of the highest ENMAX indicated in the pseudopotential files for the Projector Augmented Waves (PAW) method for each compound). We chose an electronic convergence criterion of $10^{-7}$ eV and the ionic convergence criterion of $10^{-6}$ eV. We perform the calculations of single-atom-displaced supercells with the MPStaticMaker,[306] setting the displacement amplitude to 0.01 Å and setting the minimum edge length to 20 Å, which allows us to perform the calculation with only one k-point. All data and scripts for reproducing the benchmark can be found under https://doi.org/10.5281/zenodo.15462976.

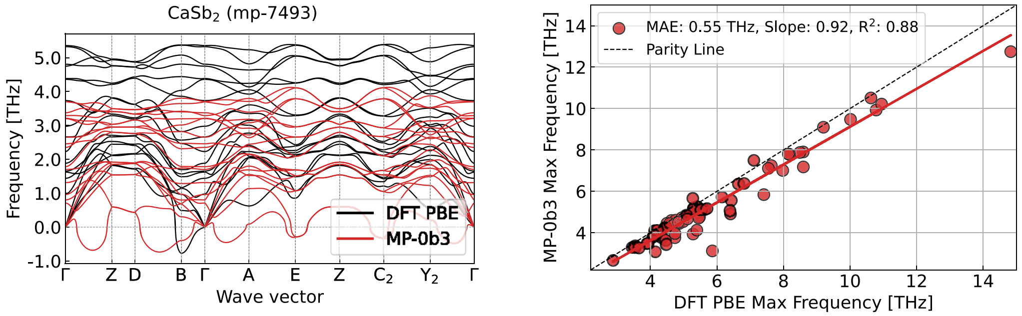
As examples, we show two compounds out of our set where DFT and MACE-MP-0b3 agree well [Figs. 60(a) and 60(b)], and one of the compounds of the largest qualitative disagreement [Fig. 60(c)]. The noticeable drop in the optical modes around Gamma for PbTe arises from neglecting the non-analytic corrections derived from Born charges, which are unavailable from electronic structure-less ML potentials. The parity plot of MACE-MP-0b3 vs PBE highest-frequency phonon modes in Fig. 60(d) reveals that MACE-MP-0b3 typically slightly underestimates the highest overall frequency in the phonon band structure (henceforth referred to as bandwidth, which correlates with the stiffness of the material's strongest bond) by a mean absolute error of 0.55 THz in comparison to DFT. MACE-MP-0b3 bandwidths have an excellent predictive power of $R^2 = 0.88$. The least squares regression (red line), with a slope close to one, indicates that the error is largely independent of the actual bandwidth. The largest absolute discrepancy between PBE vs MACE-MP-0b3 maximal phonon frequency is around 2.7 THz (47%) for orthorhombic GaAs (mp-603640). In contrast, 49 out of 70 materials show a maximal phonon frequency error of less than 10%. The violin and box plot of the root mean square errors (RMSEs) in Fig. 61(a) indicate a median error of 0.33 THz. We calculate the RMSE over the entire Brillouin zone, as the density of states (DOS) is the only invariant quantity and the choice of the *k*-path is arbitrary. Highly specialized models for one element can achieve much lower errors of 0.1–0.2 THz over the whole band structure.[303] MACE-MP-0b3 attains this level of accuracy for a few materials but does not achieve it on average. However, the deviations are better compared to results from models that specifically train only on the highest phonon peak on MatBench.[308] Here, the best models arrive at mean absolute errors of 0.8–1 THz for the last DOS peak of phonons, which is a similar but not identical quantity to the width of the band.

As shown in the confusion matrix inset to Fig. 61, the model also provides a signal on the presence of imaginary modes. The presence of such lattice vibrations indicates that the structure is dynamically unstable, meaning it can transform into a lower-energy structure by displacing atoms along the corresponding vibrational eigenvector. MACE-MP-0b3 achieves 64% accuracy on binary dynamic stability classification with a PBE unstable rate of 51.4% (a nearly balanced dataset). Although this outperforms the dummy accuracy of 50%, it also leaves room for significant improvement. A set threshold of 0.05 THz determines how large the absolute value of the imaginary mode must be to count as imaginary. For example, a mode depicted at −0.04 THz in the phonon band structure would



(a) PbTe phonon band structure

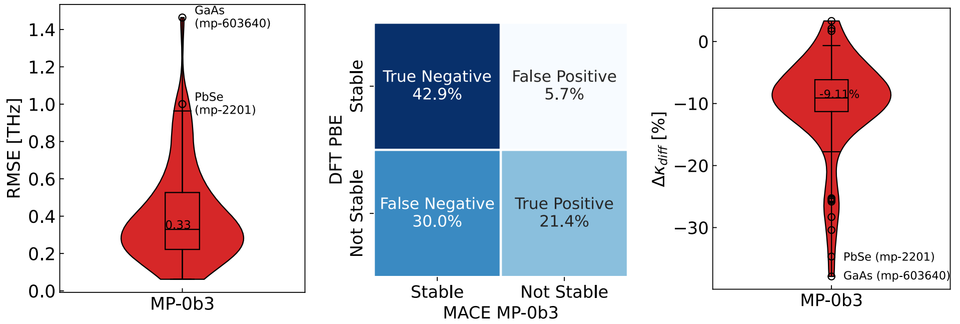
(b) $Ag_2S$ phonon band structure



(c) $CaSb_2$ phonon band structure

(d) MACE vs PBE highest phonon frequency parity plot

**FIG. 60.** Comparison of DFT PBE and MACE-MP-0b3 phonon band structures. Imaginary modes are plotted with a negative sign. (a) and (b) show examples of particularly good phonon bands. (c) It is a particularly bad example. (d) Parity plot of MACE-MP-0b3 vs PBE highest phonon band frequency showing good agreement across the chosen materials. Showing excellent agreement across diverse materials.



(a) Violin/Box plot of phonon RMSE.

(b) confusion matrix

(c) Relative error of diffusive thermal conductivity

**FIG. 61.** Comparison of DFT PBE and MACE-MP-0b3 phonons. (a) shows the violin and box plot of the RMSE over the whole Brillouin zone. (b) shows the confusion matrix of the occurrence of imaginary modes with a threshold of $|\omega_{imag}|$ < 0.05 THz. (c) Relative Error of the diffusive thermal conductivity.

not be considered dynamically unstable, while −0.06 THz would. The false negative values are strikingly high. We assume there are three reasons for this. First, a caveat to the above-mentioned analysis is that the PBE unstable labels we use as ground truth may themselves contain false negatives due to potential convergence errors. Second, the dataset contains mainly thermoelectric materials, which are more prone to the occurrence of imaginary modes. Third, some materials in the dataset have exceptionally long unit cells (20 Å) in one direction. In these nearly amorphous structures in one direction, the energy landscape becomes more complex due to the increased degrees of freedom and leads faster to false positives or false negatives.

As is, MACE-MP-0b3 can already serve as a useful prescreening filter for dynamically unstable materials, especially given the lower false positive (5.7%) than false negative rate (30.0%). This means MACE-MP-0b3 is less likely to predict materials as unstable that are stable than vice versa and, therefore, is biased toward keeping materials in the candidate pool.

We also emphasize the low computational cost of these predictions. Generating a complete phonon band structure on an Apple M2 Max CPU takes ∼30 s. Running MACE-MP-0b3 on a supercomputer could easily generate approximate phonon bands for every material in MP or other large databases. However, the accuracy of harmonic phonons of MACE-MP-0b3 is still insufficient to be a standard substitute for DFT.

However, MACE-MP-0b3 is already well suited for estimating diffusive thermal conductivity, which represents a lower limit of thermal conductivity. Following the proposed model of Agne *et al.*,[309] the diffusive thermal conductivity in the high-temperature limit only depends on the average phonon frequency. Figure 61 shows the violin and parity plot of the relative diffusive thermal conductivity error of the prediction by MACE-MP-0b3. Since the phonon bandwidth is systematically underestimated [Fig. 60(d)], a consistent downshift is observed, with a median of −9.11%. The other differences in the phonon band shapes between DFT and MACE-MP-0b3 are effectively averaged out. Imaginary modes were not included in the calculation. The long tail in the violin plot arises from a few cases where the maximum phonon frequencies are significantly underestimated. This occurs due to a softening effect,[262] which results in an effective underestimation of the PES and, consequently, an underestimation of the forces, leading to a reduced phonon mean frequency.

**Similarity statement**

The structures whose phonon predictions we analyzed are all part of the training set. However, the supercells, including single-atom perturbations performed by the finite-displacement method, are not.

## 2. Bulk and shear moduli

MACE-MP-0 was benchmarked against the elastic properties of over 8000 materials stored in the MP database. Being able to capture elastic properties such as bulk and shear modulus—which depend on the second derivatives of the energy with respect to strain—demonstrates a more precise ability to capture the potential energy surface.

Specifically, MACE-MP-0 was used to calculate the Voigt–Reuss–Hill average[310–312] bulk modulus and shear modulus as derived from stress–strain relations. The initial structures used for these calculations were the relaxed PBE[118] structures from MP; these structures were then re-relaxed using the MACE-MP-0 model and then deformed. Specifically, a total of four strain magnitudes were used along six independent strain modes (in Voigt notation): $\boldsymbol{\epsilon} \in [\epsilon_{11}, \epsilon_{22}, \epsilon_{33}, \epsilon_{44}, \epsilon_{55}, \epsilon_{66}]$. For $\epsilon_{11}$, $\epsilon_{22}$, and $\epsilon_{33}$, the strain magnitudes were ±0.01 and ±0.005. For $\epsilon_{44}$, $\epsilon_{55}$, and $\epsilon_{66}$, the strain magnitudes were ±0.06 and ±0.03. These calculations were performed using the `elasticity` module from the `MatCalc` package.[313] Hence, all of these predictions are based on equilibrium, bulk crystals alone. Moreover, to filter out likely unphysical DFT predictions, elastic properties from MP were excluded from this analysis if the DFT VRH average bulk or shear modulus was less than −50 GPa or greater than 600 GPa. Note that the data excluded due to unphysical DFT-based properties are distinct from the data not plotted in Fig. 62 due to poor MACE-MP-0 predictions.

The results comparing MP and MACE-MP-0 bulk moduli with MAE of 15.70 GPa and $R^2$ of 0.84 are shown in Fig. 62(a). This compares favorably to the $R^2$ value of 0.757 reported for M3GNet.[94] Similarly, results for shear moduli are shown in Fig. 62(b). MACE-MP-0 struggles to predict shear properties, likely due to a lack of sheared structures in MP.

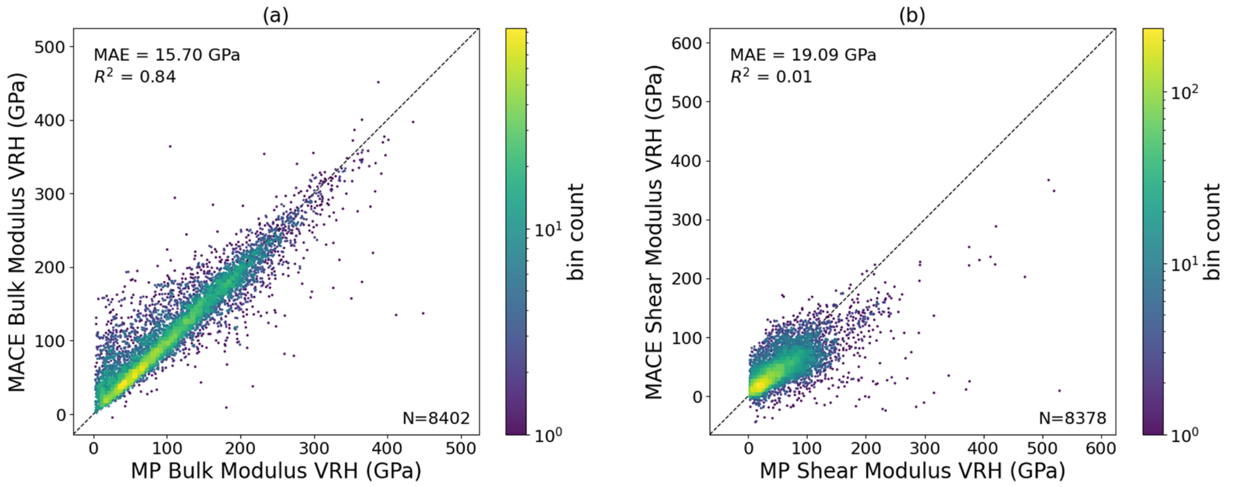


**FIG. 62.** Comparison between MACE-MP-0- and MP-calculated (a) bulk moduli and (b) shear moduli for ∼8400 materials stored in the MP database. The dashed line is a parity line (i.e., the target distribution of data). For shear moduli, note that ten points were excluded from this plot, as MACE-MP-0 predicted an unphysically high (≥600 GPa) or low (≤−50 GPa) shear modulus; this includes mp-{1007819, 1008669, 1009019, 11478, 2458, 27954, 631377, 631633, 721759, 984628}.

**Similarity statement**

None of the DFT deformation calculations contained in MP are present in the MPtrj training set used for MACE-MP-0.

### 3. Cohesive energies

In this section, we benchmark MACE-MP-0 against the cohesive energies of widely used datasets of molecules and molecular crystals, S66[88] and X23.[89] S66 is a dataset comprising 66 molecular complexes at their reference equilibrium geometries, designed to cover the most common types of noncovalent interactions in biomolecules while maintaining a balanced representation of dispersion and electrostatic contributions. X23 is a dataset of 23 organic molecular crystals. Furthermore, we analyze the relative stabilities of the ice polymorphs in DMC-ICE13.[35] The DFT calculations were performed using VASP[4,291,292] with the PBE functional and D3 dispersion correction with Becke–Johnson damping. The energy cutoff is 520 eV. Gas phase calculations were performed at the Γ point in a 25 Å cubic box. Solid phase calculations were performed with a 4 × 4 × 4 k-point grid. The results comparing MACE-MP-0 and DFT binding energies of the S66 dimers and the lattice energies of X23 and DMC-ICE13 are shown in Figs. 63–65. The MAE is ∼4 meV/atom for S66, 459 meV for X23, and 190 meV for the DMC-ICE13 absolute lattice energies. The relative stabilities of the ice polymorphs are correctly captured, with an MAE of 5 meV on the relative lattice energies.

**Similarity statement**

S66 and X23 comprise dimers and molecular crystals containing C, H, N, or O atoms. Ice polymorphs contain H and O atoms. The MP database contains 73 799 structures with O atoms, 10 312 structures with H atoms, 11 356 structures with N atoms, and 9043 structures with C atoms. The database

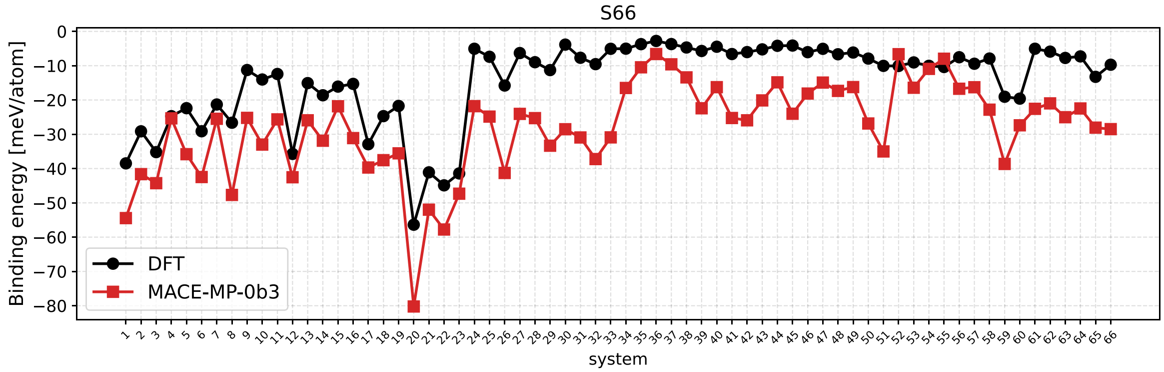


**FIG. 63.** Comparison between DFT (black) and MACE-MP-0 (red) calculated binding energies of the S66 dimers. The binding energies are divided by the number of atoms in each dimer. The lines are guides for the eye.

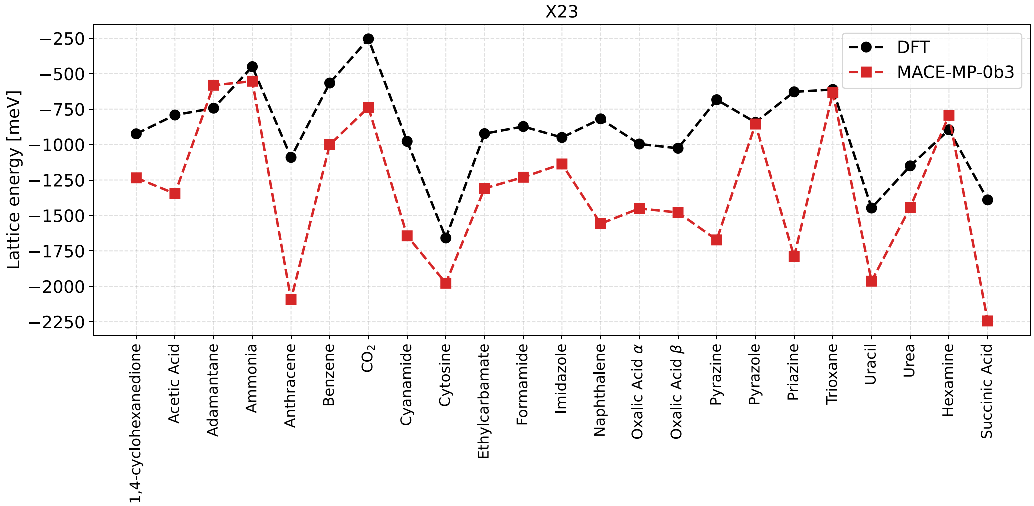


**FIG. 64.** Comparison between DFT (black) and MACE-MP-0-calculated (red) lattice energies of the X23 dataset. The lines are guides for the eye.

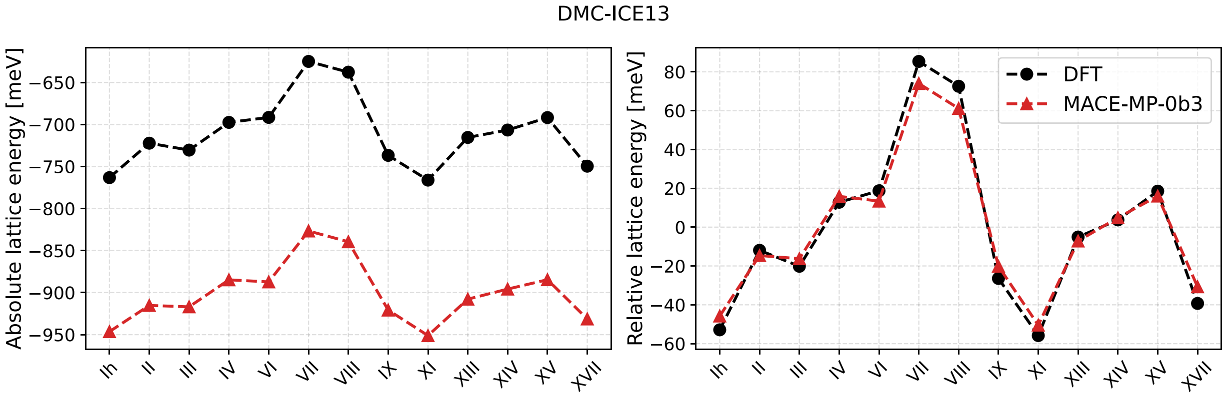


**FIG. 65.** Comparison between DFT (black) and MACE-MP-0b3 (red) lattice energies of the DMC-ICE13 dataset. We report both the absolute lattice energies (left), i.e., the energy per molecule of each crystalline phase with respect to the gas phase, and the relative lattice energies (right), i.e., the lattice energy relative to the mean value among the 13 polymorphs.

contains six structures matching an exact chemical formula in S66; these are $H_4O_2$, $C_4HO_8O_4$, $C_8H_8$, and $C_4H_4$. 16 structures match an exact chemical formula in X23; these are $C_8H_{16}O_8$, $C_{20}H_{32}$, $H_{12}N_4$, $C_4O_8$, $C_8H_{16}N_{16}$, and $C_2H_8N_4O_2$. 9 structures match an exact chemical formula in DMC-ICE13; these are $H_{24}O_{12}$ and $H_{16}O_8$. Overall, the database contains 630 structures with organic molecules and 1342 structures with water molecules. We provide `s66_chemiscope_input.json`, `x23_chemiscope_input.json`, and `dmcice13_chemiscope_input.json` to help visualize the interactive UMAP on https://chemiscope.org/.

### 4. Atomization energies and lattice constants of solids

In the following section, we benchmark MACE-MP-0 against the atomization energies and lattice constants of a set of solids. The details of the DFT calculations are the same as in Appendix B 3. Solid phase DFT total energies were computed with a 16 × 16 × 16 k-point grid. MACE-MP-0 and DFT atomization energies (in eV/atom) are reported in Table V. The MAE is 0.03 eV/atom. Lattice constants are computed on equilibrium structures obtained by geometry relaxation with a force convergence threshold of 0.03 eV/Å. MACE-MP-0 and DFT lattice constants (in Å) are reported in Table VI. The MAE is 0.03 Å.

**Similarity statement**

All the tested solids are contained in the database, with an exact matching chemical formula for 71 structures. We provide `solids_chemiscope_input.json` to help visualize the interactive UMAP on https://chemiscope.org/.

### 5. Reaction barrier heights

In the following section, we benchmark MACE-MP-0 against the reaction barrier heights of the CRBH20 database,[90] comprising

**TABLE V.** Comparison between DFT and MACE-MP-0b3-calculated atomization energies of solids. The column Δ reports the difference between DFT and MACE-MP-0 values. Energies are reported in eV/atom.

| | Atomization energy of solids | | | |
|---|---|---|---|---|
| | DFT | MACE-MP-0 | Δ | Δ/DFT (%) |
| Ag | 3.0980 | 3.0560 | 0.0420 | 1.4 |
| Pd | 4.3980 | 4.3560 | 0.0420 | 1.0 |
| Rh | 6.4040 | 6.4380 | −0.0340 | 0.5 |
| Li | 1.7820 | 1.7820 | 0.0000 | 0.0 |
| Na | 1.2440 | 1.2300 | 0.0140 | 1.1 |
| K | 0.9870 | 0.9700 | 0.0170 | 1.7 |
| Rb | 0.8810 | 0.8500 | 0.0310 | 3.5 |
| Cs | 0.8080 | 0.7770 | 0.0310 | 3.8 |
| Ca | 2.1380 | 2.1370 | 0.0010 | 0.0 |
| Sr | 1.8100 | 1.8120 | −0.0020 | 0.1 |
| Ba | 2.0790 | 2.0790 | 0.0000 | 0.0 |
| Al | 3.8910 | 3.8320 | 0.0590 | 1.5 |
| Cu | 4.0750 | 4.0840 | −0.0090 | 0.2 |
| Si | 4.9220 | 4.8680 | 0.0540 | 1.1 |
| Ge | 4.0360 | 4.0870 | −0.0510 | 1.3 |
| C | 8.0330 | 7.9910 | 0.0420 | 0.5 |
| LiF | 4.2670 | 4.1830 | 0.0840 | 2.0 |
| NaF | 3.7980 | 3.7320 | 0.0660 | 1.7 |
| NaCl | 3.1290 | 3.1080 | 0.0210 | 0.7 |
| MgO | 4.7100 | 4.7020 | 0.0080 | 0.2 |
| SiC | 5.7620 | 5.7130 | 0.0490 | 0.9 |
| GaAs | 2.5480 | 2.4720 | 0.0760 | 3.0 |
| LiCl | 3.3770 | 3.3540 | 0.0230 | 0.7 |
| | | | MAE | |
| | | | 0.03 | |

**TABLE VI.** Comparison between DFT and MACE-MP-0b3 lattice constants of solids. The column Δ reports the difference between DFT and MACE-MP-0b3 values. Lattice constants are in Å.

| | Lattice constants of solids | | | |
|---|---|---|---|---|
| | DFT | MACE-MP-0b3 | Δ | Δ/DFT (%) |
| Ag | 4.0820 | 4.0820 | 0.0000 | 0 |
| Al | 4.0020 | 3.9880 | 0.0140 | 0 |
| Ba | 4.9760 | 4.9490 | 0.0270 | 1 |
| C | 3.5620 | 3.5510 | 0.0110 | 0 |
| Ca | 5.4630 | 5.4410 | 0.0220 | 0 |
| Cs | 6.1060 | 5.9220 | 0.1840 | 3 |
| Cu | 3.5680 | 3.5480 | 0.0200 | 1 |
| GaAs | 5.6900 | 5.6760 | 0.0140 | 0 |
| Ge | 5.7190 | 5.6770 | 0.0420 | 1 |
| K | 5.1910 | 5.1910 | 0.0000 | 0 |
| Li | 3.3520 | 3.2890 | 0.0630 | 2 |
| LiCl | 5.0560 | 5.0490 | 0.0070 | 0 |
| LiF | 3.9950 | 4.0210 | −0.0260 | 1 |
| MgO | 4.2030 | 4.2070 | −0.0040 | 0 |
| Na | 4.1070 | 4.1850 | −0.0780 | 2 |
| NaCl | 5.5850 | 5.5750 | 0.0100 | 0 |
| NaF | 4.6190 | 4.5980 | 0.0210 | 0 |
| Pd | 3.8910 | 3.9120 | −0.0210 | 1 |
| Rb | 5.5720 | 5.5720 | 0.0000 | 0 |
| Rh | 3.7600 | 3.8090 | −0.0490 | 1 |
| Si | 5.4340 | 5.4170 | 0.0170 | 0 |
| Sr | 5.9080 | 5.9660 | −0.0580 | 1 |
| SiC(a) | 3.0720 | 3.0780 | −0.0060 | 0 |
| SiC(c) | 5.0290 | 5.0310 | −0.0020 | 0 |
| | | | MAE | |
| | | | 0.03 | |

20 barrier heights for the cycloreversion of heterocyclic rings. The setup of the DFT calculations is the same as in Appendix B 3. MACE-MP-0 and DFT barrier heights (in eV) are reported in Table VII. The MAE is ~0.3 eV.

**Similarity statement**

The CRBH20 dataset comprises barrier heights of organic molecules, containing H, O, C, N, S, and F atoms. We report the list of atoms with the number of structures in which they are contained in parentheses: O (73 799), S (11 972), H (10 312), N (11 356), F (11 277), and C (9043). Overall, the database contains no structures matching an exact chemical formula in CRBH20. We provide `crbh20_chemiscope_input.json` to help visualize the interactive UMAP on https://chemiscope.org/.

### 6. Homonuclear diatomics

Core repulsion is essential for stable modeling of atomic interactions, so atoms are prevented from coming too close together, especially when modeling high temperatures and pressures. The energies of all pairs of atoms were evaluated with the model in vacuum to test the 2-body interaction. The resultant curves for homonuclear diatomics are plotted in Fig. 66. All elements have repulsive potential at small distances, even elements with minimal presence in the training set (Fig. 70). As an outlier, the praseodymium pair shows an attractive, non-physically large potential with an energy gain of over 450 eV for two Pr atoms combining in a vacuum.

## APPENDIX C: TRAINING METHODS AND DATA EXPLORATION

### 1. Training protocol

The pretrained MACE-MP-0 interatomic potentials consist of many-body message passing (interaction) layers. In each layer, the message is encoded in the irreducible representation basis with $C$ channels up to an angular frequency of order $L$. This is specified as (128x0e+128x1o) for 128 channels and $L = 1$.

In each batch updating step, the weighted sum of Huber losses[314] of energy, forces, and stress incurred by all structures in a batch are averaged and backpropagated into the neural networks,

$$\begin{aligned}\mathcal{L} = {} & \frac{\lambda_E}{N_b}\sum_{b=1}^{N_b} \mathcal{L}_{\text{Huber}}\left(\frac{\hat{E}_b}{N_a}, \frac{E_b}{N_a}, \delta_E\right) \\ & + \frac{\lambda_F}{3\sum_{b=1}^{N_b} N_a}\sum_{b=1}^{N_b}\sum_{a=1}^{N_a}\sum_{i=1}^{3}\mathcal{L}^{\star}_{\text{Huber}}\left(-\frac{\partial \hat{E}_b}{\partial r_{b,a,i}}, F_{b,a,i}, \delta_F\right) \\ & + \frac{\lambda_\sigma}{9N_b}\sum_{b=1}^{N_b}\sum_{i=1}^{3}\sum_{j=1}^{3}\mathcal{L}_{\text{Huber}}\left(\frac{1}{V_b}\frac{\partial \hat{E}_b}{\partial \varepsilon_{b,ij}}, \sigma_{b,ij}, \delta_\sigma\right),\end{aligned} \tag{C1}$$

**TABLE VII.** Comparison between DFT and MACE-MP-0b3 barrier heights for CRBH20. The column Δ reports the difference between DFT and MACE-MP-0b3 values. Energies are in eV.

| | Reaction barrier heights | | | |
|---|---|---|---|---|
| | DFT | MACE-MP-0b3 | Δ | Δ/DFT (%) |
| 1 | 1.7194 | 2.0431 | −0.3237 | 19 |
| 2 | 1.9241 | 1.9988 | −0.0747 | 4 |
| 3 | 1.7499 | 1.8505 | −0.1006 | 6 |
| 4 | 1.8238 | 1.8179 | 0.0059 | 0 |
| 5 | 1.7237 | 1.8621 | −0.1384 | 8 |
| 6 | 1.5653 | 1.1341 | 0.4312 | 28 |
| 7 | 1.0911 | 1.1134 | −0.0223 | 2 |
| 8 | 1.8983 | 1.6116 | 0.2867 | 15 |
| 9 | 1.5477 | 1.7161 | −0.1684 | 11 |
| 10 | 1.7115 | 1.1390 | 0.5725 | 33 |
| 11 | 1.7379 | 1.8005 | −0.0626 | 4 |
| 12 | 2.0361 | 1.6698 | 0.3663 | 18 |
| 13 | 1.8739 | 1.5611 | 0.3128 | 17 |
| 14 | 1.9760 | 1.6166 | 0.3594 | 18 |
| 15 | 1.8865 | 1.5719 | 0.3146 | 17 |
| 16 | 1.5741 | 0.7963 | 0.7778 | 49 |
| 17 | 1.2587 | 0.8127 | 0.4460 | 35 |
| 18 | 1.7497 | 1.3373 | 0.4124 | 24 |
| 19 | 1.6989 | 1.4281 | 0.2708 | 16 |
| 20 | 1.7654 | 0.8742 | 0.8912 | 50 |
| | | | MAE<br>0.3 | |

where $\lambda_E, \lambda_F, \lambda_\sigma$ are predetermined weights of energy, forces, and stress losses. $(\lambda_E, \lambda_F, \lambda_\sigma) = (1, 10, 10)$ is adopted. Huber deltas $\delta_E = \delta_F = \delta_\sigma = 0.01$ are used. In particular, we use conditional Huber loss $\mathcal{L}^\star_{\text{Huber}}$ for forces, where the Huber delta $\delta_F$ is adaptive to the force magnitude on each atom. To be specific, the Huber delta $\delta_F$ decreases stepwise by a factor from 1.0 to 0.1 as the atomic force increases from 0 to 300 eV/Å/atom. The Huber loss for forces can therefore be equivalently represented as

$$\mathcal{L}^\star_{\text{Huber}}\left(\frac{\partial \hat{E}_b}{-\partial r_{b,a,i}}, F_{b,a,i}, \delta_F\right) = \begin{cases} \mathcal{L}_{\text{Huber}}(\ldots, \delta_F), & F_{b,a,i} < 100, \\ \mathcal{L}_{\text{Huber}}(\ldots, 0.7\delta_F), & 100 \le F_{b,a,i} < 200, \\ \mathcal{L}_{\text{Huber}}(\ldots, 0.4\delta_F), & 200 \le F_{b,a,i} < 300, \\ \mathcal{L}_{\text{Huber}}(\ldots, 0.1\delta_F), & F_{b,a,i} \ge 300. \end{cases} \tag{C2}$$

Standardization of target variables (here energies, forces, and stresses) with different scales has been proven to be important for weight initialization and training stability,[315] in that a large spread of input or output will result in large and uneven weight values and cause model instability. After each message passing layer $k$, the node energies $\epsilon_a$ are scaled and shifted before sum pooling. The energy prediction of each structure, therefore, reads

$$\hat{E} = \sum_{a=1}^{N} \left[\sigma\left(\sum_{k=1}^{K} \epsilon_a^{(k)}\right) + \mu_{Z_a}\right], \tag{C3}$$

where $K$ denotes the total number of message passing layers and $\epsilon_a^{(k)}$ is the node energy of atom $a$ at the $k$th layer. $\sigma$ is the root

**FIG. 66.** Energies of homonuclear diatomics in vacuum. Interactions are repulsive at small distances for the entire periodic table.

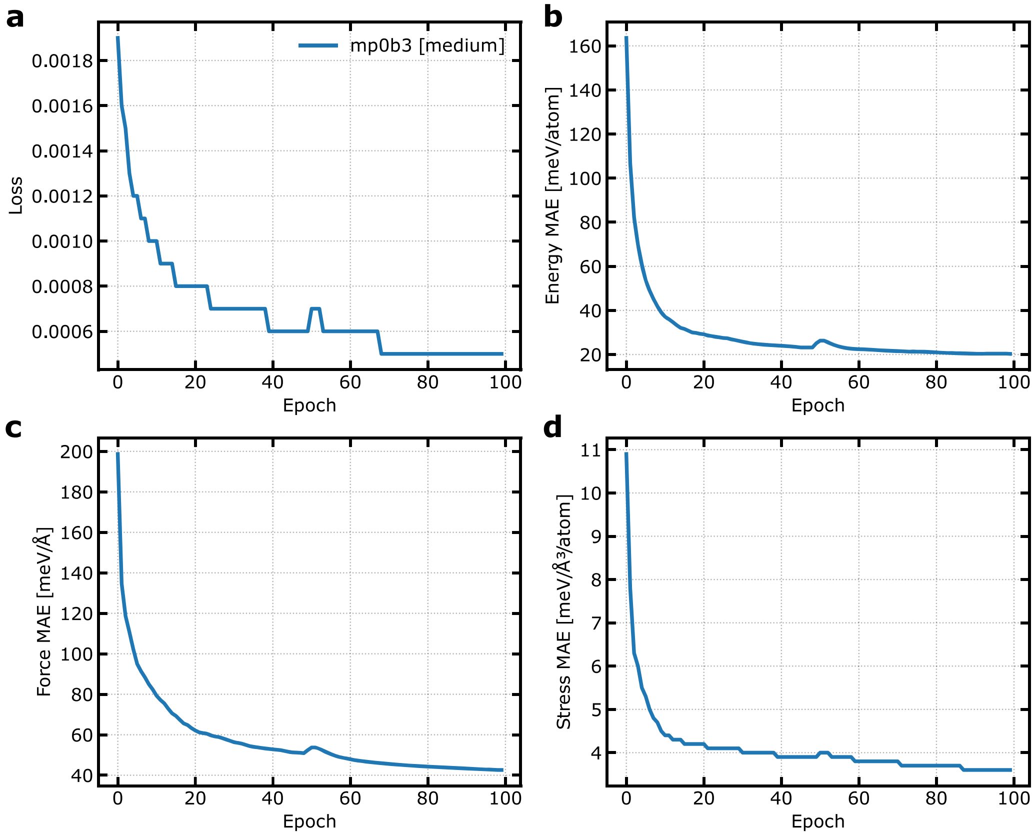


**FIG. 67.** Training curves of MACE-MP-0b3 models fitted to MPtrj data. (a) Loss [Eq. (5)]. (b)–(d) Root Mean Square Error (RMSE) of (b) energy per atom, (c) force, and (d) stress per atom. All curves are evaluated on the validation set.

mean square of the atomic forces computed over the training dataset. In order to ensure the correct limit for the dissociation of atoms, we take $\mu_Z$ as the isolated atom energy computed with the MPtrj DFT. The predicted forces and stress are computed through PyTorch's automatic differentiation `torch.autograd` of total energy with respect to atomic positions and lattice strain tensor.

The original MACE-MP-0a models were trained for 200 epochs with 40–80 NVIDIA A100 GPUs across 10–20 nodes on the HPE (Hewlett Packard Enterprise) Cray EX supercomputer Perlmutter, maintained by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory (LBNL).

The updated MACE-MP-0b3 model is trained for 99 epochs with 32 NVIDIA H100 GPUs across 8 nodes on the Jean Zay cluster managed by the GENCI–IDRIS.

All the fine-tuned models were trained on a single NVIDIA H100 GPU Jean Zay cluster managed by the GENCI–IDRIS.

## 2. Fine-tuning protocol

As described in the previous sections, MACE-MP-0 offers qualitatively good performance across a wide range of chemistry and materials at the PBE+U level of theory. However, for specific applications, it may be beneficial to fine-tune the model to improve its accuracy. There are two main motivations for fine-tuning a foundation model:

1. To reach better quantitative accuracy for a specific application, in order to compute an observable that requires high precision.
2. To increase the electronic-structure level of theory, in case a more accurate functional is required for a specific application.

The fine-tuning process involves training the model on a new dataset, called the fine-tuning dataset, which contains a limited amount of data that is only relevant to the specific application. The goals in designing a good fine-tuning protocol are threefold:

1. To ensure that the fine-tuned model achieves the desired accuracy for the specific application (no underfitting).

2. To ensure that the fine-tuned model retains the robustness of the foundation model (no catastrophic forgetting).
3. Ensure that the cost of the fine-tuning process is reasonable.

   Fine-tuning protocols have received significant attention in the machine learning community,[316,317] especially in the context of computer vision and natural language processing. The most naïve protocols consist of just continuing to train the model on the new dataset and restarting from the pre-trained model parameters. However, this approach is prone to catastrophic forgetting,[91] where the model forgets the knowledge it has learned during the pre-training phase. In the context of ML potentials, this can lead to a model that is not robust and that is more prone to explode during molecular dynamics simulations. To mitigate catastrophic forgetting, several strategies have been proposed that can be decomposed into four main categories: (1) regularization, (2) architecture, (3) data, and (4) optimization.

At present, there exists a wide variety of fine-tuning strategies in the literature that are often specific to the application, the resource constraints, and the desired accuracy. Most state-of-the-art fine-tuning protocols use a combination of regularization, architecture, data, and optimization techniques.[318] One of the most widely used approaches is the replay buffer, which consists of storing a subset of the pre-training dataset and replaying it during the fine-tuning phase.

In the context of ML potentials, the pre-training dataset is often very large, and efficient subsampling techniques are required to maximize the diversity of the replay buffer while keeping the computational cost reasonable. Moreover, the model might be fine-tuned to new data that were generated using a different level of theory or electronic structure code from the one used to generate the pre-training dataset. Therefore, the model needs to be able to learn from potentially inconsistent data effectively.

In the following, we present a fine-tuning protocol that we call *multi-head replay*, specifically designed for ML potentials, which is used to fine-tune the MACE-MP-0 model for different tasks. We demonstrate that fine-tuned models using this protocol achieve the desired accuracy for the specific application with a few configurations while retaining the robustness of the foundation model.

### a. Multi-head replay fine-tuning protocol

The multi-head replay protocol consists of two main steps: (1) the replay buffer construction and (2) the fine-tuning phase. The construction of the replay buffer consists of selecting a subset of the pre-training dataset that will maximize the robustness of the fine-tuned model while keeping the computational cost reasonable.

Let $\mathcal{D}_{\text{pre}}$ be the pre-training dataset, and $\mathcal{D}_{\text{fine}}$ be the fine-tuning dataset. In addition, let $Z_{\text{pre}}$ and $Z_{\text{fine}}$ be the set of atomic numbers present in $\mathcal{D}_{\text{pre}}$ and $\mathcal{D}_{\text{fine}}$, respectively. We first select a subset of the pre-training dataset that has at least two elements that are present in the fine-tuning dataset,

$$\mathcal{D}_{\text{Z}} = \{d \in \mathcal{D}_{\text{pre}} \mid \exists Z_1, Z_2 \in Z_{\text{fine}}, Z_1 \neq Z_2, Z_1, Z_2 \in d\}. \tag{C4}$$

If the size of $\mathcal{D}_{\text{Z}}$ is smaller than a predefined threshold, we add a random selection of configurations from $\mathcal{D}_{\text{pre}}$, $\mathcal{D}_{\text{Z}} = \mathcal{D}_{\text{Z}} \cup \text{random\_sample}(\mathcal{D}_{\text{pre}})$. If the size of $\mathcal{D}_{\text{Z}}$ is larger than a predefined threshold, we subsample it to keep the computational cost reasonable by using a farthest point sampling algorithm. We embed the configurations in a high-dimensional space using the pre-trained model descriptors, and we select the configurations that are the farthest from each other. Note that other subsampling approaches might be more efficient, and this is an ongoing research question.

A MACE model can be decomposed into two main parts, the descriptor part and the readout part. The descriptor part is the part of the model that is responsible for computing the atomic features that are invariant to rotations, and the readout maps the atomic features to the atomic energies. In order to train the model on both the pre-training and fine-tuning datasets, we use a multi-head architecture in which the model has two readouts—one for the pre-training dataset and one for the fine-tuning dataset. Since the descriptor part is shared between the two readouts, the model can learn from both datasets simultaneously.

An important detail in MACE is the normalization of the total energies. When fitting the energies, we usually subtract the sum of the isolated atomic energies, $\mu_{Z_a}$, from the total energy. This is performed to ensure that the model is able to predict the energy of the isolated atoms correctly. As the isolated atomic energies are different for the pre-training and fine-tuning datasets, we need to ensure that the model is able to predict the energy of the isolated atoms correctly for both datasets.

### b. Multi-head fine-tuning examples

**r2scan Materials Project dataset**

One application of fine-tuning is to increase the level of theory of the model. To test the performance of our fine-tuning protocol, we use the r2scan dataset, which is a subset of the Materials Project dataset that was recomputed using the r2scan functional. It contains 17 000 relaxed configurations out of the 1.5M configurations in the MPtrj dataset.

We observe in Table VIII that going from a model trained from scratch to a model fine-tuned even naively reduces the MAE energy by a factor of 3. Using the multi-head replay protocol with randomly selected 100 000 configurations from MPtraj as the replay buffer, we are able to reduce the energy MAE by a factor of 17 compared to training from scratch, reaching an energy MAE of 0.032 eV. The test set consists of the Gnome dataset, which is entirely composed of new crystals.

**TABLE VIII.** Comparison of energy MAE for different methods. Models are trained on 17 000 configurations from the r2scan Materials Project dataset and tested on the 300 000 configurations from the newly discovered Gnome dataset.[99] "Scratch" refers to a model trained from scratch on the r2scan dataset. "Fine-tune No Replay" refers to a model fine-tuned just by restarting training.

| Method | Energy MAE (eV) |
|---|---|
| Scratch ($n_{\text{r2scan}} = 17\,200$) | 0.5430 |
| MP-0 fine-tune no replay ($n_{\text{r2scan}} = 17\,200$) | 0.1620 |
| MP-0 fine-tune replay + heads ($n_{\text{r2scan}} = 300$) | 0.077 |
| MP-0 fine-tune replay + heads ($n_{\text{r2scan}} = 3000$) | 0.039 |
| MP-0 fine-tune replay + heads ($n_{\text{r2scan}} = 17\,200$) | 0.032 |

**Fine-tuning on Applications**

To further demonstrate the effectiveness of our fine-tuning protocols, we pick configurations on a variety of examples from the present paper and perform fine-tuning on them. For each selected example, we select a hundred representative configurations, and we recompute them using the MPtraj DFT. We then perform a multi-head fine-tuning using both the MP-0 and MPA-0 models. We use randomly selected 100 000 configurations from MPtraj as the replay buffer, a learning rate of 0.0001, and an exponential moving average decay of 0.9999, with the Adam optimizer and a batch size of 16. We use a 10% validation set for each task. In order to assess the performance of our fine-tuning protocol, we train models from scratch, using the same model sizes as the foundation models and the default optimization parameters.

In Fig. 68, we compare the RMSE force errors on the validation set of each task of (i) the two out-of-the-box foundation models, (ii) the equivalent models trained from scratch, and (iii) the multi-head replay fine-tuned models. The first remarkable result is that the foundation model often matches in performance the from scratch trained model, showing the good out-of-the-box performance of these models. The MACE-MPA-0 shows consistent out-of-the-box improvement. Second, the fine-tuning shows an improvement of an ×4 to ×10 in performance compared to the models trained from scratch. Fine-tuning shows impressive data efficiency and significantly reduces the cost of building a system specific model even when the out-of-the-box model is not the most accurate. We also observe only a slight improvement in accuracy for the fine-tuned MPA-0 model compared to MP-0, demonstrating that they are both good for fine-tuning purposes.

To evaluate the amount of forgetting that occurs during fine-tuning, we assess the accuracy of a model fine-tuned on one task (A01) across configurations from other tasks. We then compute the ratio of fine-tuned model accuracy compared to the original model's accuracy. We call this metric the "Forgetting ratio." In Fig. 69, we compare the forgetting ratios of two multi-head replay fine-tuned models and one naively fine-tuned model. All three models achieve similar validation RMSE on the A01 task and require similar GPU resources for the fine-tuning step. We observe minimal force forgetting in the multi-head replay fine-tuned model (forgetting ratios of ~1.1 for MPA-0 and 1.4 for MP-0), while the naively fine-tuned model (without replay) shows forgetting ratios around 4.0 for forces. For energy and stress predictions, the multi-head replay fine-tuned models demonstrate an order of magnitude improvement in forgetting ratios compared to the naively fine-tuned model. This dramatic reduction in forgetting significantly improves the stability and robustness of the fine-tuned model without compromising accuracy or performance, as all these models maintain similar performance metrics and GPU resource requirements.

### 3. Exploration of the training data

In this subsection, training data is explored and presented in figures. Figure 70 presents the structures and their counts per chemical element; Fig. 71 shows the distributions of energy of formation, forces, stresses and magnetic moments; Fig. 72 shows the distribution of magnetic moments per chemical element; Fig. 73 shows the distribution of approximative deviation of energy within each material due to variation of magnetic moments; Fig. 74 compares numbers of MP structures compared with MPTraj;

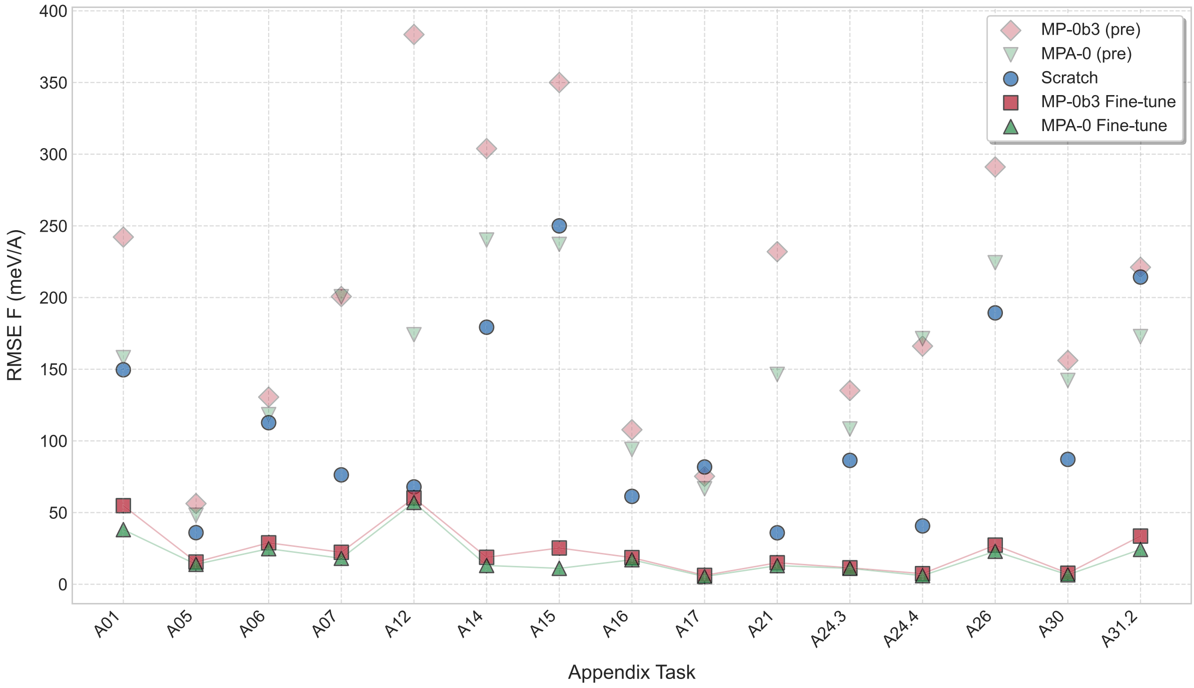


**FIG. 68.** Comparison of validation RMSE on different applications for out-of-the-box foundation models MP-0 and MPA-0, equivalent models trained from scratch, and multi-head fine-tuned models.

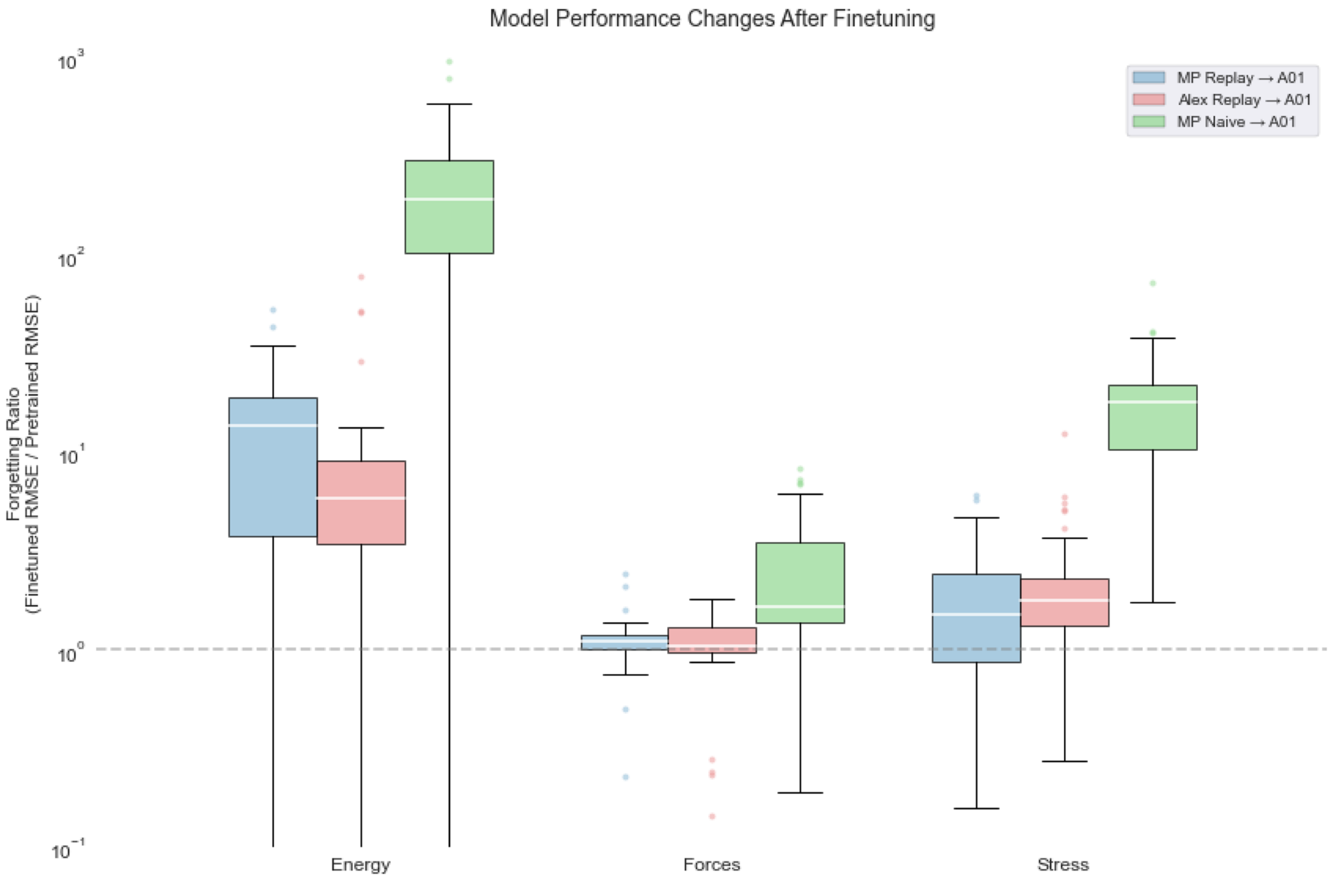


**FIG. 69.** Comparison of performance degradation after fine-tuning across three model variants: MP with multi-head replay (blue), Alexandria with multi-head replay (pink), and MP with naïve fine-tuning (green). The forgetting ratio (y axis, log scale) represents the ratio of RMSE between fine-tuned and pretrained models for energy, forces, and stress predictions. A ratio of 1.0 (dashed line) indicates no degradation in performance. The box plots show the distribution of ratios across different configurations, with outliers represented as individual points.

and Fig. 75 presents the UMAP of MACE descriptors for the MPtraj.

### 4. Similarity analysis

In the examples earlier, we have shown that MACE-MP-0 is capable of surprising degrees of extrapolation. The use of semi-local features (as a result of message passing)[23] and element mixing[24] within the MACE architecture are key components underlying MACE-MP-0's capabilities. These components allow MACE-MP-0 to extrapolate to systems that globally seem completely different from the training data but have close matches locally.

To quantify the similarity, we compare atomic environments from test systems to filtered portions of the training data using the following procedure:

1. Filter training data to a subset with elemental compositions similar to or exactly matching the test system.
2. Use MACE-MP-0 to extract invariant descriptors for all atoms in both the test system and the filtered training subset.
3. Calculate the cosine similarity between the atoms in the test system and each filtered training structure. For each atom, we use the maximum cosine similarity found this way. This is essentially a best-match structure kernel[322] that allows many-to-one mappings.
4. Average these maximum atomic similarities by element and then combine them by averaging again, yielding an element-stratified similarity.

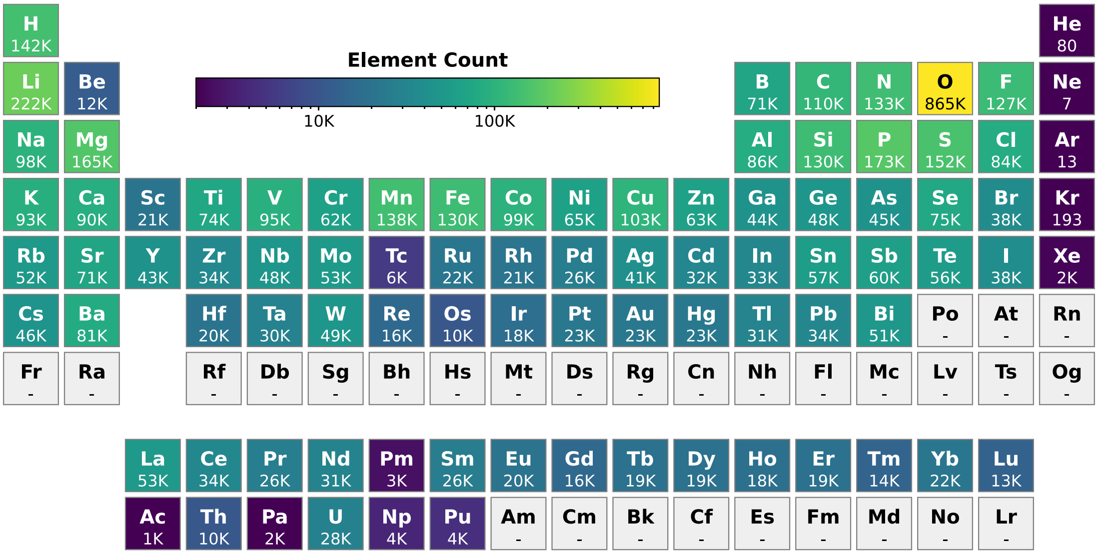
Through this procedure, we identify training set structures that contain the most similar local environments to those in any given test system. In addition, we create `chemiscope`[323] (https://chemiscope.org/) input files containing UMAP[42] projections of the atomic descriptors (fitted only on the training environments), allowing a more granular and interactive inspection of the environments in the test and training data.

The code for analyzing the data and generating `chemiscope` inputs is available as a Python package.[43]

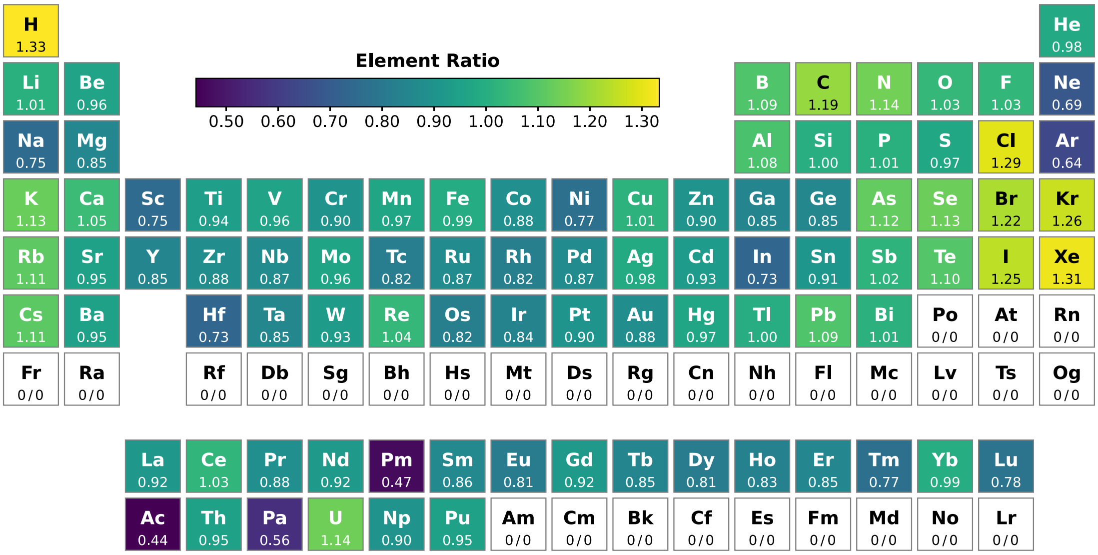
## APPENDIX D: UNCERTAINTY QUANTIFICATION

### 1. Theory and implementation

Predictive uncertainties for MACE-MP-0b3 are obtained with the last-layer prediction rigidity (LLPR) method.[324] This approach



(a) MPtrj training set element occurrence



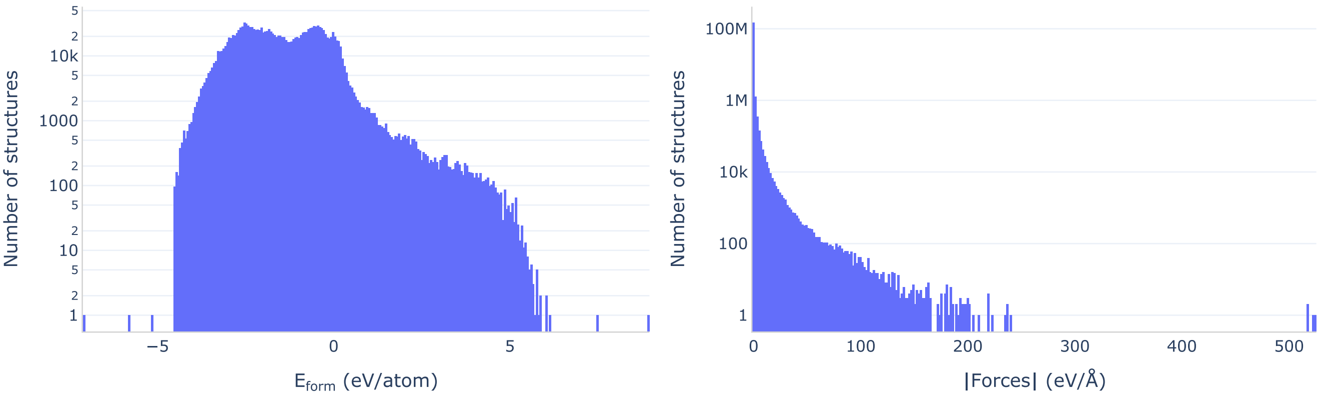
(b) Normalized ratio of elements in MPtrj to MP, $\frac{\text{MPtrj / len(MPtrj)}}{\text{MP / len(MP)}}$

**FIG. 70.** (a) Number of structures containing a given element in the MPtrj training set.[95] MPtrj consists of multiple configurations from every relaxation trajectory in MP. Some elements can require more ionic steps to relax than others. To visualize this, (b) shows the overabundance of elements relative to the number of structures containing a given element in MP ground states (after normalizing by dividing each dataset by its number of structures). That is, the factor 1.33 for hydrogen in b indicates that structures containing hydrogen were selected 33% more frequently than the base prevalence of hydrogen in MP ground states.

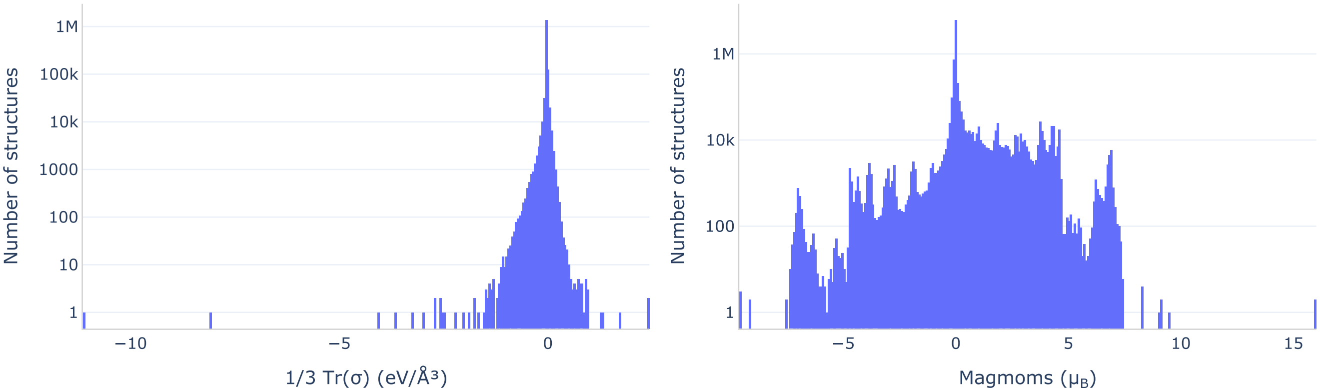
is well-suited for MACE-MP-0b3, as it is simple, scalable, and allows us to obtain uncertainty for neural networks that have already been trained. In the simplest form, LLPR uncertainties are given by

$$\sigma_i^2 = \alpha^2 \mathbf{f}_i^{\top} \mathbf{H}_o^{-1} \mathbf{f}_i, \tag{D1}$$

where $\mathbf{f}_i$ is the hidden features of structure $i$ in the last layer of the neural network (before entering the final linear transformation), $\mathbf{H}_o$ is the generalized Gauss–Newton pseudo-Hessian of the loss at the end of training, $\alpha^2$ is a calibration parameter, and $\sigma_i^2$ is the resulting LLPR uncertainty for structure $i$.



(a) MPtrj formation energies (b) MPtrj forces



(c) MPtrj stresses (d) MPtrj magnetic moments

**FIG. 71.** Distribution of energies, forces, stresses, and magnetic moments in the MPtrj dataset.[95,277] The bimodality in the formation energy distribution is due to the MP anion correction scheme,[319,320] which significantly lowers oxide formation energies.

In calculating the $\mathbf{H}_o$ for MACE-MP-0b3, only the energy part of the loss function is taken into account, while the force and stress terms are omitted. This greatly simplifies the practical calculation of the pseudo-Hessian and—as we show in Fig. 76—still allows one to obtain high-quality error estimates on forces and stresses. In simple cases, LLPR uncertainties can easily be analytically propagated to the derived targets. A simple example is given by the subtraction of two different energies, which is later used to generate uncertainties for several NEB experiments in Appendix D 3. In such cases, the uncertainty in the difference between the energies of structures $i$ and $j$ is computed as $\alpha^2(\mathbf{f}_i - \mathbf{f}_j)^\top \mathbf{H}_o^{-1}(\mathbf{f}_i - \mathbf{f}_j)$. Other cases can be treated similarly by making use of the connection[324] between the LLPR method and the Laplace approximation.[325,326]

In terms of computational cost, the energy uncertainties obtained with Eq. (D1) generate negligible overhead when compared to simple energy predictions of the MACE architecture. Predicting the error on gradients (e.g., forces and stresses) is more demanding, as it requires performing multiple backpropagation steps. However, this is only needed when explicit estimates of the gradient error, or of properties that depend explicitly on the gradient values, are desired. A LLPR-wrapped version of MACE-MP-0b3 that can output uncertainties for energy, force, and stress predictions is made available alongside the original model.

### 2. Distribution of errors in the validation set

Figure 76 shows the distributions of the predicted and actual errors on the forces and stresses of the validation set used in the training of the MACE-MP-0b3 potential. Energies are omitted, as the limited size of the validation set (160 structures) is not sufficient to gather significant statistics. Nonetheless, it is clear that the LLPR method recovers good-quality uncertainty predictions in both cases, despite the exclusion of forces and stress terms from the pseudo-Hessian calculations. See Appendix D 3 a for results on out-of-domain energy predictions.

### 3. Case studies

In this section, we show the LLPR uncertainties for a subset of the case studies found in Appendix A. When interpreting these

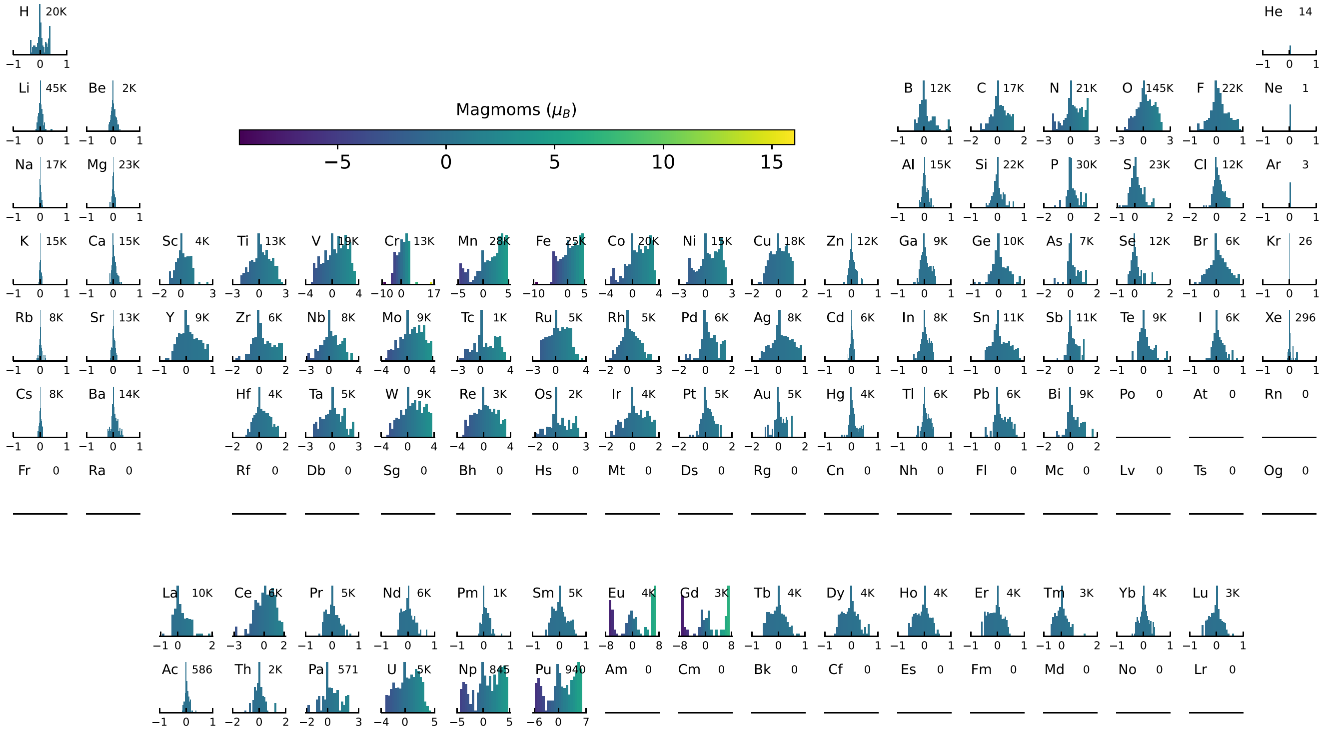


**FIG. 72.** Distribution of magnetic moments for each element in the MPtrj dataset.[95,277,321] The *y* axis is log-scaled to allow visualization of the tail of high magnetic moments in some elements with a sharp peak at 0. The number in the top right corner of each element tile counts magnetic moments for that element in the MPtrj dataset. This plot reveals rare erroneous data points in MPtrj. For instance, Cr has a single-point calculation with a highly unphysical magnetic moment of 17 $\mu_B$.

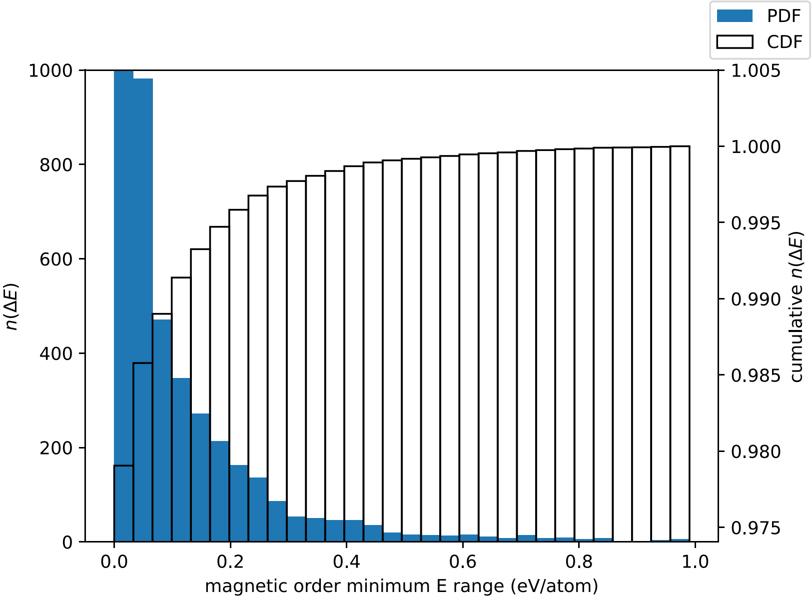
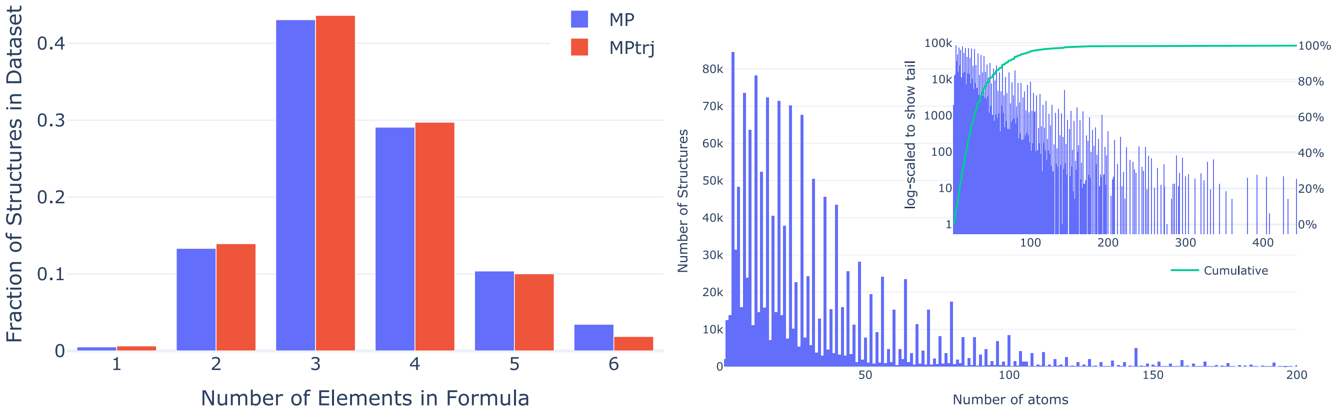


**FIG. 73.** Distribution (PDF) of approximate deviation of energy within each material due to variation in magnetic order (non-magnetic calculation, calculation converges to moment zero, ferromagnetic, and other magnetic orders), with cumulative distribution (CDF) on the right axis. The vast majority of materials have very small variation (the y axis range does not show the full extent of the first PDF bar), but a few hundred include different magnetic orders with energies that vary by more than 0.1 eV/atom.



(a) (b)

**FIG. 74.** (a) Distribution of the number of elements in the compositions of MP structures compared to MPtrj. We observe a slight overabundance of small numbers of elements in MPtrj relative to MP. (b) Distribution of a number of sites in MPtrj. The inset shows the same distribution log-scaled to visualize the tail of high site counts. The green cumulative line in the inset shows that 82% have less than 50 sites, and 97% of structures in MPtrj have less than 100 atoms.

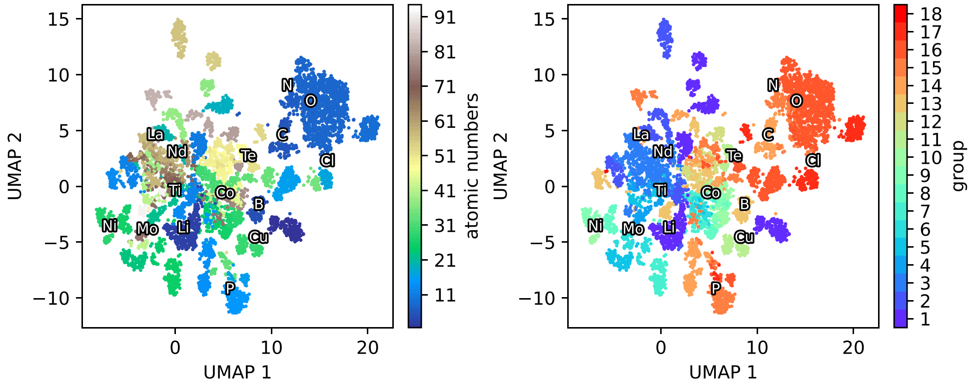


**FIG. 75.** UMAP projection of MACE descriptors for atoms in MPtrj. Each point represents the averaged feature vector of a single element in one structure and is colored by atomic number (left) and group (right) in the periodic table. The features of the MACE-MP-0 model are 256-dimensional vectors (concatenation from both the first and second layers of 128 channels). Manhattan distance is used for the construction of a high-dimensional UMAP manifold.

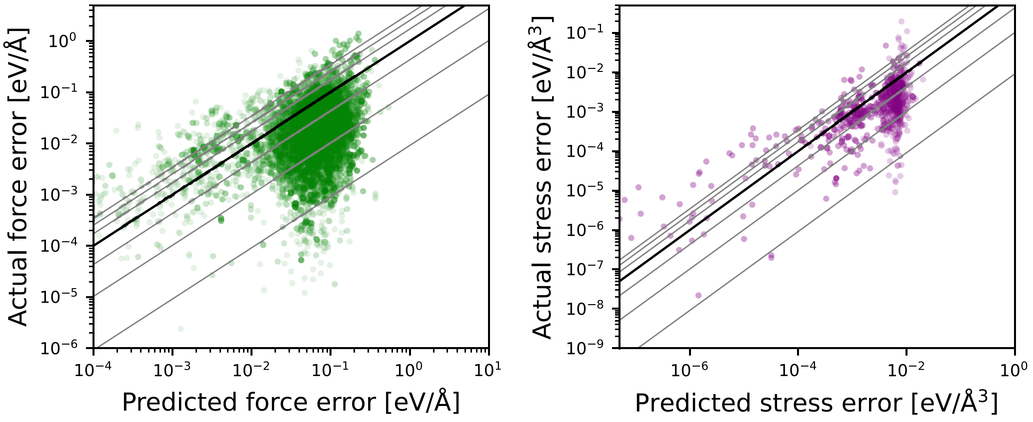


**FIG. 76.** Force (left) and stress (right) errors on the validation structures for MACE-MP-0b3. The black line represents $y = x$, while the gray lines are pairs of isolines enclosing ∼68%, 95%, and 99% of the expected distribution, respectively. Errors on very small predictions (which are present due to the symmetry of several validation structures) are cut out, but they are nonetheless well-predicted by the LLPR approach. See Appendix D 3 a for more details on these plots and their expected distribution.

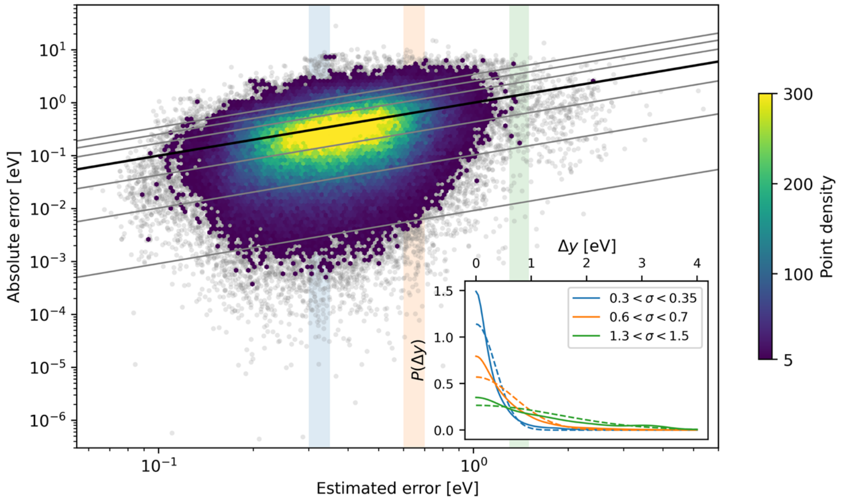


**FIG. 77.** Scatter plot that shows the absolute error vs estimated error in the total energies for the WBM dataset of Wang *et al.*[275] $y = x$ is shown in black, and isolines that successively bound 68%, 95%, and 99% of the distribution are shown in gray. The scatter plot for the entire dataset (light gray) is shown together with its 2D histogram. The inset shows the probability densities of points within three "slices" (highlighted in the main panel) of estimated errors. The solid line shows the probability density within a given slice estimated by Gaussian kernel density estimation. The dashed line shows the actual Gaussian distribution using the mean estimated error of the slice.

results, it is important to keep in mind that only the WBM dataset (Appendix D 3 a) provides reference DFT energies that are consistent with the dataset MACE-MP-0b3 has been trained on. This is not the case for the other case studies considered, which makes it difficult to make any quantitative assessments, and in particular to pinpoint whether the discrepancy between the DFT reference data and the MACE-MP-0b3 predictions stems from the error in model prediction or the inconsistencies in the DFT methods employed. Nonetheless, the following subsections demonstrate that performing uncertainty quantification can still be insightful in detecting where the model succeeds or fails, even in the absence of consistent DFT calculations. Note that uncertainties for the NEB case studies are demonstrated using the NEB frames acquired with a previous generation of the model (MACE-MP-0a).

### *a. Uncertainty quantification for the WBM dataset*

As discussed in Appendix A 29, the WBM dataset[275] can be regarded as an out-of-distribution test set for MACE-MP-0b3. The WBM dataset also provides reference DFT energies that are consistent with those of the MPtrj dataset, which the MACE-MP-0b3 has been trained on. We hence demonstrate the efficacy of the LLPR-based error estimates by performing uncertainty quantification for the total energies of the structures in the WBM dataset. Figure 77 shows the distribution of the absolute errors vs the estimated errors.

It should be understood that the estimated error corresponds to the width of the distribution of the prediction errors and that it does not directly correlate with the actual absolute error. This is apparent from the probability densities of the vertical "slices" (the inset of Fig. 77) in the distribution, where the observed probability densities closely match that of the expected Gaussian distributions with $\sigma$ values corresponding to each slice. One can see that a majority of the empirical errors lie within $3\sigma$ of the estimated error, and there is a limited number of statistically significant underestimations.

### *b. Uncertainty quantification for elemental defects in alumina*

In Appendix A 14 b [Figs. 28(d) and 28(e)], it is shown that MACE-MP-0b3 accurately recovers the reference NEB barrier path

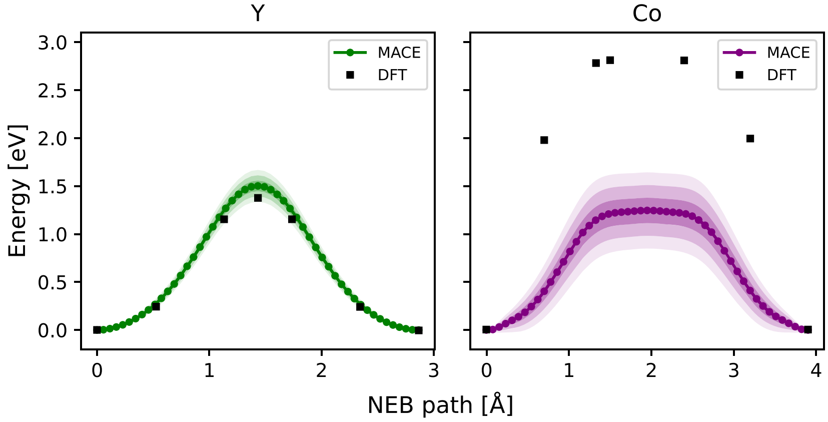


**FIG. 78.** MACE-MP-0b3-predicted relative energies along the NEB barrier paths of elemental defect diffusion in alumina, for Y (left) and Co (right). Estimated error bounds are shaded along the plot, where $1\sigma$, $2\sigma$, and $3\sigma$ bounds are shown in successively lighter shades. Available DFT data (which are not entirely consistent with the MPtrj dataset) are shown with black square markers.

of an elemental defect diffusion for Y, while failing to do so for Co. In Fig. 78, we show that LLPR-based uncertainty quantification can correctly discern the model accuracy between these two cases. The uncertainties for the case of Y remain small at the beginning and end of the NEB path and only show notable uncertainty in the middle of the path. On the contrary, the estimated uncertainties for the Co NEB barrier path are considerably large throughout the entire NEB path, even close to the beginning and end. In fact, they are generally larger by a factor of 4 when compared with those of Y.

### *c. Uncertainty quantification for CO oxidation on Cu*

Section II B and Appendix A 24 demonstrate the application of MACE-MP-0b3 for CO oxidation on several different surfaces of Cu. Figure 79 shows the results for the first half of the pathway accompanied by the LLPR-based uncertainties. In all cases, LLPR uncertainties generally remain small for the reaction pathway up to coupling at (i). Despite inconsistencies in the DFT details, this is indeed where the MACE-MP-0b3 predictions remain close to the available DFT data. Past this range in the reaction coordinate, the uncertainties become larger in all cases, and the actual errors with respect to the available DFT data also become more severe.

### *d. Uncertainty quantification for carborane rearrangement*

Appendix A 25 discusses the capability of MACE-MP-0b3 to qualitatively capture the energetics of carborane rearrangement. Figure 80 shows that LLPR-based uncertainty quantification can further shed light on the accuracy of the model along the NEB pathways. In both cases, the estimated uncertainties grow quickly along the NEB path, especially toward the transition state, but diminish toward the end when the system reaches the *meta* isomer products. These uncertainties are in good alignment with the actual discrepancies between the MACE-MP-0b3 predictions and the available DFT reference data, where the MACE-MP-0b3 predictions are found to notably underestimate the energies in the middle of the NEB paths but become comparable when predicting the overall reaction energy.

### 4. Uncertainty calibration

Post-processing techniques offer a way to calibrate inaccurate uncertainty estimates. A common approach involves introducing a calibration set, $\mathcal{D}_{\text{cal}} := \{(\mathbf{X}_n, \mathbf{Y}_n)\}_{n=1}^{N}$, to calibrate the original uncertainty. Among these techniques, conformal prediction (CP) has gained popularity for its simplicity and effectiveness in both classification and regression tasks.[327–329]

To calibrate a heuristic uncertainty $\sigma(\mathbf{X})$ using CP, we first define a score function: $s(\mathbf{X},\mathbf{Y}) := \|\widetilde{\mathbf{Y}} - \mathbf{Y}\|/\sigma(\mathbf{X})$, where $\widetilde{\mathbf{Y}}$ represents the model's prediction (e.g., energy, forces, or stress). For each data point in the calibration set $\mathcal{D}_{\text{cal}}$, we compute the corresponding score $s_i = s(\mathbf{X}_i, \mathbf{Y}_i)$. Next, let $\alpha \in (0,1)$, and define $q_\alpha$ as the empirical quantile of $\{s_i\}_{i=1}^{N}$, given by

$$q_\alpha := \text{quantile}\left(\{s_i\}_{i=1}^{N}, \frac{\lceil (N+1)(1-\alpha)\rceil}{N}\right).$$

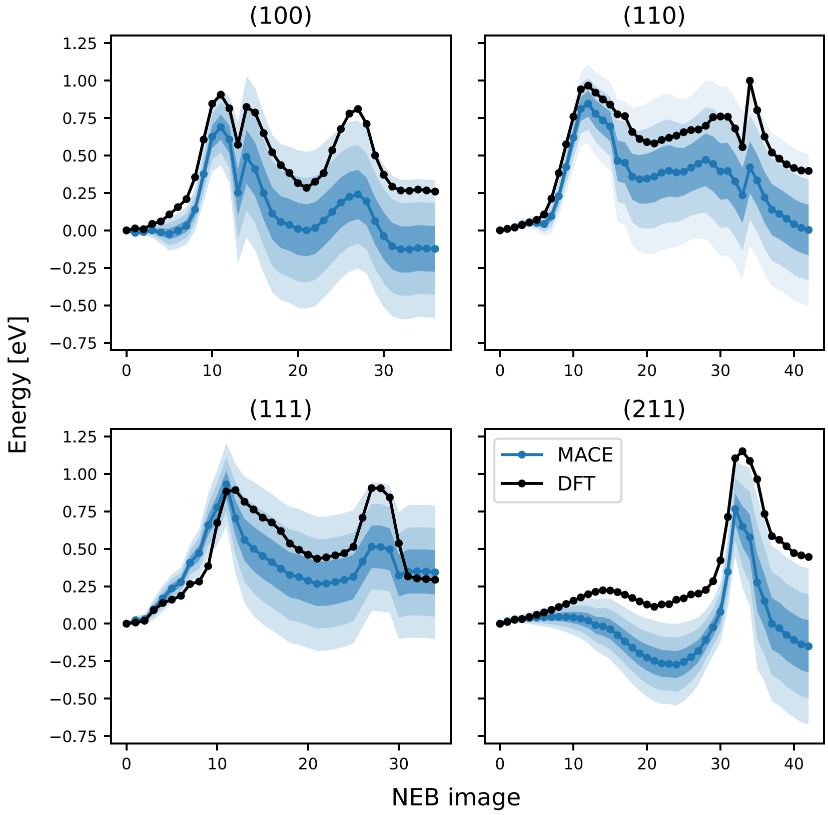


**FIG. 79.** MACE-MP-0b3-predicted relative energies along the NEB barrier paths of CO oxidation on four different surfaces of Cu. Estimated error bounds are shaded along the plot, where $1\sigma$, $2\sigma$, and $3\sigma$ bounds are shown in successively lighter shades. The DFT reference data (which are not entirely consistent with the MPtrj dataset) are shown in black.

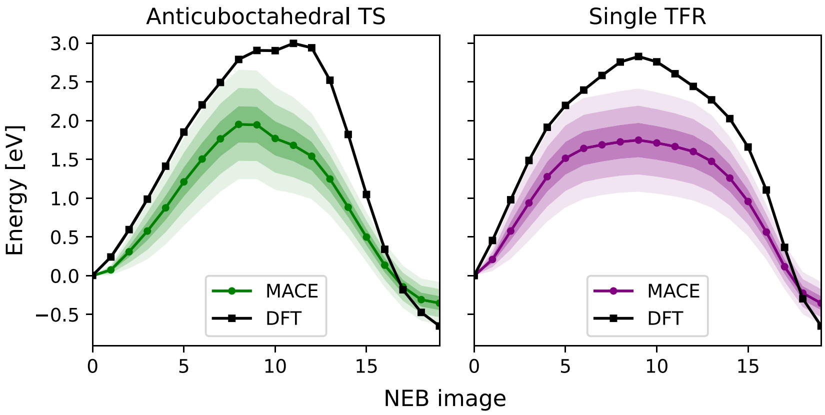


**FIG. 80.** Relative energies predicted by MACE-MP-0b3 along the NEB barriers obtained for the anticuboctahedral transition state and single triangular face rotation pathways suggested for carborane rearrangement. Estimated error bounds are shaded along the plot, where $1\sigma$, $2\sigma$, and $3\sigma$ bounds are shown in successively lighter shades. DFT reference data (which are not entirely consistent with the MPtrj dataset) are shown in black.

The calibrated uncertainty is then obtained as $q_\alpha \sigma(\mathbf{X})$. This adjustment guarantees that the calibrated uncertainty provides a probabilistic bound on the true error, $\|\widetilde{\mathbf{Y}} - \mathbf{Y}\|$, with at least $1 - \alpha$ confidence.[327,330]

A recent extension of the standard CP method is to incorporate the dependence of the quantile value on the local atomic environment (LAE).[331] Specifically, the quantile forms a step function, denoted as $\hat{q}_\alpha(\mathcal{G}(\mathbf{X}_i))$, where $\mathcal{G}$ is a classification model that assigns each LAE $\mathbf{X}_i$ to a distinct class. The calibration $\hat{q}_\alpha$ is then applied independently within each class, a technique referred to as class-based CP.

The hyperparameter $\alpha$ is a crucial component of this framework and must be carefully fine-tuned. In our numerical experiments, we set $\alpha = 0.5$ by default unless stated otherwise, as it consistently yields stable results across different scenarios. Alternatively, cross-validation on various calibration sets can be employed to determine the optimal $\alpha$. The classification model $\mathcal{G}$ is constructed by clustering invariant descriptors generated by MACE-MP-0B3 using a Bayesian variant of the Gaussian Mixture Model (GMM).[332]

Next, we examine the feasibility of uncertainty calibration by applying CP and its classification-based extensions to three case studies: (1) validation error on the MP dataset, (2) NEB transition paths for CO oxidation on Cu with different surface orientations, and (3) NEB transitions in carborane rearrangement. In this section, force errors are measured using the magnitude of the force on each atom rather than considering force components, as performed in Appendix D 3.

### a. Calibrating uncertainty for the MP validation set

We begin by demonstrating the improvements in predicted error using CP and class-based CP. In this example, 10% of the data were randomly selected for calibration. As shown in Fig. 81, the predicted force error from LLPR (left panel) overestimates the true force error. The original CP without classification (middle panel) provides a modest improvement over LLPR. Notably, applying class-based CP (right panel) results in a further slight enhancement, primarily in terms of point density. However, increasing the number of classes in GMM clustering does not lead to any signifi-

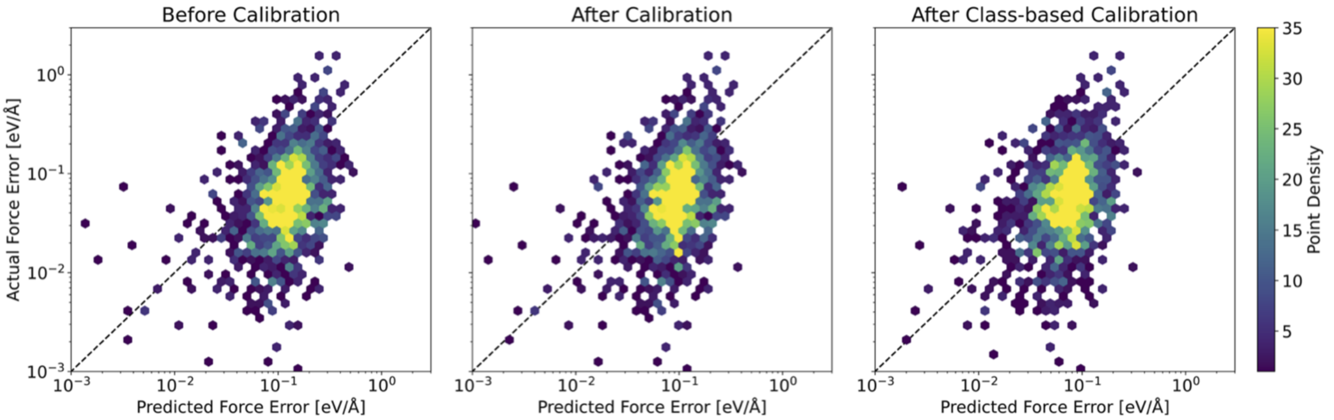


**FIG. 81.** Force errors on the validation set for Large MACE-MP0B3. Quantile in simple CP is $\hat{q} = 0.76$. Quantiles for class-based CP with six classes range from 0.5 to 2.0 and are of a similar range with an increasing number of clusters.

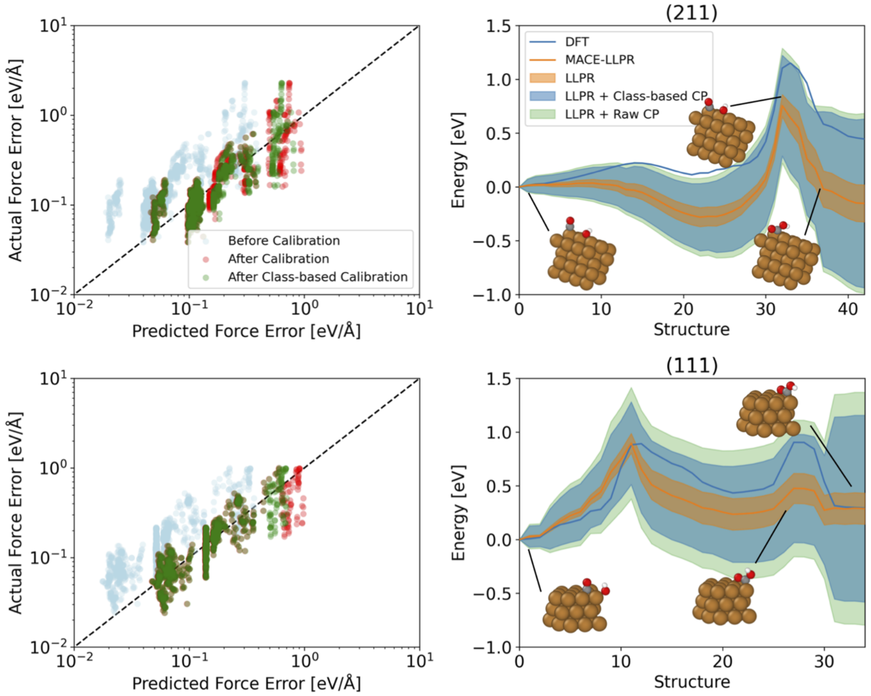


**FIG. 82.** MACE-predicted forces and relative energies along the NEB barrier paths for CO oxidation on the Cu(211) (top) and Cu(111) (bottom) surfaces. The estimated force and NEB barrier path errors from MACE-LLPR uncertainty, after calibration and class-based calibration, are shown in different colors. The DFT reference data for the NEB barrier path are shown in blue.

cant improvement. These findings suggest that while CP effectively improves uncertainty calibration, the additional benefit of class-based CP is marginal for large, diverse datasets. This outcome is expected, given the inherent difficulty of classifying complex atomic environments.

### b. Calibrating uncertainty for CO oxidation on Cu

To highlight the potential of classification on calibrating force uncertainty in specific cases, we evaluate the uncertainty of forces and energy differences along the NEB paths for CO oxidation on the (111) and (211) surfaces, where LLPR provides only qualitative uncertainty estimates. For calibration, we select 10 uniformly spaced images along each NEB path rather than using random sampling.

Figure 82 compares the predicted force errors against DFT references and illustrates the NEB barrier paths for two different surfaces, with error bars derived from LLPR uncertainty, original CP, and class-based CP.

After calibration, the predicted force errors align more closely with the ground truth for both surfaces. However, when applying original CP to all sites on the[111] surface, force uncertainties remain inaccurate due to the predominance of "bulk" copper surface sites in the calibration set, as indicated by the brown atoms in the right panels of Fig. 82. This imbalance leads to overestimated quantiles and inflated errors for high-error sites. By incorporating classification, the predicted errors for outlier sites better match the true errors, as shown in the left panel of the figure.

For uncertainties in energy differences along the NEB paths (right panel of Fig. 82), the original CP provides accurate calibration based on LLPR uncertainties. Class-based CP offers a slightly more refined calibration, particularly for the (111) surface.

### c. Calibrating uncertainty for carborane rearrangement

Next, we further evaluate the robustness of the proposed method on a more challenging system: carborane rearrangement. We selected Anticuboctahedral TS and single TFR cases, as their results are representative. Calibration samples are selected using the same scheme as in the previous example. We set $\alpha = 0.2$, as this choice enables meaningful variation in quantiles across distinct classes during class-based CP calibration.

As shown in the left panels of Fig. 83, the LLPR method slightly underestimates the actual force error in both cases, while calibration improves alignment with the true error. Notably, classification plays a crucial role in achieving robust calibration, as it accounts for the coexistence of both accurate and underestimated LLPR uncertainties within the system. This effect is particularly evident in the isomerization reaction of TFR (bottom), where separate calibration for different subsets of observations is necessary. This underscores the potential application of class-based calibration in active learning for chemically diverse systems, where both severe under- and overestimation of force uncertainties occur during iterative fine-tuning.

However, as shown in the right panels of Fig. 83, while calibration significantly improves force error estimates and classification enhances robustness, this effect does not directly carry over to the calibration of energy differences along the NEB path. This discrepancy arises from the strong symmetry in system movement, which leads to error cancellation when summing uncertainty contributions from individual atoms. We speculate that this issue could be mitigated by performing calibration directly on energy rather than force, particularly for energy-related properties.

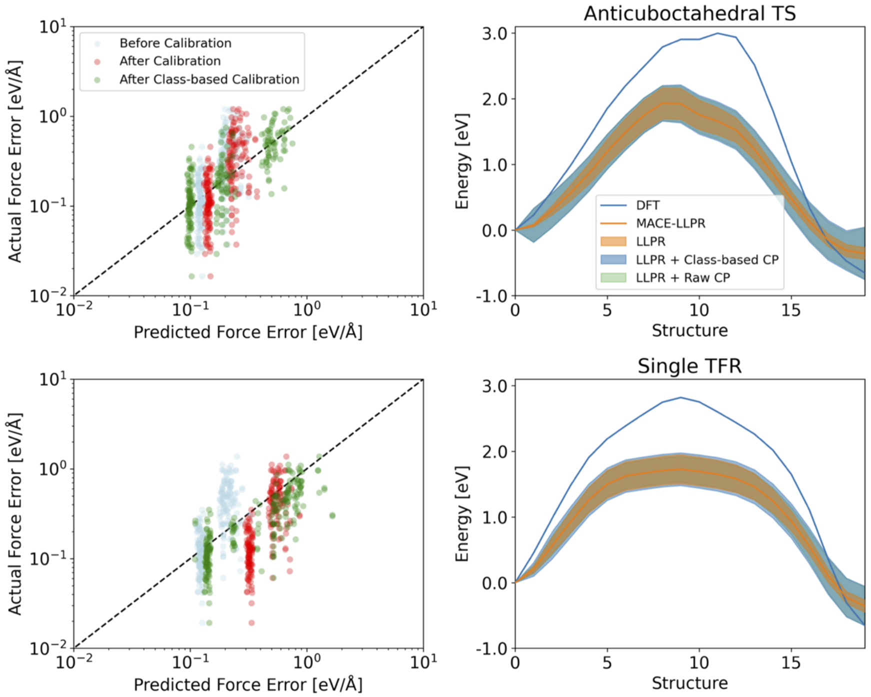


**FIG. 83.** MACE-predicted forces and relative energies along the NEB barrier paths for carborane rearrangement in anticuboctahedral TS and single TFR. The estimated force and NEB barrier path errors from MACE-LLPR uncertainty, after calibration and class-based calibration, are shown in different colors. The DFT reference data for the NEB barrier path are shown in blue.

## REFERENCES

[1] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. **140**, A1133–A1138 (1965).

[2] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru, and S. Baroni, "Quantum ESPRESSO toward the exascale," J. Chem. Phys. **152**, 154105 (2020).

[3] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann *et al.*, "CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations," J. Chem. Phys. **152**(19), 194103 (2020).

[4] G. Kresse and J. Hafner, "*Ab initio* molecular dynamics for liquid metals," Phys. Rev. B **47**, 558 (1993).

[5] P. J. Hasnip, K. Refson, M. I. J. Probert, J. R. Yates, S. J. Clark, and C. J. Pickard, "Density functional theory in the solid state," Philos. Trans. R. Soc., A **372**, 20130270 (2014).

[6] A. Jain, Y. Shin, and K. A. Persson, "Computational predictions of energy materials using density functional theory," Nat. Rev. Mater. **1**, 15004 (2016).

[7] J. Neugebauer and T. Hickel, "Density functional theory in materials science," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **3**, 438–448 (2013).

[8] M. Finnis, *Interatomic Forces in Condensed Matter*, *Oxford Series on Materials Modelling* (Oxford University Press, 2003), Vol. 1, ISBN-10: 0198509774, ISBN-13: 978-0198509776.

[9] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," Phys. Rev. Lett. **98**, 146401 (2007).

[10] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," Phys. Rev. Lett. **104**, 136403 (2010).

[11] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, "Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials," J. Comput. Phys. **285**, 316–330 (2015).

[12] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet – A deep learning architecture for molecules and materials," J. Chem. Phys. **148**, 241722 (2018).

[13] V. L. Deringer, M. A. Caro, and G. Csányi, "Machine learning interatomic potentials as emerging tools for materials science," Adv. Mater. **31**(46), 1902765 (2019).

[14] R. Drautz, "Atomic cluster expansion for accurate and transferable interatomic potentials," Phys. Rev. B **99**, 014104 (2019).

[15] O. A. von Lilienfeld and K. Burke, "Retrospective on a decade of machine learning for chemical discovery," Nat. Commun. **11**(1), 4895 (2020).

[16] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, "E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials," Nat. Commun. **13**(1), 2453 (2022).

[17] T. W. Ko and S. P. Ong, "Recent advances and outstanding challenges for machine learning interatomic potentials," Nat. Comput. Sci. **3**, 998–1000 (2023).

[18] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, "Gaussian process regression for materials and molecules," Chem. Rev. **121**(16), 10073–10141 (2021).

[19] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csányi, "MACE: Higher order equivariant message passing neural networks for fast and accurate force fields," in *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, Inc., 2022), ISBN: 9781713871088.

[20] G. Dusson, M. Bachmayr, G. Csányi, R. Drautz, S. Etter, C. van der Oord, and C. Ortner, "Atomic cluster expansion: Completeness, efficiency and stability," J. Comput. Phys. **454**, 110946 (2022).

[21] Y. Lysogorskiy, C. v. d. Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner, and R. Drautz, "Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon," npj Comput. Mater. **7**(1), 97 (2021).

[22] W. C. Witt, C. van der Oord, E. Gelžinytė, T. Järvinen, A. Ross, J. P. Darby, C. H. Ho, W. J. Baldwin, M. Sachs, J. Kermode, N. Bernstein, G. Csányi, and C. Ortner, "ACEpotentials.jl: A Julia implementation of the atomic cluster expansion," J. Chem. Phys. **159**(16), 164101 (2023).

[23] I. Batatia, S. Batzner, D. P. Kovács, A. Musaelian, G. N. C. Simm, R. Drautz, C. Ortner, B. Kozinsky, and G. Csányi, "The design space of E(3)-equivariant atom-centred interatomic potentials," Nat. Mach. Intell. **7**, 56–67 (2025).

[24] J. P. Darby, D. P. Kovács, I. Batatia, M. A. Caro, G. L. W. Hart, C. Ortner, and G. Csányi, "Tensor-reduced atomic density representations," Phys. Rev. Lett. **131**, 028001 (2023).

[25] V. Delle Rose, A. Kozachinskiy, C. Rojas, M. Petrache, and P. Barceló, "Three iterations of ($d-1$)-WL Test distinguish non isometric clouds of $d$-dimensional points," in *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, edited by A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Curran Associates, Inc., 2023), ISBN: 9781713899921.

[26] L. B. Skinner, C. Huang, D. Schlesinger, L. G. M. Pettersson, A. Nilsson, and C. J. Benmore, "Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range," J. Chem. Phys. **138**, 074506 (2013).

[27] J. E. Bertie and Z. Lan, "Infrared intensities of liquids XX: The intensity of the OH stretching band of liquid water revisited, and the best current values of the optical constants of $H_2O$(l) at 25 °C between 15,000 and 1 cm$^{-1}$," Appl. Spectrosc. **50**(8), 1047–1057 (1996).

[28] D. R. Moberg, S. C. Straight, C. Knight, and F. Paesani, "Molecular origin of the vibrational structure of ice Ih," J. Phys. Chem. Lett. **8**(12), 2579–2583 (2017).

[29] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," J. Chem. Phys. **132**(15), 154104 (2010).

[30] N. O'Neill, C. Schran, S. J. Cox, and A. Michaelides, "Crumbling crystals: On the dissolution mechanism of NaCl in water," Phys. Chem. Chem. Phys. **26**, 26933 (2024).

[31] M. J. Gillan, D. Alfè, and A. Michaelides, "Perspective: How good is DFT for water?," J. Chem. Phys. **144**(13), 130901 (2016).

[32] D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, "The nature of the hydrated excess proton in water," Nature **397**, 601–604 (1999).

[33] M. E. Tuckerman, D. Marx, and M. Parrinello, "The nature and transport mechanism of hydrated hydroxide ions in aqueous solution," Nature **417**, 925–929 (2002).

[34] N. Agmon, H. J. Bakker, R. K. Campen, R. H. Henchman, P. Pohl, S. Roke, M. Thämer, and A. Hassanali, "Protons and hydroxide ions in aqueous systems," Chem. Rev. **116**, 7642–7672 (2016).

[35] F. Della Pia, A. Zen, D. Alfè, and A. Michaelides, "DMC-ICE13: Ambient and high pressure polymorphs of ice from diffusion Monte Carlo and density functional theory," J. Chem. Phys. **157**, 134701 (2022).

[36] G. Algara-Siller, O. Lehtinen, F. C. Wang, R. R. Nair, U. Kaiser, H. A. Wu, A. K. Geim, and I. V. Grigorieva, "Square ice in graphene nanocapillaries," Nature **519**, 443–445 (2015).

[37] L. Fumagalli, A. Esfandiar, R. Fabregas, S. Hu, P. Ares, A. Janardanan, Q. Yang, B. Radha, T. Taniguchi, K. Watanabe, G. Gomila, K. S. Novoselov, and A. K. Geim, "Anomalously low dielectric constant of confined water," Science **360**, 1339–1342 (2018).

[38] V. Kapil, C. Schran, A. Zen, J. Chen, C. J. Pickard, and A. Michaelides, "The first-principles phase diagram of monolayer nanoconfined water," Nature **609**(7927), 512–516 (2022).

[39] H. A. Hansen, J. Rossmeisl, and J. K. Nørskov, "Surface pourbaix diagrams and oxygen reduction activity of Pt, Ag and Ni(111) surfaces studied by DFT," Phys. Chem. Chem. Phys. **10**, 3722–3730 (2008).

[40] S. Vijay, G. Kastlunger, K. Chan, and J. K. Nørskov, "Limits to scaling relations between adsorption energies?," J. Chem. Phys. **156**, 231102 (2022).

[41] L. L. Schaaf, E. Fako, S. De, A. Schäfer, and G. Csányi, "Accurate energy barriers for catalytic reaction pathways: An automatic training protocol for machine learning force fields," npj Comput. Mater. **9**(1), 180 (2023).

[42] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," arXiv:1802.03426 (2018).

[43] R. Elijošius (2023). "MACE-MP UMAP analysis," Zenodo; github.com. https://github.com/RokasEl/mace-mp-umap; https://doi.org/10.5281/zenodo.10426282

[44] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, "Towards the computational design of solid catalysts," Nat. Chem. **1**, 37–46 (2009).

[45] A. J. Medford, A. Vojvodic, J. S. Hummelshøj, J. Voss, F. Abild-pedersen, F. Studt, T. Bligaard, A. Nilsson, and J. K. Nørskov, "From the Sabatier principle to a predictive theory of transition-metal heterogeneous catalysis," J. Catal. **328**, 36–42 (2015).

[46] A. Bruix, J. T. Margraf, M. Andersen, and K. Reuter, "First-principles-based multiscale modelling of heterogeneous catalysis," Nat. Catal. **2**, 659–670 (2019).

[47] X. Qin, T. Vegge, and H. A. Hansen, "Cation-coordinated inner-sphere $CO_2$ electroreduction at Au–water interfaces," J. Am. Chem. Soc. **145**, 1897–1905 (2023).

[48] I. C. Man, H. Su, F. Calle-vallejo, H. A. Hansen, J. I. Martínez, N. G. Inoglu, J. Kitchin, T. F. Jaramillo, J. K. Nørskov, and J. Rossmeisl, "Universality in oxygen evolution electrocatalysis on oxide surfaces," ChemCatChem **3**, 1159–1165 (2011).

[49] A. Auer, M. Andersen, E.-M. Wernig, N. G. Hörmann, N. Buller, K. Reuter, and J. Kunze-liebhäuser, "Self-activation of copper electrodes during co electro-oxidation in alkaline electrolyte," Nat. Catal. **3**, 797–803 (2020).

[50] Q. Wang, J. Pan, J. Guo, H. A. Hansen, H. Xie, L. Jiang, L. Hua, H. Li, Y. Guan, P. Wang, W. Gao, L. Liu, H. Cao, Z. Xiong, T. Vegge, and P. Chen, "Ternary ruthenium complex hydrides for ammonia synthesis via the associative mechanism," Nat. Catal. **4**, 959–967 (2021).

[51] J. K. Nørskov, F. Studt, F. Abild-Pedersen, and T. Bligaard, *Fundamental Concepts in Heterogeneous Catalysis* (John Wiley and Sons, Inc., 2014).

[52] J. T. Margraf, H. Jung, C. Scheurer, and K. Reuter, "Exploring catalytic reaction networks with machine learning," Nat. Catal. **6**, 112–121 (2023).

[53] X. Yang, A. Bhowmik, T. Vegge, and H. A. Hansen, "Neural network potentials for accelerated metadynamics of oxygen reduction kinetics at Au–water interfaces," Chem. Sci. **14**, 3913–3922 (2023).

[54] K. Tran and Z. W. Ulissi, "Active learning across intermetallics to guide discovery of electrocatalysts for $CO_2$ reduction and $H_2$ evolution," Nat. Catal. **1**, 696–703 (2018).

[55] L. Foppa, C. Sutton, L. M. Ghiringhelli, S. De, P. Löser, S. A. Schunk, A. Schäfer, and M. Scheffler, "Learning design rules for selective oxidation catalysts from high-throughput experimentation and artificial intelligence," ACS Catal. **12**(4), 2223–2232 (2022).

[56] M. Khatamirad, E. Fako, C. Boscagli, M. Müller, F. Ebert, R. Naumann d'Alnoncourt, A. Schaefer, S. A. Schunk, I. Jevtovikj, F. Rosowski, and S. De, "A data-driven high-throughput workflow applied to promoted in-oxide catalysts for $CO_2$ hydrogenation to methanol," Catal. Sci. Technol. **13**(9), 2656–2661 (2023).

[57] S. Stocker, H. Jung, G. Csányi, C. F. Goldsmith, K. Reuter, and J. T. Margraf, "Estimating free energy barriers for heterogeneous catalytic reactions with machine learning potentials and umbrella integration," J. Chem. Theory Comput. **19**, 6796–6804 (2023).

[58] R. Tran, J. Lan, M. Shuaibi, B. M. Wood, S. Goyal, A. Das, J. Heras-domingo, A. Kolluru, A. Rizvi, N. Shoghi, A. Sriram, F. Therrien, J. Abed, O. Voznyy, E. H. Sargent, Z. Ulissi, and C. L. Zitnick, "The Open Catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts," ACS Catal. **13**, 3066–3084 (2023).

[59] M. Pourbaix, *ATLAS of Electrochemical Equilibria in Aqueous Solutions* (Pergamon Press, Oxford, 1966).

[60] M. Pourbaix, *Lectures on Electrochemical Corrosion* (Springer Science and Business Media, 1973).

[61] J. K. Nørskov, J. Rossmeisl, A. Logadottir, L. Lindqvist, J. R. Kitchin, T. Bligaard, and H. Jónsson, "Origin of the overpotential for oxygen reduction at a fuel-cell cathode," J. Phys. Chem. B **108**, 17886–17892 (2004).

[62] K. A. Persson, B. Waldwick, P. Lazic, and G. Ceder, "Prediction of solid-aqueous equilibria: Scheme to combine first-principles calculations of solids with experimental aqueous states," Phys. Rev. B **85**, 235438 (2012).

[63] A. K. Singh, L. Zhou, A. Shinde, S. K. Suram, J. H. Montoya, D. Winston, J. M. Gregoire, and K. A. Persson, "Electrochemical stability of metastable materials," Chem. Mater. **29**, 10159–10167 (2017).

[64] F. Abild-Pedersen, J. Greeley, F. Studt, J. Rossmeisl, T. R. Munter, P. G. Moses, E. Skúlason, T. Bligaard, and J. K. Nørskov, "Scaling properties of adsorption energies for hydrogen-containing molecules on transition-metal surfaces," Phys. Rev. Lett. **99**, 016105 (2007).

[65] J. Pérez-Ramírez and N. López, "Strategies to break linear scaling relationships," Nat. Catal. **2**, 971–976 (2019).

[66] R. García-Muelas and N. López, "Statistical learning goes beyond the $d$-band model providing the thermochemistry of adsorbates on transition metals," Nat. Commun. **10**, 4687 (2019).

[67] A. Tiwari, H. H. Heenen, A. S. Bjørnlund, D. Hochfilzer, K. Chan, and S. Horch, "Electrochemical oxidation of CO on Cu single crystals under alkaline conditions," ACS Energy Lett. **5**(11), 3437–3442 (2020).

[68] S. Dang, B. Qin, Y. Yang, H. Wang, J. Cai, Y. Han, S. Li, P. Gao, and Y. Sun, "Rationally designed indium oxide catalysts for $CO_2$ hydrogenation to methanol with high activity and selectivity," Sci. Adv. **6**(25), eaaz2060 (2020).

[69] A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein, and R. Q. Snurr, "Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery," Matter **4**(5), 1578–1597 (2021).

[70] A. S. Rosen, V. Fung, P. Huck, C. T. O'Donnell, M. K. Horton, D. G. Truhlar, K. A. Persson, J. M. Notestein, and R. Q. Snurr, "High-throughput predictions of metal–organic framework electronic properties: Theoretical challenges, graph neural networks, and data exploration," npj Comput. Mater. **8**(1), 112 (2022).

[71] J. Zeng, D. Zhang, D. Lu, P. Mo, Z. Li, Y. Chen, M. Rynik, L. Huang, Z. Li, S. Shi, Y. Wang, H. Ye, P. Tuo, J. Yang, Y. Ding, Y. Li, D. Tisi, Q. Zeng, H. Bao, Y. Xia, J. Huang, K. Muraoka, Y. Wang, J. Chang, F. Yuan, S. L. Bore, C. Cai, Y. Lin, B. Wang, J. Xu, J.-X. Zhu, C. Luo, Y. Zhang, R. E. A. Goodall, W. Liang, A. K. Singh, S. Yao, J. Zhang, R. Wentzcovitch, J. Han, J. Liu, W. Jia, D. M. York, W. E, R. Car, L. Zhang, and H. Wang, "DeePMD-kit v2: A software package for deep potential models," J. Chem. Phys. **159**, 054801 (2023).

[72] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff, "Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations," J. Am. Chem. Soc. **114**(25), 10024–10035 (1992).

[73] T. A. Manz and N. G. Limas, "Introducing DDEC6 atomic population analysis: Part 1. Charge partitioning theory and methodology," RSC Adv. **6**(53), 47771–47801 (2016).

[74] J. J. Potoff and J. I. Siepmann, "Vapor–liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen," AIChE J. **47**(7), 1676–1682 (2001).

[75] O. M. Yaghi, M. J. Kalmutzki, and C. S. Diercks, *Introduction to Reticular Chemistry: Metal-Organic Frameworks and Covalent Organic Frameworks* (John Wiley and Sons, 2019).

[76] D. Britt, H. Furukawa, B. Wang, T. G. Glover, and O. M. Yaghi, "Highly efficient separation of carbon dioxide by a metal-organic framework replete with open metal sites," Proc. Natl. Acad. Sci. U. S. A. **106**(49), 20637–20640 (2009).

[77] W. L. Queen, M. R. Hudson, E. D. Bloch, J. A. Mason, M. I. Gonzalez, J. S. Lee, D. Gygi, J. D. Howe, K. Lee, T. A. Darwish *et al.*, "Comprehensive study of carbon dioxide adsorption in the metal–organic frameworks $M_2$(dobdc) (M = Mg, Mn, Fe, Co, Ni, Cu, Zn)," Chem. Sci. **5**(12), 4569–4581 (2014).

[78] J. H. Choe, H. Kim, and C. S. Hong, "MOF-74 type variants for $CO_2$ capture," Mater. Chem. Front. **5**(14), 5172–5185 (2021).

[79] Ü. Kökçam-Demir, A. Goldman, L. Esrafili, M. Gharib, A. Morsali, O. Weingart, and C. Janiak, "Coordinatively unsaturated metal sites (open metal sites) in metal–organic frameworks: Design and applications," Chem. Soc. Rev. **49**(9), 2751–2798 (2020).

[80] B. Zheng, F. L. Oliveira, R. Neumann Barros Ferreira, M. Steiner, H. Hamann, G. X. Gu, and B. Luan, "Quantum informed machine-learning potentials for molecular dynamics simulations of $CO_2$'s chemisorption and diffusion in Mg-MOF-74," ACS Nano **17**(6), 5579–5587 (2023).

[81] L. Valenzano, B. Civalleri, S. Chavan, G. T. Palomino, C. O. Areán, and S. Bordiga, "Computational and experimental studies on the adsorption of CO, $N_2$, and $CO_2$ on Mg-MOF-74," J. Phys. Chem. C **114**(25), 11185–11191 (2010).

[82] M. Witman, S. Ling, S. Anderson, L. Tong, K. C. Stylianou, B. Slater, B. Smit, and M. Haranczyk, "*In silico* design and screening of hypothetical MOF-74 analogs and their experimental synthesis," Chem. Sci. **7**(9), 6263–6272 (2016).

[83] T. D. Bennett and S. Horike, "Liquid, glass and amorphous solid states of coordination polymers and metal–organic frameworks," Nat. Rev. Mater. **3**(11), 431–440 (2018).

[84] N. Castel and F.-X. Coudert, "Atomistic models of amorphous metal–organic frameworks," J. Phys. Chem. C **126**(16), 6905–6914 (2022).

[85] J. D. Evans, V. Bon, I. Senkovska, H.-C. Lee, and S. Kaskel, "Four-dimensional metal-organic frameworks," Nat. Commun. **11**(1), 2690 (2020).

[86] M. K. Taylor, T. Runcevski, J. Oktawiec, J. E. Bachman, R. L. Siegelman, H. Jiang, J. A. Mason, J. D. Tarver, and J. R. Long, "Near-perfect $CO_2/CH_4$ selectivity achieved through reversible guest templating in the flexible metal–organic framework Co(bdp)," J. Am. Chem. Soc. **140**(32), 10324–10331 (2018).

[87] V. Van Speybroeck, S. Vandenhaute, A. E. J. Hoffman, and S. M. J. Rogge, "Towards modeling spatiotemporal processes in metal–organic frameworks," Trends Chem. **3**, 605–619 (2021).

[88] J. Řezáč, K. E. Riley, and P. Hobza, "S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures," J. Chem. Theory Comput. **7**(8), 2427–2438 (2011).

[89] A. M. Reilly and A. Tkatchenko, "Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals," J. Chem. Phys. **139**, 024705 (2013).

[90] T. W. Keal and D. J. Tozer, "Semiempirical hybrid functional with improved performance in an extensive chemical assessment," J. Chem. Phys. **123**, 121103 (2005).

[91] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," Psychol. Learn. Motiv. **24**, 109–165 (1989).

[92] C. Chen, W. Ye, Y. Zuo, C. Zheng, and S. P. Ong, "Graph networks as a universal machine learning framework for molecules and crystals," Chem. Mater. **31**(9), 3564–3572 (2019).

[93] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, "Commentary: The materials project: A materials genome approach to accelerating materials innovation," APL Mater. **1**(1), 011002 (2013).

[94] C. Chen and S. P. Ong, "A universal graph deep learning interatomic potential for the periodic table," Nat. Comput. Sci. **2**(11), 718–728 (2022).

[95] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, "CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling," Nat. Mach. Intell. **5**(9), 1031–1041 (2023).

[96] K. Choudhary, B. DeCost, L. Major, K. Butler, J. Thiyagalingam, and F. Tavazza, "Unified graph neural network force-field for the periodic table: Solid state applications," Digital Discovery **2**(2), 346–355 (2023).

[97] K. Choudhary, K. F. Garrity, A. C. E. Reid, B. DeCost, A. J. Biacchi, A. R. Hight Walker, Z. Trautt, J. Hattrick-Simpers, A. G. Kusne, A. Centrone, A. Davydov, J. Jiang, R. Pachter, G. Cheon, E. Reed, A. Agrawal, X. Qian, V. Sharma, H. Zhuang, S. V. Kalinin, B. G. Sumpter, G. Pilania, P. Acar, S. Mandal, K. Haule, D. Vanderbilt, K. Rabe, and F. Tavazza, "The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design," npj Comput. Mater. **6**(1), 173 (2020).

[98] J. Klimeš, D. R. Bowler, and A. Michaelides, "Chemical accuracy for the van der Waals density functional," J. Phys.: Condens. Matter **22**(2), 022201 (2010).

[99] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling deep learning for materials discovery," Nature **624**, 80–85 (2023).

[100] D. Zhang, H. Bi, F.-Z. Dai, W. Jiang, L. Zhang, and H. Wang, "Pretraining of attention-based deep learning potential model for molecular simulation," npj Comput. Mater. **10**, 94 (2024).

[101] D. Zhang, X. Liu, X. Zhang, C. Zhang, C. Cai, H. Bi, Y. Du, X. Qin, A. Peng, J. Huang, B. Li, Y. Shan, J. Zeng, Y. Zhang, S. Liu, Y. Li, J. Chang, X. Wang, S. Zhou, J. Liu, X. Luo, Z. Wang, W. Jiang, J. Wu, Y. Yang, J. Yang, M. Yang, F.-Q. Gong, L. Zhang, M. Shi, F.-Z. Dai, D. M. York, S. Liu, T. Zhu, Z. Zhong, J. Lv, J. Cheng, W. Jia, M. Chen, G. Ke, W. E, L. Zhang, and H. Wang, "DPA-2: A large atomic model as a multi-task learner," npj Comput. Mater. **10**, 293 (2024).

[102] S. Takamoto, C. Shinagawa, D. Motoki, K. Nakago, W. Li, I. Kurata, T. Watanabe, Y. Yayama, H. Iriguchi, Y. Asano *et al.*, "Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements," Nat. Commun. **13**(1), 2991 (2022).

[103] S. Takamoto, S. Izumi, and J. Li, "TeaNet: Universal neural network interatomic potential inspired by iterative electronic relaxations," Comput. Mater. Sci. **207**, 111280 (2022).

[104] S. Takamoto, D. Okanohara, Q.-J. Li, and J. Li, "Towards universal neural network interatomic potential," J. Materiomics **9**(3), 447–454 (2023).

[105] J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost," Chem. Sci. **8**(4), 3192–3203 (2017).

[106] J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev, and A. E. Roitberg, "Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning," Nat. Commun. **10**(1), 2903 (2019).

[107] R. Zubatyuk, J. S. Smith, J. Leszczynski, and O. Isayev, "Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network," Sci. Adv. **5**(8), eaav6490 (2019).

[108] D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, Y. Pu, V. Kapil, W. C. Witt, I.-B. Magdău, D. J. Cole, and G. Csányi, "MACE-OFF: Short-range transferable machine learning force fields for organic molecules," J. Am. Chem. Soc. **147**(21), 17598–17611 (2025).

[109] N. Lopanitsyna, G. Fraux, M. A. Springer, S. De, and M. Ceriotti, "Modeling high-entropy transition metal alloys with alchemical compression," Phys. Rev. Mater. **7**(4), 045802 (2023).

[110] J. Schmidt, T. F. T. Cerqueira, A. H. Romero, A. Loew, F. Jäger, H.-C. Wang, S. Botti, and M. A. L. Marques, "Improving machine-learning models in materials science through large datasets," Mater. Today Phys. **48**, 101560 (2024).

[111] L. Barroso-Luque, M. Shuaibi, X. Fu, B. M. Wood, M. Dzamba, M. Gao, A. Rizvi, C. L. Zitnick, and Z. W. Ulissi, "Open Materials 2024 (OMat24) inorganic materials dataset and models," arXiv:2410.12771 (2024).

[112] J. Kim, J. Kim, J. Kim, J. Lee, Y. Park, Y. Kang, and S. Han, "Data-efficient multifidelity training for high-fidelity machine learning interatomic potentials," J. Am. Chem. Soc. **147**(1), 1042–1054 (2024).

[113] A. Bochkarev, Y. Lysogorskiy, and R. Drautz, "Graph atomic cluster expansion for semilocal interactions beyond equivariant message passing," Phys. Rev. X **14**, 021036 (2024).

[114] M. Neumann, J. Gin, B. Rhodes, S. Bennett, Z. Li, H. Choubisa, A. Hussey, and J. Godwin, "Orb: A fast, scalable neural network potential," arXiv:2410.22570v1 (2024).

[115] B. Rhodes, S. Vandenhaute, V. Šimkus, J. Gin, J. Godwin, T. Duignan, and M. Neumann, "Orb-v3: Atomistic simulation at scale," arXiv:2504.06231 (2025).

[116] H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen, C. Zeni, M. Horton, R. Pinsler, A. Fowler, D. Zügner, T. Xie, J. Smith, L. Sun, Q. Wang, L. Kong, C. Liu, H. Hao, and Z. Lu, "MatterSim: A deep learning atomistic model across elements, temperatures and pressures," arXiv:2405.04967v2 (2024).

[117] X. Fu, B. M. Wood, L. Barroso-Luque, D. S. Levine, M. Gao, M. Dzamba, and C. L. Zitnick, "Learning smooth and expressive interatomic potentials for physical property prediction," arXiv:2502.12147 (2025).

[118] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**(18), 3865 (1996).

[119] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, "Accurate and numerically efficient r$^2$SCAN meta-generalized gradient approximation," J. Phys. Chem. Lett. **11**, 8208–8215 (2020).

[120] R. Kingsbury, A. S. Gupta, C. J. Bartel, J. M. Munro, S. Dwaraknath, M. Horton, and K. A. Persson, "Performance comparison of $r^2$SCAN and SCAN metaGGA density functionals for solid materials via an automated, high-throughput computational workflow," Phys. Rev. Mater. **6**, 013801 (2022).

[121] T. M. Henderson, J. Paier, and G. E. Scuseria, "Accurate treatment of solids with the HSE screened hybrid," Phys. Status Solidi B **248**(4), 767–774 (2011).

[122] J. Harl, L. Schimka, and G. Kresse, "Assessing the quality of the random phase approximation for lattice constants and atomization energies of solids," Phys. Rev. B **81**, 115126 (2010).

[123] A. D. Kaplan, R. Liu, J. Qi, T. W. Ko, B. Deng, J. Riebesell, G. Ceder, K. A. Persson, and S. P. Ong, "A foundational potential energy surface dataset for materials," arXiv:2503.04070 (2025).

[124] S. A. Ghasemi, A. Hofstetter, S. Saha, and S. Goedecker, "Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network," Phys. Rev. B **92**, 045131 (2015).

[125] A. Grisafi and M. Ceriotti, "Incorporating long-range physics in atomic-scale machine learning," J. Chem. Phys. **151**, 204105 (2019).

[126] T. W. Ko, J. A. Finkler, S. Goedecker, and J. Behler, "A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer," Nat. Commun. **12**, 398 (2021).

[127] M. Vondrák, K. Reuter, and J. T. Margraf, "q-pac: A Python package for machine learned charge equilibration models," J. Chem. Phys. **159**, 054109 (2023).

[128] I. Novikov, B. Grabowski, F. Körmann, and A. Shapeev, "Magnetic moment tensor potentials for collinear spin-polarized materials reproduce different magnetic states of bcc Fe," npj Comput. Mater. **8**(1), 13 (2022).

[129] M. Rinaldi, M. Mrovec, A. Bochkarev, Y. Lysogorskiy, and R. Drautz, "Non-collinear magnetic atomic cluster expansion for iron," npj Comput. Mater. **10**, 12 (2023).

[130] J. Byggmästar, A. Hamedani, K. Nordlund, and F. Djurabekova, "Machine-learning interatomic potential for radiation damage and defects in tungsten," Phys. Rev. B **100**, 144105 (2019).

[131] J. L. A. Gardner, K. T. Baker, and V. L. Deringer, "Synthetic pre-training for neural-network interatomic potentials," Mach. Learn.: Sci. Technol. **5**, 015003 (2024).

[132] C. Ben Mahmoud, J. L. A. Gardner, and V. L. Deringer, "Data as the next challenge in atomistic machine learning," Nat. Comput. Sci. **4**, 384–387 (2024).

[133] A. P. Bartók, M. J. Gillan, F. R. Manby, and G. Csányi, "Machine-learning approach for one- and two-body corrections to density functional theory: Applications to molecular and condensed water," Phys. Rev. B **88**, 054104 (2013).

[134] P. O. Dral, A. Owens, A. Dral, and G. Csányi, "Hierarchical machine learning of potential energy surfaces," J. Chem. Phys. **152**, 204110 (2020).

[135] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), ISBN: 9781713807933.

[136] M. Geiger and T. Smidt, "e3nn: Euclidean neural networks," arXiv:2207.09453 (2022).

[137] D. P. Kovács, I. Batatia, E. S. Arany, and G. Csányi, "Evaluation of the MACE force field architecture: From medicinal chemistry to materials science," J. Chem. Phys. **159**, 044118 (2023).

[138] J. Ziegler, J. Biersack, and U. Littmark, *The Stopping and Range of Ions in Solids, Stopping and Ranges of Ions in Matter* (Pergamon, 1985).

[139] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in International Conference on Learning Representations, 2017.

[140] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," in *6th International Conference on Learning Representations, Vancouver, BC, Canada, April 30-May 3, 2018, Conference Track Proceedings* (OpenReview.net, 2018), https://openreview.net/forum?id=ryQu7f-RZ.

[141] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in 3rd International Conference on Learning Representations, San Diego, CA, USA, May 7–9, 2015.

[142] See https://docs.materialsproject.org/methodology/materials-methodology/calculation-details for Materials project calculation details; accessed 18 December 2023.

[143] See https://docs.materialsproject.org/methodology/materials-methodology/calculation-details/gga+u-calculations/hubbard-u-values for Materials project; accessed 12 December 2023.

[144] A. Jain, G. Hautier, S. P. Ong, C. J. Moore, C. C. Fischer, K. A. Persson, and G. Ceder, "Formation enthalpies by mixing GGA and GGA $+U$ calculations," Phys. Rev. B **84**, 045115 (2011).

[145] M. K. Horton, J. H. Montoya, M. Liu, and K. A. Persson, "High-throughput prediction of the ground-state collinear magnetic order of inorganic materials using density functional theory," npj Comput. Mater. **5**, 64 (2019).

[146] R. Sabatini, T. Gorni, and S. de Gironcoli, "Nonlocal van der Waals density functional made simple and efficient," Phys. Rev. B **87**, 041108 (2013).

[147] I.-C. Lin, A. P. Seitsonen, I. Tavernelli, and U. Rothlisberger, "Structure and dynamics of liquid water from ab initio molecular dynamics—Comparison of BLYP, PBE, and revPBE density functionals with and without van der Waals corrections," J. Chem. Theory Comput. **8**(10), 3902–3910 (2012).

[148] A. V. Terentjev, L. A. Constantin, and J. M. Pitarke, "Dispersion-corrected PBEsol exchange-correlation functional," Phys. Rev. B **98**, 214108 (2018).

[149] F. Formalik, M. Fischer, J. Rogacka, L. Firlej, and B. Kuchta, "Benchmarking of GGA density functionals for modeling structures of nanoporous, rigid and flexible MOFs," J. Chem. Phys. **149**, 064110 (2018).

[150] B. M. Axilrod and E. Teller, "Interaction of the van der Waals type between three atoms," J. Chem. Phys. **11**, 299–300 (1943).

[151] S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," J. Comput. Chem. **32**(7), 1456–1465 (2011).

[152] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, "Machine learning a general-purpose interatomic potential for silicon," Phys. Rev. X **8**(4), 041048 (2018).

[153] Y. Yoshida and G. Langouche, *Defects and Impurities in Silicon Materials* (Springer, 2015).

[154] D. A. Richie, J. Kim, S. A. Barr, K. R. A. Hazzard, R. Hennig, and J. W. Wilkins, "Complexity of small silicon self-interstitial defects," Phys. Rev. Lett. **92**(4), 045501 (2004).

[155] B. Sahli and W. Fichtner, "*Ab initio*molecular dynamics simulation of self-interstitial diffusion in silicon," Phys. Rev. B **72**(24), 245210 (2005).

[156] M. Posselt, F. Gao, and D. Zwicker, "Migration of di- and tri-interstitials in silicon," Nucl. Instrum. Methods Phys. Res., Sect. B **228**(1–4), 212–217 (2005).

[157] Y. A. Du, S. A. Barr, K. R. A. Hazzard, T. J. Lenosky, R. G. Hennig, and J. W. Wilkins, "Fast diffusion mechanism of silicon tri-interstitial defects," Phys. Rev. B **72**(24), 241306 (2005).

[158] P. Pichler, *Intrinsic Point Defects, Impurities, and Their Diffusion in Silicon* (Springer Science and Business Media, 2012).

[159] F. Dorner, Z. Sukurma, C. Dellago, and G. Kresse, "Melting Si: Beyond density functional theory," Phys. Rev. Lett. **121**(19), 195701 (2018).

[160] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, and S. R. Elliott, "Origins of structural and electronic transitions in disordered silicon," Nature **589**(7840), 59–64 (2021).

[161] N. Bernstein, B. Bhattarai, G. Csányi, D. A. Drabold, S. R. Elliott, and V. L. Deringer, "Quantifying chemical structure and machine-learned atomic energies in amorphous and liquid silicon," Angew. Chem., Int. Ed. **58**, 7057–7061 (2019).

[162] J. D. Morrow and V. L. Deringer, "Indirect learning and physically guided validation of interatomic potential models," J. Chem. Phys. **157**(10), 104105 (2022).

[163] V. L. Deringer, N. Bernstein, A. P. Bartók, M. J. Cliffe, R. N. Kerber, L. E. Marbella, C. P. Grey, S. R. Elliott, and G. Csányi, "Realistic atomistic structure of amorphous silicon from machine-learning-driven molecular dynamics," J. Phys. Chem. Lett. **9**(11), 2879–2885 (2018).

[164] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," J. Comput. Phys. **117**(1), 1–19 (1995).

[165] M. W. C. Dharma-wardana, D. D. Klug, and R. C. Remsing, "Liquid-liquid phase transitions in silicon," Phys. Rev. Lett. **125**, 075702 (2020).

[166] A. Hedler, S. L. Klaumünzer, and W. Wesch, "Amorphous silicon exhibits a glass transition," Nat. Mater. **3**, 804–809 (2004).

[167] P. F. McMillan, "Relaxing times for silicon," Nat. Mater. **3**, 755–756 (2004).

[168] J. S. Custer, M. O. Thompson, D. C. Jacobson, J. M. Poate, S. Roorda, W. C. Sinke, and F. Spaepen, "Density of amorphous Si," Appl. Phys. Lett. **64**, 437–439 (1994).

[169] S. Roorda, W. C. Sinke, J. M. Poate, D. C. Jacobson, S. Dierker, B. S. Dennis, D. J. Eaglesham, F. Spaepen, and P. Fuoss, "Structural relaxation and defect annihilation in pure amorphous silicon," Phys. Rev. B **44**(8), 3702 (1991).

[170] P. Grigorev, L. Frérot, F. Birks, A. Gola, J. Golebiowski, J. Grießer, J. L. Hörmann, A. Klemenz, G. Moras, W. G. Nöhring *et al.*, "matscipy: Materials science at the atomic scale with Python," J. Open Source Software **9**(93), 5668 (2024).

[171] V. L. Deringer and G. Csányi, "Machine learning based interatomic potential for amorphous carbon," Phys. Rev. B **95**, 094203 (2017).

[172] R. Jana, D. Savio, V. L. Deringer, and L. Pastewka, "Structural and elastic properties of amorphous carbon from simulated quenching at low rates," Modell. Simul. Mater. Sci. Eng. **27**, 085009 (2019).

[173] M. Qamar, M. Mrovec, Y. Lysogorskiy, A. Bochkarev, and R. Drautz, "Atomic cluster expansion for quantum-accurate large-scale simulations of carbon," J. Chem. Theory Comput. **19**, 5151–5167 (2023).

[174] T. K. Stenczel, Z. El-Machachi, G. Liepuoniute, J. D. Morrow, A. P. Bartók, M. I. J. Probert, G. Csányi, and V. L. Deringer, "Machine-learned acceleration for molecular dynamics in CASTEP," J. Chem. Phys. **159**, 044803 (2023).

[175] C. de Tomas, A. Aghajamali, J. L. Jones, D. J. Lim, M. J. López, I. Suarez-Martinez, and N. A. Marks, "Transferability in interatomic potentials for carbon," Carbon **155**, 624–634 (2019).

[176] G. A. Marchant, M. A. Caro, B. Karasulu, and L. B. Pártay, "Exploring the configuration space of elemental carbon with empirical and machine learned interatomic potentials," npj Comput. Mater. **9**, 131 (2023).

[177] M. A. Caro, V. L. Deringer, J. Koskinen, T. Laurila, and G. Csányi, "Growth mechanism and origin of high $sp^3$ content in tetrahedral amorphous carbon," Phys. Rev. Lett. **120**, 166101 (2018).

[178] P. Broqvist, J. Kullgren, M. J. Wolf, A. C. T. van Duin, and K. Hermansson, "ReaxFF force-field for ceria bulk, surfaces, and nanoparticles," J. Phys. Chem. C **119**(24), 13598–13609 (2015).

[179] T. P. Senftle, S. Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, T. Verstraelen, A. Grama, and A. C. T. van Duin, "The ReaxFF reactive force-field: Development, applications and future directions," npj Comput. Mater. **2**, 15011 (2016).

[180] J. Kullgren, K. Hermansson, and P. Broqvist, "Supercharged low-temperature oxygen storage capacity of ceria at the nanoscale," J. Phys. Chem. Lett. **4**(4), 604–608 (2013).

[181] A. Marronnier, G. Roma, S. Boyer-Richard, L. Pedesseau, J. M. Jancu, Y. Bonnassieux, C. Katan, C. C. Stoumpos, M. G. Kanatzidis, and J. Even, "Anharmonicity and disorder in the black phases of cesium lead iodide used for stable inorganic perovskite solar cells," ACS Nano **12**, 3477–3486 (2018).

[182] W. J. Baldwin, X. Liang, J. Klarbring, M. Dubajic, D. Dell'Angelo, C. Sutton, C. Caddeo, S. D. Stranks, A. Mattoni, A. Walsh, and G. Csányi, "Dynamic local structure in caesium lead iodide: Spatial correlation and transient domains," Small **20**(3), 2303565 (2023).

[183] E. Fransson, J. Wiktor, and P. Erhart, "Phase transitions in inorganic halide perovskites from machine-learned potentials," J. Phys. Chem. C **127**(28), 13773–13781 (2023).

[184] R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse, and M. Bokdam, "Phase transitions of hybrid perovskites simulated by machine-learning force fields trained on the fly with Bayesian inference," Phys. Rev. Lett. **122**, 225701 (2019).

[185] L. Mao, Y. Wu, C. C. Stoumpos, B. Traore, C. Katan, J. Even, M. R. Wasielewski, and M. G. Kanatzidis, "Tunable white-light emission in single-cation-templated three-layered 2D perovskites $(\mathrm{CH_3CH_2NH_3})_4\mathrm{Pb_3Br}_{10-x}\mathrm{Cl}_x$," J. Am. Chem. Soc. **139**(34), 11956–11963 (2017).

[186] E. R. Dohner, A. Jaffe, L. R. Bradshaw, and H. I. Karunadasa, "Intrinsic white-light emission from layered hybrid perovskites," J. Am. Chem. Soc. **136**(38), 13154–13157 (2014).

[187] C. Lermer, S. P. Harm, S. T. Birkhold, J. A. Jaser, C. M. Kutz, P. Mayer, L. Schmidt-Mende, and B. V. Lotsch, "Benzimidazolium lead halide perovskites: Effects of anion substitution and dimensionality on the bandgap," Z. Anorg. Allg. Chem. **642**(23), 1369–1376 (2016).

[188] Z. Tang, J. Guan, and A. M. Guloy, "Synthesis and crystal structure of new organic-based layered perovskites with 2,2′-biimidazolium cations," J. Mater. Chem. **11**, 479–482 (2001).

[189] D. Weber, "$\mathrm{CH_3NH_3PbX_3}$, ein Pb(II)-system mit kubischer perowskitstruktur/$\mathrm{CH_3NH_3PbX_3}$, a Pb(II)-system with cubic perovskite structure," Z. Naturforsch., B **33**(12), 1443–1445 (1978).

[190] A. K. Jena, A. Kulkarni, and T. Miyasaka, "Halide perovskite photovoltaics: Background, status, and future prospects," Chem. Rev. **119**(5), 3036–3103 (2019).

[191] D. H. Cao, C. C. Stoumpos, O. K. Farha, J. T. Hupp, and M. G. Kanatzidis, "2D homologous perovskites as light-absorbing materials for solar cell applications," J. Am. Chem. Soc. **137**(24), 7843–7850 (2015).

[192] N. Karimitari, W. J. Baldwin, E. W. Muller, Z. J. L. Bare, W. J. Kennedy, G. Csányi, and C. Sutton, "Accurate crystal structure prediction of new 2D hybrid organic–inorganic perovskites," J. Am. Chem. Soc. **146**(40), 27392–27404 (2024).

[193] T. Wang, X. He, M. Li, B. Shao, and T.-Y. Liu, "AIMD-Chig: Exploring the conformational space of a 166-atom protein *Chignolin* with *ab initio* molecular dynamics," Sci. Data **10**, 549 (2023).

[194] S. Agrawalla and A. C. T. van Duin, "Development and application of a reaxff reactive force field for hydrogen combustion," J. Phys. Chem. A **115**(6), 960–972 (2011).

[195] D. L. Baulch, C. T. Bowman, C. J. Cobos, R. A. Cox, T. Just, J. A. Kerr, M. J. Pilling, D. Stocker, J. Troe, W. Tsang, R. W. Walker, and J. Warnatz, "Evaluated kinetic data for combustion modeling: Supplement II," J. Phys. Chem. Ref. Data **34**, 757–1397 (2005).

[196] A. Ishikawa and Y. Tateyama, "Reaction energy benchmarks of hydrocarbon combustion by Gaussian basis and plane wave basis approaches," J. Comput. Chem. **40**(20), 1866–1873 (2019).

[197] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, "PACKMOL: A package for building initial configurations for molecular dynamics simulations," J. Comput. Chem. **30**(13), 2157–2164 (2009).

[198] A. Stukowski, "Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool," Modell. Simul. Mater. Sci. Eng. **18**(1), 015012 (2009).

[199] R. Steudel, *Elemental Sulfur and Sulfur-Rich Compounds II* (Springer Science and Business Media, 2003), Vol. 2.

[200] P. Lindgren, G. Kastlunger, and A. A. Peterson, "Scaled and dynamic optimizations of nudged elastic bands," J. Chem. Theory Comput. **15**(11), 5787–5793 (2019).

[201] S. Makri, C. Ortner, and J. R. Kermode, "A preconditioning scheme for minimum energy path finding methods," J. Chem. Phys. **150**, 094109 (2019).

[202] M. Trachta, O. Bludský, J. Vaculík, R. Bulánek, and M. Rubeš, "Investigation of Brønsted acidity in zeolites through adsorbates with diverse proton affinities," Sci. Rep. **13**, 12380 (2023).

[203] S. Melchionna, G. Ciccotti, and B. Lee Holian, "Hoover NPT dynamics for systems varying in shape and size," Mol. Phys. **78**(3), 533–544 (1993).

[204] S. Melchionna, "Constrained systems and statistical distribution," Phys. Rev. E **61**, 6165–6170 (2000).

[205] C. S. Cucinotta, A. Ruini, A. Catellani, and A. Stirling, "Ab initio molecular dynamics study of the keto–enol tautomerism of acetone in solution," ChemPhysChem **7**(6), 1229–1234 (2006).

[206] A. M. Elena (2023). "md driver," Zenodo; gitlab.com. https://gitlab.com/drFaustroll/lavello; https://doi.org/10.5281/zenodo.10432005

[207] D. A. Stevens and J. R. Dahn, "The mechanisms of lithium and sodium insertion in carbon materials," J. Electrochem. Soc. **148**, A803 (2001).

[208] J.-X. Huang, G. Csányi, J.-B. Zhao, J. Cheng, and V. L. Deringer, "First-principles study of alkali-metal intercalation in disordered carbon anode materials," J. Mater. Chem. A **7**, 19070–19080 (2019).

[209] M. Babar, H. L. Parks, G. Houchins, and V. Viswanathan, "An accurate machine learning calculator for the lithium-graphite system," J. Phys.: Energy **3**, 014005 (2020).

[210] A. R. Genreith-Schriever, A. Alexiu, G. S. Phillips, C. S. Coates, L. A. V. Nagle-Cocco, J. D. Bocarsly, F. N. Sayed, S. E. Dutton, and C. P. Grey, "Jahn-Teller distortions and phase transitions in $LiNiO_2$: Insights from *ab initio* molecular dynamics and variable-temperature x-ray diffraction," Chem. Mater. **36**, 2289–2303 (2024).

[211] L. A. V. Nagle-Cocco, A. R. Genreith-Schriever, J. M. A. Steele, C. Tacconis, J. D. Bocarsly, O. Mathon, J. C. Neuefeind, J. Liu, C. A. O'Keefe, A. L. Goodwin, C. P. Grey, J. S. O. Evans, and S. E. Dutton, "Displacive Jahn−Teller transition in $NaNiO_2$," J. Am. Chem. Soc. **146**(43), 29560–29574 (2024).

[212] D. Rodney, L. Ventelon, E. Clouet, L. Pizzagalli, and F. Willaime, "Ab initio modeling of dislocation core properties in metals and semiconductors," Acta Mater. **124**, 633–659 (2017).

[213] A. M. Goryaeva, J. Dérès, C. Lapointe, P. Grigorev, T. D. Swinburne, J. R. Kermode, L. Ventelon, J. Baima, and M.-C. Marinica, "Efficient and transferable machine learning potentials for the simulation of crystal defects in bcc Fe and W," Phys. Rev. Mater. **5**, 103803 (2021).

[214] P. Hiremath, S. Melin, E. Bitzek, and P. A. T. Olsson, "Effects of interatomic potential on fracture behaviour in single- and bicrystalline tungsten," Comput. Mater. Sci. **207**, 111283 (2022).

[215] D. Smirnova, S. Starikov, G. D. Leines, Y. Liang, N. Wang, M. N. Popov, I. A. Abrikosov, D. G. Sangiovanni, R. Drautz, and M. Mrovec, "Atomistic description of self-diffusion in molybdenum: A comparative theoretical study of non-Arrhenius behavior," Phys. Rev. Mater. **4**(1), 013605 (2020).

[216] C. Yang and L. Qi, "Modified embedded-atom method potential of niobium for studies on mechanical properties," Comput. Mater. Sci. **161**, 351–363 (2019).

[217] M. Čák, T. Hammerschmidt, J. Rogal, V. Vitek, and R. Drautz, "Analytic bond-order potentials for the bcc refractory metals Nb, Ta, Mo and W," J. Phys.: Condens. Matter **26**(19), 195501 (2014).

[218] P. W. Ma and S. L. Dudarev, "Universality of point defect structure in body-centered cubic metals," Phys. Rev. Mater. **3**(1), 013605 (2019).

[219] S. Starikov, P. Grigorev, and P. A. T. Olsson, "Angular-dependent interatomic potential for large-scale atomistic simulation of W-Mo-Nb ternary alloys," Comput. Mater. Sci. **233**, 112734 (2024).

[220] L. Ventelon and F. Willaime, "Generalized stacking-faults and screw-dislocation core-structure in bcc iron: A comparison between *ab initio* calculations and empirical potentials," Philos. Mag. **90**, 1063 (2010).

[221] L. Dezerald, L. Ventelon, E. Clouet, C. Denoual, D. Rodney, and F. Willaime, "*Ab initio* modeling of the two-dimensional energy landscape of screw dislocations in bcc transition metals," Phys. Rev. B **89**, 024104 (2014).

[222] P. Grigorev, A. M. Goryaeva, M.-C. Marinica, J. R. Kermode, and T. D. Swinburne, "Calculation of dislocation binding to helium-vacancy defects in tungsten using hybrid ab initio-machine learning methods," Acta Mater. **247**, 118734 (2023).

[223] T. D. Swinburne and J. R. Kermode, "Computing energy barriers for rare events from hybrid quantum/classical simulations through the virtual work principle," Phys. Rev. B **96**, 144102 (2017).

[224] P. Grigorev, T. D. Swinburne, and J. R. Kermode, "Hybrid quantum/classical study of hydrogen-decorated screw dislocations in tungsten: Ultrafast pipe diffusion, core reconstruction, and effects on glide mechanism," Phys. Rev. Mater. **4**(2), 023601 (2020).

[225] R. H. Doremus, "Diffusion in alumina," J. Appl. Phys. **100**, 101301 (2006).

[226] E. L. Kolsbjerg, M. N. Groves, and B. Hammer, "An automated nudged elastic band method," J. Chem. Phys. **145**, 094107 (2016).

[227] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, "First principles methods using CASTEP," Z. Kristallogr. - Cryst. Mater. **220**(5–6), 567–570 (2005).

[228] C. J. Pickard and R. J. Needs, "*Ab initio* random structure searching," J. Phys.: Condens. Matter **23**(5), 053201 (2011).

[229] N. Bernstein, G. Csányi, and V. L. Deringer, "De novo exploration and self-guided learning of potential-energy surfaces," npj Comput. Mater. **5**(1), 99 (2019).

[230] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, and A. R. Oganov, "Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning," Phys. Rev. B **99**(6), 064114 (2019).

[231] C. J. Pickard, "Ephemeral data derived potentials for random structure search," Phys. Rev. B **106**(1), 014102 (2022).

[232] D. Vanderbilt, "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism," Phys. Rev. B **41**(11), 7892 (1990).

[233] M. Hart, J. Chen, A. Michaelides, A. Sella, M. S. P. Shaffer, and C. G. Salzmann, "One-dimensional arsenic allotropes: Polymerization of yellow arsenic inside single-wall carbon nanotubes," Angew. Chem. **130**(36), 11823–11827 (2018).

[234] D. Schiferl and C. S. Barrett, "The crystal structure of arsenic at 4.2, 78 and 299 °K," J. Appl. Crystallogr. **2**(1), 30–36 (1969).

[235] P. M. Smith, A. J. Leadbetter, and A. J. Apling, "The structures of orthorhombic and vitreous arsenic," Philos. Mag. **31**(1), 57–64 (1975).

[236] P. Silas, J. R. Yates, and P. D. Haynes, "Density-functional investigation of the rhombohedral to simple-cubic phase transition of arsenic," Phys. Rev. B **78**(17), 174101 (2008).

[237] M. Southard and D. Green, *Perry's Chemical Engineers' Handbook*, 9th ed. (McGraw-Hill Education, 2018).

[238] L. Liu, M. Krack, and A. Michaelides, "Density oscillations in a nanoscale water film on salt: Insight from ab initio molecular dynamics," J. Am. Chem. Soc. **130**, 8572–8573 (2008).

[239] Y. Chiang (2023), "Muse: A Python package for fast building amorphous solids and liquid mixtures," Zenodo; github.com. https://doi.org/10.5281/zenodo.10369245

[240] G. I. L. van Oudenaren, J. A. Ocadiz-Flores, and A. L. Smith, "Coupled structural-thermodynamic modelling of the molten salt system NaCl-$UCl_3$," J. Mol. Liq. **342**, 117470 (2021).

[241] D. A. Andersson and B. W. Beeler, "*Ab initio* molecular dynamics (AIMD) simulations of nacl, $UCl_3$ and NaCl-$UCl_3$ molten salts," J. Nucl. Mater. **568**, 153836 (2022).

[242] I.-B. Magdău, D. J. Arismendi-Arrieta, H. E. Smith, C. P. Grey, K. Hermansson, and G. Csányi, "Machine learning force fields for molecular liquids: Ethylene carbonate/ethyl methyl carbonate binary solvent," npj Comput. Mater. **9**(1), 146 (2023).

[243] L. P. N. Rebelo, V. Najdanovic-Visak, Z. P. Visak, M. Nunes da Ponte, J. Szydlowski, C. A. Cerdeiriña, J. Troncoso, L. Romaní, J. M. S. S. Esperança, H. J. R. Guedes, and H. C. de Sousa, "A detailed thermodynamic analysis of [C4mim][BF4] + water as a case study to model ionic liquid aqueous solutions," Green Chem. **6**(8), 369–381 (2004).

[244] E. Perlt, P. Ray, A. Hansen, F. Malberg, S. Grimme, and B. Kirchner, "Finding the best density functional approximation to describe interaction energies and structures of ionic liquids in molecular dynamics studies," J. Chem. Phys. **148**(19), 193835 (2018).

[245] F. Zills, M. R. Schäfer, N. Segreto, J. Kästner, C. Holm, and S. Tovey, "Collaboration on machine-learned potentials with IPSuite: A modular framework for learning-on-the-fly," J. Phys. Chem. B **128**, 3662–3676 (2024).

[246] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," J. Chem. Inf. Comput. Sci. **28**, 31–36 (1988).

[247] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1–2**, 19–25 (2015).

[248] J. N. Canongia Lopes, J. Deschamps, and A. A. H. Pádua, "Modeling ionic liquids using a systematic all-atom force field," J. Phys. Chem. B **108**, 2038–2047 (2004).

[249] E. Gregoryanz, C. Ji, P. Dalladay-Simpson, B. Li, R. T. Howie, and H.-K. Mao, "Everything you always wanted to know about metallic hydrogen but were afraid to ask," Matter Radiat. Extremes **5**(3), 038101 (2020).

[250] I. B. Magdău, M. Marqués, B. Borgulya, and G. J. Ackland, "Simple thermodynamic model for the hydrogen phase diagram," Phys. Rev. B **95**(9), 094107 (2017).

[251] I. B. Magdău, F. Balm, and G. J. Ackland, "Theory of high pressure hydrogen, made simple," J. Phys.: Conf. Ser. **950**(4), 042059 (2017).

[252] C. J. Pickard, M. Martinez-Canales, and R. J. Needs, "Density functional theory study of phase IV of solid hydrogen," Phys. Rev. B **85**(21), 214114 (2012).

[253] I. B. Magdău and G. J. Ackland, "Identification of high-pressure phases III and IV in hydrogen: Simulating Raman spectra using molecular dynamics," Phys. Rev. B **87**(17), 174110 (2013).

[254] R. T. Howie, I. B. Magdău, A. F. Goncharov, G. J. Ackland, and E. Gregoryanz, "Phonon localization by mass disorder in dense hydrogen-deuterium binary alloy," Phys. Rev. Lett. **113**(17), 175501 (2014).

[255] I. B. Magdău and G. J. Ackland, "Infrared peak splitting from phonon localization in solid hydrogen," Phys. Rev. Lett. **118**(14), 145701 (2017).

[256] P. I. C. Cooke, I. B. Magdău, M. Peña-Alvarez, V. Afonina, P. Dalladay-Simpson, X.-D. Liu, R. T. Howie, E. Gregoryanz, and G. J. Ackland, "Raman signal from a hindered hydrogen rotor," Phys. Rev. B **102**(6), 064102 (2020).

[257] B. Cheng, G. Mazzola, C. J. Pickard, and M. Ceriotti, "Evidence for supercritical behaviour of high-pressure liquid hydrogen," Nature **585**(7824), 217–220 (2020).

[258] H. Zong, H. Wiebe, and G. J. Ackland, "Understanding high pressure molecular hydrogen with a hierarchical machine-learned potential," Nat. Commun. **11**(1), 5014 (2020).

[259] T. Bischoff, B. Jäckl, and M. Rupp, "Hydrogen under pressure as a benchmark for machine-learning interatomic potentials," arXiv:2409.13390 (2024).

[260] G. J. Ackland and J. S. Loveday, "Structures of solid hydrogen at 300 K," Phys. Rev. B **101**(9), 094104 (2020).

[261] S. Frueh, R. Kellett, C. Mallery, T. Molter, W. S. Willis, C. King'ondu, and S. L. Suib, "Pyrolytic decomposition of ammonia borane to boron nitride," Inorg. Chem. **50**, 783–792 (2011).

[262] B. Deng, Y. Choi, P. Zhong, J. Riebesell, S. Anand, Z. Li, K. Jun, K. A. Persson, and G. Ceder, "Systematic softening in universal machine learning interatomic potentials," npj Comput. Mater. **11**(1), 9 (2025).

[263] A. Marfavi, P. Kavianpour, and L. M. Rendina, "Carboranes in drug discovery, chemical biology and molecular imaging," Nat. Rev. Chem. **6**, 486–504 (2022).

[264] I. B. Sivaev and V. I. Bregadze, "Chemistry of nickel and iron bis(dicarbollides). A review," J. Organomet. Chem. **614-615**, 27–36 (2000).

[265] C. A. Brown and M. L. McKee, "Rearrangements in icosahedral boranes and carboranes revisited," J. Mol. Model. **12**, 653–664 (2006).

[266] F. Neese, "Software update: The orca program system—Version 5.0," Wiley Interdiscip. Rev.:Comput. Mol. Sci. **12**, e1606 (2022).

[267] A. S. Rosen, J. M. Notestein, and R. Q. Snurr, "Comprehensive phase diagrams of MoS$_2$ edge sites using dispersion-corrected DFT free energy calculations," J. Phys. Chem. C **122**(27), 15318–15329 (2018).

[268] D. Kieczka, T. Durrant, K. Milton, K. E. J. Goh, M. Bosman, and A. Shluger, "Defects in WS$_2$ monolayer calculated with a nonlocal functional: Any difference from GGA?," Electron. Struct. **5**(2), 024001 (2023).

[269] Z. Fang, M. P. Confer, Y. Wang, Q. Wang, M. R. Kunz, E. J. Dufek, B. Liaw, T. M. Klein, D. A. Dixon, and R. Fushimi, "Formation of surface impurities on lithium–nickel–manganese–cobalt oxides in the presence of CO$_2$ and H$_2$O," J. Am. Chem. Soc. **143**(27), 10261–10274 (2021).

[270] S. P. Niblett, P. Kourtis, I.-B. Magdău, C. P. Grey, and G. Csányi, "Transferability of data sets between machine-learned interatomic potential algorithms," J. Chem. Theory Comput. **21**(12), 6096–6112 (2025).

[271] S. Ju, J. You, G. Kim, Y. Park, H. An, and S. Han, "Application of pretrained universal machine-learning interatomic potential for physicochemical simulation of liquid electrolytes in Li-ion battery," Digital Discovery **4**, 1544–1559 (2025).

[272] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus *et al.*, "The atomic simulation environment—A Python library for working with atoms," J. Phys.: Condens. Matter **29**(27), 273002 (2017).

[273] E. Vanden-Eijnden and G. Ciccotti, "Second-order integrators for Langevin equations with holonomic constraints," Chem. Phys. Lett. **429**(1–3), 310–316 (2006).

[274] E. Kasoar, P. Austin, H. Devereux, K. Harris, D. Mason, J. Wilkins, F. Zanca, and A. Elena (2025). "janus-core (v0.8.5)," Zenodo; github.com. https://github.com/stfc/janus-core; https://doi.org/10.5281/zenodo.17259290.

[275] H.-C. Wang, S. Botti, and M. A. L. Marques, "Predicting stable crystalline compounds using chemical similarity," npj Comput. Mater. **7**(1), 12 (2021).

[276] H. Glawe, A. Sanna, E. K. U. Gross, and M. A. L. Marques, "The optimal one dimensional periodic table: A modified Pettifor chemical scale from data mining," New J. Phys. **18**(9), 093011 (2016).

[277] J. Riebesell, R. E. A. Goodall, P. Benner, Y. Chiang, B. Deng, G. Ceder, M. Asta, A. A. Lee, A. Jain, and K. A. Persson, "A framework to evaluate machine learning crystal stability predictions," Nat. Mach. Intell. **7**, 836–847 (2025).

[278] D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, and S. Rehme, "Recent developments in the inorganic crystal structure database: Theoretical crystal structure data and related features," J. Appl. Crystallogr. **52**, 918–925 (2019).

[279] N. E. R. Zimmermann and A. Jain, "Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity," RSC Adv. **10**, 6063–6081 (2020).

[280] Y. S. Al-Hamdani, M. Rossi, D. Alfè, T. Tsatsoulis, B. Ramberger, J. G. Brandenburg, A. Zen, G. Kresse, A. Grüneis, A. Tkatchenko, and A. Michaelides, "Properties of the water to boron nitride interaction: From zero to two dimensions with benchmark accuracy," J. Chem. Phys. **147**, 044710 (2017).

[281] J. G. Brandenburg, A. Zen, M. Fitzner, B. Ramberger, G. Kresse, T. Tsatsoulis, A. Grüneis, A. Michaelides, and D. Alfè, "Physisorption of water on graphene: Subchemical accuracy from many-body electronic structure methods," J. Phys. Chem. Lett. **10**, 358–368 (2019).

[282] C. Ehlert, A. Piras, and G. Gryn'ova, "CO$_2$ on graphene: Benchmarking computational approaches to noncovalent interactions," ACS Omega **8**, 35768–35778 (2023).

[283] T. Tsatsoulis, S. Sakong, A. Groß, and A. Grüneis, "Reaction energetics of hydrogen on Si(100) surface: A periodic many-electron theory study," J. Chem. Phys. **149**, 244105 (2018).

[284] T. Tsatsoulis, F. Hummel, D. Usvyat, M. Schütz, G. H. Booth, S. S. Binnie, M. J. Gillan, D. Alfè, A. Michaelides, and A. Grüneis, "A comparison between quantum chemistry and quantum Monte Carlo techniques for the adsorption of water on the (001) LiH surface," J. Chem. Phys. **146**, 204108 (2017).

[285] H.-Z. Ye and T. C. Berkelbach, "Ab initio surface chemistry with chemical accuracy," arXiv:2309.14640 (2023).

[286] P. G. Lustemberg, P. N. Plessow, Y. Wang, C. Yang, A. Nefedov, F. Studt, C. Wöll, and M. V. Ganduglia-Pirovano, "Vibrational frequencies of cerium-oxide-bound CO: A challenge for conventional DFT methods," Phys. Rev. Lett. **125**, 256101 (2020).

[287] B. X. Shi, A. Zen, V. Kapil, P. R. Nagy, A. Grüneis, and A. Michaelides, "Many-body methods for surface chemistry come of age: Achieving consensus with experiments," J. Am. Chem. Soc. **145**, 25372–25381 (2023).

[288] N. Hanikel, X. Pei, S. Chheda, H. Lyu, W. Jeong, J. Sauer, L. Gagliardi, and O. M. Yaghi, "Evolution of water structures in metal-organic frameworks for improved atmospheric water harvesting," Science **374**, 454–459 (2021).

[289] F. Berger, M. Rybicki, and J. Sauer, "Molecular dynamics with chemical accuracy–alkane adsorption in acidic zeolites," ACS Catal. **13**, 2011–2024 (2023).

[290] F. Berger and J. Sauer, "Dimerization of linear butenes and pentenes in an acidic zeolite (H-MFI)," Angew. Chem., Int. Ed. **60**, 3529–3533 (2021).

[291] G. Kresse and J. Hafner, "*Ab initio* molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium," Phys. Rev. B **49**, 14251 (1994).

[292] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," Comput. Mater. Sci. **6**, 15 (1996).

[293] J. K. Nørskov, F. Abild-Pedersen, F. Studt, and T. Bligaard, "Density functional theory in surface chemistry and catalysis," Proc. Natl. Acad. Sci. U. S. A. **108**, 937–943 (2011).

[294] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, and Z. Ulissi, "Open Catalyst 2020 (OC20) dataset and community challenges," ACS Catal. **11**, 6059–6072 (2021).

[295] B. Cantor, I. T. H. Chang, P. Knight, and A. J. B. Vincent, "Microstructural development in equiatomic multicomponent alloys," Mater. Sci. Eng.: A **375–377**, 213–218 (2004).

[296] A. Mazitov, M. A. Springer, N. Lopanitsyna, G. Fraux, S. De, and M. Ceriotti, "Surface segregation in high-entropy alloys from alchemical machine learning," J. Phys.: Mater. **7**, 025007 (2024).

[297] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen *et al.*, "LAMMPS - A flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," Comput. Phys. Commun. **271**, 108171 (2022).

[298] C. R. Trott, D. Lebrun-Grandié, D. Arndt, J. Ciesko, V. Dang, N. Ellingwood, R. Gayatri, E. Harvey, D. S. Hollman, D. Ibanez, N. Liber, J. Madsen, J. Miles, D. Poliakoff, A. Powell, S. Rajamanickam, M. Simberg, D. Sunderland, B. Turcksin, and J. Wilke, "Kokkos 3: Programming model extensions for the exascale era," IEEE Trans. Parallel Distrib. Syst. **33**(4), 805–817 (2022).

[299] L. Barroso-Luque, M. Shuaibi, X. Fu, B. M. Wood, M. Dzamba, M. Gao, A. Rizvi, C. L. Zitnick, and Z. W. Ulissi, "Open Materials 2024 (OMat24) inorganic materials dataset and models," arXiv:2410.12771v1 (2024).

[300] R. P. Stoffel, C. Wessel, M.-W. Lumey, and R. Dronskowski, "Ab initio thermochemistry of solid-state materials," Angew. Chem., Int. Ed. **49**(31), 5242 (2010).

[301] C. J. Bartel, "Review of computational approaches to predict the thermodynamic stability of inorganic solids," J. Mater. Sci. **57**, 10475 (2022).

[302] J. D. Morrow, J. L. A. Gardner, and V. L. Deringer, "How to validate machine-learned interatomic potentials," J. Chem. Phys. **158**, 121501 (2023).

[303] J. George, G. Hautier, A. P. Bartók, G. Csányi, and V. L. Deringer, "Combining phonon accuracy with high transferability in Gaussian approximation potential models," J. Chem. Phys. **153**, 044104 (2020).

[304] A. Togo, "First-principles phonon calculations with phonopy and phono3py," J. Phys. Soc. Jpn. **92**(1), 012001 (2023).

[305] A. Togo, L. Chaput, T. Tadano, and I. Tanaka, "Implementation strategies in phonopy and phono3py," J. Phys.: Condens. Matter **35**(35), 353001 (2023).

[306] A. M. Ganose, H. Sahasrabuddhe, M. Asta, K. Beck, T. Biswas, A. Bonkowski, J. Bustamante, X. Chen, Y. Chiang, D. C. Chrzan *et al.*, "Atomate2: Modular workflows for materials science," Digital Discovery **4**, 1944–1973 (2025).

[307] See https://github.com/materialsproject/atomate2 for atomate2.

[308] A. Dunn, Q. Wang, A. Ganose, D. Dopp, and A. Jain, "Benchmarking materials property prediction methods: The Matbench test set and Automatminer reference algorithm," npj Comput. Mater. **6**(1), 138 (2020).

[309] M. T. Agne, R. Hanus, and G. J. Snyder, "Minimum thermal conductivity in the context of diffuson-mediated thermal transport," Energy Environ. Sci. **11**(3), 609–616 (2018).

[310] W. Voigt, *Lehrbuch Der Kristallphysik: (Mit Ausschluss Der Kristalloptik)* (B.G. Teubner, 1910).

[311] A. Reuss, "Berechnung der fließgrenze von mischkristallen auf grund der plastizitätsbedingung für einkristalle," ZAMM - J. Appl. Math. Mech./Z. Angew. Math. Mech. **9**(1), 49–58 (1929).

[312] R. Hill, "The elastic behaviour of a crystalline aggregate," Proc. Phys. Soc. A **65**(5), 349 (1952).

[313] J. Riebesell, E. Liu, J. Qi, T. W. Ko, and S. P. Ong, "MatCalc: A Python library for calculating materials properties," github.com (2023), released July 2023, https://github.com/materialsvirtuallab/matcalc.

[314] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in Statistics: Methodology and Distribution* (Springer, 1992), pp. 492–518.

[315] C. M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, 1995).

[316] J. Dodge, G. Ilharco, R. Schwartz, A. Farhadi, H. Hajishirzi, and N. Smith, "Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping," arXiv:2002.06305 (2020).

[317] H. Liu, D. Tam, M. Mohammed, J. Mohta, T. Huang, M. Bansal, and C. Raffel, "Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning," in *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, Inc., 2022), ISBN: 9781713871088.

[318] T. Lesort, H. Caselles-Dupré, M. Garcia-Ortiz, A. Stoian, and D. Filliat, "Generative models from the perspective of continual learning," in 2019 International Joint Conference on Neural Networks (IJCNN), 2019.

[319] A. Wang, R. Kingsbury, M. McDermott, M. Horton, A. Jain, S. P. Ong, S. Dwaraknath, and K. A. Persson, "A framework for quantifying uncertainty in DFT energy corrections," Sci. Rep. **11**, 15496 (2021).

[320] R. S. Kingsbury, A. S. Rosen, A. S. Gupta, J. M. Munro, S. P. Ong, A. Jain, S. Dwaraknath, M. K. Horton, and K. A. Persson, "A flexible and scalable scheme for mixing computed formation energies from different levels of theory," npj Comput. Mater. **8**(1), 195 (2022).

[321] J. Riebesell (2022). "Pymatviz: Visualization toolkit for materials informatics," Zenodo; github.com. https://github.com/janosh/pymatviz; https://doi.org/10.5281/zenodo.7486816

[322] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, "Comparing molecules and solids across structural and alchemical space," Phys. Chem. Chem. Phys. **18**(20), 13754–13769 (2016).

[323] G. Fraux, R. K. Cersonsky, and M. Ceriotti, "Chemiscope: Interactive structure-property explorer for materials and molecules," J. Open Source Software **5**(51), 2117 (2020).

[324] F. Bigi, S. Chong, M. Ceriotti, and F. Grasselli, "A prediction rigidity formalism for low-cost uncertainties in trained neural networks," Mach. Learn.: Sci. Technol. **5**, 045018 (2024).

[325] P. S. Laplace, "Mémoire de mathématique et de physique," in *Tome Sixième* (A Paris De L'Imprimerie Royale, 1774), Vol. 6.

[326] E. Daxberger, A. Kristiadi, A. Immer, R. Eschenhagen, M. Bauer, and P. Hennig, "Laplace redux – effortless Bayesian deep learning," in *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, edited by M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, and J. Wortman Vaughan (Curran Associates, Inc., 2021), ISBN: 9781713845393.

[327] A. N. Angelopoulos and S. Bates, "A gentle introduction to conformal prediction and distribution-free uncertainty quantification," arXiv:2107.07511 (2021).

[328] Y. Bai, S. Mei, H. Wang, Y. Zhou, and C. Xiong, "Efficient and differentiable conformal prediction with general function classes," arXiv:2202.11091 (2022).

[329] Y. Hu, J. Musielewicz, Z. W. Ulissi, and A. J. Medford, "Robust and scalable uncertainty estimation with conformal prediction for machine-learned interatomic potentials," Mach. Learn.: Sci. Technol. **3**(4), 045028 (2022).

[330] G. Shafer and V. Vovk, "A tutorial on conformal prediction," J. Mach. Learn. Res. **9**(3), 371–421 (2008).

[331] C. H. Ho, C. Ortner, and Y. Wang, "Flexible uncertainty calibration for machine-learned interatomic potentials," arXiv:2510.00721 (2025).

[332] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning* (Springer, 2006), Vol. 4.