



BENCHMARK • OPEN ACCESS

Benchmarking universal machine learning interatomic potentials for rapid analysis of inelastic neutron scattering data^{*}

To cite this article: Bowen Han and Yongqiang Cheng 2025 *Mach. Learn.: Sci. Technol.* **6** 030504

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

You may also like

- [Complex optimization for big computational and experimental neutron datasets](#)

Feng Bao, Richard Archibald, Jennifer Niedziela et al.

- [Structure and magnetism of the triangular lattice material YbBO₃](#)

G Sala, Matthew B Stone, Seung-Hwan Do et al.

- [Roadmap for the development of machine learning-based interatomic potentials](#)

Yong-Wei Zhang, Viacheslav Sorkin, Zachary H Aitken et al.



BENCHMARK

OPEN ACCESS

RECEIVED
1 May 2025REVISED
8 July 2025ACCEPTED FOR PUBLICATION
11 August 2025PUBLISHED
21 August 2025

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.



Benchmarking universal machine learning interatomic potentials for rapid analysis of inelastic neutron scattering data*

Bowen Han** and Yongqiang Cheng**

Neutron Scattering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, United States of America

**Authors to whom any correspondence should be addressed.

E-mail: hanb1@ornl.gov and chengy@ornl.gov**Keywords:** phonon, inelastic neutron scattering, machine learning, interatomic potentialSupplementary material for this article is available [online](#)

Abstract

The accurate calculation of phonons and vibrational spectra remains a significant challenge, requiring highly precise evaluations of interatomic forces. Traditional methods based on the quantum description of the electronic structure, while widely used, are computationally expensive and demand substantial expertise. Emerging universal machine learning interatomic potentials (uMLIPs) offer a transformative alternative by employing pre-trained neural network surrogates to predict interatomic forces directly from atomic coordinates. This approach dramatically reduces computation time and minimizes the need for technical knowledge. In this paper, we produce a phonon database comprising nearly 5000 inorganic crystals to benchmark the performance of several leading uMLIPs. We further assess these models in real-world applications by using them to analyze experimental inelastic neutron scattering data collected on a variety of materials. Through detailed comparisons, we identify the strengths and limitations of these uMLIPs, providing insights into their accuracy and suitability for fast calculations of phonons and related properties, as well as the potential for real-time interpretation of neutron scattering spectra. Our findings highlight how the rapid advancement of AI in science is revolutionizing experimental research and data analysis.

1. Introduction

Atomistic modeling is an indispensable tool to connect theory and experiment and to illustrate fundamental mechanisms behind the structure-property relationships in materials. An essential step in atomistic modeling is the evaluation of interatomic interactions, which can be derived from the potential energy as a function of the atomic coordinates [1]. Such calculations can be performed at different levels of accuracy and with different efficiency. A simple model may describe interatomic interactions in various ways, such as using a parameterized formula to calculate forces from the bond length/angle. Empirical models like these are quick to compute, but cannot describe complex many-body interactions. The potential energy and forces can also be calculated from density functional theory (DFT), by solving the electron density and wave functions of the ground state, as well as their dependence on the atomic displacements [2]. These calculations based on quantum mechanics can be very accurate, but they are also computationally expensive. The accuracy-efficiency tradeoff leads to many practical challenges. For example, the analysis of vibrational or phonon spectroscopy, such as those measured from inelastic neutron scattering (INS) [3], requires high-precision calculations of interatomic interactions, which is usually nontrivial, time-consuming, and needs special

* This manuscript has been authored by UT-Battelle, LLC. under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

computing resources and expertise including proficiency in DFT and electronic-structure software. The analysis usually cannot be performed in real-time or on-demand, which is not only a bottleneck for productivity but also a barrier for future autonomous experiments, where understanding the results on-the-fly is a prerequisite [4–6].

The recent development of neural networks and machine learning for materials science has led to a possible solution that enables a new approach to the calculation of interatomic forces: training a neural network surrogate to predict energies and forces of a given atomic structure using available DFT data [7]. These neural network surrogates are usually called machine learning force fields or machine learning interatomic potentials (MLIPs). Neural networks excel at capturing complex, high-dimensional, nonlinear mappings, making it possible for MLIPs to achieve both accuracy and efficiency. Although early MLIPs were developed for specific compositions or materials, the availability of larger databases with more comprehensive coverage, combined with the rapid growth of computing power, makes it possible to train foundation models, or universal MLIPs (uMLIPs), that can describe a wide range of materials. The first uMLIPs are benchmarked for energies and forces, as well as some other global properties such as modulus [8]. However, studies found that accurate descriptions of phonons could be more challenging, and most early models systematically underestimate the frequencies [9]. Several models published very recently seem to be better and could be ready for phonon calculations according to Loew *et al* [10]. Can we now use uMLIPs for the analysis of phonons and interpretation of vibrational spectral data?

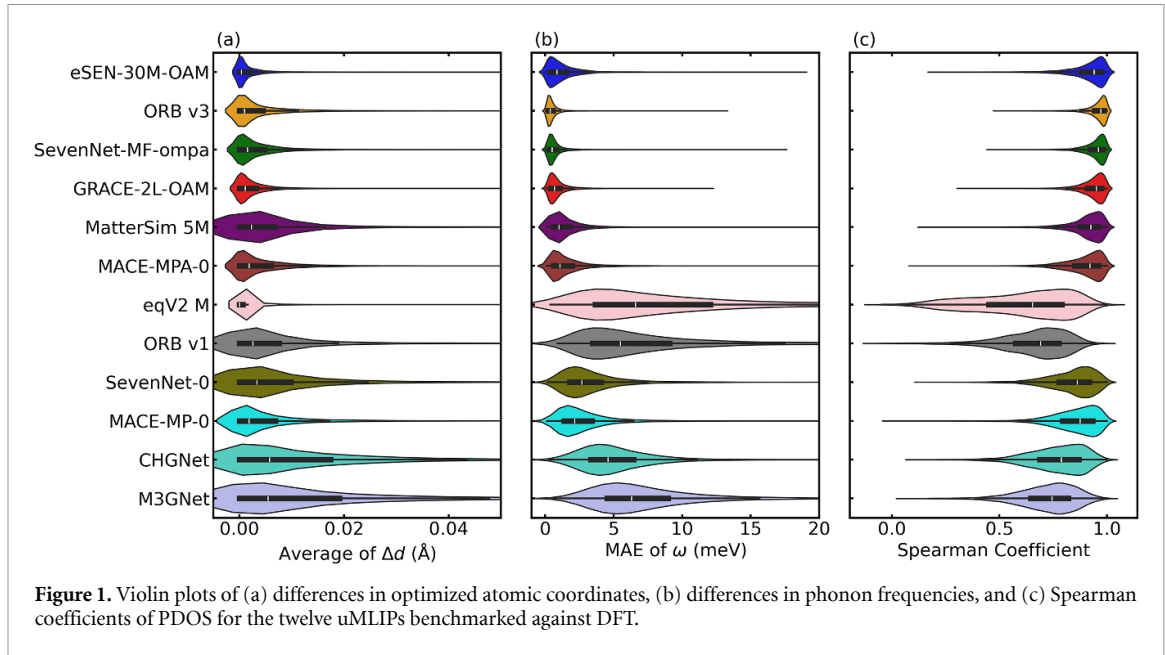
In this work, we produce a new phonon database using DFT. The database covers nearly 5000 crystals with unit cells containing up to 12 atoms. Compared to existing phonon databases (such as the one by Togo [11]), we focus on a more comprehensive coverage of crystals with relatively simple crystal structures, large Brillouin zones, and rich phonon dispersion features that reflect the subtle interatomic interactions. We then benchmark the best-performing uMLIPs against our DFT database. Our application scenario focuses on the analysis of INS data, for which modeling of phonons has been the primary bottleneck. We compare phonons and simulated INS spectra obtained from the uMLIPs with available experimental data. Finally, we include some of the latest uMLIPs in the INSPIRED software [12], which has a simple graphical user interface to allow easy and quick simulations of INS spectra from a structure model. The results of this work illustrate the capability of real-time analysis and interpretation of INS data enabled by uMLIPs.

2. Results and discussion

The crystals included in the benchmarking database are chosen from the Materials Project [13]. Only stable and realistic materials are considered. We build our database starting from smaller unit cells (ranked by the number of atoms in the unit cell, in ascending order, up to 12 atoms). Details on the production of the database can be found in Methods. The final database contains 4869 crystals that cover 86 elements.

We include 12 uMLIPs from the top of the MatBench Discovery leaderboard by April 2025 [8]. They are eSEN-30 M-OAM [14], ORB v3 [15], SevenNet-MF-ompa [16, 17], GRACE-2 L-OAM [18], MatterSim v1 5 M [19], MACE-MPA-0 [20–22], eqV2 M [23], ORB v1 [24], SevenNet-0 [16], MACE-MP-0 [20–22], CHGNet [25], and M3GNet [26]. For each crystal, with each uMLIP, we first relax the atomic coordinates to find the optimized structure with the minimum potential energy. This structure is compared with the DFT-relaxed structure, and differences in atomic coordinates are averaged to account for discrepancies in the ground-state structure (figure 1(a)). Phonon calculations are then performed on a uniform mesh sampling of the first Brillouin zone, and the average differences in the phonon frequencies are compared in figure 1(b). Finally, the phonon density of states (PDOS) is compared for spectral similarity, quantified by Spearman coefficients [27, 28] (figure 1(c)). The Spearman coefficient ranges between 0 and 1, with higher values indicating better alignment between the two PDOS compared. We also calculate the vibrational entropy (S), Helmholtz free energy (F), and constant volume heat capacity (C_V) from the phonon results, and they are compared in the supplementary information (figure S1). Detailed statistics of the comparison among all uMLIPs are shown in tables S1–S6. These benchmarking results suggest that the most accurate uMLIPs for phonon calculations are ORB v3, SevenNet-MP-ompa and GRACE-2 L-OAM, with MatterSim 5 M, MACE-MPA-0, and eSEN-30 M-OAM following closely. Our benchmark also suggests that the latest uMLIPs outperform the top ones in an earlier benchmark performed by Loew *et al* [10], when these potentials had not been published.

The above analysis shows statistically how well each uMLIP performs against DFT. However, the ultimate benchmark should be the comparison with the experimental data. For example, in a real application scenario, we would compare the uMLIP calculations with the collected experimental data and use the model to provide quick analysis and interpretation by matching the simulated spectral features with the experimentally measured ones. The validation of the phonon calculation results with experimental data is



also crucial for the prediction of phonon-based thermodynamic and kinetic properties. In the following, we will compare the performance of the uMLIPs using INS data as benchmarks.

INS measures the dynamical structure factor $S(\mathbf{Q}, E)$, which is a function of momentum $\mathbf{Q}$ (as a vector in the three-dimensional reciprocal space) and energy E . With orientational averaging for powder samples, the 4-dimensional (4D) $S(\mathbf{Q}, E)$ reduces to 2D $S(|\mathbf{Q}|, E)$, which is a function of the magnitude of $\mathbf{Q}$ (or $|\mathbf{Q}|$) and energy E [29]. Neutron scattering from a material can be coherent or incoherent. Some elements (such as H and V) [30] have intrinsically high incoherent neutron scattering cross-sections; thus materials containing such elements will have INS intensities mostly from the incoherent scattering. Some elements are strong coherent scatterers (e.g. C, O, and Al) [30]; thus scattering from these elements interferes with each other, exhibiting significant $\mathbf{Q}$ -dependent patterns in the $S(\mathbf{Q}, E)$ data. Note that when a crystal structure is complex with a large unit cell, even if the constituent elements are coherent scatterers, the Brillouin zone may be so small that the $\mathbf{Q}$ dependence cannot be resolved after powder averaging. As a result, the overall scattering may appear to be incoherent. This is the so-called incoherent approximation [29]. In the following, we will show examples for each scenario.

We start with a simple crystal of a coherent neutron scatterer: graphite. As a reference, the phonon dispersion curves calculated for graphite using various uMLIPs are compared with DFT in figures 2(a)–(f). A powder sample of graphite was measured at SEQUOIA [31], a direct geometry spectrometer at the Spallation Neutron Source, to obtain $S(|\mathbf{Q}|, E)$ [32]. Strong effects of coherent inelastic scattering can be seen even in the powder data, where various phonon dispersion curves can be resolved (figure 2(g)). Compared with the experimental and DFT-calculated INS spectra, it is clear that ORB v3 and MatterSim are among the best to capture the details in the powder $S(|\mathbf{Q}|, E)$. However, ORB v3 produces ZA branches with slight negative frequencies, whereas MatterSim predicts stable ZA modes (here ZA refers to out-of-plane transverse acoustic phonons in graphite). Negative frequency modes are also observed with SevenNet and GRACE, but not for the two MACE potentials. The negative frequencies indicate phonon instability and can be detrimental to the simulated INS spectra at low energy. We also note that the phonon frequencies calculated by MACE-MP-0 are significantly underestimated. The observations in the simulated INS spectra are consistent with the comparisons of calculated phonon dispersion.

Next, we move to a more complex crystal, $\text{Ba}_3\text{ZnRu}_2\text{O}_9$, [33] that has 30 atoms in the unit cell. Phonon dispersion plots in figures 3(a)–(f) show more complex features and obvious discrepancies between uMLIP and DFT in all models. Specifically, the main issue with most uMLIPs is still the systematic softening of the calculated frequencies in the high-frequency optical band. Two models, ORB v3 and MatterSim, outperform others in this regard, reproducing the highest-frequency branches with better agreement. Comparison of powder INS spectra in figures 3(g)–(n) reveals the same trend. The more complex crystal structure results in weaker coherent scattering patterns in the powder $S(|\mathbf{Q}|, E)$, preventing a direct comparison of specific phonon dispersion curves. However, the energy levels of multiple phonon bands are clearly visible and can be used to evaluate the accuracy of different models. Note that the feature at below $1 \text{ \AA}^{-1}$ in the experimental spectrum is due to magnetic scattering, which is not included in the simulated ones where only intensities from phonons are calculated.

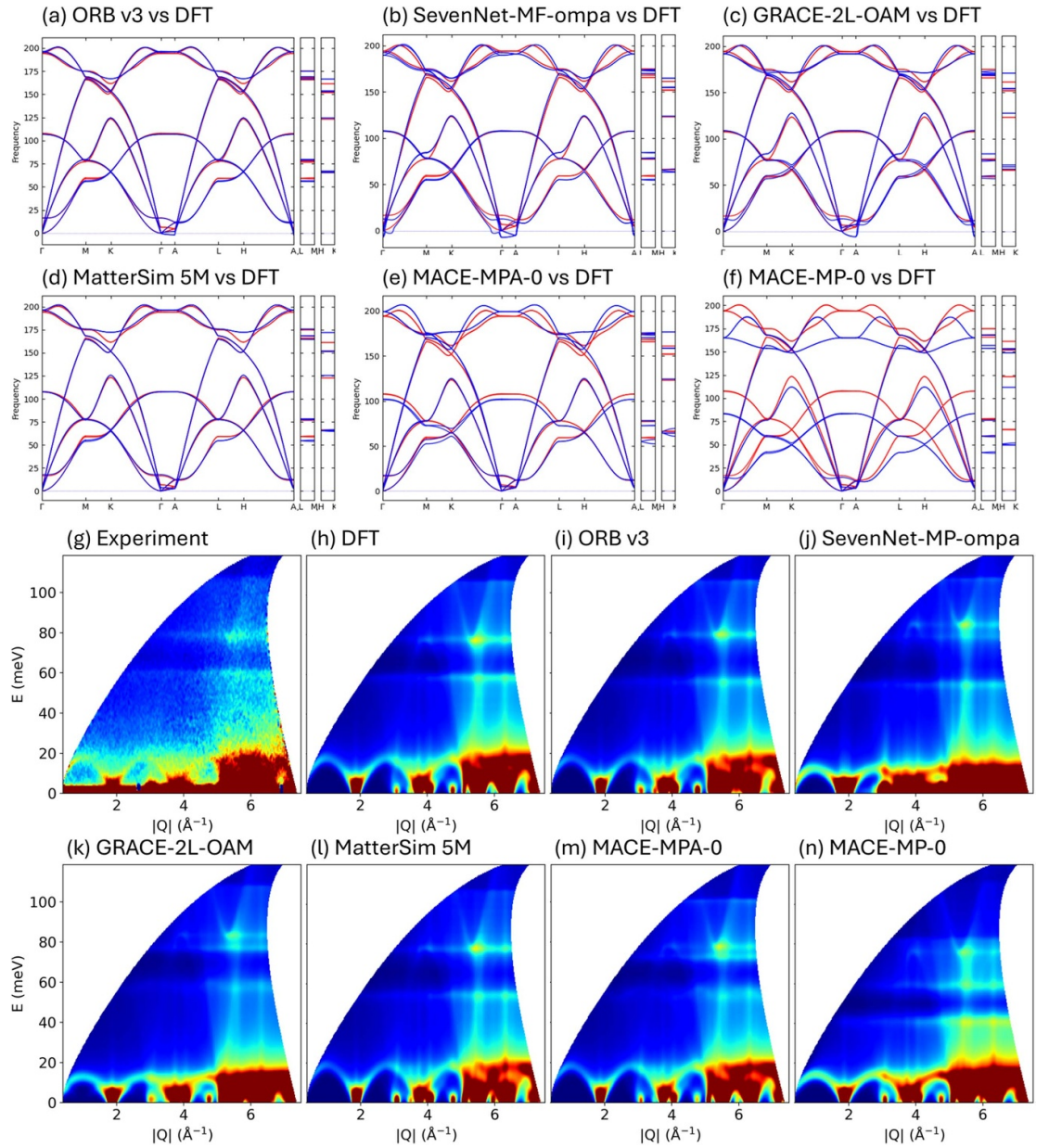


Figure 2. (a)–(f) Phonon dispersion of graphite obtained from uMLIPs (blue) and DFT (red). The powder INS spectra measured at SEQUAIA is shown in (g) [32], with the corresponding simulations in (h)–(n).

A key advantage of INS is its unique capability in accurately measuring the complete phonon dispersion in the 4D ($\mathbf{Q}, E$) space. Data collected on a single crystal Cu_2O (cuprite) [34] are presented in figure 4. Such measurements provide a detailed comparison of theoretically predicted and experimentally measured phonon dispersion along any direction of interest in the Brillouin zone. This can be considered the golden standard in the benchmarking of phonon calculations. By overlapping the calculated phonon dispersion on the experimentally measured spectra, we can clearly see that, for cuprite, MatterSim outperforms all other uMLIPs, with ORB v3 also showing excellent agreement. Underestimation of the optical band frequencies remains an issue, especially for the MACE-MP-0 potential. Another new observation is the underestimation of the acoustic branch below 10 meV, which is also present even in the DFT result. However, MACE-MPA-0 is in better agreement with the experiment on this acoustic branch.

The 4D $S(\mathbf{Q}, E)$ measured on single crystals is very sensitive to the details in phonon dispersion, and one can choose to examine the data along various directions, in multiple Brillouin zones. Figure 5 provides another example of RuCl_3 , [35] where we directly compare experimental and simulated INS spectra. Such a direct comparison of the spectra illustrates the accuracy in not only the calculated phonon frequencies but also the vibrational modes. This is because the INS scattering ‘intensity’ is determined by the phonon polarization vectors, thus matching the simulated and experimental INS intensity profile requires not only

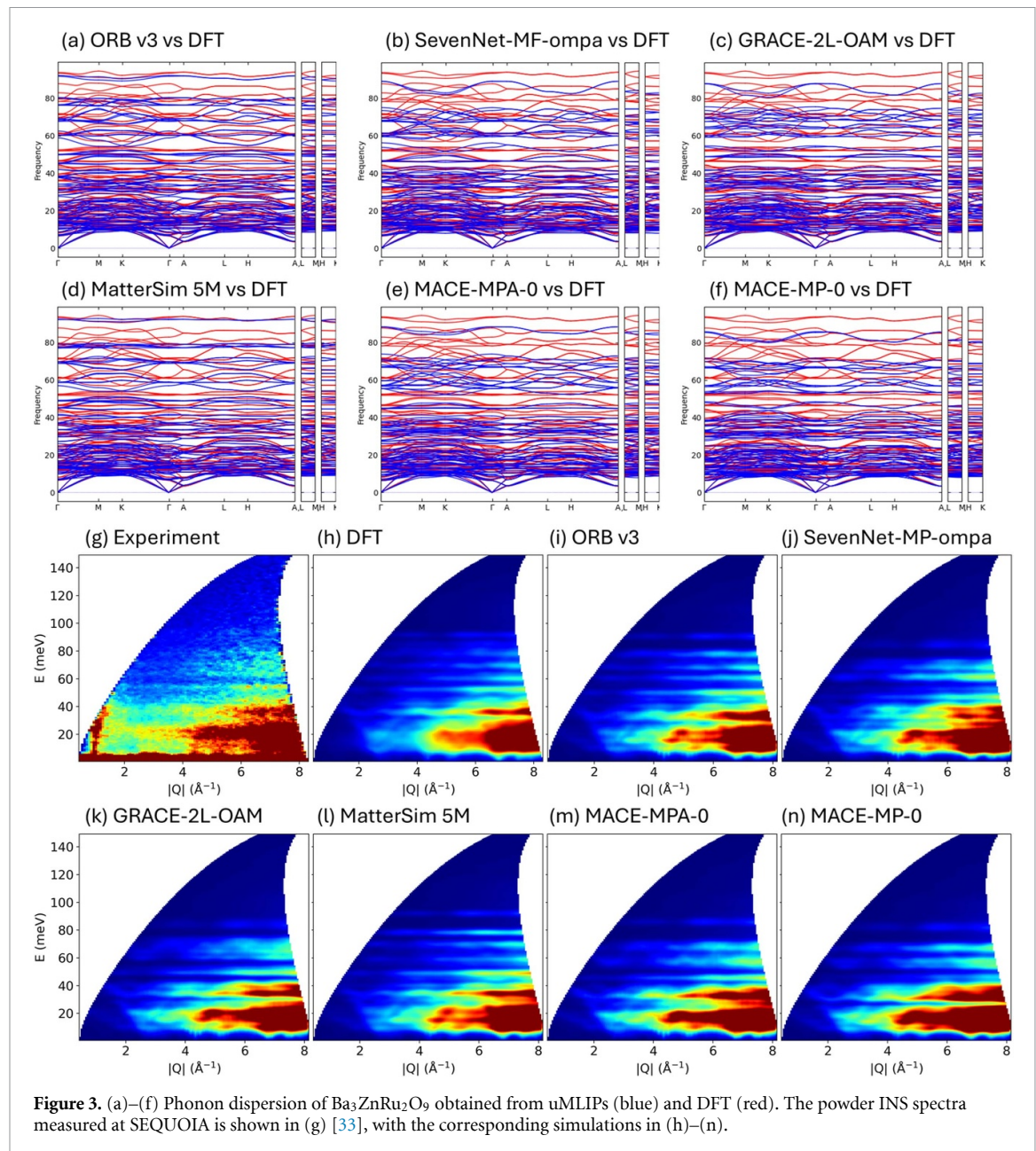


Figure 3. (a)–(f) Phonon dispersion of $\text{Ba}_3\text{ZnRu}_2\text{O}_9$ obtained from uMLIPs (blue) and DFT (red). The powder INS spectra measured at SEQUOIA is shown in (g) [33], with the corresponding simulations in (h)–(n).

correct frequencies, but also correct polarization vectors. In the case of RuCl_3 , figure 5 shows that MatterSim achieves near-DFT accuracy in both vibrational frequencies and modes. However, if examined closely, there are still fine details that MatterSim cannot reproduce accurately, such as the faint band at about 15 meV near the Zone center, as well as the side branches near 35 meV. Even with these discrepancies, the similarities are sufficient to allow unambiguous assignments of the modes. Other models, with varying degrees of agreement, exhibit significant discrepancies in certain parts of the spectra.

We now set a higher standard to examine some more difficult cases. The modeling of hydrogen-containing materials has posed significant challenges for uMLIPs because hydrogen can form various bonds with different elements in different compounds, exhibiting drastically different vibrational properties. This is particularly relevant for INS analysis because neutrons scatter strongly from hydrogen with a large incoherent cross-section. Since the neutron scattering from hydrogen-containing material is predominantly incoherent, the Q dependence is much less relevant, and one can focus on the energy spectra in 1D. VISION [36] is an indirect geometry neutron spectrometer optimized to measure such energy spectra from hydrogen-containing materials. It can be considered a neutron scattering analogue of Raman or infrared spectroscopy. In the following, we will show a few examples in figure 6, each representing a group of hydrogen-containing materials. Specifically, we will study (a) ZrH_2 , an inorganic hydride with a simple crystal structure, (b) ZIF-8, a metal-organic framework, (c) polyethylene, an organic polymer, (d) toluene, an aromatic molecular crystal, (e) butyric acid, an aliphatic hydrocarbon, and (f) remdesivir, a more complex

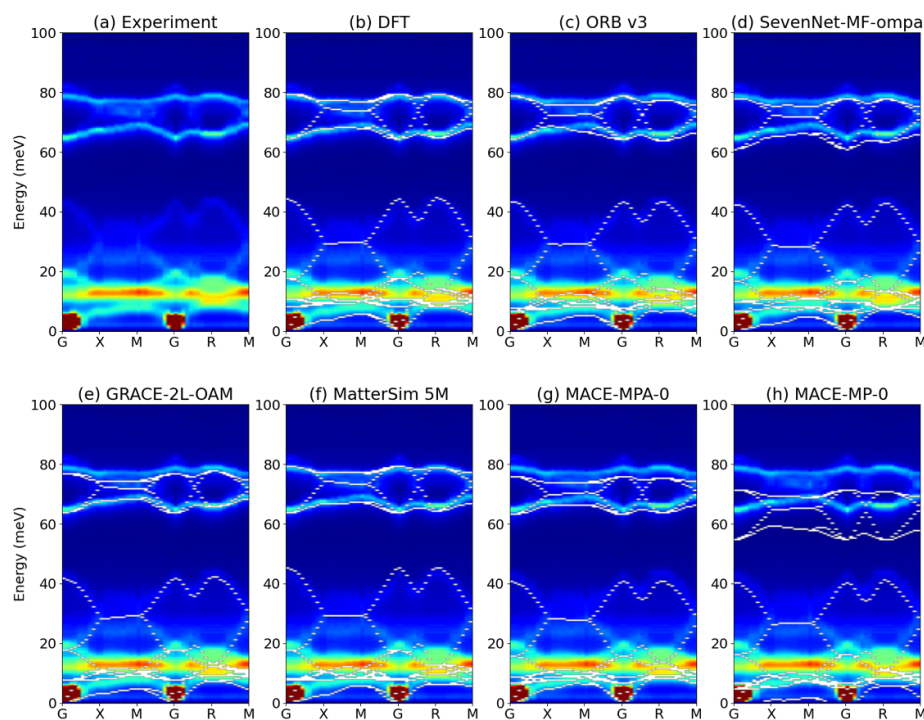


Figure 4. Phonon dispersion of Cu_2O along representative paths. (a) Shows the INS spectrum obtained from a single-crystal INS experiment by Saunders *et al* [34]. In (b)–(h), the calculated phonon dispersion from DFT and various uMLIPs are overlapped (as white dots) on the experimental spectrum for comparison.

crystal of drug molecules. In addition to the previously mentioned uMLIPs, we also include MACE-OFF (MACE-OFF23_medium) [37] to replace MACE-MP-0 when applicable. MACE-OFF is a derivative of MACE fine-tuned for organic crystals made of one or more elements in H, C, N, O, F, P, S, Cl, Br, and I.

The VISION spectrum from ZrH_2 (figure 6(a)) is relatively simple, showing a strong Zr–H libration band at 150 meV and its higher-order overtones. Although all uMLIPs reproduced this general feature, MatterSim and MACE-MP-0 underestimate frequencies significantly. The fundamental excitation peaks from MACE-MP-0 appear to be broader than they should be, possibly due to an overestimated anisotropy in the potential energy profile for hydrogen, causing broader splitting of H libration in different directions. ZIF-8 is a hybrid material that contains organic ligands anchored by metal centers. The hybrid nature makes it challenging for uMLIPs to describe. Indeed, none of the potentials tested is good enough for ZIF-8 (figure 6(b)), and ORB v3 is the only one that is reasonably close. Polyethylene is a polymer with a simple composition and crystal structure. The INS peaks are mainly caused by the deformation of the molecular chain and local vibrations of the bonds in methylene groups ($-\text{CH}_2-$). The relative motions between chains and long-range phonon modes contribute to the band below 30 meV (figure 6(c)). Not surprisingly, MACE-OFF produces the best results because it has been specifically fine-tuned to describe such interactions. ORB v3 also gives satisfactory results. Toluene has a simple molecular structure, but a relatively complex crystal structure (figure 6(d)). The methyl group and the aromatic ring each produce some characteristic sharp peaks in the INS spectra, which are determined by multiple types of bonds and interactions. Figure 6(d) shows that only the spectrum simulated by MACE-OFF is close enough to the experiment for peak assignment and spectral interpretation. ORB v3 and SevenNet-MF-ompa also capture the lower energy features (<75 meV) well. The observation is similar for butyric acid (figure 6(e)), where vibrations from the OH group and the short C–C chain dominate the INS intensities. To further illustrate the usefulness of uMLIPs in more complex molecular crystals, we also examined remdesivir, an antiviral medication used to treat COVID-19. The spectra are much more complex with more peaks (figure 6(f)), making peak assignment and spectral interpretation more challenging. The spectrum from MACE-OFF is reasonably close to that of DFT, but this is not sufficient because here even DFT fails to reproduce all the peaks.

Overall, the above analysis shows that the best performing uMLIPs can be an accurate alternative to DFT when the latter is not available or cannot meet the requirements of time-sensitive tasks. There are still certain areas where these potentials tend to be less accurate, such as inorganic-organic hybrid materials or complex molecular crystals, but they are ready to play an increasingly important role in the analysis of phonons, thermodynamic properties, and rapid spectral analysis. Specifically, we have found MatterSim 5 M and ORB

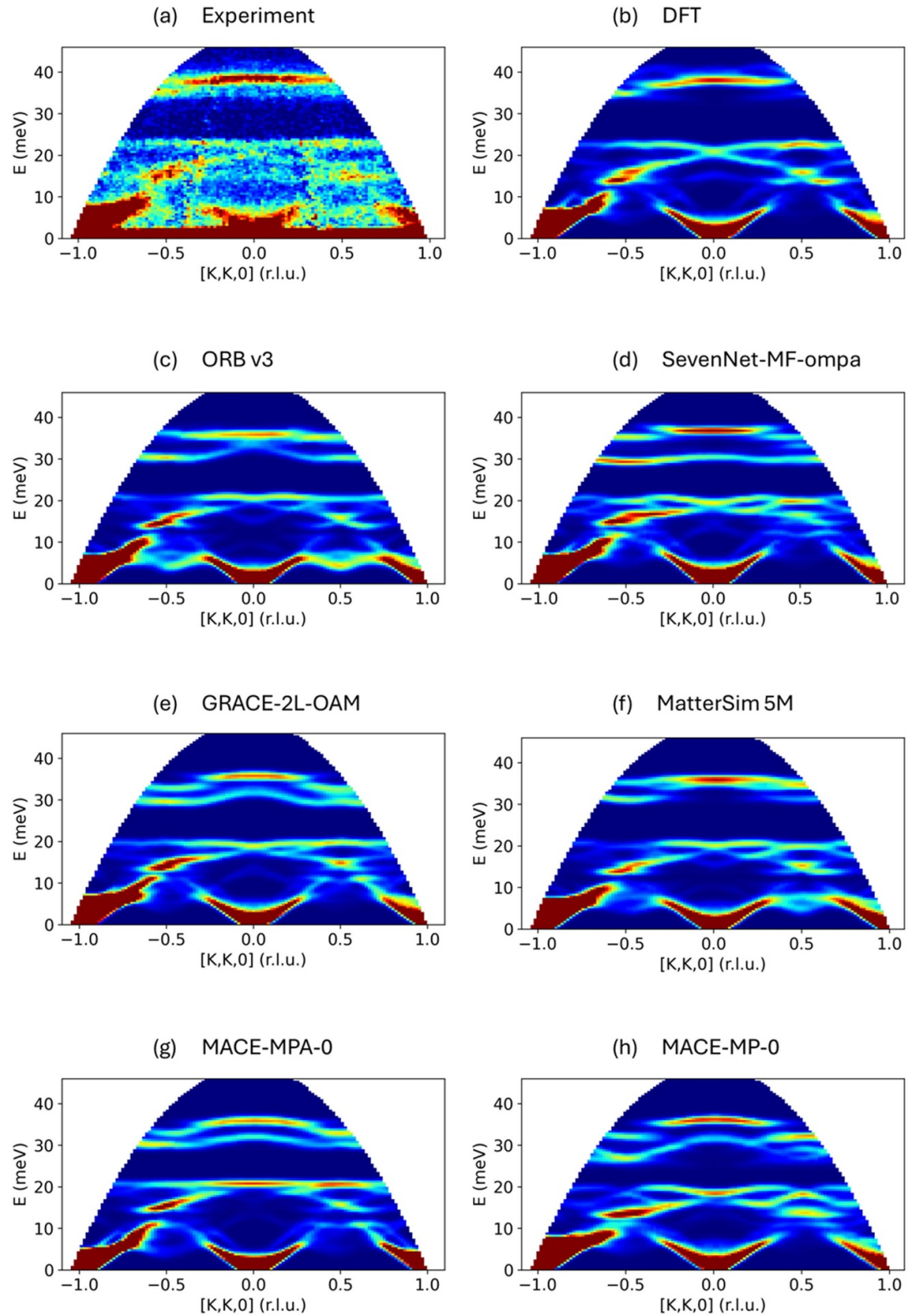


Figure 5. Single crystal INS spectra from RuCl_3 along the $[K,K,0]$ direction from (a) experiment by Mu *et al* [35] and (b)–(h) simulations.

v3 are constant top performers in reproducing features in INS spectra. For organic materials, MACE-OFF usually performs better than others because it has been fine-tuned with databases including organic materials. The general observation is that uMLIPs trained on all three major databases (OMat24, MPtrj, and Alexandria) tend to perform better, and ORB v3 stands out in many tests, probably because it is the only one that included the complete Alexandria database with more than 30 million structures (with a total of 133 million structures). MatterSim was trained on its own database of 17 million structures. Its outstanding

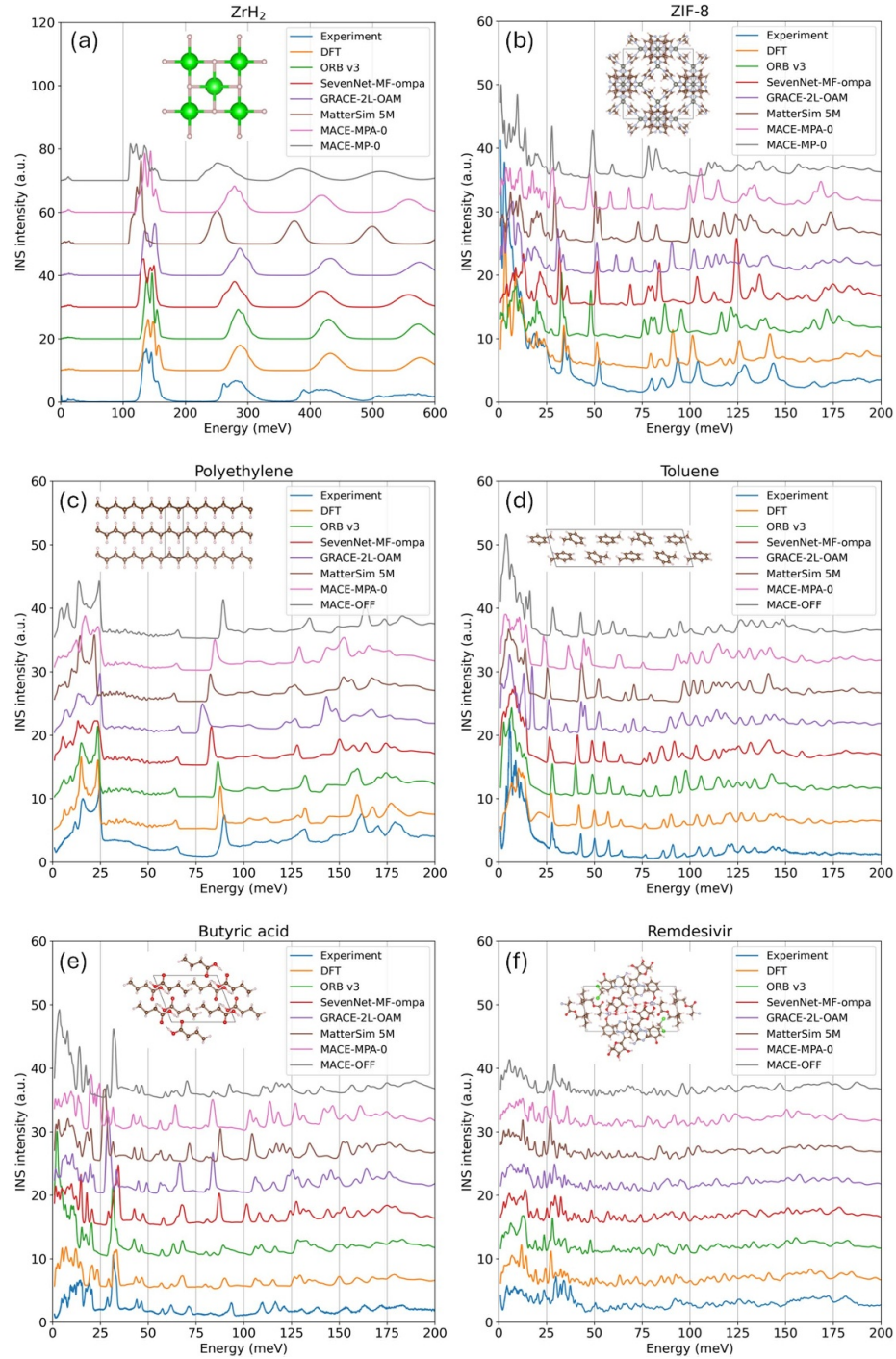


Figure 6. Experimental (measured at VISION) and simulated INS spectra for (a) ZrH_2 [38] (b) ZIF-8 [39] (c) polyethylene (d) toluene [33] (e) butyric acid (f) remdesivir [40] Insets show the crystal structure of each material.

performance is a demonstration of the efficiency of its data selection. The coverage of a wide temperature and pressure range in their database is likely the key. Note that all uMLIPs benchmarked in this work are published versions in literature without further fine-tuning. While the selection of training datasets certainly has a major impact on the performance, the architecture used in different models may also play a significant role. For uMLIPs trained on different datasets, it would be important to perform fine-tuning on a designated dataset to better assess the architecture itself. Our benchmark with INS spectra also highlights that similar performance in traditional metrics (errors in energies and forces) does not necessarily translate to similar performance in a specific application. In this case, we clearly show that, although high-ranking uMLIPs generally perform better in simulating INS, significant variations exist from case to case, and the best uMLIP for INS simulation may not have the highest benchmarking score with traditional metrics.

Based on the findings in this work, we have included several of the top performing uMLIPs in the new version of INSPIRED [12], which has a graphical user interface for users to perform fast and efficient simulations of the INS spectra. Compared to conventional DFT calculations, which take hours to days on a computer cluster, the same simulations with uMLIPs only take seconds to minutes on a single CPU (see tables S8–S10). Given the typical timescale of a neutron scattering experiment, the analysis and interpretation can now be performed on-the-fly, as the experimental data are being collected, without special computing resources, and the results can be used for immediate peak assignment, spectral interpretation, and steering of the experiment. This can have significant implications on the future of INS experiments and vibrational spectroscopy in general.

3. Conclusion

We have produced a high-quality DFT-based phonon database focusing on crystals with small unit cells and rich phonon dispersion features. By benchmarking uMLIPs against this database, we have demonstrated that the most recent uMLIPs, such as MatterSim, ORB v3, MACE-OFF, SevenNet-MF-ompa, GRACE-2 L-OAM, and MACE-MPA-0, can potentially be used to perform calculations of phonon dispersion and vibrational spectra with near DFT accuracy. We further illustrate how the simulated INS spectra from the uMLIPs compare with experiments in various scenarios, including direct and indirect geometry spectrometers, powder and single crystal measurements, coherent and incoherent scatterers, inorganic and organic materials. The side-by-side comparisons verify the applicability of the uMLIPs for rapid analysis of INS spectra and lay the foundation for future autonomous experiments.

The rapid development of uMLIPs has been fueled by the growth of computing power, the availability of new databases, and accelerated machine learning algorithms. It took only a few years for the early models to evolve into powerful ones that seem to be sufficient for phonon calculations. The next challenge would be the higher-order force constants, phonon anharmonicity, and properties away from equilibrium. Such properties are even more difficult to model with DFT, and the production of large-scale ground truth datasets is the key. Nonetheless, the lesson we learned from the development of uMLIPs is that AI in science is opening a new era of materials research and revolutionizing how experiments and data analysis are performed at an unprecedented speed.

4. Methods

Plane-wave DFT calculations were performed using the Vienna *Ab initio* Simulation Package (VASP) [41, 42]. The crystal structure from the MP database [13] was taken as the initial structure, and the atomic coordinates are relaxed before phonon calculation. Perdew–Burke–Ernzerhof implementation of the generalized gradient approximation [43] was used for all calculations. The cutoff energy is 1.6 times the maximum ENMAX in the POTCAR. The density of k-points per volume is $160 \text{ \AA}^{-3}$. VASP precision level was set as ‘accurate’. The energy tolerance for electronic minimization was 10^{-8} eV . The maximum force after relaxation was $0.002 \text{ eV \AA}^{-1}$. For nonmetals with more than one element, Born effective charges were calculated, and non-analytical calculations were automatically performed to describe the LO-TO splitting. Spin-polarized calculations were performed for magnetic materials. The dimensions of the supercells were automatically determined so that the minimum dimension in a supercell is at least $12 \text{ \AA}$. Phonopy [44, 45] was used to calculate the dispersion and uniform sampling of phonons in the Brillouin zone. Finally, crystals with significant phonon instability after such calculations (phonons with negative/imaginary frequencies) were not included in the database.

The specific versions of used uMLIPs are tabulated in table S7. To calculate the phonon properties of the materials with selected uMLIPs in the database, the geometry relaxations were performed with the atomic simulation environment (ASE) [46], using the fast inertial relaxation engine with force convergence criteria set to $0.005 \text{ (eV \AA}^{-1})$. We did not relax the volume or shape of the unit cells in this study because our goal was to benchmark and compare the phonon properties calculated on the same crystal structures (unit cells). We do not evaluate the capability of the potential to predict the volume/density of the material, which corresponds to a different dimension in the potential energy profile. A proper prediction of volume/density (to be compared with experiments) also requires consideration of the temperature-dependent vibrational entropy and is deferred to future studies.

The average difference of interatomic distance between the initial and optimized atomic coordinates for each material is calculated as

$$\Delta d = \frac{\sum_{i=1}^n |x_i - \hat{x}_i|}{n}, \quad (1)$$

where $\mathbf{x}_i$ is the optimized atomic coordinate of atom i based on DFT, $\hat{\mathbf{x}}_i$ is that based on uMLIP, and n is the total number of atoms in the supercell of the material. Furthermore, force calculations were conducted with the ASE and Phonopy packages using finite displacement methods. The displacement is set to 0.03 Å.

The $\mathbf{q}$ -points for each material are generated by the pymatgen package [47], and the density of the grid is 1000 per Å⁻³ in the reciprocal unit cell with Γ -point at the center. The phonon frequencies are calculated on each $\mathbf{q}$ -point for all uMLIPs and DFT. The mean absolute errors (MAEs) of the phonon frequencies of each material and each uMLIP compared to that of DFT are calculated as

$$\text{MAE}_\omega = \frac{\sum w_j \cdot \left(\frac{\sum_{i=1}^n |\omega_i - \hat{\omega}_i|}{n} \right)}{\sum w_j}, \quad (2)$$

where w_j is the weight of $\mathbf{q}$ -point j , n is the number of phonon modes at each $\mathbf{q}$ -point, $\hat{\omega}_i$ is the frequency of phonon mode i from a uMLIP and ω_i is the frequency of phonon mode i from the DFT. The phonon density of states and the thermodynamical properties including Helmholtz free energy, heat capacity, and entropy are calculated based on the phonon properties using Phonopy. For the thermodynamic properties, a temperature scan from 0 K to 1000 K is performed (every 10 K). The MAEs of the thermodynamical properties are calculated as

$$\text{MAE}_X = \frac{1}{T_{\max}} \int_0^{T_{\max}} |X(T) - \hat{X}(T)| dT, \quad (3)$$

where the X and $\hat{X}$ represent one of the thermodynamic properties from DFT and uMLIPs, respectively, T is the temperature and $T_{\max}$ is the maximum temperature, 1000 K.

To compare the PDOS obtained from DFT calculations with those generated from uMLIP, we use Spearman's rank correlation coefficient as the statistical measurement. Spearman's rank correlation coefficient is expressed as

$$\rho = 1 - \frac{6 \sum_i d_i^2}{n(n^2 - 1)}, \quad (4)$$

where d_i is the difference between the ranks of y_i and $\hat{y}_i$ in DFT PDOS and uMLIP PDOS, respectively, and n is the number of elements in each vector.

For the computing time listed in tables S8–S10, all DFT calculations were performed on a computer cluster, with each crystal running on 32 CPU cores/threads. All uMLIP calculations were performed on a single CPU with `OMP_NUM_THREADS = 2`. Note that various factors, including hardware and software configurations, can influence the exact computing time of each uMLIP. In this work, we did not attempt to optimize the performance for individual uMLIPs. Rather, the primary purpose of the comparisons in tables S8–S10 is to demonstrate the overall orders-of-magnitude speedup that uMLIPs offer over DFT, enabling rapid and effortless phonon calculations.

Data availability statement

The code and data that support the findings of this study are openly available at the following URL/DOI: https://github.com/maplewen4/phonon_uMLIP and <https://doi.org/10.5281/zenodo.15298436>.

Acknowledgments

The authors were supported by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy (DOE), under Contract No. DE-AC0500OR22725 with UT Battelle, LLC. This research was partially sponsored by the Artificial Intelligence Initiative as part of the Laboratory Directed Research and Development (LDRD) program of Oak Ridge National Laboratory (ORNL). Computing resources were made available through the VirtuES and ICE-MAN projects, funded by the LDRD program and the Compute and Data Environment for Science (CADES) at ORNL, as well as resources of 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, operated under Contract No. DE-AC02-05CH11231 using NERSC award ERCAP0024340. INS spectra not previously published were obtained at the VISION beamline of the Spallation Neutron Source, a DOE Office of Science User Facility operated by ORNL.

Supporting information available

The Supporting Information contains additional figure and tables.

ORCID iDs

Bowen Han  0000-0002-1979-2402

Yongqiang Cheng  0000-0002-3263-4812

References

- [1] Reissland J A 1973 *The Physics of Phonons* (Wiley)
- [2] Togo A and Tanaka I 2015 First principles phonon calculations in materials science *Scr. Mater.* **108** 1–5
- [3] Fultz B 2020 *Inelastic Scattering* revision 1.0 edn (California Institute of Technology)
- [4] Liang H, Wang C, Yu H, Kirsch D, Pant R, McDannald A, Kusne A G, Zhao J-C, Takeuchi I 2024 Real-time experiment-theory closed-loop interaction for autonomous materials science (arXiv:2410.17430)
- [5] Xie Y, Sattari K, Zhang C and Lin J 2023 Toward autonomous laboratories: convergence of artificial intelligence and experimental automation *Prog. Mater. Sci.* **132** 101043
- [6] Tang S *et al* 2024 A machine learning decision criterion for reducing scan time for hyperspectral neutron computed tomography systems *Sci. Rep.* **14** 15171
- [7] Unke O T, Chmiela S, Sauceda H E, Gastegger M, Poltavsky I, Schütt K T, Tkatchenko A and Müller K-R 2021 Machine learning force fields *Chem. Rev.* **121** 10142–86
- [8] Riebesell J, Goodall R E A, Benner P, Chiang Y, Deng B, Ceder G, Asta M, Lee A A, Jain A and Persson K A 2024 Matbench discovery—a framework to evaluate machine learning crystal stability predictions (arXiv:2308.14920)
- [9] Deng B, Choi Y, Zhong P, Riebesell J, Anand S, Li Z, Jun K, Persson K A and Ceder G 2025 Systematic softening in universal machine learning interatomic potentials *npj Comput. Mater.* **11** 1–9
- [10] Loew A, Sun D, Wang H-C, Botti S and Marques M A L 2024 Universal machine learning interatomic potentials are ready for phonons (arXiv:2412.16551)
- [11] Togo A 2025 Atztogo/Phonondb (available at: <https://github.com/atztogo/phonondb>)
- [12] Han B, Savici A T, Li M and Cheng Y 2024 INSPIRED: inelastic neutron scattering prediction for instantaneous results and experimental design *Comput. Phys. Commun.* **304** 109288
- [13] Jain A *et al* 2013 Commentary: The materials project: a materials genome approach to accelerating materials innovation *APL Mater.* **1** 011002
- [14] Fu X, Wood B M, Barroso-Luque L, Levine D S, Gao M, Dzamba M and Zitnick C L 2025 Learning smooth and expressive interatomic potentials for physical property prediction (arXiv:2502.12147)
- [15] Rhodes B, Vandenhaute S, Šimkus V, Gin J, Godwin J, Duignan T and Neumann M 2025 Orb-v3: atomistic simulation at scale (arXiv:2504.06231)
- [16] Park Y, Kim J, Hwang S and Han S 2024 Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations *J. Chem. Theory Comput.* **20** 4857–68
- [17] Kim J, Kim J, Kim J, Lee J, Park Y, Kang Y and Han S 2024 Data-efficient multifidelity training for high-fidelity machine learning interatomic potentials *J. Am. Chem. Soc.* **147** 1042–54
- [18] Bochkarev A, Lysogorskiy Y and Drautz R 2024 Graph atomic cluster expansion for semilocal interactions beyond equivariant message passing *Phys. Rev. X* **14** 021036
- [19] Yang H *et al* 2024 MatterSim: a deep learning atomistic model across elements, temperatures and pressures (arXiv:2405.04967)
- [20] Batatia I *et al* 2024 A foundation model for atomistic materials chemistry (arXiv:2401.00096)
- [21] Batatia I, Kovacs D P, Simm G N C, Ortner C and Csanyi G 2022 MACE: higher order equivariant message passing neural networks for fast and accurate force fields *Advances in Neural Information Processing Systems*
- [22] Batatia I, Batzner S, Kovács D P, Musaelian A, Simm G N C, Drautz R, Ortner C, Kozinsky B and Csányi G 2022 The design space of E(3)-equivariant atom-centered interatomic potentials (arXiv:2205.06643)
- [23] Barroso-Luque L, Shuaibi M, Fu X, Wood B M, Dzamba M, Gao M, Rizvi A, Zitnick C L and Ulissi Z W 2024 Open Materials 2024 (OMat24) inorganic materials dataset and models (arXiv:2410.12771)
- [24] Neumann M, Gin J, Rhodes B, Bennett S, Li Z, Choubisa H, Hussey A and Godwin J 2024 Orb: a fast, scalable neural network potential (arXiv:2410.22570)
- [25] Deng B, Zhong P, Jun K, 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** 1031–41
- [26] Chen C and Ong S P 2022 A universal graph deep learning interatomic potential for the periodic table *Nat. Comput. Sci.* **2** 718–28
- [27] Baumann K and Clerc J 1997 Computer-assisted IR spectra prediction—linked similarity searches for structures and spectra *Anal. Chim. Acta* **348** 327–43
- [28] Zou Z, Zhang Y, Liang L, Wei M, Leng J, Jiang J, Luo Y and Hu W 2023 A deep learning model for predicting selected organic molecular spectra *Nat. Comput. Sci.* **3** 957–64
- [29] Squires G L 2012 *Introduction to the Theory of Thermal Neutron Scattering* 3rd edn (Cambridge University Press)
- [30] Sears V F 1992 Neutron scattering lengths and cross sections *Neutron News* **3** 26–37
- [31] Granroth G E, Kolesnikov A I, Sherline T E, Clancy J P, Ross K A, Ruff J P C, Gaulin B D and Nagler S E 2010 SEQUOIA: a newly operating chopper spectrometer at the SNS *J. Phys.: Conf. Ser.* **251** 012058
- [32] Cheng Y Q, Daemen L L, Kolesnikov A I and Ramirez-Cuesta A J 2019 Simulation of inelastic neutron scattering spectra using OCLIMAX *J. Chem. Theory Comput.* **15** 1974–82
- [33] Cheng Y, Wu G, Pajewski D M, Stone M B, Savici A T, Li M and Ramirez-Cuesta A J 2023 Direct prediction of inelastic neutron scattering spectra from the crystal structure *Mach. Learn.: Sci. Technol.* **4** 015010
- [34] Saunders C N *et al* 2022 Thermal expansion and phonon anharmonicity of cuprite studied by inelastic neutron scattering and *ab initio* calculations *Phys. Rev. B* **105** 174308
- [35] Mu S *et al* 2022 Role of the third dimension in searching for Majorana fermions in α -RuCl₃ via phonons *Phys. Rev. Res.* **4** 013067

- [36] Seeger P A, Daemen L L and Larese J Z 2009 Resolution of VISION, a crystal-analyzer spectrometer *Nucl. Instrum. Methods Phys. Res.* **604** 719–28
- [37] Kovács D P *et al* 2025 MACE-OFF: transferable short range machine learning force fields for organic molecules (arXiv:2312.15211)
- [38] Zhang J, Cheng Y, Kolesnikov A I, Bernholc J, Lu W and Ramirez-Cuesta A J 2021 Study of anharmonicity in zirconium hydrides using inelastic neutron scattering and *ab-initio* computer modeling *Inorganics* **9** 29
- [39] Casco M E, Fernández-Catalá J, Cheng Y, Daemen L, Ramirez-Cuesta A J, Cuadrado-Collados C, Silvestre-Albero J and Ramos-Fernandez E V 2017 Understanding ZIF-8 performance upon gas adsorption by means of inelastic neutron scattering *Chem. Select* **2** 2750–3
- [40] Mamontov E, Cheng Y, Daemen L L, Kolesnikov A I, Ramirez-Cuesta A J, Ryder M R and Stone M B 2021 Low rotational barriers for the most dynamically active methyl groups in the proposed antiviral drugs for treatment of SARS-CoV-2, apilimod and tetrandrine *Chem. Phys. Lett.* **777** 138727
- [41] 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
- [42] Blöchl P E 1994 Projector augmented-wave method *Phys. Rev. B* **50** 17953–79
- [43] Perdew J P, Burke K and Ernzerhof M 1996 Generalized gradient approximation made simple *Phys. Rev. Lett.* **77** 3865–8
- [44] Togo A, Chaput L, Tadano T and Tanaka I 2023 Implementation strategies in phonopy and phono3py *J. Phys.: Condens. Matter* **35** 353001
- [45] Togo A 2023 First-principles phonon calculations with phonopy and phono3py *J. Phys. Soc. Japan* **92** 012001
- [46] Larsen A H *et al* 2017 The atomic simulation environment-a python library for working with atoms *J. Phys.: Condens. Matter* **29** 273002
- [47] Ong S P, Richards W D, Jain A, Hautier G, Kocher M, Cholia S, Gunter D, Chevrier V L, Persson K A and Ceder G 2013 Python materials genomics (pymatgen): a robust, open-source python library for materials analysis *Comput. Mater. Sci.* **68** 314–9