

PAPER • OPEN ACCESS

Generalization of graph-based active learning relaxation strategies across materials

To cite this article: Xiaoxiao Wang *et al* 2024 *Mach. Learn.: Sci. Technol.* **5** 025018

View the [article online](#) for updates and enhancements.

You may also like

- [Actively learning costly reward functions for reinforcement learning](#)
André Eberhard, Houssam Metni, Georg Fahland et al.
- [Efficient interpolation of molecular properties across chemical compound space with low-dimensional descriptors](#)
Yun-Wen Mao and Roman V Krems
- [Laziness, barren plateau, and noises in machine learning](#)
Junyu Liu, Zexi Lin and Liang Jiang



PAPER

OPEN ACCESS

RECEIVED

9 November 2023

REVISED

9 March 2024

ACCEPTED FOR PUBLICATION

26 March 2024

PUBLISHED

22 April 2024

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Generalization of graph-based active learning relaxation strategies across materials

Xiaoxiao Wang^{1,6} , Joseph Musielewicz^{1,6} , Richard Tran¹ , Sudheesh Kumar Ethirajan² ,
Xiaoyan Fu³ , Hilda Mera¹ , John R Kitchin^{1,*} , Rachel C Kurchin^{4,*} and Zachary W Ulissi^{1,5,*}

¹ Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA, United States of America

² Department of Chemical Engineering, University of California, Davis, CA, United States of America

³ Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, People's Republic of China

⁴ Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA, United States of America

⁵ Fundamental Artificial Intelligence Research, Meta Platforms, Inc., Menlo Park, CA, United States of America

⁶ These authors contributed equally to this work.

* Authors to whom any correspondence should be addressed.

E-mail: jkitchin@andrew.cmu.edu, rkurchin@cmu.edu and zulissi@meta.com

Keywords: machine learned potential, geometry optimization, graph neural network

Supplementary material for this article is available [online](#)

Abstract

Although density functional theory (DFT) has aided in accelerating the discovery of new materials, such calculations are computationally expensive, especially for high-throughput efforts. This has prompted an explosion in exploration of machine learning (ML) assisted techniques to improve the computational efficiency of DFT. In this study, we present a comprehensive investigation of the broader application of Finetuna, an active learning framework to accelerate structural relaxation in DFT with prior information from Open Catalyst Project pretrained graph neural networks. We explore the challenges associated with out-of-domain systems: alcohol ($C_{>2}$) on metal surfaces as larger adsorbates, metal oxides with spin polarization, and three-dimensional (3D) structures like zeolites and metal organic frameworks. By pre-training ML models on large datasets and fine-tuning the model along the simulation, we demonstrate the framework's ability to conduct relaxations with fewer DFT calculations. Depending on the similarity of the test systems to the training systems, a more conservative querying strategy is applied. Our best-performing Finetuna strategy reduces the number of DFT single-point calculations by 80% for alcohols and 3D structures, and 42% for oxide systems.

1. Introduction

The urgent need to address climate and societal challenges has prompted the search for innovative solutions, and new material design and discovery have emerged as critical pathways. For instance, new catalysts are being explored for various crucial chemical reactions such as nitrogen reduction to replace the energy-intensive Haber-Bosch process [1, 2], and alcohol dehydrogenation for energy storage applications [3]. Metal oxides are being investigated as catalysts for oxygen reduction and evolution reactions in fuel cells [4, 5]. With the continued advancements in simulation tools and increased computing power, we can harness the power of computational chemistry to further accelerate the process of materials discovery. By utilizing computational techniques, researchers can rapidly screen a vast space of chemical compositions, structures, and properties that would otherwise be prohibitively expensive and time-consuming to explore experimentally [6, 7]. Simulations can also provide valuable insights into the underlying reaction mechanisms, allowing for the optimization of material performance.

However, the computational cost of first-principles simulations becomes infeasible for large chemical spaces. In these cases, one prominent approach in accelerating material discovery is through the use of machine learning (ML) methods, which leverage large databases and advanced algorithms to predict the properties and performance of materials. A properly-trained ML model is vastly lower in computational cost

than computational tools such as density functional theory (DFT) and can have comparable accuracy, provided it is being applied on sufficiently similar systems to the training data [8]. The development of large material datasets, such as the Open Catalyst 2020 (OC20) Dataset and Open Catalyst 2022 (OC22) Dataset for metal and metal-oxide materials [9, 10], the CoRE [11], CSD MOF [12], QMOF [13], MOFX-DB [14] for metal organic frameworks (MOFs), and Database of Zeolite Structures [15] and Zeo-1 [16] for zeolites, has enabled data-driven materials modeling research on these materials classes. Remarkable progress has been made in the development of ML models for accurate calculation of interatomic potentials and prediction of materials properties. Graph neural network (GNN) models are specifically designed to handle the inherent graph-like structure of molecules and crystals, where atoms and bonds are represented as nodes and edges in a graph. This allows GNN models to capture the complex interactions between atoms. GNN models such as GemNet [17], GemNet-OC [18], SCN [19], eSCN [20], and M3GNet [21], can learn vast amounts of chemical information from training data, facilitating the efficient screening of materials and identifying promising candidates for specific applications. For example, the present leading ML model on the Open Catalyst Project (OCP) Leaderboard can achieve a total energy mean absolute error (MAE) as low as 0.22 eV on OC20-like catalyst structures.

Training ML models on large datasets can be time-consuming and computationally demanding, especially in the field of molecular modeling, where diverse chemical structures and properties are involved. One approach to mitigate this challenge is to utilize transfer learning. Transfer learning allows a model to leverage the knowledge and representations learned during pretraining to initialize its parameters for a new dataset, significantly reducing the need for extensive data and computational resources in training [22]. In our prior work, we proposed the Fine-Tuning Accelerated molecular simulations framework (Finetuna) [23] as a promising implementation of online transfer learning with pre-trained ML models. Finetuna utilizes an active querying strategy to determine when to perform a DFT calculation and fine-tune the ML model using the DFT results. More specifically, we benchmarked Finetuna performance on a set of catalyst systems from the OC20 validation set with a GemNet model trained on the OC20 training data. Finetuna demonstrated that local optimizations of OC20-like systems can be greatly accelerated without sacrificing accuracy, as evidenced by a reduction of 90% DFT calculations compared to a baseline approach. We note that the Finetuna framework is one of many online active learning approaches for accelerating atomistic simulations. Other on-the-fly active learning frameworks typically start from scratch, using Gaussian process models or simpler neural network potentials [24–29].

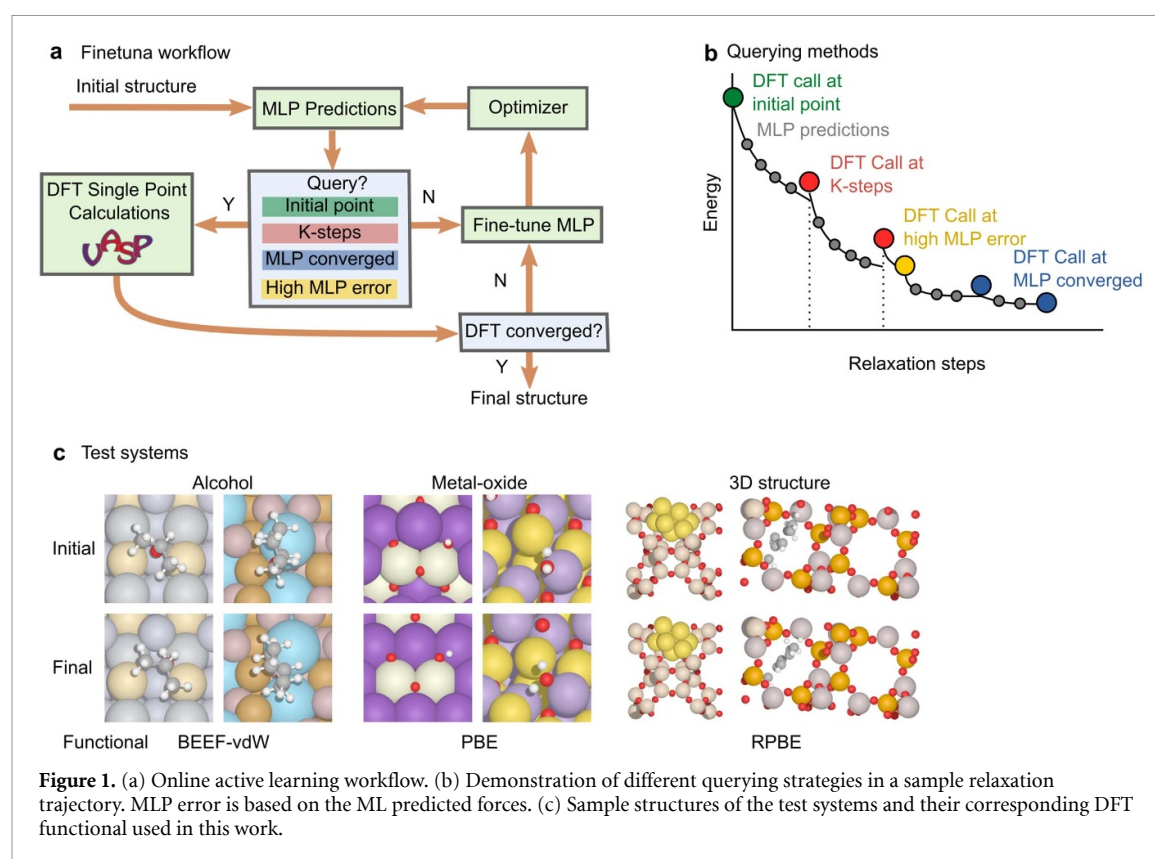
In this work, we seek to extend Finetuna to different out-of-domain chemical systems and gain insights into the capabilities and limitations of the workflow. We conduct a case study of the Finetuna workflow on three groups of materials, namely ($C_{>2}$) alcohols, metal oxides, and three-dimensional (3D) structures, for different applications. While the OC20 dataset is limited to the exploration of small ($C_{\leq 2}$) adsorbates, we explore the adsorption of $C_{>2}$ alcohol-to-ketone intermediates in this work. We also explore metal-oxide catalysts and the structures of zeolites and MOFs which significantly deviate from the catalyst surfaces in the original OC20 dataset. Finally we examine the effect of spin polarization and systems with significantly more atoms than those in the training dataset, both factors that significantly decrease computational speed of DFT. We introduce new querying strategies for these systems with our best-performing Finetuna strategy capable of reducing DFT calls by 80% for both ($C_{>2}$) alcohol and 3D structure systems and 42% for oxide systems.

2. Methods

2.1. Finetuna workflow

The active learning workflow accelerates geometric optimizations using a pre-trained GNN for fast force estimations. As shown in figure 1, the atomic structure is evaluated by the machine learning potential (MLP), and if a querying criterion is met, such as force convergence is reached or a step threshold is exceeded, a DFT single-point calculation will be triggered (also referred to as a ‘parent call’). The DFT-calculated forces are used to fine-tune the MLP. The optimizer uses the atomic forces to update the structure, and the MLP is used for force prediction. Similar to other relaxation processes, the convergence criterion is based on the maximum force in the system.

In the Finetuna workflow, many components, e.g. the ML model, the DFT functional and software, the optimizer, and the querying strategy, can be modified to suit specific systems. The choice of the MLP includes all the available pre-trained models, such as the GemNet model trained on OC20 dataset, the GemNet model trained on OC22 dataset, the GemNet-OC model trained on both OC20 and OC22 datasets, etc. We refer to these models as GN-OC20, GN-OC22, GNOC-OC20+22, respectively. The selection of the DFT functional depends mostly on the chemical system of interest; we discuss this further in section 2.2. We limit the DFT code to the Vienna *Ab initio* Simulation Package (VASP) for the majority of this work to be



consistent with the OC20 and OC22 dataset [30–34], and the optimizer to Broyden-Fletcher-Goldfarb-Shanno (BFGS) in the Atomic Simulation Environment Python package (ASE) in order to eliminate variations arising from differences between optimizers [35]. Interfaces to other DFT codes, i.e. Quantum Espresso [36, 37], can be found in the GitHub repository [38], and a detailed comparison of different optimizers can be found in the prior work [23].

The querying strategy determines when a DFT calculation is performed. The four main strategies are demonstrated in figure 1(c). ‘Initial point’ queries at the initial structure. ‘K-steps’ determines the lower bound of querying frequency, similar to [23]. The ‘MLP converged’ strategy queries when the maximum force from ML prediction is less than the convergence criterion. We also introduce a new strategy, labeled as ‘high MLP error’ in figure 1. After every DFT evaluation and retraining, we use the finetuned MLP to evaluate the current image. The training error is defined as the L2 norm of the difference between the DFT forces and the retrained MLP prediction. If the training error exceeds a set threshold, DFT will be queried for the following frame again. This approach is needed for high-error systems because training on a single frame may not improve the model’s prediction enough to allow the framework to correctly converge. This increases the overall frequency with which the model is retrained for challenging systems, when a single DFT query does not sufficiently increase model accuracy, but otherwise does not affect training hyperparameters.

The baseline Finetuna framework uses the GN-OC20 model as the MLP, and queries the parent calculator at the initial point, every 30 ML steps, or when the ML-predicted maximum force is below the convergence criterion (in this work, $0.05 \text{ eV } \text{\AA}^{-1}$). The baseline strategy is applied to all systems discussed herein. However, the GN-OC20 model exhibits different levels of accuracy on the different types of systems tested in this work. We expect these differences to be caused by how well-represented those systems are in the OC20 training dataset. This, in turn, should affect how conservative the user ought to be in setting the Finetuna algorithm querying strategy. Systems that lie further outside the domain of the training data should require more training, and predictions by the model should be less trustworthy. Therefore, we also test different pre-trained MLPs for the oxide systems, and apply the additional query by high MLP error to the 3D structure systems. The detailed strategies will be further explained in section 2.2.

2.2. Chemical systems

2.2.1. Alcohol dehydrogenation

Alcohol dehydrogenation has been widely explored as a means to catalytically produce and store hydrogen gas (H_2) in fuel cells. Modeling the dehydrogenation of relatively large ($\text{C}_{>2}$) alcohol molecules to ketones is

a complex process as it involves varying possible surface intermediates at every step of the reaction network. Many useful industrial applications in catalysis involve reactions that either lead to the formation or breakdown of alcohols, such as the dehydrogenation of alcohols, the formation of fuel from CO₂ reduction reactions, or the catalytic formation of plastics. Here, we assess the performance of Finetuna in modeling a two-step alcohol dehydrogenation reaction. We investigated three out-of-domain adsorbates of different sizes: C₃ 2-propanol or isopropyl alcohol (IPA) (C₃H₈O), C₄ 2-butanol (C₄H₁₀O), and C₆ cyclohexanol (C₆H₁₂O). The surface intermediates investigated are similar to the original OC20 training dataset, but these systems are also considered out-of-domain because they are much larger than anything in the pretraining dataset. Because of this similarity, we expect training to be relatively easy and for the Finetuna algorithm to converge relatively quickly compared to more out-of-domain systems.

For large organic molecules (C_{>2}), the contributions to the binding energy come from a mixture of physisorption and chemisorption. This is due to the large dipole moments inherent to these molecules which results in significant van der Waals contributions to binding energy. Modeling these molecules using DFT typically requires the addition of dispersion corrections. Here we use the Bayesian error estimation functional with van der Waals correlation (BEEF-vdW) functional to account for these dispersion interactions [39]. The OC20 dataset was calculated using the revised Perdew–Burke–Ernzerhof (RPBE) functional with carbon-based adsorbates limited to C₁ and C₂ molecules without accounting for any dispersion interactions. To assess the influence of these dispersion interactions, we perform single-point DFT calculations with RPBE and BEEF-vdW functionals on the initial structures of alcohol systems. We find that the force differences between the two DFT functionals are very small compared to the difference between DFT forces and MLP forces. Details can be found in figure S2.

2.2.2. Metal oxides

Metal oxides have been extensively studied as candidate catalysts for crucial electrochemical reactions, e.g. the oxygen evolution reaction (OER), oxygen reduction reaction (ORR), and hydrogen evolution reaction (HER). However, oxide systems can be challenging to study using DFT calculations due to their complex electronic structure. The release of the OC22 dataset, which focused on metal oxides, has provided a valuable resource to advance the understanding of metal-oxide systems. Nevertheless, it is important to note that the OC22 dataset represents only a small subset of the extensive design space of oxides, underscoring the need for efficient exploration of these systems. Metal-oxide systems tested in this work are taken from the OC22 validation set. These are also adsorbate/slab configurations similar in form to training data in the OC20 dataset. These systems often exhibit spin ordering, so spin-polarized calculations are generally required. In addition, a Hubbard U correction is applied to certain elements in these systems to improve the description of localized electron states [40]. We suspect this can result in a potential energy surface with multiple nearby local minima, and thus give different outcomes when performing geometric optimizations [41]. For the selected testing systems, we used the same DFT settings from the OC22 dataset [10].

Evaluating pre-trained graph models like GemNet on oxide systems results in higher errors than non-oxide materials, due to the complexity of oxides. Training on both oxide and non-oxide materials improves the model accuracy, but energy and force prediction on oxide materials is still a more difficult task. Graph model featurization schemes may fail to capture some complex properties that are more common in, or specific to, oxide materials, such as magnetic effects and long-range charge effects [42]. Unlike metallic systems, semiconductors can have essentially identical structures, but relax to significantly different final geometries due to long-range effects of the number of shared electrons across the entire system, or exhibiting different magnetic configurations of the same structure. It has been shown that the GemNet-OC model trained only on OC20 dataset performs especially poorly on OC22 prediction tasks, e.g. the force MAE is 0.384 eV Å⁻¹ [10]. Even when given many training examples like those in the OC22 dataset, the accuracy of GNNs like GemNet on oxide systems does not approach the accuracy on non-oxide systems. As shown in the OCP leaderboard, the best GemNet model for out-of-distribution force predictions results in a force MAE of 0.031 eV Å⁻¹ for OC22, versus 0.023 eV Å⁻¹ force MAE for OC20 at the time of writing [9, 10]. In addition to the baseline Finetuna strategy that uses GN-OC20, we also test the performance of GN-OC22 and GNOC-OC20+22 as the underlying MLP, due to the anticipated difficulty of this task. The interaction between magnetic spins in systems with spin polarization is one of the contributing factors to the complexity of oxide materials. Spin polarization effects are typically long-range. However, GNN models like GemNet are developed under the assumption that local interactions dominate [10].

2.2.3. 3D structures

Zeolites and MOFs are two common materials with porous structures. They have gained significant attention in recent decades due to their versatile properties and wide-ranging applications in gas separation, water treatment, and catalysis [43–45]. One advantage of these 3D structures as catalysts is their tunability. By

modifying the composition, structure, and pore size, their selectivity and performance can be calibrated for specific applications [46, 47].

The test 3D structures cover a wide range of systems, including copper-doped zeolites, MOFs, and zeolites with a variety of adsorbates such as gold nanoparticles and aromatics. These catalysts are modeled very differently from OC20-like metal surfaces. Rather than adsorbing onto the surface of a dimensionally confined slab structure, which is the core focus of the OC20 dataset, these structures are fully 3D, and adsorbates, if present, are incorporated into pores in the 3D structure. Structural relaxations then take place throughout the simulation cell (rather than only in a few atomic layers near the surface/adsorbate). Examples of such complex structures are not found within the OC20 training dataset. While there may be numerous examples of certain environments containing similar metal-oxygen configurations, there should be many differences due to the addition of 3D surroundings for most atomic neighborhoods. The metal shells of these 3D structures are also effectively made up of metal-oxide configurations, adding to the difficulty. We anticipate this kind of structure to be very difficult to adapt to for a GNN model like GemNet, trained only on OC20-like structures.

To compensate for this, we test another, more conservative, querying strategy for Finetuna, in addition to the baseline approach. We measure the mean error of the retrained forces after each parent call during the Finetuna loop; by taking the L2 norm of the difference between the parent forces and the retrained model prediction. If the retrained mean force error (after fine-tuning) is still above some threshold, this signals that training was insufficient, and the parent DFT call is triggered on the next step of the relaxation as well. This process repeats at every parent call until the model error drops below a threshold (in this case $0.05 \text{ eV } \text{\AA}^{-1}$). This approach aims to ensure that the model is sufficiently trained on the new system to make force predictions that will take the relaxation in a reasonable direction to reduce the forces and energy.

3. Results and discussion

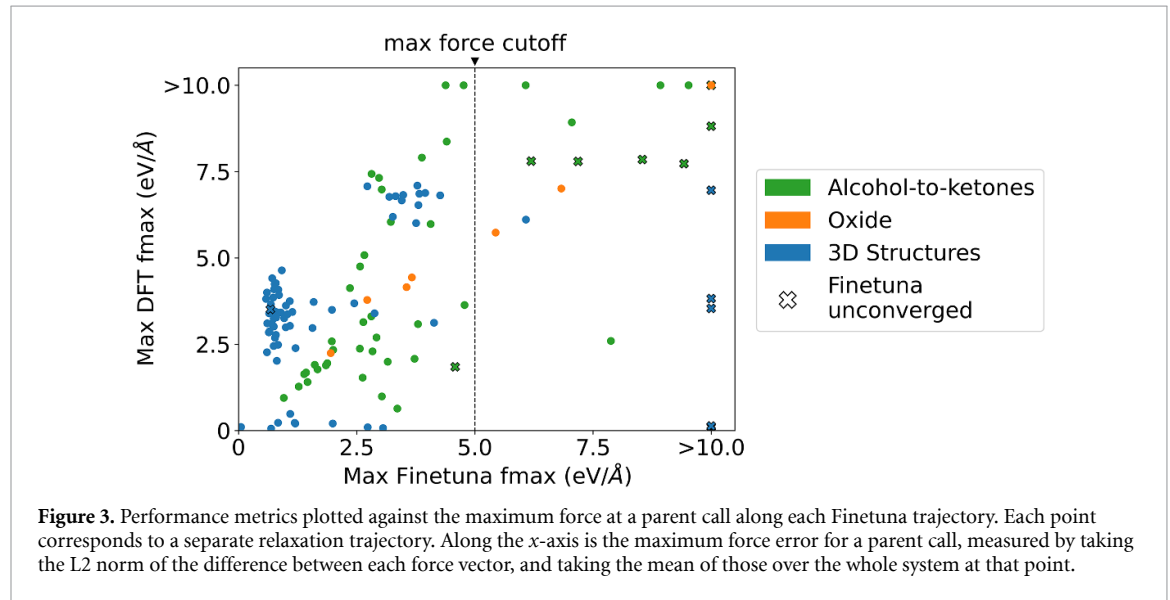
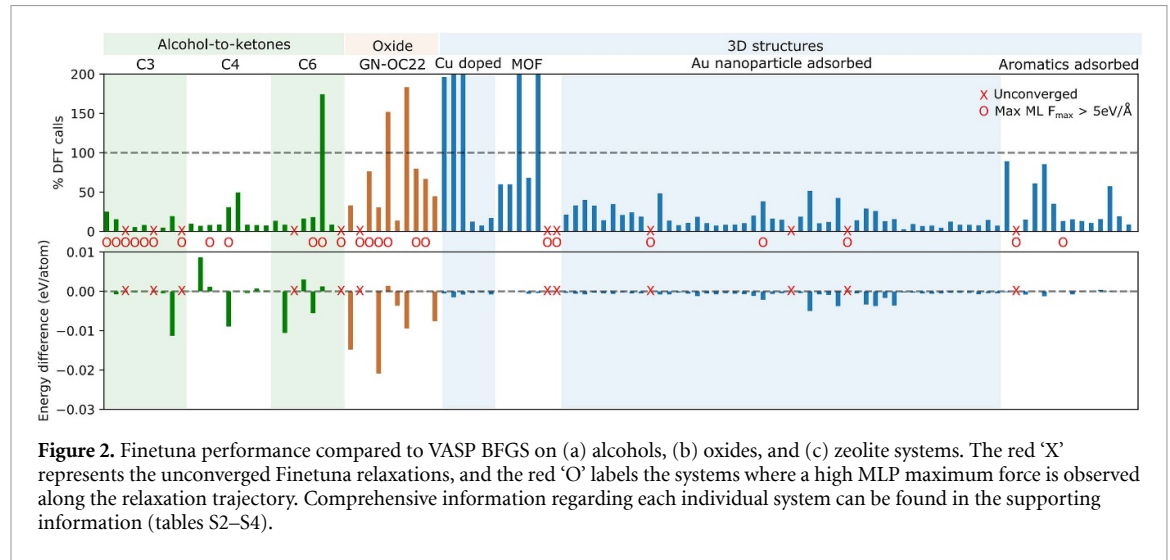
We selected 27 alcohol-to-ketone systems, 10 oxide systems, and 81 3D structure systems and performed both VASP BFGS and Finetuna relaxations. The systems that experienced convergence issues with VASP BFGS and the systems that failed due to memory limitations are excluded from this section. A full list of the tested systems and results can be found in the Supporting Information. The size of the chemical systems in the experiment varied from 50 to 250 atoms. We report the energy difference between the VASP BFGS and Finetuna relaxations in eV/atom to facilitate comparison. We also perform a related benchmark of the Finetuna algorithm using Quantum Espresso instead of VASP as the parent calculator in section S3 of the supporting information.

A summary of the overall performance of Finetuna compared to VASP BFGS can be found in table 1, and a plot of the individual runs is shown in figure 2. For each class of systems, we evaluate the efficiency of the Finetuna workflow by calculating the ratio between the total number of DFT calls with Finetuna relaxation and that with VASP BFGS. This metric is referred to as the percentage DFT calls and represents the overall saving of DFT calculations. The unconverged Finetuna runs are labeled in red crosses in figure 2. We found that none of the metal-oxide systems converged with the baseline GN-OC20 model. This is not surprising due to the lack of oxide systems and spin polarization information in the OC20 training data. For simplicity, the results from Finetuna with GN-OC22 model are shown in figure 2, and the comparisons with other models can be found in the supporting information figure S1. In general, Finetuna underperforms on both accuracy and speed when compared to the results in [23] across all three system types. We believe this is the result of a significant domain shift, which should make realigning the GNN model with fine-tuning slower and less accurate. We consider magnetic effects and 3D structural effects to be the most significant causes of domain shift. By increasing the conservativeness in the Finetuna algorithm through repeated training when errors are high, the overall parent calls can be brought below 20% for alcohol-to-ketones and the 3D structures, and 58% for oxide systems.

Across alcohol-to-ketones, metal oxides, and 3D structure systems, there are a number of cases where Finetuna does not converge. As shown in figure 3, most of the unconverged cases have observed high ML predicted forces ($> 5.0 \text{ eV } \text{\AA}^{-1}$) across the entire relaxation trajectory. This implies that for trajectories where the MLP made a large force prediction, they generally should not be trusted to converge, and should be terminated early. In addition, we would expect that (provided initial structure guesses are reasonable), the occurrence of such a large force would indicate that the system has moved to a less realistic region of the configuration space and any local minimum it finds would likely be less physically meaningful. The only exceptions to this are with a few of the alcohol-to-ketones systems, where the trajectory slowly converged (exceeding 100% of the DFT parent calls) despite the high error. Interestingly, these systems do not appear to find significantly different local minima from the original DFT relaxation, so in these cases, it would appear that Finetuna took a somewhat roundabout route to find a very similar result.

Table 1. Summary of the Finetuna performance on alcohol, oxide, and 3D structure systems. The baseline querying strategy calls a DFT calculation at the initial point, every 30 MLP steps, and when MLP predicted forces meet the convergence criterion.

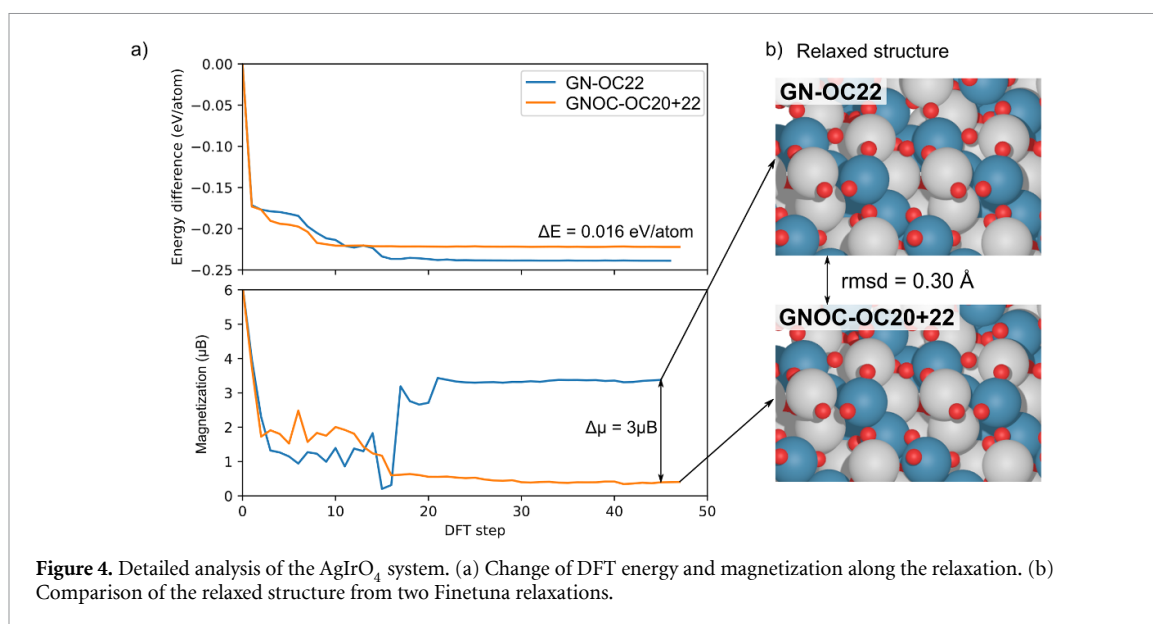
System	ML model	Querying strategy	Functional	% Converged	% DFT calls
Alcohols	GN-OC20	Baseline	BEEF-vdW	81%	18%
Oxides	GN-OC20	Baseline	PBE	0%	N/A
	GN-OC22			90%	58%
	GNOC-OC20+22			70%	36%
Zeolite	GN-OC20	Baseline + High MLP error	RPBE	92%	19%



For the oxide systems, we found that the potential energy surfaces are generally more complicated, as evidenced by the discrepancies in the relaxed energy from the two DFT calculations (VASP BFGS and VASP conjugate gradient (CG) in figure S1). The total magnetization (M) is given by the net spin of the electrons in the system:

$$M = \mu_B \int (n_{\uparrow} - n_{\downarrow}) d^3r, \quad (1)$$

where μ_B is the Bohr magneton, and n_{\uparrow} and n_{\downarrow} represent spin densities. For our purposes, M represents the extent of spin polarization in a given system. Figure 4 shows a case where relaxation with two different models led to nearly identical sets of atomic positions but considerably different magnetizations and



energies. The spin polarization adds another degree of freedom to the local minimization process and breaks the one-to-one mapping between atomic structure and potential energy presumed by GemNet. Additionally, as previously mentioned, the ML model we use in this work sets a local cutoff when representing the atomic structure, and is unable to capture long-range interaction from magnetic effects. This suggests that the ML potential and the performance of the algorithm may be compromised when applied to spin-polarized systems, and highlights the importance of developing MLPs, such as CHGNET [48], that effectively capture magnetic information for these applications.

4. Conclusion

In this study, we conducted experiments to assess the performance of the Finetuna workflow on various chemical systems including alcohol-to-ketones, metal oxides, and MOFs. Our findings revealed that, while showing promising results in the original domain (OC20-like systems) with a significant reduction of DFT calls by 90%, Finetuna was less performant on these out-of-domain systems. The results show that we were able to reduce the number of DFT calls by 85% for alcohol-to-ketones systems, 49% for oxide systems, and 82% for MOF systems, with a few unconverged cases. We believe that the complex electronic interactions in these chemical systems, particularly the oxides with spin polarization, limit Finetuna's performance. The long-range magnetic interaction introduces more complexity to the potential energy surface. While this has affected the efficiency of the Finetuna workflow, it also presents opportunities for further exploration. We have established a heuristic criterion to predict the success of a Finetuna run. Specifically, if the maximum force predicted by the ML model exceeds $5 \text{ eV } \text{\AA}^{-1}$, we recommend using a different initialization strategy, or terminating the simulation and performing DFT calculations instead.

As the accuracy of ML models improves and more material datasets become available, we believe that the Finetuna workflow can be applied to a broader range of chemical systems. We anticipate the rapid emergence of new ML models and look forward to their integration into the workflow. However, to effectively utilize these new models, further research is needed to explore transfer learning strategies and develop fine-tuning techniques specific to the models. The choice of a pre-trained model with suitable architectural features and trained on relevant data can significantly impact the model's performance and generalizability. Investigating out-of-domain detection methods can enhance the robustness of our approach. Additionally, the inclusion of uncertainty quantification in ML models would be valuable, as it could be leveraged as a querying strategy to guide the selection of informative data points for further exploration. These future directions hold promise for advancing the efficiency and accuracy of material discovery applications. The active learning and finetuning approach has the potential to greatly accelerate molecular simulations, leading to the development of more efficient and sustainable materials with a wide range of applications.










Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/ulissigroup/finetuna-2023-manuscript>.

Acknowledgments

The authors acknowledge support from Meta Platforms, Inc. via the Open Catalyst Project collaboration.

ORCID iDs

Xiaoxiao Wang  <https://orcid.org/0000-0001-8587-8610>
Joseph Musielewicz  <https://orcid.org/0000-0002-5296-9177>
Richard Tran  <https://orcid.org/0000-0002-0308-2182>
Sudheesh Kumar Ethirajan  <https://orcid.org/0000-0002-8144-2657>
Xiaoyan Fu  <https://orcid.org/0009-0007-9084-9815>
Hilda Mera  <https://orcid.org/0009-0000-5615-8702>
John R Kitchin  <https://orcid.org/0000-0003-2625-9232>
Rachel C Kurchin  <https://orcid.org/0000-0002-2147-4809>
Zachary W Ulissi  <https://orcid.org/0000-0002-9401-4918>

References

- [1] Li W, Li K, Ye Y, Zhang S, Liu Y, Wang G, Liang C, Zhang H and Zhao H 2021 Efficient electrocatalytic nitrogen reduction to ammonia with aqueous silver nanodots *Commun. Chem.* **4** 1–11
- [2] Foster S L, Bakovic S I, Duda R D, Maheshwari S, Milton R D, Minteer S D, Janik M J, Renner J N and Greenlee L F 2018 Catalysts for nitrogen reduction to ammonia *Nat. Catal.* **1** 490–500
- [3] Bonitatibus P J, Chakraborty S, Doherty M D, Siclován O, Jones W D and Soloveichik G L 2015 Reversible catalytic dehydrogenation of alcohols for energy storage *Proc. Natl Acad. Sci. USA* **112** 1687–92
- [4] Ge X, Sumboja A, Wu D, An T, Li B, Goh F W, Hor T S, Zong Y and Liu Z 2015 Oxygen reduction in alkaline media: from mechanisms to recent advances of catalysts *ACS Catal.* **5** 4643–67
- [5] Beall C E, Fabbri E and Schmidt T J 2021 Perovskite oxide based electrodes for the oxygen reduction and evolution reactions: the underlying mechanism *ACS Catal.* **11** 3094–114
- [6] Petousis I, Mrdjenovich D, Ballouz E, Liu M, Winston D, Chen W, Graf T, Schladt T D, Persson K A and Prinz F B 2017 Data descriptor: high-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials *Sci. Data* **4** 1–12
- [7] Lu Z 2021 Computational discovery of energy materials in the era of big data and machine learning: a critical review *Mater. Rep.: Energy* **1** 100047
- [8] Schmidt J, Marques M R, Botti S and Marques M A 2019 Recent advances and applications of machine learning in solid-state materials science *npj Comput. Mater.* **5** 83
- [9] Chanussot L et al 2021 Open catalyst 2020 (OC20) dataset and community challenges *ACS Catal.* **11** 6059–72
- [10] Tran R et al 2023 The open catalyst 2022 (OC22) dataset and challenges for oxide electrocatalysts *ACS Catal.* **13** 3066–84
- [11] Chung Y G et al 2019 Advances updates and analytics for the computation-ready, experimental metal-organic framework database: CoRE MOF 2019 *J. Chem. Eng. Data* **64** 5985–98
- [12] Moghadam P Z, Li A, Wiggins S B, Tao A, Maloney A G, Wood P A, Ward S C and Fairen-Jimenez D 2017 Development of a cambridge structural database subset: a collection of metal-organic frameworks for past, present and future *Chem. Mater.* **29** 2618–25
- [13] Rosen A S, Iyer S M, Ray D, Yao Z, Aspuru-Guzik A, Gagliardi L, Notestein J M and Snurr R Q 2021 Machine learning the quantum-chemical properties of metal-organic frameworks for accelerated materials discovery *Matter* **4** 1578–97
- [14] Bobbitt N S et al 2023 Mofx-db: an online database of computational adsorption data for nanoporous materials *J. Chem. Eng. Data* **68** 483–98
- [15] Baerlocher L M C 2017 Database of zeolite structures (available at: www.iza-structure.org/databases/)
- [16] Komissarov L and Verstraeten T 2022 Zeo-1, a computational data set of zeolite structures *Sci. Data* **9** 1–7
- [17] Klicpera J, Becker F and Günnemann S 2021 GemNet: universal directional graph neural networks for molecules *Advances in Neural Information Processing Systems* vol 9 pp 6790–802
- [18] Gasteiger J, Shuaibi M, Sriram A, Günnemann S, Ulissi Z, Zitnick C L and Das A 2022 GemNet-OC: developing graph neural networks for large and diverse molecular simulation datasets (arXiv:2204.02782)
- [19] Zitnick C L, Das A, Kolluru A, Lan J, Shuaibi M, Sriram A, Ulissi Z and Wood B 2022 Spherical channels for modeling atomic interactions (arXiv:2206.14331)
- [20] Passaro S and Zitnick C L 2023 Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs (arXiv:2302.03655)
- [21] Chen C and Ong S P 2022 A universal graph deep learning interatomic potential for the periodic table *Nat. Comput. Sci.* **2** 718–28
- [22] Kolluru A, Shoghi N, Shuaibi M, Goyal S, Das A, Zitnick C L and Ulissi Z 2022 Transfer learning using attentions across atomic systems with graph neural networks (TAAG) *J. Chem. Phys.* **156** 18
- [23] Musielewicz J, Wang X, Tian T and Ulissi Z 2022 FINETUNA: fine-tuning accelerated molecular simulations *Mach. Learn.: Sci. Technol.* **3** 03LT01
- [24] Tran K and Ulissi Z W 2018 Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution *Nat. Catal.* **1** 696–703
- [25] Zhong M et al 2020 Accelerated discovery of CO₂ electrocatalysts using active machine learning *Nature* **581** 178–83
- [26] Vandermause J, Torrisi S B, Batzner S, Xie Y, Sun L, Kolpak A M and Kozinsky B 2020 On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events *npj Comput. Mater.* **6** 20
- [27] Vandermause J, Xie Y, Lim J S, Owen C J and Kozinsky B 2021 Active learning of reactive bayesian force fields: application to heterogeneous hydrogen-platinum catalysis dynamics (arXiv:2106.01949)
- [28] Yang Y, Jiménez-Negrón O A and Kitchin J R 2021 Machine-learning accelerated geometry optimization in molecular simulation *J. Chem. Phys.* **154** 234704

- [29] Shuaibi M, Sivakumar S, Chen R Q and Ulissi Z W 2021 Enabling robust offline active learning for machine learning potentials using simple physics-based priors *Mach. Learn.: Sci. Technol.* **2** 025007
- [30] Kresse G and Furthmüller J 1996 Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set *Comput. Mater. Sci.* **6** 15–50
- [31] Kresse G 1995 *Ab initio* molecular dynamics for liquid metals *J. Non-Cryst. Solids* **192–193** 222–9
- [32] Kresse G and Hafner J 1994 *Ab initio* molecular-dynamics simulation of the liquid-metalamorphous- semiconductor transition in germanium *Phys. Rev. B* **49** 14251–69
- [33] Kresse G and Furthmüller J 1996 Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set *Phys. Rev. B* **54** 11169–86
- [34] Tian T and Musielewicz J 2022 Vaspinteractive: Interactive vasp calculator (available at: <https://github.com/ulissigroup/vasp-interactive>)
- [35] Hjorth Larsen A et al 2017 The atomic simulation environment - a python library for working with atoms *J. Phys.: Condens. Matter* **29** 273002
- [36] Giannozzi P et al 2009 QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials *J. Phys.: Condens. Matter* **21** 395502
- [37] Giannozzi P et al 2017 Advanced capabilities for materials modelling with quantum ESPRESSO *J. Phys.: Condens. Matter* **29** 465901
- [38] Musielewicz J, Wang X, Tian T and Ulissi Z 2022 Finetuna: fine-tuning accelerated molecular simulations (available at: <https://github.com/ulissigroup/finetuna>)
- [39] Wellendorff J, Lundgaard K T, Møgelhøj A, Petzold V, Landis D D, Nørskov J K, Bligaard T and Jacobsen K W 2012 Density functionals for surface science: exchange-correlation model development with Bayesian error estimation *Phys. Rev. B* **85** 235149
- [40] Jain A et al 2013 Commentary: the materials project: a materials genome approach to accelerating materials innovation *APL Mater.* **1** 011002
- [41] Meredig B, Thompson A, Hansen H A, Wolverton C and Van De Walle A 2010 Method for locating low-energy solutions within DFT+ U *Phys. Rev. B* **82**
- [42] Akpabio I I and Savari S A 2021 Uncertainty quantification of machine learning models: on conformal prediction *J. Micro/Nanopatterning Mater. Metrol.* **20** 1–14
- [43] Van Speybroeck V, Hemelsoet K, Joos L, Waroquier M, Bell R G and Catlow C R A 2015 Advances in theory and their application within the field of zeolite chemistry *Chem. Soc. Rev.* **44** 7044–111
- [44] Lee J, Farha O K, Roberts J, Scheidt K A, Nguyen S T and Hupp J T 2009 Metal-organic framework materials as catalysts *Chem. Soc. Rev.* **38** 1450–9
- [45] Yang D and Gates B C 2019 Catalysis by metal organic frameworks: perspective and suggestions for future research *ACS Catal.* **9** 1779–98
- [46] Bacariza M C, Graça I, Lopes J M and Henriques C 2019 Tuning zeolite properties towards CO₂ methanation: an overview *ChemCatChem* **11** 2388–400
- [47] Pascanu V, González Miera G, Inge A K and Martín-Matute B 2019 Metal-organic frameworks as catalysts for organic synthesis: a critical perspective *J. Am. Chem. Soc.* **141** 7223–34
- [48] Deng B, Zhong P, Jun K J, Riebesell J, Han K, Bartel C J and Ceder G 2023 CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling *Nat. Mach. Intell.* **5** 1031–1041