

# MatterSim-MT: A multi-task foundation model for *in silico* materials characterization

Han Yang<sup>1\*</sup>, Xixian Liu<sup>2†</sup>, Chenxi Hu<sup>2†</sup>, Yichi Zhou<sup>2†</sup>, Yu Shi<sup>2†</sup>,  
Chang Liu<sup>2†</sup>, Junfu Tan<sup>2†</sup>, Jielan Li<sup>2\*†</sup>, Guanzhi Li<sup>2†</sup>, Qian Wang<sup>2†</sup>,  
Yu Zhu<sup>2†</sup>, Zekun Chen<sup>2†</sup>, Shuizhou Chen<sup>2†</sup>, Fabian Thiemann<sup>1</sup>,  
Claudio Zeni<sup>1</sup>, Matthew Horton<sup>2</sup>, Robert Pinsler<sup>1</sup>, Andrew Fowler<sup>1</sup>,  
Daniel Zügner<sup>1</sup>, Tian Xie<sup>2</sup>, Lixin Sun<sup>1</sup>, Yicheng Chen<sup>2</sup>, Lingyu Kong<sup>2</sup>,  
Yeqi Bai<sup>1</sup>, Deniz Gunceler<sup>1</sup>, Frank Noé<sup>1</sup>, Hongxia Hao<sup>2\*</sup>, Ziheng Lu<sup>2\*</sup>

<sup>1</sup>Microsoft Research AI for Science.

<sup>2</sup>This work was carried out while the author was affiliated with Microsoft Research.

\*Corresponding author(s). E-mail(s): [yang.han@microsoft.com](mailto:yang.han@microsoft.com); [jielanlee@outlook.com](mailto:jielanlee@outlook.com);  
[hongxiahao19@gmail.com](mailto:hongxiahao19@gmail.com); [zluag@connect.ust.hk](mailto:zluag@connect.ust.hk);

†These authors contributed equally to this work.

## Abstract

Accurate property characterization is a major bottleneck in materials design. While first-principles methods and task-specific machine-learning models have driven important progress, they remain fundamentally limited in scalability and generalizability across the vast space of structures and properties relevant to real-world materials design. We present **MatterSim-MT**, a multi-task foundation model for *in silico* materials simulation and property characterization. The model is pretrained on over 35 million first-principles-labeled structures covering 89 elements, temperatures up to 5000 K and pressures up to 1000 GPa, and is fine-tuned on various properties including Bader charges, magnetic moments, Born effective charges, and dielectric matrices. Out of the box, **MatterSim-MT** not only serves as a foundation model for predicting material structure, dynamics and thermodynamics, its multi-task architecture also enables a wide range of complex simulations that cannot be captured by potential energy surfaces alone. For example, we demonstrate pressure-dependent LO-TO phonon splitting in SiC with close agreement with experiment, electric hysteresis in ferroelectric BaTiO<sub>3</sub>, and the cationic-to-anionic redox transition during delithiation of a Li-rich cathode material. Finally, we show that **MatterSim-MT** scales well with more data and parameters, can be efficiently fine-tuned to higher levels of theory, and can be efficiently extended to new systems via active learning. Overall, we believe this approach provides a scalable route to accurate *in silico* materials characterization.

# 1 Introduction

Materials design drives a wide range of technological advances, for example in nanoelectronics,[1, 2] energy storage,[3, 4] biomedicine,[5] and environmental sustainability[6, 7]. However, the enormous space of potential candidates makes it challenging to find materials with the required properties under desired operating conditions. This problem is further exacerbated by the slow pace of traditional experimental synthesis and characterization. Recent advances in large-scale materials databases[8–18] and generative models for materials[19–24] have accelerated the generation of hypothetical structures with desired target properties. Yet, the ability to rapidly generate candidates has only increased the burden on experimental synthesis and property characterization.

Universal machine learning interatomic potentials (uMLIPs) aim to accelerate the materials design process by providing accurate stability and property predictions for a wide range of materials, improving candidate selection. Inspired by the success of foundation models in areas such as natural language processing[25–31] and image generation,[32, 33] a growing number of deep-learning-based uMLIPs trained on large-scale atomistic data for materials have been proposed.[15, 16, 34–46] Similar to foundation models in other fields, atomistic foundation models should support a wide range of downstream tasks. Yet, the dominant paradigm is to train uMLIPs exclusively on potential energy surface (PES) data.[34–37] While such uMLIPs accelerate structure optimization and molecular dynamics simulations,[47] many materials problems—such as predicting the dielectric response or polarization[48, 49]—cannot be solved by computing energies, forces, or stresses alone. To address this, a few efforts have sought to move beyond MLIPs by predicting additional material properties.[38, 50–53] However, these approaches are often limited to expert models tailored to specific systems and tasks, or evaluated only on machine learning benchmarks rather than realistic simulation workflows.

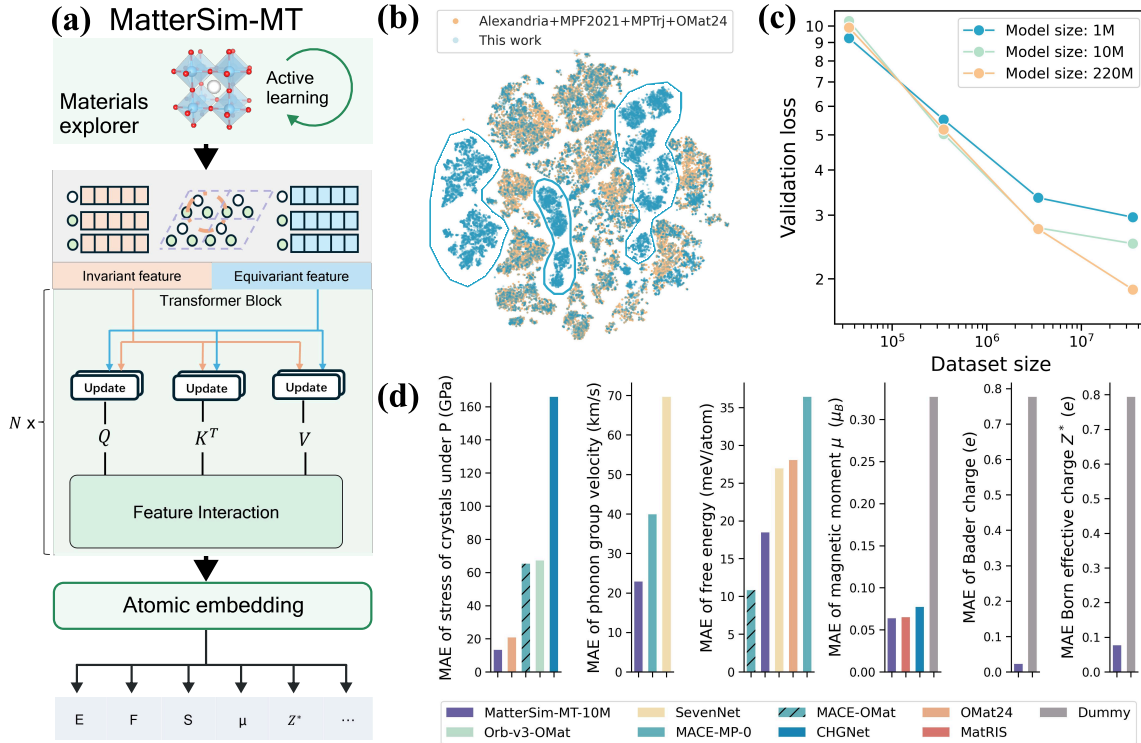
We present **MatterSim-MT**, a multi-task (MT) foundation model for *in silico* materials simulation and property characterization. The model learns a unified atomic representation trained to jointly predict multiple physical properties derived from first-principles calculations across a large range of materials. By leveraging active learning and integrating molecular dynamics simulations with large-scale first-principles computations,[54–56] we construct a dataset of atomic structures covering 89 elements, temperatures up to 5000 K, and pressures up to 1000 GPa. We pretrain **MatterSim-MT** on this data to enable first-principles-quality characterization of material stability, phonon spectra, mechanical and transport properties under wide temperature and pressure ranges. For example, **MatterSim-MT** predicts maximum phonon frequencies with 1 THz error and provides more accurate predictions of phase stability for binary phase diagrams than previous methods, especially at higher temperatures and pressures. We further fine-tune the model on existing first-principles calculations of various materials properties to enable accurate property prediction. Beyond energies, forces and

stresses, **MatterSim-MT** natively supports predictions for magnetic moments, Bader charges, Born effective charges (BEC), and dielectric matrices. For instance, the model achieves Born effective charges with 0.08 e error on inorganic solids relative to first-principles calculations. More importantly, our multi-task architecture enables complex simulations of materials beyond the PES using a single unified model, including the simulation of materials under external conditions such as electrical fields. As a demonstration, we reproduce pressure-dependent LO-TO phonon splitting in SiC with close agreement with experiment, electric hysteresis in ferroelectric BaTiO<sub>3</sub>, and the cationic-to-anionic redox transition during delithiation of a Li-rich cathode material. Finally, we show that **MatterSim-MT** predictably improves with more data and model parameters, can be efficiently fine-tuned to higher levels of theory, and can be efficiently extended to new systems via active learning. Together, these capabilities establish **MatterSim-MT** as a versatile foundation model for materials simulation and property characterization across a wide range of operating conditions.

## 2 Learning the materials space

**MatterSim-MT** learns universal representations of materials across a wide range of temperatures and pressures by pretraining on large-scale atomistic data to reproduce a unified first-principles PES. The model is further fine-tuned on additional properties including magnetic moments ( $\mu$ ; assuming ferromagnetic ordering), Bader charges, dielectric matrices ( $\epsilon_\infty$ ), and Born effective charges ( $Z^*$ ) (see [Section S3](#)). Together, these tasks ensure that the model learns a unified representation of atomistic systems for predicting a wide range of materials properties, including structures, energetics, dynamics and thermodynamics, while also supporting multi-task simulations that requires prediction of properties beyond potential energy surfaces. This makes it possible to transfer knowledge across materials classes, properties, and operating conditions at scale.

**MatterSim-MT** builds on the iterative active learning framework introduced in Ref. [46](#) to generate a large, diverse first-principles dataset spanning ground-state and off-equilibrium structures over wide temperature and pressure ranges ([Fig. 1\(a\)](#), top). Underexplored structures are selectively labeled with first-principles calculations and used to continue improving the model. This process yields a dataset with broad chemical and configurational coverage, comprising 35 M atomic structures and approximately 450 M atomic force vector labels. To illustrate the diversity of the sampled configurational space, [Fig. 1\(b\)](#) shows a t-SNE analysis of atomic-level embeddings for the first 20 elements (see [Fig. S7](#) for an analysis across all 89 elements). Compared to datasets constructed from near-equilibrium relaxations [[9](#), [14](#), [34](#), [38](#)] or short *ab initio* molecular dynamics simulations [[16](#)], the dataset used in this work exhibits broader configurational coverage (see circled regions in [Fig. 1\(b\)](#)). Compared to **MatterSim-v1** in Ref. [46](#), the dataset size increases from 6 M to 35 M structures and



**Fig. 1: MatterSim-MT is a scalable materials foundation model trained on a large, diverse dataset, achieving state-of-the-art performance on a variety of prediction tasks.** (a) Illustration of MatterSim-MT architecture: The dataset is enriched by an active-learning-based materials explorer (top), which is used to train a transformer architecture with invariant and equivariant features (center), providing an atomic embedding from which various materials properties are predicted (bottom); (b) t-SNE analysis of the atomic embedding of the first 20 elements selected from the dataset used in this work compared with the combination of MPF2021, MPTrj, Alexandria and OMat24 datasets; (c) Validation loss of energy prediction with MatterSim-MT given different model and dataset sizes; (d) Performance comparison of MatterSim-MT-10M, OMat24 (eqV2-M OMat MPTrj-sAlex),[16] MACE-MP-0,[35], MACE-OMat, CHGNet,[38] Orb-v3,[37] SevenNet,[43], MatRIS,[51], and a dummy predictor that predicts the mean value of the dataset.

covers structures under a more continuous range of temperatures of 0–5000 K and pressures of 0–1000 GPa. This extension facilitates a wide range of downstream tasks from materials synthesis and reactions to high-pressure applications (see Fig. S6).

The model architecture of MatterSim-MT is based on GeoMFormer[57], a scalable, transformer-based architecture for atomistic simulation. We further incorporate physical priors to enforce energy conservation, respect periodic boundary conditions, and preserve translational and rotational invariance (Fig. 1(a); see Section S1 for details). In addition, we ensure that the model is able to make effective use of as much data as possible given that first-principles labels are expensive to generate. In Fig. 1(c), we show that the validation loss of MatterSim-MT keeps decreasing as the dataset size and model size increase. In practice, scaling the model size is limited by the speed requirements of the downstream task. We find that a 10 M parameter model strikes a balance between accuracy and

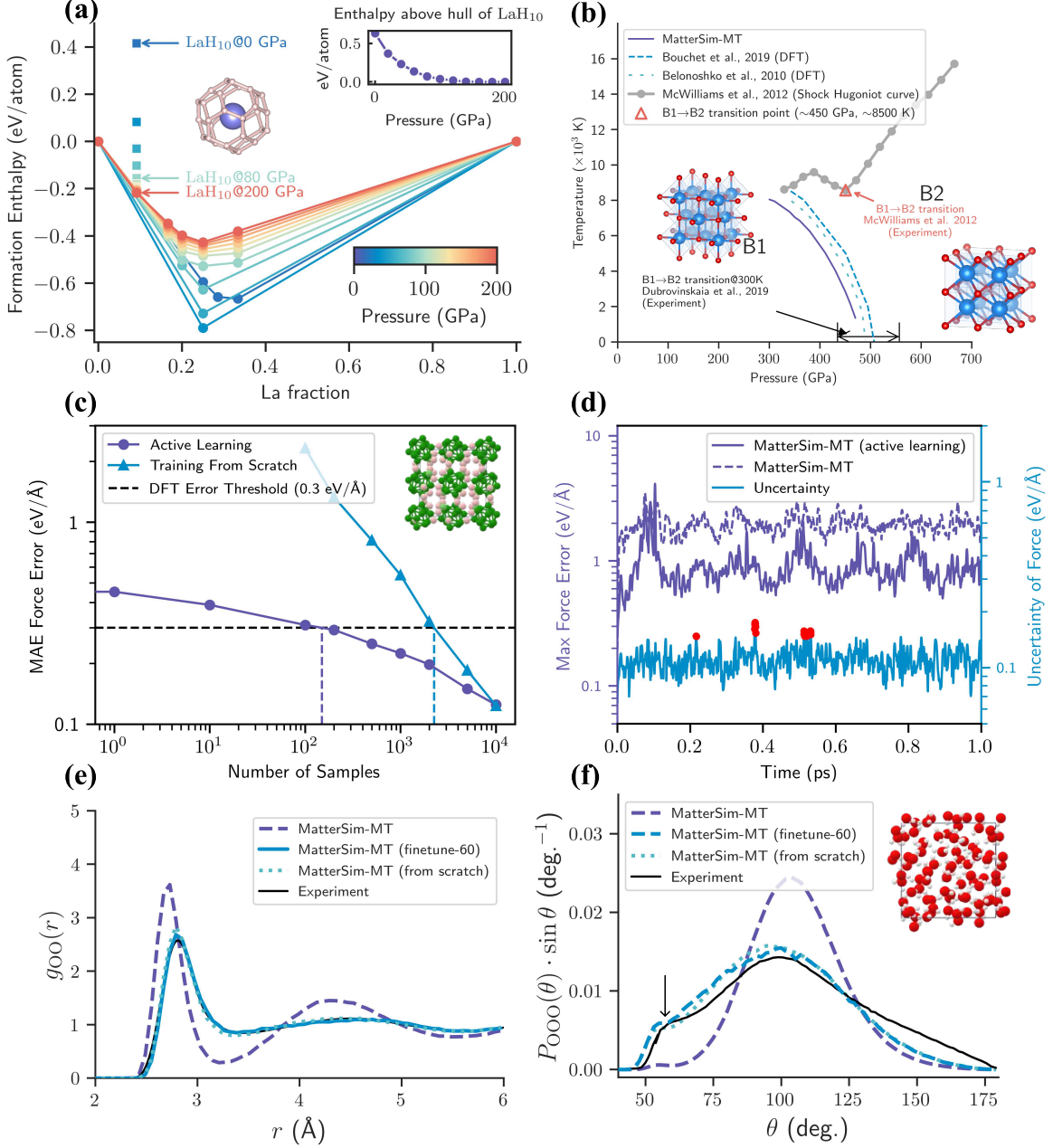
throughput for computationally intensive tasks such as running molecular dynamics simulations. In the following, we use `MatterSim-MT-10M` with 10 M parameters for predictions and simulations.

### 3 MatterSim-MT as a machine learning interatomic potential

As a materials foundation model pretrained on a large-scale, first-principles PES, `MatterSim-MT` delivers accurate predictions of energies, forces, and stresses for atomic structures across a wide range of chemical compositions, temperatures, and pressures. As an illustrative example, [Fig. 1\(d\)](#) presents a comparison of predicted stress errors for materials<sup>[61]</sup> that are predicted to be stable across a broad pressure range according to first-principles calculations. `MatterSim-MT` reproduces these stress states with a mean error of 8.5 GPa, demonstrating its ability to accurately capture structural responses from ambient to high-pressure conditions. In this section, we evaluate the model’s ability to predict material stability and phase diagrams under finite temperature and pressure conditions. Additional benchmarks on phonon spectra, mechanical properties, and molecular dynamics simulations are provided in [Section S4](#).

First, we evaluate the model’s ability to predict material stability. `MatterSim-MT` demonstrates strong performance in predicting phonon properties, achieving errors of approximately 1 THz for max phonon frequencies ([Table S4](#)) and 23 km/s for group velocities ([Fig. 1\(d\)](#)). These capabilities enable accurate estimation of vibrational free energies. In [Fig. 1\(d\)](#), we further demonstrate the model’s ability to predict free energy differences between 300–900 K for inorganic crystalline materials listed in [Table S5](#), achieving a mean absolute error of 18.50 meV/atom compared to first-principles calculations. Notably, the error drops to 9.62 meV/atom when increasing the number of parameters up to 1.3 B (`MatterSim-MT-1.3B`), further validating the scalability of the model. This ability enables rapid *ab initio* prediction of material stability under practical conditions.

Next, we evaluate the model’s ability to predict phase diagrams. In [Fig. 2\(a\)](#), we focus on the La–H chemical system and study the phase stability of LaH<sub>10</sub>, a well-known high-temperature superconductive hydride previously synthesized under extreme pressure.<sup>[62]</sup> A comparison with several publicly available foundational uMLIPs on the La–H convex hull ([Fig. S11](#)) shows that `MatterSim-MT` faithfully reproduces stabilization of LaH<sub>10</sub> under pressure at ~150 GPa, while others either predict premature stabilization or exhibit qualitatively different behavior at elevated pressures. Interestingly, we also observe a destabilization range of a few tens of gigapascals for the LaH<sub>2</sub> phase, which has been verified through first-principles computations.<sup>[62]</sup> Traditionally, deriving the pressure-dependent stability range of binary systems via experiments or conventional first-principles computations would take months. In this example, `MatterSim-MT` can predict phase stability within minutes.



**Fig. 2: MatterSim-MT is a foundation model for predicting material structure, dynamics and thermodynamics.** (a) Enthalpy above hull analysis of the La-H chemical system at pressures up to 200 GPa. Inset shows the stabilization of LaH<sub>10</sub> with increasing pressure. (b) Predicted B1-B2 phase boundary of MgO, benchmarked against first-principles calculations and experimental data. (c) Mean absolute error (MAE) of forces against first-principles results in the simulation trajectory of Li<sub>2</sub>B<sub>12</sub>H<sub>12</sub> (crystal structure shown in top right) from an actively learned model and a model trained from scratch, with increasing number of data samples used in active learning. (d) Force error and prediction uncertainty along the *ab initio* molecular dynamics (MD) trajectory for Li<sub>2</sub>B<sub>12</sub>H<sub>12</sub>. Red points represent structures with above-threshold uncertainty taken for active learning. (e) Oxygen-oxygen radial distribution functions (RDF) of bulk water from MD simulations using the pretrained MatterSim-MT model (zero-shot), MatterSim-MT finetuned on 60 water snapshots and MatterSim-MT trained from scratch on 900 water snapshots. Black line indicates the experimental reference.[58, 59] (f) Oxygen-oxygen-oxygen-angular distribution functions,  $P_{OOO}(\theta)$ , from MD simulations using the pretrained MatterSim-MT model, MatterSim-MT finetuned on 60 water snapshots and MatterSim-MT trained from scratch on 900 water snapshots. Black line represents the empirical potential structural refinement (EPSR) of joint X-ray and Neutron scattering measurements of bulk water at 298 K.[60]

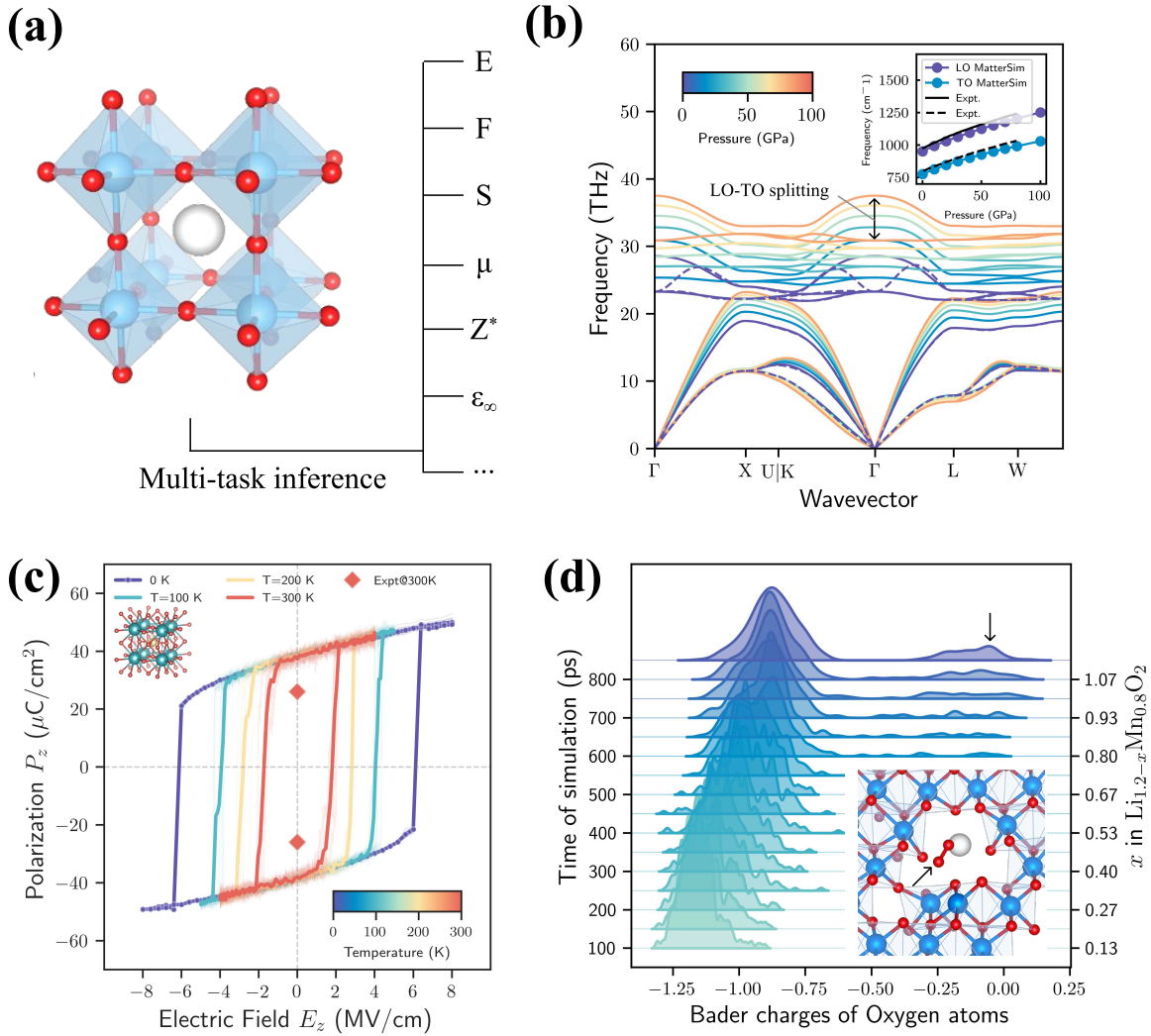
We also predict the temperature- and pressure-dependent phase diagram of MgO,[63, 64] a mineral that shapes the convection of Earth’s mantle[65, 66]. Our predictions extend to extreme conditions of 10,000 K and 500 GPa, with phase boundaries showing good agreement compared to both experimental measurements[63, 67] and first-principles density functional theory (DFT) calculations (Fig. 2(b)).[68, 69] A comparison with several publicly available foundational uMLIPs on the MgO B1–B2 phase boundary (Fig. S12) confirms that `MatterSim-MT` better reproduces the DFT reference, while others fail to capture the temperature dependence and shape of the phase boundary. These results demonstrate that `MatterSim-MT` serves as a high-accuracy uMLIP for energy-based simulations, especially at higher temperatures and pressures.

As a materials foundation model, `MatterSim-MT` can be further extended to new systems via active learning. For complex systems where out-of-distribution configurations may arise during long molecular dynamics simulations,[70–72] `MatterSim-MT` employs uncertainty-driven active learning to selectively label and incorporate the most informative configurations. We benchmark this approach on molten phosphorus, molten boron, and lithium dodecahydro-closo-dodecaborate ( $\text{Li}_2\text{B}_{12}\text{H}_{12}$ ) (see details in Section S6). As shown in Fig. 2(c,d), the actively learned model matches the accuracy of a model trained from scratch while requiring only a fraction of the data—just 6.6% for  $\text{Li}_2\text{B}_{12}\text{H}_{12}$ —confirming the broad configurational coverage of the pretrained model.

Similarly, the model can be fine-tuned to higher levels of theory beyond the Perdew–Burke–Ernzerhof (PBE) functional[73] used during pretraining (detailed in Section S2.3). As a demonstration, we fine-tune `MatterSim-MT` to the rev-PBE0-D3 level of theory[74] for liquid water (see Section S7.1 for details). With only 60 configurations, the fine-tuned model predicts the structural and dynamical properties of water (Fig. 2(e,f)) as accurate as a model trained from scratch on 900 configurations. The fine-tuned model also predicts the self-diffusion coefficient  $D = 0.193 \pm 0.011 \text{ \AA}^2 \text{ ps}^{-1}$  within 20% of the experimental value range  $0.23\text{--}0.24 \text{ \AA}^2 \text{ ps}^{-1}$ . [75, 76] This corresponds to a 92% reduction in data generation cost (see Fig. S14).

## 4 `MatterSim-MT` as a multi-task materials foundation model

Many important physical phenomena cannot be captured by PES predictions alone. `MatterSim-MT` natively predicts several other fundamental material properties, including atomic Bader charges, magnetic moments (assuming ferromagnetic ordering), Born effective charges ( $Z^*$ ), and dielectric matrices ( $\epsilon_\infty$ ). The ability to accurately predict these properties is crucial for applications such as catalysis, energy storage, and planetary sciences. On held-out test sets, the model achieves mean absolute errors of 0.023 e for Bader charges,  $0.064 \mu_B$  for magnetic moments,  $0.076 e$  for Born effective charges, and 0.25 for dielectric tensor elements (see Fig. 1(d) and Table S6). These additional outputs



**Fig. 3: MatterSim-MT as a multi-task emulator beyond the potential energy surface.** (a) Illustration of the multi-task inference capabilities of MatterSim-MT, including predictions of energy (E), forces (F), stress (S), magnetic moments ( $\mu$ ), Born effective charges ( $Z^*$ ), and dielectric matrices ( $\epsilon_\infty$ ) from atomic structures. (b) Pressure-dependent phonon spectrum of SiC up to 100 GPa, with inset comparing MatterSim’s predicted longitudinal optical (LO) and transverse optical (TO) splitting against experimental measurements, and the dashed curve illustrating the phonon dispersion under 0 GPa *without* LO-TO splitting for comparison. (c) Predicted hysteresis curve of polarization density as a function of the electrical field along z direction in the ferroelectric tetragonal BaTiO<sub>3</sub> material. (d) Evolution of oxygen Bader charge distributions in Li<sub>1.2-x</sub>Mn<sub>0.8</sub>O<sub>2</sub> during delithiation, with arrows indicating the formation of an O<sub>2</sub> molecule.

enable simulations of physical phenomena that are inaccessible to models trained on only PES data while spanning a much wider range of tasks than specialized expert models. We illustrate these multi-task capabilities through three case studies spanning vibrational spectroscopy, ferroelectric switching, and electrochemical redox, each requiring a distinct combination of property predictions.

First, we demonstrate how predictions of Born effective charges and dielectric properties enable the computation of phonon spectra in polar crystals, where long-range Coulomb interaction between ions splits the longitudinal optical (LO) and transverse optical (TO) phonon modes at the Brillouin

zone center.[49] LO-TO splitting is a hallmark of polar crystals but is absent in standard force-constant approaches. That is because it requires knowledge of the Born effective charge tensor  $Z^*$  and the electronic dielectric matrix  $\varepsilon_\infty$  to construct the non-analytical correction to the dynamical matrix.[49, 77, 78] As an illustration, we simulate the LO-TO splitting of phonon modes in 3c-SiC, a promising material for high-temperature and high-power electronics, under pressures up to 100 GPa.[79] As shown in Fig. 3(b) and Table S7, MatterSim-MT predicts a Born effective charge  $Z_{\text{Si}}^* = 2.71 e$  and a dielectric constant  $\varepsilon_\infty = 7.31$  at zero pressure, in close agreement with both theoretical ( $Z_{\text{Si}}^* = 2.72 e$ ,  $\varepsilon_\infty = 7.02$ )[80] and experimental ( $Z_{\text{Si}}^* = 2.70 e$ ,  $\varepsilon_\infty = 6.52$ )[81, 82] values. The resulting LO-TO splitting of 5.26 THz deviates by only 0.06 THz from *ab initio* calculations[80] and 0.03 THz from experimental measurements.[83] The inset of Fig. 3(b) further shows that the predicted pressure dependence of the LO and TO phonon frequencies agrees well with experiments across the full pressure range. Notably, MatterSim-MT accurately predicts Born effective charges and dielectric matrices for pressures up to 100 GPa, even though the training data did not contain any labels at such pressure levels. This result highlights the model’s ability to extrapolate beyond its training data.

The predicted Born effective charges also enable the simulation of polarization-driven phenomena under external electric fields. In ferroelectric materials, the coupling between atomic displacements and the macroscopic polarization gives rise to switchable polarization states. Simulating this behavior requires accurate interatomic forces and Born effective charges that relate atomic displacements to the macroscopic polarization. For BaTiO<sub>3</sub>, we incorporate the contribution of an external electric field into molecular dynamics simulations and reproduce the polarization-electric field hysteresis curve (Fig. 3(c)). The simulations correctly predict that finite-temperature effects lead to a reduced coercive field, even though the spontaneous polarization of 38  $\mu\text{C}/\text{cm}^2$  predicted at 300 K is slightly higher than the experimental value (26  $\mu\text{C}/\text{cm}^2$ ), likely due to the well-known underbinding at the PBE level of theory (see Ref. 50 and 53). More broadly, the ability to simulate polarization hysteresis without material-specific fitting suggests that this approach could be readily utilized to study complex ferroelectric phenomena such as ferroelectric-paraelectric phase transitions across diverse chemical systems.

Finally, we simultaneously predict atomic charges and magnetic moments to study the electronic degrees of freedom in chemical bonding and redox processes. Here, we examine the behavior of the cathode material Li<sub>1.2-x</sub>Mn<sub>0.8</sub>O<sub>2</sub> under electrochemical charging conditions. These lithium-rich transition-metal oxides are promising next-generation batteries due to their high energy density, but suffer from irreversible capacity loss associated with the anionic oxygen redox mechanism.[84] We perform molecular dynamics simulations of progressive lithium extraction at 1000 K, removing

one lithium atom every 5 ps until complete delithiation, while monitoring ion dynamics, magnetic moments, and atomic Bader charges predicted by **MatterSim-MT** (see Fig. S9 for details).

Our analysis shows a clear transition from cationic to anionic redox during lithium extraction. At low degrees of delithiation, Mn serves as the primary redox center, with magnetic moment changes of approximately  $3\text{--}4\mu_{\text{B}}$ . Upon reaching  $x \approx 0.5$  in  $\text{Li}_{1.2-x}\text{Mn}_{0.8}\text{O}_2$ , Mn ions begin migrating from octahedral sites in the transition-metal layers to tetrahedral sites in the lithium layers, accompanied by distortions of the oxygen sublattice. Further delithiation to  $x \approx 0.9$  triggers the formation of oxygen dimers, typically following Mn migration events. The evolution of these unstable dimers leads to the formation of molecular  $\text{O}_2$  within the lattice (Fig. 3(d), inset), a hallmark of irreversible anionic redox. This is corroborated by the Bader charge analysis in Fig. 3(d), where the  $\text{O}_2$  molecules exhibit near-zero charges, confirming their molecular rather than ionic character. Continued delithiation destabilizes the lattice further, resulting in increased molecular oxygen formation and an eventual transition toward nanofluidic oxygen states.[84] Notably, this comprehensive picture of the cationic-to-anionic redox transition and lattice degradation naturally emerges from the multi-task predictions, without any task-specific training on battery materials.

## 5 Discussion

With the advent of efficient generative models for materials exploration, accurate property characterization is becoming a major bottleneck in computational materials design. Atomistic foundation models offer a promising route to accelerate property characterization, but previous models have mostly focused on modeling the PES. In this work, we introduce **MatterSim-MT**, a multi-task foundation model for *in silico* materials property characterization beyond the PES. The model enables atomistic simulations for materials composed of arbitrary elemental combinations, for temperatures up to 5000 K and pressures up to 1000 GPa, providing accurate predictions for energetics, vibrational properties, and thermal transport. **MatterSim-MT** also natively predicts atomic Bader charges, magnetic moments, Born effective charges, and dielectric matrices, enabling simulations under external conditions such as electric fields using a single model. Although we demonstrate multi-task prediction for four properties beyond the PES, the underlying framework is general: Additional properties computed from first principles calculations can be incorporated through the same strategy. As the availability of high-quality synthetic data continues to grow, this multi-task paradigm allows us to greatly expand the range of applications that atomistic foundation models can address.

Despite these advances, several limitations remain. Scaling the model size increases computational cost, making system-level solutions such as model compression and efficient system engineering increasingly important.[85–87] More fundamentally, the physical reliability of predicted properties is

limited by the underlying level of theory. For example, Born effective charges and dielectric tensors computed via Berry-phase approaches[48, 88] become ill-defined in narrow-gap or metallic systems,[89] and a multi-task model may inadvertently learn these nonphysical divergences. Relatedly, some properties may not be well-defined for some systems. For example, we predict magnetic density assuming ferromagnetic ordering, but the material may not be ferromagnetic in practice. These examples highlight the need for careful data curation, awareness of method failure modes, and task-dependent validation as the number of supported properties grows.

## Acknowledgements

We thank Chris Bishop, Tao Qin, Tie-Yan Liu, Haiguang Liu, Bin Shao, Jia Zhang, and Karin Strauss for discussions that helped shape this work; Prof. Davide Donadio for comments on applications of MatterSim; and Bichlien Nguyen, Yu Xie, and Jonas Köhler for manuscript feedback. We also thank Ryota Tomioka for DFT computation pipelines, and Maik Riechert, Hannes Schulz, Stefano Battaglia, and Thijs Vogels for computational infrastructure support. We appreciate Shoko Ueda, Peggy Dai, and Roberto Sordillo for project management, and Lina Lu, Yang Ou, and Jingyun Bai for graphic design. We are grateful to the Microsoft Research AI for Science team for ongoing discussions. Z.L. thanks Pascal Salzbrener and Prof. Chris Pickard for discussions on high-pressure crystal structures.

## Contributions

**Conceptualization:** H.Y., J.L., H.H., Z.L.; **Data curation:** H.Y., J.L., G.L., Q.W., Y.Zhu, Z.C., S.C.; **Methodology:** H.Y., X.L., C.H., Y.Zhou, Y.S., C.Liu, J.T., J.L., H.H., Z.L.; **Validation:** H.Y., J.L., G.L., Q.W., Y.Zhu, Z.C., S.C., F.T., C.Z., M.H., R.P., A.F., D.Z., T.X., J.S., L.S., Y.Chen, L.K., H.H., Z.L.; **Software:** Y.B., D.G.; **Supervision:** H.Y., F.N., H.H., Z.L.; **Resources:** Z.L.; **Writing – review & editing:** All authors.

## Inclusion and ethics declarations

**Competing interests:** The authors declare no competing interests.

## Data availability

The benchmark datasets will be made publicly available following editorial request upon publication.

## Code availability

We will open-source the code and weights following editorial request upon publication.

## References

- [1] G. Fiori, F. Bonaccorso, G. Iannaccone, T. Palacios, D. Neumaier, A. Seabaugh, S.K. Banerjee, L. Colombo, Electronics based on two-dimensional materials. *Nature nanotechnology* **9**(10), 768–779 (2014)
- [2] T. Li, G. Galli, Electronic properties of mos2 nanoparticles. *The Journal of Physical Chemistry C* **111**(44), 16192–16196 (2007)
- [3] K. Mizushima, P. Jones, P. Wiseman, J.B. Goodenough,  $\text{Li}_x\text{CoO}_2$  ( $0 < x \leq 1$ ): A new cathode material for batteries of high energy density. *Materials Research Bulletin* **15**(6), 783–789 (1980)
- [4] G. Ceder, Y.M. Chiang, D. Sadoway, M. Aydinol, Y.I. Jang, B. Huang, Identification of cathode materials for lithium batteries guided by first-principles calculations. *Nature* **392**(6677), 694–696 (1998)
- [5] M.W. Tibbitt, C.B. Rodell, J.A. Burdick, K.S. Anseth, Progress in material design for biomedical applications. *Proceedings of the National Academy of Sciences* **112**(47), 14444–14451 (2015)
- [6] X. Li, J. Xie, C. Jiang, J. Yu, P. Zhang, Review on design and evaluation of environmental photocatalysts. *Frontiers of Environmental Science & Engineering* **12**, 1–32 (2018)
- [7] X. Hu, G. Li, J.C. Yu, Design, fabrication, and modification of nanostructured semiconductor materials for environmental and energy applications. *Langmuir* **26**(5), 3031–3039 (2010)
- [8] M. Hellenbrandt, The inorganic crystal structure database (icsd)—present and future. *Crystallography Reviews* **10**(1), 17–22 (2004)
- [9] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, et al., Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials* **1**(1) (2013)
- [10] S. Curtarolo, G.L. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design. *Nature materials* **12**(3), 191–201 (2013)
- [11] S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton, The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *npj Computational Materials* **1**(1), 1–15 (2015)

- [12] C. Draxl, M. Scheffler, The nomad laboratory: from data sharing to artificial intelligence. *Journal of Physics: Materials* **2**(3), 036001 (2019)
- [13] J. Schmidt, N. Hoffmann, H.C. Wang, P. Borlido, P.J. Carriço, T.F. Cerqueira, S. Botti, M.A. Marques, Large-scale machine-learning-assisted exploration of the whole materials space. *arXiv preprint arXiv:2210.00579* (2022)
- [14] J. Schmidt, H.C. Wang, T.F. Cerqueira, S. Botti, M.A. Marques, A dataset of 175k stable and metastable materials calculated with the pbesol and scan functionals. *Scientific Data* **9**(1), 64 (2022)
- [15] A. Merchant, S. Batzner, S.S. Schoenholz, M. Aykol, G. Cheon, E.D. Cubuk, Scaling deep learning for materials discovery. *Nature* pp. 1–6 (2023)
- [16] L. Barroso-Luque, M. Shuaibi, X. Fu, B.M. Wood, M. Dzamba, M. Gao, A. Rizvi, C.L. Zitnick, Z.W. Ulissi, Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771* (2024)
- [17] M.K. Horton, P. Huck, R.X. Yang, J.M. Munro, S. Dwaraknath, A.M. Ganose, R.S. Kingsbury, M. Wen, J.X. Shen, T.S. Mathis, et al., Accelerated data-driven materials science with the materials project. *Nature Materials* **24**, 1522–1532 (2025). <https://doi.org/10.1038/s41563-025-02272-0>
- [18] T. Cavignac, J. Schmidt, P.P. De Breuck, A. Loew, T.F.T. Cerqueira, H.C. Wang, A. Bochkarev, Y. Lysogorskiy, A.H. Romero, R. Drautz, S. Botti, M.A.L. Marques, Ai-driven expansion and application of the alexandria database. *arXiv preprint arXiv:2512.09169* (2025)
- [19] C. Zeni, R. Pinsler, D. Zügner, A. Fowler, M. Horton, X. Fu, S. Shysheya, J. Crabbé, L. Sun, J. Smith, et al., A generative model for inorganic materials design. *Nature* (2025). <https://doi.org/10.1038/s41586-025-08628-5>
- [20] S. Yang, K. Cho, A. Merchant, P. Abbeel, D. Schuurmans, I. Mordatch, E.D. Cubuk, Scalable diffusion for materials generation. *arXiv preprint arXiv:2311.09235* (2023)
- [21] S. Yang, S. Batzner, R. Gao, M. Aykol, A.L. Gaunt, B. McMorrow, D.J. Rezende, D. Schuurmans, I. Mordatch, E.D. Cubuk, Generative hierarchical materials search. *arXiv preprint arXiv:2409.06762* (2024)

- [22] B.K. Miller, M. Geiger, T.E. Smidt, F. Noé, FlowMM: Generating Materials with Riemannian Flow Matching, in International Conference on Machine Learning (2024)
- [23] Z. Cao, X. Luo, J. Lv, L. Wang, Space group informed transformer for crystalline materials generation. *Science Bulletin* (2025). <https://doi.org/10.1016/j.scib.2025.09.035>
- [24] A. Sriram, A.J. Bose, E. Dupont, D. Torber, L. Barroso-Luque, X. Fu, C.L. Zitnick, Z.W. Ulissi, Flowllm: Flow matching for material generation with large language models as base distributions. arXiv preprint arXiv:2410.23405 (2024)
- [25] T.B. Brown, Language models are few-shot learners. arXiv preprint arXiv:2005.14165 (2020)
- [26] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, et al., Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023)
- [27] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H.W. Chung, C. Sutton, S. Gehrmann, et al., Palm: Scaling language modeling with pathways. arxiv 2022. arXiv preprint arXiv:2204.02311 10 (2022)
- [28] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F.L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al., Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
- [29] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, et al., The llama 3 herd of models. arXiv preprint arXiv:2407.21783 (2024)
- [30] Gemini Team, R. Anil, S. Borgeaud, J.B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A.M. Dai, A. Hauth, et al., Gemini: A family of highly capable multimodal models. arXiv preprint arXiv:2312.11805 (2024)
- [31] D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, et al., Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948 (2025)
- [32] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, M. Chen, Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125 1(2), 3 (2022)
- [33] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-Resolution Image Synthesis with Latent Diffusion Models, in IEEE/CVF Conference on Computer Vision and Pattern

Recognition (2022), pp. 10684–10695

- [34] C. Chen, S.P. Ong, A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science* **2**(11), 718–728 (2022)
- [35] I. Batatia, P. Benner, Y. Chiang, A.M. Elena, D.P. Kovács, J. Riebesell, X.R. Advincula, M. Asta, W.J. Baldwin, N. Bernstein, et al., A foundation model for atomistic materials chemistry. arXiv preprint arXiv:2401.00096 (2023)
- [36] M. Neumann, J. Gin, B. Rhodes, S. Bennett, Z. Li, H. Choubisa, A. Hussey, J. Godwin, Orb: A fast, scalable neural network potential. arXiv preprint arXiv:2410.22570 (2024)
- [37] B. Rhodes, S. Vandenhoute, V. Šimkus, J. Gin, J. Godwin, T. Duignan, M. Neumann, Orb-v3: atomistic simulation at scale. arXiv preprint arXiv:2504.06231 (2025)
- [38] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C.J. Bartel, G. Ceder, Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **5**(9), 1031–1041 (2023)
- [39] D. Zhang, X. Liu, X. Zhang, C. Zhang, C. Cai, H. Bi, Y. Du, X. Qin, J. Huang, B. Li, et al., Dpa-2: Towards a universal large atomic model for molecular and material simulation. arXiv preprint arXiv:2312.15492 (2023)
- [40] F. Xie, T. Lu, S. Meng, M. Liu, Gptff: A high-accuracy out-of-the-box universal ai force field for arbitrary inorganic materials. arXiv preprint arXiv:2402.19327 (2024)
- [41] N. Shoghi, A. Kolluru, J.R. Kitchin, Z.W. Ulissi, C.L. Zitnick, B.M. Wood, From molecules to materials: Pre-training large generalizable models for atomic property prediction. arXiv preprint arXiv:2310.16802 (2023)
- [42] K. Choudhary, B. DeCost, Atomistic line graph neural network for improved materials property predictions. *npj Computational Materials* **7**(1), 185 (2021)
- [43] Y. Park, J. Kim, S. Hwang, S. Han, Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations. *Journal of Chemical Theory and Computation* (2024)
- [44] Y.L. Liao, B.M. Wood, A. Das, T.E. Smidt, EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations, in International Conference on Learning Representations (2024)

- [45] A. Mazitov, F. Bigi, M. Kellner, P. Pegolo, D. Tisi, G. Fraux, S. Pozdnyakov, P. Loche, M. Ceriotti, Pet-mad, a lightweight universal interatomic potential for advanced materials modeling. *Nature Communications* **16**, 10653 (2025). <https://doi.org/10.1038/s41467-025-65662-7>
- [46] H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen, C. Zeni, et al., Mattersim: A deep learning atomistic model across elements, temperatures and pressures. arXiv preprint arXiv:2405.04967 (2024)
- [47] O.T. Unke, S. Chmiela, H.E. Sauceda, M. Gastegger, I. Poltavsky, K.T. Schütt, A. Tkatchenko, K.R. Müller, Machine learning force fields. *Chemical Reviews* **121**(16), 10142–10186 (2021)
- [48] R. Resta, Macroscopic polarization in crystalline dielectrics: the geometric phase approach. *Reviews of modern physics* **66**(3), 899 (1994)
- [49] S. Baroni, S. De Gironcoli, A. Dal Corso, P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory. *Reviews of modern Physics* **73**(2), 515 (2001)
- [50] S. Falletta, A. Cepellotti, A. Johansson, C.W. Tan, M.L. Descoteaux, A. Musaelian, C.J. Owen, B. Kozinsky, Unified differentiable learning of electric response. *Nature communications* **16**(1), 4031 (2025)
- [51] Y. Zhou, S. Hu, X. Zhang, H. Wang, G. Tan, W. Jia, MatRIS: Toward Reliable and Efficient Pretrained Machine Learning Interatomic Potentials, in The Fourteenth International Conference on Learning Representations (2026)
- [52] P. Zhong, D. Kim, D.S. King, B. Cheng, Machine learning interatomic potential can infer electrical response. *npj Computational Materials* (2025)
- [53] B.A. Martin, A.M. Ganose, V. Kapil, T. Li, K.T. Butler, General learning of the electric response of inorganic materials. arXiv preprint arXiv:2508.17870 (2025)
- [54] P. Hohenberg, W. Kohn, Inhomogeneous electron gas. *Physical review* **136**(3B), B864 (1964)
- [55] W. Kohn, L.J. Sham, Self-consistent equations including exchange and correlation effects. *Physical review* **140**(4A), A1133 (1965)
- [56] A.M. Ganose, H. Sahasrabudde, 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**(7), 1944–1973 (2025)

- [57] T. Chen, S. Luo, D. He, S. Zheng, T.Y. Liu, L. Wang, Geomformer: A general architecture for geometric molecular representation learning (2023)
- [58] L.B. Skinner, C. Benmore, J.C. Neufeind, J.B. Parise, The structure of water around the compressibility minimum. *The Journal of chemical physics* **141**(21) (2014)
- [59] W. Chen, F. Ambrosio, G. Miceli, A. Pasquarello, Ab initio electronic structure of liquid water. *Physical review letters* **117**(18), 186401 (2016)
- [60] A. Soper, C. Benmore, Quantum differences between heavy and light water. *Physical review letters* **101**(6), 065502 (2008)
- [61] F. Giannessi, S. Di Cataldo, S. Saha, L. Boeri, A database of high-pressure crystal structures from hydrogen to lanthanum. *Scientific Data* **11**(1), 766 (2024)
- [62] A. Drozdov, P. Kong, V. Minkov, S. Besedin, M. Kuzovnikov, S. Mozaffari, L. Balicas, F. Balakirev, D. Graf, V. Prakapenka, et al., Superconductivity at 250 k in lanthanum hydride under high pressures. *Nature* **569**(7757), 528–531 (2019)
- [63] N. Dubrovinskaia, S. Petitgirard, S. Chariton, R. Tucoulou, J. Garrevoet, K. Glazyrin, H.P. Liermann, V.B. Prakapenka, L. Dubrovinsky, B1-b2 phase transition in mgo at ultra-high static pressure. *arXiv preprint arXiv:1904.00476* (2019)
- [64] S. Zhang, R. Paul, S. Hu, M.A. Morales, Toward an accurate equation of state and b1-b2 phase boundary for magnesium oxide up to terapascal pressures and electron-volt temperatures. *Physical Review B* **107**(22), 224109 (2023)
- [65] X. Tang, J. Dong, Lattice thermal conductivity of mgo at conditions of earth’s interior. *Proceedings of the National Academy of Sciences* **107**(10), 4539–4543 (2010)
- [66] T. Lay, Q. Williams, E.J. Garnero, The core–mantle boundary layer and deep earth dynamics. *Nature* **392**(6675), 461–468 (1998)
- [67] R.S. McWilliams, D.K. Spaulding, J.H. Eggert, P.M. Celliers, D.G. Hicks, R.F. Smith, G.W. Collins, R. Jeanloz, Phase transformations and metallization of magnesium oxide at high pressure and temperature. *Science* **338**(6112), 1330–1333 (2012)
- [68] A.B. Belonoshko, S. Arapan, R. Martonak, A. Rosengren, Mgo phase diagram from first principles in a wide pressure-temperature range. *Physical Review B* **81**(5), 054110 (2010)

- [69] J. Bouchet, F. Bottin, V. Recoules, F. Remus, G. Morard, R. Bolis, A. Benuzzi-Mounaix, Ab initio calculations of the b1-b2 phase transition in mgo. *Physical Review B* **99**(9), 094113 (2019)
- [70] X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli, T. Jaakkola, Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *arXiv preprint arXiv:2210.07237* (2022)
- [71] Y. Xie, J. Vandermause, S. Ramakers, N.H. Protik, A. Johansson, B. Kozinsky, Uncertainty-aware molecular dynamics from bayesian active learning for phase transformations and thermal transport in sic. *npj Computational Materials* **9**(1), 36 (2023)
- [72] L. Kong, J. Li, L. Sun, H. Yang, H. Hao, C. Chen, N. Artrith, J.A.G. Torres, Z. Lu, Y. Zhou, Overcoming the size limit of first principles molecular dynamics simulations with an in-distribution substructure embedding active learner. *arXiv preprint arXiv:2311.08177* (2023)
- [73] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple. *Physical review letters* **77**(18), 3865 (1996)
- [74] B. Cheng, E.A. Engel, J. Behler, C. Dellago, M. Ceriotti, Ab initio thermodynamics of liquid and solid water. *Proceedings of the National Academy of Sciences* **116**(4), 1110–1115 (2019)
- [75] K. Krynicki, C.D. Green, D.W. Sawyer, Pressure and temperature dependence of self-diffusion in water. *Faraday Discussions of the Chemical Society* **66**, 199–208 (1978)
- [76] E.H. Hardy, A. Zygar, M.D. Zeidler, M. Holz, F.D. Sacher, Isotope effect on the translational and rotational motion in liquid water and ammonia. *The Journal of Chemical Physics* **114**(7), 3174–3181 (2001)
- [77] W. Cochran, R. Cowley, Dielectric constants and lattice vibrations. *Journal of Physics and Chemistry of Solids* **23**(5), 447–450 (1962)
- [78] P. Giannozzi, S. De Gironcoli, P. Pavone, S. Baroni, Ab initio calculation of phonon dispersions in semiconductors. *Physical Review B* **43**(9), 7231 (1991)
- [79] P.G. Neudeck, Silicon carbide technology. Tech. rep., CRC Press (2006)
- [80] K. Karch, F. Bechstedt, P. Pavone, D. Strauch, Pressure-dependent dynamical and dielectric properties of cubic sic. *Journal of Physics: Condensed Matter* **8**(17), 2945 (1996)

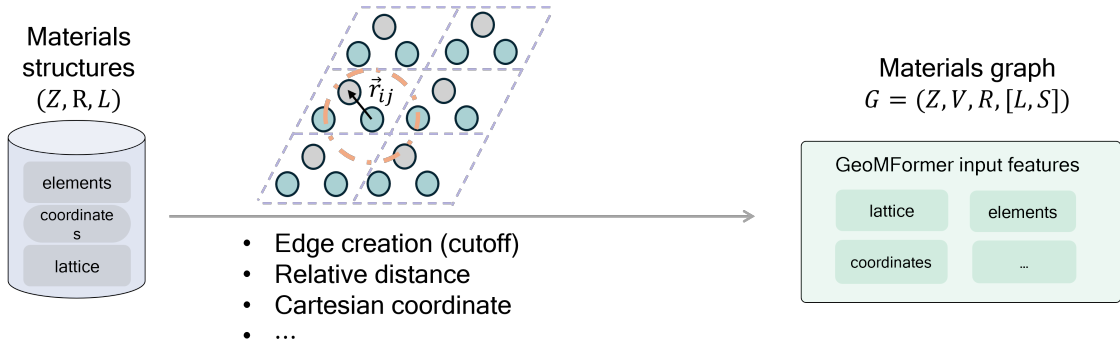
- [81] D. Olego, M. Cardona, P. Vogl, Pressure dependence of the optical phonons and transverse effective charge in 3 c-sic. *Physical Review B* **25**(6), 3878 (1982)
- [82] P. Yu, Fundamentals of semiconductors (Springer, 2005)
- [83] D. Olego, M. Cardona, Pressure dependence of raman phonons of ge and 3 c-sic. *Physical Review B* **25**(2), 1151 (1982)
- [84] K. McColl, S.W. Coles, P. Zarabadi-Poor, B.J. Morgan, M.S. Islam, Phase segregation and nanoconfined fluid o2 in a lithium-rich oxide cathode. *Nature Materials* pp. 1–8 (2024)
- [85] T. Choudhary, V. Mishra, A. Goswami, J. Sarangapani, A comprehensive survey on model compression and acceleration. *Artificial Intelligence Review* **53**, 5113–5155 (2020)
- [86] H. Wang, S. Ma, L. Dong, S. Huang, H. Wang, L. Ma, F. Yang, R. Wang, Y. Wu, F. Wei, Bitnet: Scaling 1-bit transformers for large language models. *arXiv preprint arXiv:2310.11453* (2023)
- [87] J. Ding, S. Ma, L. Dong, X. Zhang, S. Huang, W. Wang, N. Zheng, F. Wei, Longnet: Scaling transformers to 1,000,000,000 tokens. *arXiv preprint arXiv:2307.02486* (2023)
- [88] R. King-Smith, D. Vanderbilt, Theory of polarization of crystalline solids. *Physical Review B* **47**(3), 1651 (1993)
- [89] I. Souza, J. Íniguez, D. Vanderbilt, First-principles approach to insulators in finite electric fields. *Physical review letters* **89**(11), 117602 (2002)

# Supplementary Information

## Contents

<b>S1 Model architecture and training details</b>	<b>3</b>
S1.1 Materials Graphs . . . . .	3
S1.2 Model Architecture . . . . .	3
S1.3 Training details . . . . .	7
<b>S2 Materials explorer</b>	<b>8</b>
S2.1 Data exploration . . . . .	8
S2.2 Uncertainty evaluation . . . . .	9
S2.3 First-principles computation details . . . . .	10
<b>S3 Training Data</b>	<b>10</b>
S3.1 Off-equilibrium materials data . . . . .	10
S3.2 Bader charges . . . . .	11
S3.3 Born effective charges and dielectric matrices . . . . .	13
S3.4 Magnetic moments . . . . .	13
<b>S4 Benchmarks</b>	<b>13</b>
S4.1 Benchmark datasets and results . . . . .	13
S4.2 Phonons and group velocities . . . . .	15
S4.2.1 Computational setup . . . . .	16
S4.3 Free energy prediction with QHA . . . . .	16
S4.3.1 Computational setup . . . . .	17
S4.4 Results of Multiple Tasks . . . . .	18
<b>S5 Multi-task Case studies</b>	<b>19</b>
S5.1 Formation enthalpy of LaH <sub>10</sub> . . . . .	19
S5.2 Pressure dependent LO–TO splitting of 3c-SiC . . . . .	19
S5.3 Phase boundary of B1–B2 MgO . . . . .	21

S5.4 Hysteresis curve of BaTiO <sub>3</sub> . . . . .	21
S5.5 Delithiation of Li <sub>1.2-x</sub> Mn <sub>0.8</sub> O <sub>2</sub> . . . . .	21
S5.6 Comparison to other foundational interatomic potentials . . . . .	23
<b>S6 MatterSim as a continual active learner</b>	<b>25</b>
<b>S7 Finetuning and molecular dynamics on liquid water</b>	<b>25</b>
S7.1 Parameter settings for finetuning . . . . .	26
S7.2 Simulation settings for molecular dynamics . . . . .	26
S7.3 ADF of Liquid Water . . . . .	27
S7.4 Deriving Diffusion Coefficients from Mean Squared Displacements . . . . .	27
S7.5 Data efficiency of finetuning . . . . .	27



**Fig. S1:** MatterSim leverages materials graphs built upon point clouds to represent atomic interactions and geometric features in Euclidean space.

## S1 Model architecture and training details

### S1.1 Materials Graphs

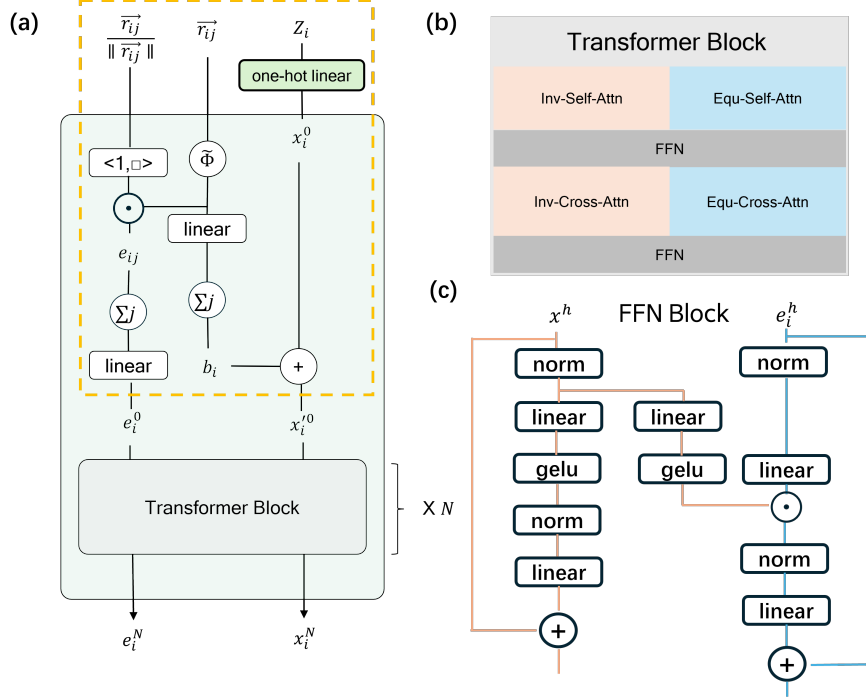
MatterSim operates on materials graphs constructed from three-dimensional point clouds under periodic boundary conditions. We define a materials graph  $\mathcal{G} = (\mathbf{Z}, \mathbf{V}, \mathbf{R}, [\mathbf{L}, \mathbf{S}])$  (see Fig. S1), where  $\mathbf{Z}$  denotes atomic numbers and additional per-atom features,  $\mathbf{R}$  contains atomic positions  $\mathbf{r}_i \in \mathbb{R}^3$ ,  $\mathbf{V}$  encodes relative displacement vectors between atom pairs,  $\mathbf{L}$  is the  $3 \times 3$  lattice matrix, and  $\mathbf{S}$  represents global scalar state variables such as temperature and pressure. Nodes correspond to atoms and edges are formed between all pairs within a radial cutoff distance  $r_c$ . Fractional coordinates are converted to Cartesian before graph construction. By design, MatterSim preserves roto-translational invariance for scalar properties (e.g., total energy) and equivariance for vectorial properties (e.g., forces).

### S1.2 Model Architecture

MatterSim-MT adopts a transformer-based architecture inspired by GeoMFormer.[1] A separate M3GNet-based MatterSim-v1[2] model is also utilized to drive the materials explorer during structure sampling.

The model is built on the Transformer architecture[3] and adapted to leverage SE(3)-equivariant vectors for both molecular and materials systems. It employs two parallel streams that maintain and learn invariant and equivariant representations, respectively. Following GeoMFormer,[1] a cross-attention module bridges the two streams, enabling information exchange and improving geometric modeling.

To handle the periodic boundary conditions (PBC) of crystal structures, we incorporate the multi-graph construction from Ref. 4, representing atoms in the unit cell via periodic graphs that include image atoms from neighboring cells up to a predefined cutoff distance. A smooth cutoff function



**Fig. S2:** (a) The overall architecture of the model, which extracts atomic embeddings using the embedding block and Transformer block. The yellow line highlights the embedding block. (b) The Transformer block comprises Self-Inv-Attn, Self-Equ-Attn, Cross-Inv-Attn, Cross-Equ-Attn, and the FFN module. (c) Implementation of the FFN module.

based on interatomic distances gradually attenuates long-range interactions, ensuring translational invariance.

**Embedding Block.** The model architecture is shown in Fig. S2(a), with the Embedding Block highlighted in yellow. It consists of three components: atom embedding, spatial relationship embedding, and equivariant feature embedding. Atom embedding maps atomic species  $Z_i$  and positions  $\mathbf{r}_i$  into feature vectors  $\mathbf{x}_i$ , initialized via a one-hot linear layer.

The spatial relationship embedding captures interatomic relationships through centrality encoding:

$$\mathbf{b}_i = \sum_{j \in \mathcal{N}(i)} \text{Linear} \left( m_{ij} \cdot \tilde{\Phi}(\|\mathbf{r}_{ij}\|) \right) \quad (1)$$

where  $\mathcal{N}(i)$  denotes the neighbors of atom  $i$  within cutoff radius  $r_c$ ,  $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$  is the relative displacement, and  $\tilde{\Phi}(\cdot)$  is a Gaussian basis kernel. The smooth cutoff mask  $m_{ij}$  is defined as:

$$m_{ij} = 1 - 6 \left( \frac{\|\mathbf{r}_{ij}\|}{r_c} \right)^5 + 15 \left( \frac{\|\mathbf{r}_{ij}\|}{r_c} \right)^4 - 10 \left( \frac{\|\mathbf{r}_{ij}\|}{r_c} \right)^3 \quad (2)$$

The centrality encoding updates the initial embedding as  $\mathbf{x}_i^{r0} = \mathbf{x}_i^0 + \mathbf{b}_i$ , which is then passed to the attention modules.

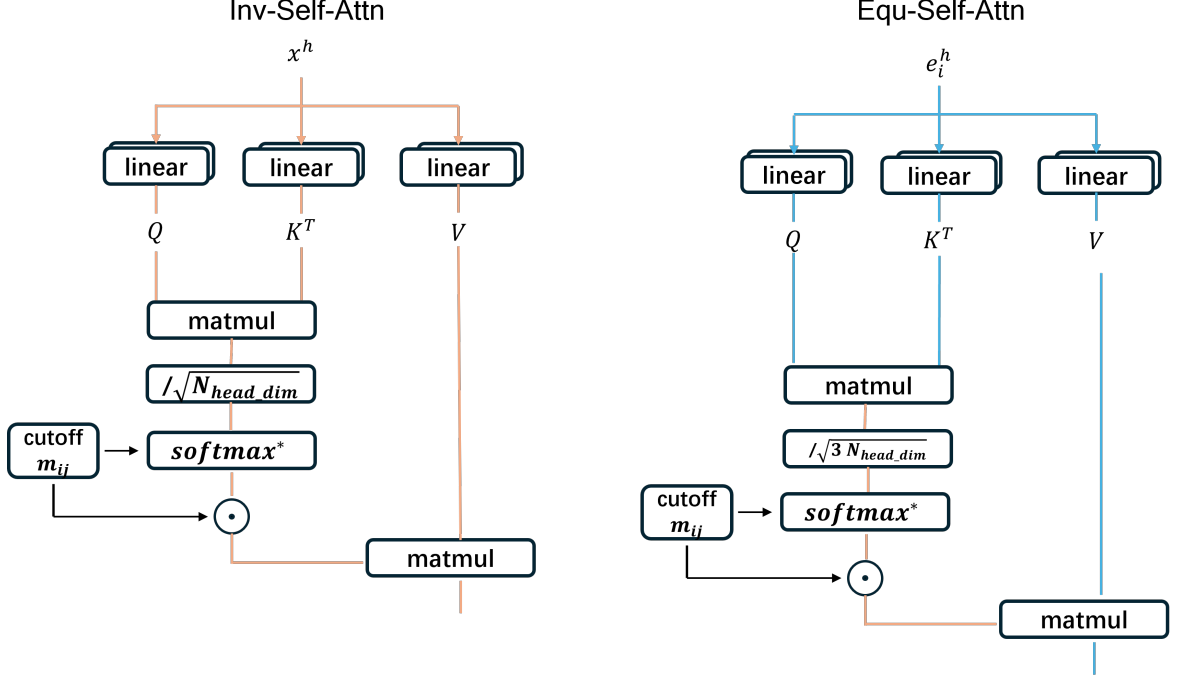


Fig. S3: Implementation of the Inv-Attention module.

The equivariant feature embedding captures relative positional information for the equivariant stream. It is initialized as:

$$\mathbf{e}_i^0 = \sum_j m_{ij} \tilde{\Phi}(\|\mathbf{r}_{ij}\|) \cdot \text{concat} \left( 1, \frac{\mathbf{r}_{ij}}{\|\mathbf{r}_{ij}\|} \right) \quad (3)$$

where  $\text{concat} \left( 1, \frac{\mathbf{r}_{ij}}{\|\mathbf{r}_{ij}\|} \right)$  concatenates a scalar 1 with the unit displacement vector.

**Transformer Block.** Each Transformer block contains four attention modules—Inv-Self-Attn, Equ-Self-Attn, Inv-Cross-Attn, and Equ-Cross-Attn—as shown in Fig. S2(b). Self-attention modules derive key–value pairs from their own stream, while cross-attention modules exchange information between the two streams.

The FFN module is shown in Fig. S2(c). Invariant features  $x_i$  are updated via a standard feed-forward network with residual connections. Because non-linear activations would break equivariance, equivariant features  $e_i$  are instead gated by invariant features, following the GeoMFormer design.

For the self-attention modules at the  $h$ -th layer (see Fig. S3), we illustrate with Inv-Self-Attn. The query  $\mathbf{Q}$ , key  $\mathbf{K}$ , and value  $\mathbf{V}$  matrices are obtained via linear projections of  $\mathbf{x}_i^h$ :

$$\mathbf{a}_{ij} = \left( \frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}} \right)_{i,j} \quad (4)$$

$$\text{Softmax}^*(\mathbf{a}_{ij}) = \frac{e_{ij}^{a_{ij}} \cdot m_{ij}}{\sum_{k \in \mathcal{N}(i)} e_{ik}^{a_{ik}} \cdot m_{ik}} \quad (5)$$

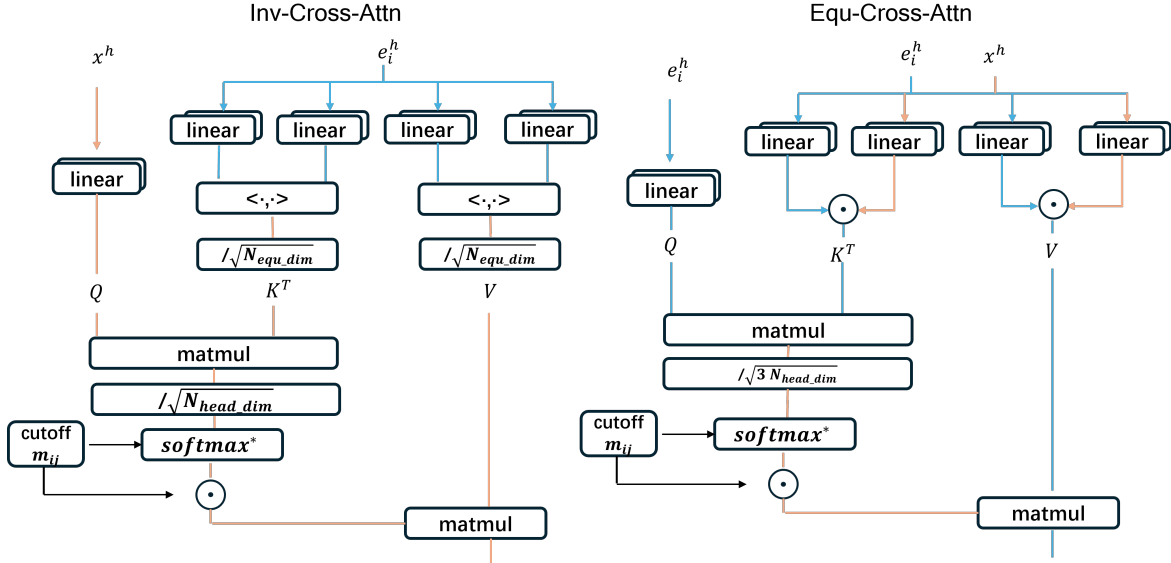


Fig. S4: Implementation of the Cross-Attention module.

$$\mathbf{x}_i^{h+1} = \sum_{j \in \mathcal{N}(i)} \text{Softmax}^*(\mathbf{a}_{ij}) \cdot m_{ij} \cdot \mathbf{V}_j \quad (6)$$

Here,  $d$  is the hidden dimension. Note that  $\mathbf{K}$  and  $\mathbf{V}$  for periodic image atoms are copied from the original atoms to ensure consistent representations between an atom and its images.

The cross-attention module (see Fig. S4) enables interaction between the two streams. Its  $\mathbf{Q}$ ,  $\mathbf{K}$ ,  $\mathbf{V}$  computation follows the GeoMFormer design, and the attention mechanism is identical to that of Inv-Self-Attn.

**Task Head.** The Transformer block produces both invariant and equivariant per-atom features, which are used by task-specific heads to predict target properties.

Each task head receives the atomic-level invariant features  $x_i^N$  and equivariant features  $e_i^N$  from the final Transformer block. These are processed through an Inv-Cross-Attn layer to produce updated scalar features  $x_i^{N+1}$ . For scalar predictions (energy, Bader charge, magnetic moments):

$$\mathbf{x}_i^{N+1} = \text{Inv-Cross-Attn}(x_i^N, e_i^N) \quad (7)$$

$$\mathbf{p}_i = W_2 f_{LN}(\sigma(W_1 x_i^{N+1} + b_1)) + b_2 \quad (8)$$

Here,  $f_{LN}$  is layer normalization and  $\sigma$  is the GELU activation. For energy prediction, mean pooling is applied over atoms; for Bader charge and magnetic moment predictions, the per-atom values  $p_i$  are used directly.

For Born effective charges and dielectric properties, we follow the ETGNN design (Ref. 5). The dielectric tensor is:

$$\varepsilon = \frac{1}{N} \sum_{i=1}^N \sum_{j \in \mathcal{N}(i)} (\mathbf{p}_i \mathbf{p}_j) \frac{\vec{r}_{ji}}{\|\vec{r}_{ji}\|} \otimes \frac{\vec{r}_{ji}}{\|\vec{r}_{ji}\|} \quad (9)$$

where  $\mathbf{p}_j$  is computed using separate linear layer parameters from  $\mathbf{p}_i$ .

The Born effective charge is decomposed into symmetric and non-symmetric contributions:

$$\mathbf{Z}_i = \mathbf{Z}_i^{sym} + \mathbf{Z}_i^{non-sym} \quad (10)$$

$$\mathbf{Z}_i^{sym} = \sum_{j \in \mathcal{N}(i)} (\mathbf{p}_i \mathbf{p}_j) \frac{\vec{r}_{ji}}{\|\vec{r}_{ji}\|} \otimes \frac{\vec{r}_{ji}}{\|\vec{r}_{ji}\|} \quad (11)$$

$$\mathbf{Z}_i^{non-sym} = \sum_{k, j \in \mathcal{N}(i), k \neq j} (\mathbf{p}_i \mathbf{p}_j \mathbf{p}_k) \frac{\vec{r}_{ki}}{\|\vec{r}_{ki}\|} \otimes \frac{\vec{r}_{ij}}{\|\vec{r}_{ij}\|} \quad (12)$$

### S1.3 Training details

The loss function for the GeoMFormer model is:

$$L = l(e, e_{\text{DFT}}) + \omega_f l(\mathbf{f}, \mathbf{f}_{\text{DFT}}) + \omega_\sigma l(\boldsymbol{\sigma}, \boldsymbol{\sigma}_{\text{DFT}}) + \sum_{i=1}^4 \omega_{t_i} l(\mathbf{p}_{t_i}, \mathbf{p}_{t_i, \text{DFT}})$$

where  $l(\cdot, \cdot)$  is the mean absolute error (MAE),  $e$  is the energy per atom,  $\mathbf{f}$  is the per-atom force vector,  $\boldsymbol{\sigma}$  is the stress tensor, and  $\omega_f$ ,  $\omega_\sigma$  are their respective loss weights. The terms  $p_{t_i}$  correspond to four auxiliary targets: magnetic moments, Bader charges, Born effective charges, and dielectric matrices ( $i \in \{1, \dots, 4\}$ ).

Table S1 lists the model hyperparameters for the 1 M, 10 M, 220 M, and 1.3 B parameter variants. The FFN embedding size refers to the dimensionality within the FFN block. To limit memory consumption during training, `pbs_expanded_num_cell_per_direction` and `pbs_expanded_token_cutoff` control the number of unit cell expansions and the maximum number of atoms after expansion, respectively.

Table S2 presents the pretraining hyperparameters. After warm-up, the learning rate decays linearly to zero.

Fig. S5 shows the validation loss as a function of training data size for models of varying capacity. Models with fewer than 10M parameters saturate quickly, whereas larger models continue to improve as the dataset scales up to 35M samples. We further demonstrate scaling to 1.3B parameters; however, given the trade-off between computational cost and accuracy, all predictions and simulations in the main text use the 10M-parameter model.

**Table S1:** Hyperparameters for training MatterSim with GeoMFormer architecture and parameters of 1 M, 10 M, 220 M and 1.3 B.

Hyperparameters	1 M	10 M	220 M	1.3 B
Number of Gaussian Basis	64	128	128	128
Number of Transformer Blocks	2	3	6	9
Number of Attention Blocks	8	12	24	36
Atom Embedding Size	128	320	768	1536
Hidden Embedding Size	128	320	768	1536
FFN Embedding Size	128	320	3078	6144
Number of Attention Heads	8	16	32	32
Cutoff Radius (Å)	5.0	5.0	5.0	5.0
pbs_expanded_num_cell_per_direction	5	5	5	5
pbs_expanded_token_cutoff	256	256	256	256

**Table S2:** Pretraining Hyperparameters

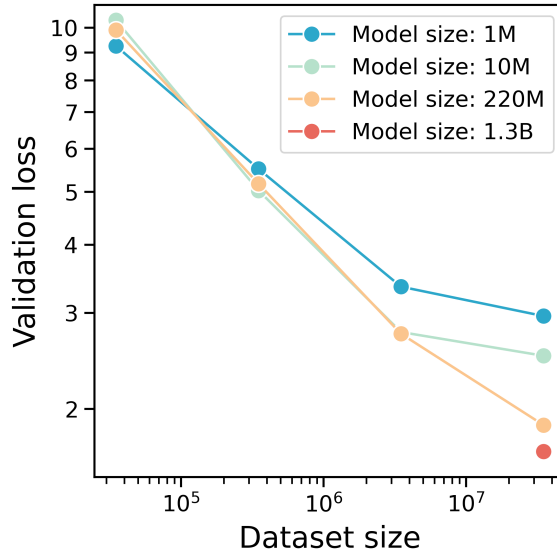
Pretraining parameters	
optimizer	AdamW[6]
max learning rate	5e-5
weight decay	0.0
$\beta_1$	0.9
$\beta_2$	0.999
$\epsilon$	1e-8
warmup epochs	1
training epochs	23
early stop epochs	3
$\omega_f$	10.0
$\omega_\sigma$	0.07
$\omega_{t_1}$	1.0
$\omega_{t_2}$	1.0
$\omega_{t_3}$	1.0
$\omega_{t_4}$	1.0
batch size	1024
EMA	0.99

## S2 Materials explorer

### S2.1 Data exploration

MatterSim’s predictive capabilities are underpinned by a two-part materials structure explorer developed in Ref. 2 that enhances the training datasets through both equilibrium and off-equilibrium structural data.

The **ground-state explorer** focuses on materials at or near atomistic equilibrium positions. It primarily utilizes an uncertainty-based method with ensemble models to selectively incorporate data from both public repositories and internally generated datasets into the database. This approach ensures the inclusion of the most informative structures for model training, enhancing the accuracy of property predictions at equilibrium. The **off-equilibrium explorer** targets materials with off-equilibrium atomistic positions. It conducts molecular dynamics (MD) simulations under a wide range



**Fig. S5:** Caption

of pressures, including 0, 200, 500, 800, and 1000 GPa. Under each pressure, we simulate each material at 300, 1000, 2000, and 5000 K. For each temperature and pressure, the NPT simulation runs for 10 ps.

These simulations are crucial for sampling a wide range of atomic configurations, allowing MatterSim to learn and predict material properties under high-pressure and high-temperature scenarios that deviate significantly from equilibrium states.

## S2.2 Uncertainty evaluation

Uncertainty quantification is essential in predictive modeling for materials properties and simulations, particularly those involving MLFFs. Accurate uncertainty assessment provides insights into the reliability of model outputs and informs decision-making processes. In materials science, where innovation potential is vast but the cost of errors is high, it is essential to estimate the uncertainty of model predictions to ensure reliable outcomes. Current uncertainty quantification methods often involve statistical techniques, such as Bayesian approaches, bootstrapping, and ensemble methods,[\[7–10\]](#) to estimate confidence intervals or prediction errors. These approaches help understand the limits of model predictions and highlight areas needing further refinement or training.

In the case of MatterSim, uncertainty quantification is addressed through an ensemble approach. By training a set of five distinct models with different random initializations, the ensemble of models gives an estimation of the uncertainty on both the energies and forces. The forces offer insight into the dynamical behavior of atoms and can be particularly revealing in scenarios where the predictions for only a small fraction of the atoms within the simulation cell are deemed unreliable. Conversely, energy predictions are often more informative in cases of crystals with a small number of atoms in the cell. By integrating both energies and forces into the uncertainty analysis, MatterSim ensures a

robust and reliable assessment of uncertainties, enhancing the confidence in its predictive capabilities for material properties and simulations.

### S2.3 First-principles computation details

The DFT parameters employed in this work are generated with the `MPreLaxSet` class defined in the `pymatgen` library [11] and the calculations are conducted with Vienna Ab-initio Simulation Package (VASP) [12, 13] using the projector augmented wave (PAW) method [14] and Perdew-Burke-Ernzerhof (PBE) [15] exchange-correlation functional. To account for on-site electronic repulsion, Hubbard U parameters were applied to the following elements in oxides and fluorides:

- Co: 3.32 eV
- Cr: 3.7 eV
- Fe: 5.3 eV
- Mn: 3.9 eV
- Mo: 4.38 eV
- Ni: 6.2 eV
- V: 3.25 eV
- W: 6.2 eV

The cutoff of the plane-wave basis set is 520 eV and the convergence threshold for total energy is  $5 \times 10^{-5}$  eV/atom. For each material, the total energy, forces on each atom and the stress are computed, stored and used for training. We encountered convergence difficulties with elements such as Gd and Eu, particularly in off-equilibrium structures where self-consistent cycles failed to converge, or energies varied significantly for two structures with nearly identical atomic positions. Such calculations are consequently excluded from the study.

## S3 Training Data

### S3.1 Off-equilibrium materials data

In this section, we analyze the distribution of the dataset across the temperature and pressure space, and compare it with public datasets, including MPF2021, [16] MPtrj, [17], Alexandria [18] and OMat24. [19]

The dataset used to train MatterSim contains 35M structures labeled with energies, forces, and stresses, 172,488 with Bader charges, 3,051 with Born effective charges and dielectric matrices, and 284,195 with magnetic moments under ferromagnetic ordering. To straightforwardly illustrate the

distribution, we have included a two-dimensional histogram of the effective temperature and stress of the main energetics-related dataset in Fig. S6, where the effective temperature is defined as follows,

1. For each material from a given dataset, we evaluate its total energy per atom ( $\varepsilon$ ) with MatterSim;
2. Then, we optimize the atomic positions with fixed lattice parameters for at most 500 steps until the max forces converge to  $0.01 \text{ eV}/\text{\AA}$ , and evaluate the relaxed total energy per atom ( $\varepsilon_0$ );
3. Finally, the effective temperature of this given material is evaluated by

$$T_{\text{eff}} = \frac{\varepsilon - \varepsilon_0}{k_{\text{B}}},$$

where  $k_{\text{B}}$  is the Boltzmann’s constant.

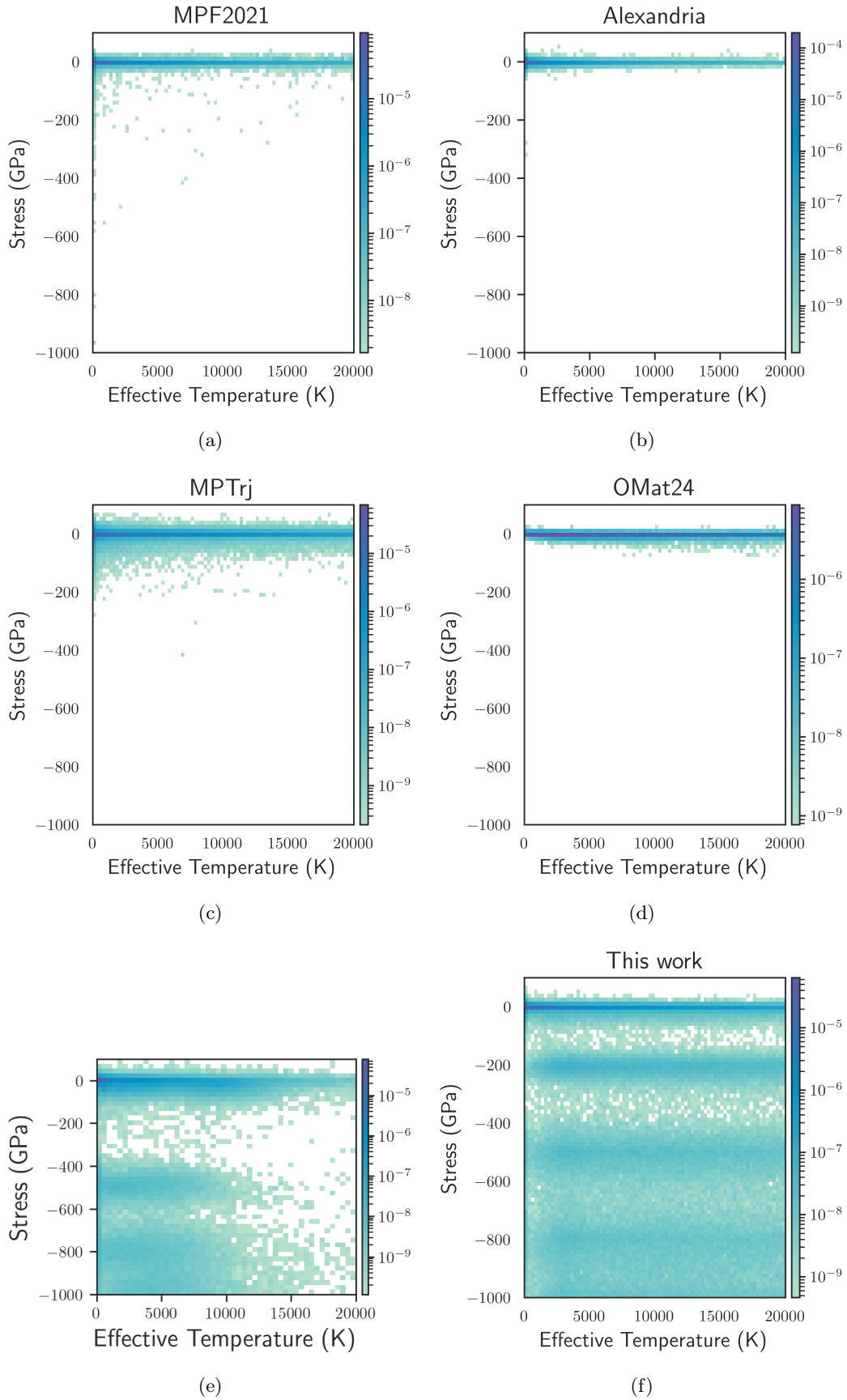
With the effective temperature, we compare the distribution of the MPF2021 dataset and 1M randomly sampled structures from Alexandria, MPTrj and OMat24, as shown in Fig. S6. Since the structures are relaxed to their corresponding local minima, it is not surprising to find that they are densely packed around the 0 GPa in stress, with very scattered data points of high effective temperature and high pressure. The dataset generated in this work, however, has a much wider coverage over the effective temperature space ( $0 - 2 \times 10^4 \text{ K}$ ) and stress space ( $0 - 1000 \text{ GPa}$  in magnitude).

We note that the effective temperature should not be directly interpreted as the physical temperature or the temperature employed in the simulations; instead, it is an intuitive metric to measure the energy distribution of the dataset.

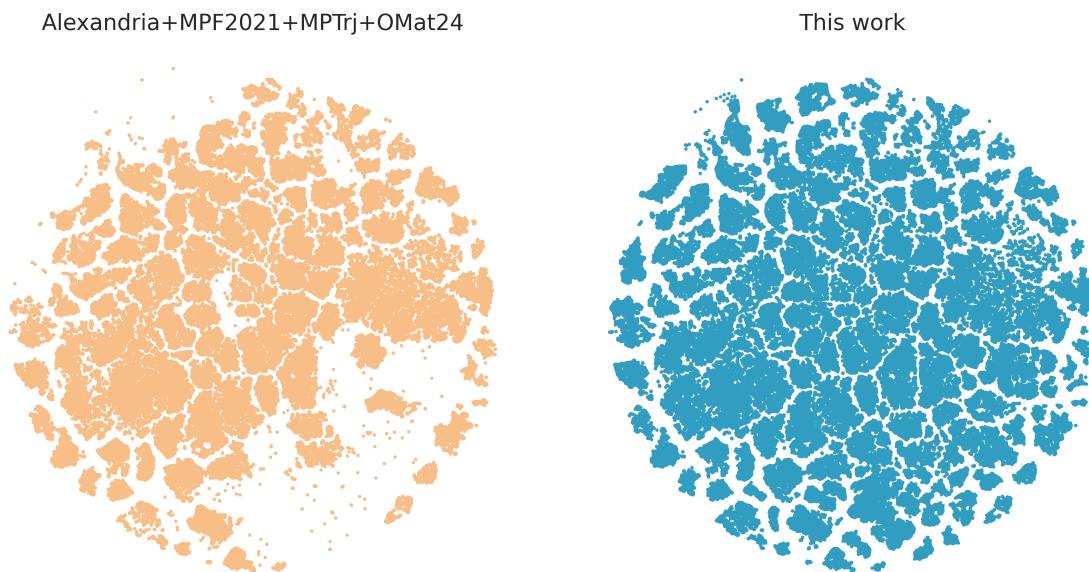
In addition to comparing the overall temperature and pressure distributions, we analyzed the atomic embeddings to further characterize the configurational space covered in this work. As shown in Fig. S7, we randomly selected 4,000 atomic embeddings for each element from our dataset, and 1,000 atomic embeddings from each of the MPF2021, MPTrj, Alexandria, and OMat24 datasets, which were then combined for comparison. In Fig. 1(b), we performed the same analysis but restricted it to the first 20 elements to provide a clearer view of configurational coverage. Both analyses demonstrate that the dataset presented in this work not only spans the configurational space covered by the four existing datasets, but also substantially extends beyond their combined coverage.

### S3.2 Bader charges

The Bader charge data are retrieved from the AFLOW[20] database with the AFLUX API. In total, the data contains the atomic net charges of 172,488 periodic structures. The atomic net charges are the number of valence electrons associated with each atom after Bader decomposition and subtraction of the number of valence electrons in the pseudopotential. The sum of all net atomic charges is zero for every structure.



**Fig. S6:** Pressure and effective temperature distribution of (a) the entire MPF2021 dataset; and 1M materials randomly sampled from (b) Alexandria, (c) MPTrj, (d) OMat24 dataset, (e) MatterSim-v1[2] and (f) this work.



**Fig. S7:** t-SNE analysis of atomic embeddings from the datasets in this work and the comparison with the combination of the Alexandria, MPF2021, MPTrj, and OMat24 datasets.

### S3.3 Born effective charges and dielectric matrices

Born effective charges and dielectric matrices were sourced from PhononDB.[21] After systematic screening to eliminate structures exhibiting dynamical instabilities (characterized by imaginary frequencies), the final dataset comprises 3,051 viable periodic structures.

### S3.4 Magnetic moments

The magnetic moments are extracted from the MPTrj dataset and are computed under ferromagnetic ordering.[17]

## S4 Benchmarks

The summary of the out-of-box performance of MatterSim-MT-10M on selected tasks is shown in Fig. 1(d). In this section, we describe in more detail the benchmark tasks, datasets used for the benchmarks, and the computational settings employed in the prediction or simulation.

### S4.1 Benchmark datasets and results

The out-of-box performance of MatterSim is benchmarked on a few datasets computed using the same level of DFT as the training data of MatterSim, including

- In-house generated benchmark datasets inherited from MatterSim-v1:
  - MPF-Alkali-TP

- MPF-TP
- Random-TP
- Extended-TP
- Public dataset
  - High-pressure Elemental Xstals (HEX) [22]

The MPF-Alkali-TP, MPF-TP, Random-TP benchmark sets are created with increasing complexity to evaluate the models’ performance on materials under finite temperature and pressure conditions with far-from-equilibrium atomic positions. All of these benchmark sets are created from first-principles molecular dynamics trajectories initialized with corresponding structures. The MPF-Alkali-TP dataset is sampled from AIMD trajectories of materials that contain alkali metals in Materials Project, and this dataset serves to assess the performance of the model for predicting ionic conductors. The selection rule of elements is that the compound should contain at least one alkali metal and at least one element from N, O, P, S, Se, F, Cl, Br, I. In total, 50 compounds are selected randomly from Materials Project following the selection rule. Similarly, MPF-TP contains molecular dynamics trajectories on 50 randomly selected compounds from the MPF2021[16] dataset without elemental constraints. For Random-TP, the initial structures are created by randomly placing 20 atoms with random elements in a simulation box. Again, 50 random structures are used for subsequent molecular dynamics simulation. During collection of the molecular dynamics simulation trajectories, each starting compound is first relaxed followed by running an NPT simulation using VASP, in which the cutoff energy of the plane wave is set to 520 eV and only the gamma point is sampled in reciprocal space to ensure the speed of sampling. The simulation is carried out for 100 ps for each material, and during the last 20 ps, 5 frames are uniformly collected for VASP calculation under MPRelaxSet settings, which are used in the final benchmark set. The temperatures and pressures are all randomly sampled. For the temperature, it is uniformly sampled between 0 and 5000 K. For the pressure, we use a log scale when carrying out the sampling. The pressure range is from 0 to 1000 GPa. These three datasets are designed to test the accuracy of emulators for finite-temperature and pressure simulations with increasing difficulty in generalizability from simple ionic compounds to complex random hypothetical structures. Since the temperature and pressure ranges are wide, these benchmark sets are also reflective of model performance on crystalline materials, amorphous materials, liquids, and pressurized materials. The Extended-TP dataset is generated in the same way as MPF-TP but with the base structures derived from beyond Materials Project. The used base structures include Materials

Test Set	MAE	OMat24	ORB-V2	SevenNet	M3GNet	CHGNet	MACE-MP-0	MatterSim-MT-10M
<b>In-house generated benchmark datasets</b>								
MPF-Alkali-TP	Energy	<b>0.040</b>	0.142	0.258	0.165	0.250	1.351	0.045
	Force	<b>0.157</b>	0.458	4.495	1.139	1.636	15.819	0.488
	Stress	<b>1.086</b>	6.625	25.039	4.911	12.625	25.723	1.252
MPF-TP	Energy	0.126	0.267	0.321	0.207	0.254	256.340	<b>0.042</b>
	Force	<b>0.233</b>	0.613	4.686	1.224	3.313	1506.854	0.530
	Stress	2.408	10.699	22.174	5.575	25.208	202.093	<b>1.368</b>
Random-TP	Energy	0.491	0.822	0.716	0.537	0.506	9.184	<b>0.259</b>
	Force	<b>0.578</b>	1.058	8.170	1.789	3.950	88.327	1.200
	Stress	<b>2.579</b>	7.254	13.696	3.216	7.230	19.224	2.715
Extended-TP	Energy	2.883	4.135	$4.5 \times 10^{10}$	0.781	4.526	$2.6 \times 10^5$	<b>0.054</b>
	Force	2.015	3.314	$3.0 \times 10^{12}$	3.543	5.972	$1.2 \times 10^6$	<b>0.780</b>
	Stress	65.847	136.058	$2.9 \times 10^{13}$	$4.3 \times 10^4$	194.145	$3.4 \times 10^7$	<b>3.274</b>
<b>Public dataset</b>								
HEX	Energy	0.528	0.532	1.463	0.436	0.972	454.462	<b>0.106</b>
	Force	<b>0.043</b>	0.074	14.638	0.699	0.196	575.443	0.118
	Stress	20.902	37.826	325.500	89.833	166.001	2597.080	<b>8.536</b>

**Table S3:** Performance of OMat24, Orb-V2, SevenNet, M3GNet, CHGNet, MACE-MP-0, and MatterSim on benchmark datasets. The mean absolute errors (MAE) of energy, force, and stress are in eV/atom, eV/Å, and GPa, respectively.

Project, Alexandria, the relaxed structures from random structure search, and the relaxed structures from MatterGen. This dataset represents the widest coverage of reasonable materials under all temperatures and pressures in this study.

In addition, we also evaluated the performance of MatterSim on the publicly available High-pressure Elemental Xstals (HEX) database,[22] which covers the ground-state structures of elemental crystals for the first 57 elements in the periodic table under 0–300 GPa.

To evaluate the performance, the per-atom mean absolute energy error, the mean error on forces, and mean error on stress are computed for each benchmark set. The results are shown in Table S3, where the comparison is carried out between MatterSim and a few open-source universal MLFFs, including M3GNet[16], CHGNet[17], MACE-MP-0[23], SevenNet,[24] Orb-v2,[25] OMat24.[19]

Here, we list the checkpoint of each model used in this benchmark:

- M3GNet: M3GNet-MP-2021.2.8-PES;
- CHGNet: 0.3.0;
- MACE-MP-0: large version of the model defined in the commit 4d2d1c4 in the [repo](#);
- SevenNet: SevenNet-0\_11July2024;
- Orb-V2: orb-v2-20241011.ckpt;
- OMat24: eqV2-M OMat MPtrj-sAlex.

## S4.2 Phonons and group velocities

We benchmark against Materials Data Repository (MDR) phonon calculation database (also known as PhononDB),[21] a database of phonon properties derived from first-principles calculations. PhononDB

MAE of	OMat24	ORB-V2	SevenNet	M3GNet	CHGNet	MACE-MP-0	MatterSim (10M)
Max phonon frequency (THz)	7.007	3.060	1.412	4.248	2.397	1.821	<b>1.016</b>
Phonon group velocity (km/s)	168.386	50.959	69.663	110.073	109.495	39.930	<b>22.895</b>
$\Delta\Delta G$ vs DFT (meV/atom)	28.064	207.715	26.959	47.333	166.099	35.433	<b>18.504</b>

**Table S4:** Performance of OMat24, Orb-V2, SevenNet, M3GNet, CHGNet, MACE-MP-0, and MatterSim on prediction of max phonon frequency, max phonon group velocity, and Gibbs free energy.

encompasses various materials, each characterized by phonon properties computed using the finite displacement method via the Phonopy software package.[26, 27] The force constants for these calculations are obtained through VASP[12, 13]. Furthermore, the Perdew-Burke-Ernzerhof for solids (PBEsol) exchange-correlation functional[28, 29] is utilized within the DFT framework. MatterSim’s performance is rigorously assessed against the entire PhononDB database.

Phonon group velocities are the derivatives of phonon frequencies with respect to phonon wavevectors,

$$\mathbf{v}_{\mathbf{q}\nu} = \frac{\partial\omega_{\mathbf{q}\nu}}{\partial\mathbf{q}}, \quad (13)$$

which characterize heat transport in crystals and thus play a significant role in solid state physics.

In Table S4, we list the results of the prediction errors for max phonon frequency and phonon group velocities. For phonon frequency prediction, MatterSim has achieved an accuracy of slightly above 1 THz. In addition, the phonon group velocities predicted with MatterSim are 3–5-fold better than those of other models.

### S4.2.1 Computational setup

Here the phonon frequencies and group velocities are computed using the finite displacement method implemented in the Phonopy package and the Phono3py packages, respectively. Before the phonon calculation, we first relax the structure of the materials with the force tolerance of 0.02 eV/Å for up to 500 steps, which relaxes both the cell size and the atomic positions of the initial structure. To generate forces required for the phonon and group velocity predictions, displacements compatible with space group are introduced to atomic positions within the supercell, and the size of the supercell is parametrized to be the same as in the PhononDB database. With a magnitude of 0.03 Å, consistent with the settings in PhononDB, the forces acting on each displaced atom are then predicted using either MatterSim or other models.

### S4.3 Free energy prediction with QHA

To construct an accurate phase diagram, we calculate the free energies of materials using the quasi-harmonic approximation (QHA) implemented in Phonopy,[30] which is a theoretical framework that allows for the calculation of free energies at different temperatures and pressures.

The quasi-harmonic approximation is an extension of the harmonic approximation. The harmonic approximation assumes that atoms in a crystal vibrate about their equilibrium positions and the potential energy can be approximated by a quadratic function of the atomic displacements. This model is accurate at low temperatures where anharmonic effects are negligible. However, as the temperature increases, anharmonic contributions become significant, and the harmonic approximation fails to predict the correct thermodynamic behavior. This is where the QHA becomes valuable.

QHA takes into account the volume dependence of the vibrational frequencies. It assumes that while the shape of the potential energy well changes with volume, each volume can still be described by a harmonic potential, but with different force constants. QHA allows the phonon frequencies to depend on the volume, thereby capturing the thermal expansion effect.

The free energy within the quasi-harmonic approximation is calculated by considering both the vibrational and static lattice contributions. The Helmholtz free energy,  $F$ , at a given temperature ( $T$ ) and volume ( $V$ ) can be expressed as:

$$F(T, V) = U(V) - TS(T, V), \quad (14)$$

where  $U$  is the internal lattice energy and  $S$  is the entropy due to the vibrational degrees of freedom.

The entropy term equals

$$S(T, V) = -\frac{1}{N} \sum_{\mathbf{q}\nu} k_B \ln [1 - \exp(-\hbar\omega_{\mathbf{q}\nu}(V)/k_B T)] + \frac{1}{NT} \sum_{\mathbf{q}\nu} \frac{\hbar\omega_{\mathbf{q}\nu}(V)}{\exp(\hbar\omega_{\mathbf{q}\nu}(V)/k_B T) - 1}, \quad (15)$$

where  $\omega_{\mathbf{q}\nu}$  is the  $\nu$ -th phonon mode at  $\mathbf{q}$  wavevector in the Brillouin zone,  $\hbar$  is the reduced Planck constant,  $k_B$  is the Boltzmann constant and  $T$  is the temperature.

The Gibbs free energy  $G$  is obtained by

$$G(T, P) = \min_V [U(V) - TS(T, V) + PV] \quad (16)$$

### S4.3.1 Computational setup

The Gibbs free energy of ordered crystalline materials is computed using MatterSim via the quasi-harmonic approximation (QHA) implemented in Phonopy. [Table S5](#) lists the materials for which the Gibbs free energies are computed. For each material, we change the volume by -2.5% to 2.5% and relax the structure with a force tolerance of 0.02 eV/Å. The phonon calculations are performed with a  $4 \times 4 \times 4$  supercell.

**Table S5:** Summary of materials and their corresponding ID in Materials Project used for the benchmarks of Gibbs free energy prediction.

mp-id	formula	mp-id	formula	mp-id	formula	mp-id	formula
mp-1000	BaTe	mp-1002164	CGe	mp-10044	AsB	mp-1007652	As <sub>2</sub> In <sub>2</sub>
mp-1007661	In <sub>2</sub> Sb <sub>2</sub>	mp-1008786	MgTe	mp-1017439	O <sub>3</sub> SiSr	mp-1018059	Ga <sub>2</sub> Sb <sub>2</sub>
mp-1018100	Al <sub>2</sub> Sb <sub>2</sub>	mp-10250	BaF <sub>3</sub> Li	mp-1039	Mg <sub>2</sub> Te <sub>2</sub>	mp-10695	SZn
mp-1070	Cd <sub>2</sub> Se <sub>2</sub>	mp-10760	MgSe	mp-1087	SSr	mp-1156	GaSb
mp-11718	FRb	mp-1184046	Cl <sub>2</sub> Cu <sub>2</sub>	mp-1190	SeZn	mp-1253	BaSe
mp-1265	MgO	mp-12779	Cd <sub>2</sub> Te <sub>2</sub>	mp-13031	MgSe	mp-13033	MgTe
mp-1315	MgS	mp-1342	BaO	mp-1367	Mg <sub>2</sub> Si	mp-1415	CaSe
mp-1479	BP	mp-149	Si <sub>2</sub>	mp-1500	BaS	mp-1519	CaTe
mp-1541	BeSe	mp-1550	AlP	mp-1639	BN	mp-1672	CaS
mp-1700	AlN	mp-1778	BeO	mp-1883	SnTe	mp-1958	SrTe
mp-1960	Li <sub>2</sub> O	mp-19717	PbTe	mp-1986	OZn	mp-20012	InSb
mp-20194	CeO <sub>2</sub>	mp-20305	AsIn	mp-20351	InP	mp-20411	InN
mp-20724	Mg <sub>2</sub> Pb	mp-2133	O <sub>2</sub> Zn <sub>2</sub>	mp-2172	AlAs	mp-2201	PbSe
mp-22205	In <sub>2</sub> N <sub>2</sub>	mp-22865	ClCs	mp-22894	Ag <sub>2</sub> I <sub>2</sub>	mp-22905	CLi
mp-22906	BrCs	mp-22913	BrCu	mp-22914	ClCu	mp-23193	ClK
mp-23209	Cl <sub>2</sub> Sr	mp-23251	BrK	mp-23259	BrLi	mp-241	CdF <sub>2</sub>
mp-2469	CdS	mp-2472	OSr	mp-2490	GaP	mp-252	BeTe
mp-2534	AsGa	mp-2542	Be <sub>2</sub> O <sub>2</sub>	mp-2597	InTe	mp-2605	CaO
mp-2624	AlSb	mp-2653	B <sub>2</sub> N <sub>2</sub>	mp-2691	CdSe	mp-2730	HgTe
mp-2758	SeSr	mp-32	Ge <sub>2</sub>	mp-3448	F <sub>3</sub> KMg	mp-380	Se <sub>2</sub> Zn <sub>2</sub>
mp-406	CdTe	mp-408	GeMg <sub>2</sub>	mp-422	BeS	mp-463	FK
mp-560588	S <sub>2</sub> Zn <sub>2</sub>	mp-561947	CsF <sub>3</sub> Hg	mp-567636	FeSbV	mp-568560	BrTl
mp-569346	Cu <sub>2</sub> I <sub>2</sub>	mp-5878	F <sub>3</sub> KZn	mp-5967	CoSbTi	mp-643	O <sub>2</sub> Th
mp-66	C <sub>2</sub>	mp-661	Al <sub>2</sub> N <sub>2</sub>	mp-672	Cd <sub>2</sub> S <sub>2</sub>	mp-682	FNa
mp-7104	CaCsF <sub>3</sub>	mp-7140	C <sub>2</sub> Si <sub>2</sub>	mp-804	Ga <sub>2</sub> N <sub>2</sub>	mp-8062	CSi
mp-820	HgSe	mp-830	GaN	mp-8399	CdCsF <sub>3</sub>	mp-8402	F <sub>3</sub> MgRb
mp-8880	Al <sub>2</sub> P <sub>2</sub>	mp-8881	Al <sub>2</sub> As <sub>2</sub>	mp-8882	Ga <sub>2</sub> P <sub>2</sub>	mp-8883	As <sub>2</sub> Ga <sub>2</sub>
mp-8884	Te <sub>2</sub> Zn <sub>2</sub>	mp-924129	NiSnZr	mp-924130	NiSnTi	mp-966800	In <sub>2</sub> P <sub>2</sub>
mp-978489	O <sub>3</sub> PbSi	mp-997618	BSb	mp-998739	F <sub>3</sub> MgTl		

**Table S6:** Multi-task Result

Multi-task Valid MAE	
magnetic moments ( $\mu_B$ )	0.0640
Bader charges ( $e$ )	0.0233
dielectric matrices	0.2478
Born effective charges ( $e$ )	0.0756

For benchmark, we first compute the change in Gibbs free energy  $\Delta G$  between 300 K and 900 K, and then compute the difference ( $\Delta\Delta G$ ) between the MatterSim and DFT results. The DFT reference Gibbs free energies are computed at the DFT/PBE level of theory, and the results are listed in [Table S4](#).

#### S4.4 Results of Multiple Tasks

The results of different tasks on the test set are shown in [Table S6](#).

## S5 Multi-task Case studies

### S5.1 Formation enthalpy of LaH<sub>10</sub>

The phase stability of the La–H system is investigated by constructing convex hulls based on the enthalpy of potential phases under pressures ranging from 0–200 GPa. All calculations are performed using MatterSim-MT 10 M parameters. The analysis involves 21 distinct phases:

- seven La-H phases from the Materials Project with the MP IDs: mp-1104116, mp-1103808, mp-24153, mp-1018144, mp-1103947, mp-973064, mp-2647067;
- four ground-state Hydrogen phases under 0, 100, 200, and 300 GPa from Ref. 22;
- four ground-state Lanthanum phases under 0, 100, 200, and 300 GPa from Ref. 22;
- six high-pressure La-H phases identified through random structure search LaH<sub>10</sub> (Fm $\bar{3}$ m), LaH<sub>3</sub> (Cmcm), LaH<sub>5</sub> (P $\bar{1}$ ), LaH<sub>2</sub> (C2/m), LaH<sub>4</sub> (I4/mmm), and LaH<sub>8</sub> (C2/m).

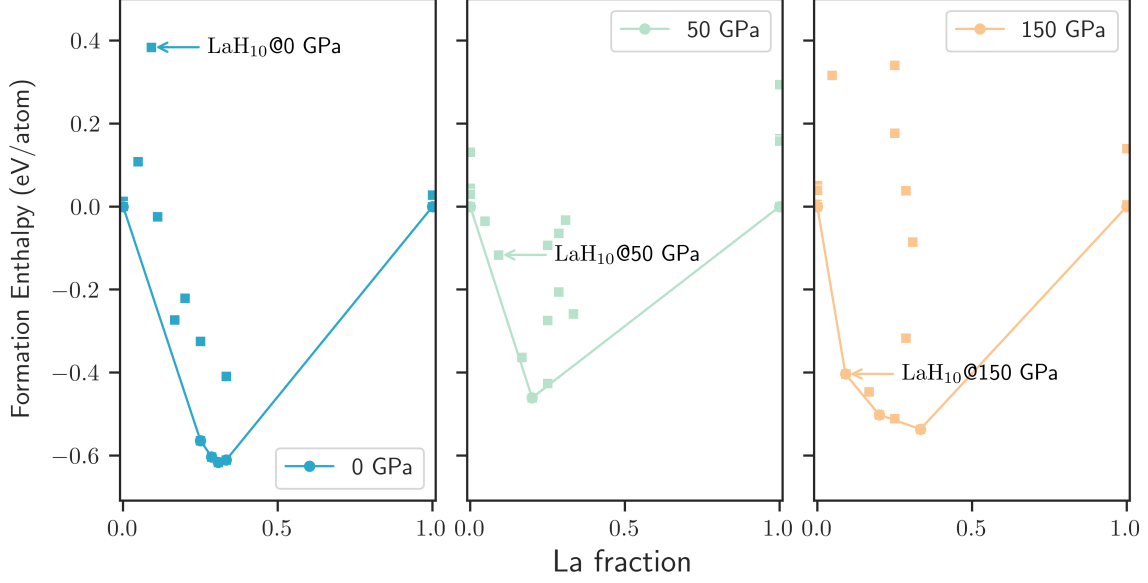
To determine the pressure-dependent enthalpy of each phase, structures are initially relaxed at 200 GPa, allowing optimization of both atomic coordinates and cell parameters. Subsequent relaxations are performed with decreasing pressure at 20 GPa intervals down to 0 GPa. The enthalpy is computed following each relaxation. Convex hulls are constructed using the PhaseDiagram package from pymatgen[11] with enthalpies normalized such that the lowest-enthalpy end members (H and La) are set to zero.

To validate the MatterSim results, density functional theory (DFT) calculations are performed using identical protocols at selected pressures: 0, 50, and 150 GPa, as presented in Fig. S8. The hydrogen-rich phase LaH<sub>10</sub> exhibits pressure-dependent stability: it is thermodynamically unstable at ambient pressure but becomes stabilized at 150 GPa, consistent with the MatterSim predictions shown in Fig. 2(a). This computational finding aligns with previous first-principles simulations that predict LaH<sub>10</sub> stability at 150 GPa[31] and experimental observations of LaH<sub>10</sub> formation within a pressure range of 137–218 GPa[32].

### S5.2 Pressure dependent LO–TO splitting of 3c-SiC

Due to its high thermal conductivity and wide band gap, 3c-SiC is believed to have potential applications in the semiconductor industry. As a polar material, 3c-SiC has a finite LO–TO splitting that cannot be predicted using traditional force fields because the prediction of Born effective charges and dielectric matrices is required for this purpose.

To showcase the multi-task capability of MatterSim, we predict the pressure-dependent LO–TO splitting of 3c-SiC as illustrated in Fig. 3(b), by adding the non-analytical correction to the force



**Fig. S8:** The formation enthalpy of the La–H chemical system under the pressure of 0, 50, and 150 GPa.

**Table S7:** Lattice constant  $a$ , Born effective charge on Silicon  $Z_{\text{Si}}^*$ , dielectric constant  $\varepsilon_\infty$ , longitudinal optical phonon frequency at  $\Gamma$  point  $\omega_{\text{LO}}(\Gamma)$ , transverse optical phonon frequency at  $\Gamma$  point  $\omega_{\text{TO}}(\Gamma)$  and LO–TO splitting predicted with MatterSim of 3c-SiC at zero pressure in comparison with literature.

	MatterSim	Theoretical	Experimental
$a$ (Å)	4.377	4.385 (PBE)[38], 4.38 (PBE)[39]	4.360[40]
$Z_{\text{Si}}^*$ (a.u.)	2.71	2.72(LDA)[36]	2.70[41]
$\varepsilon_\infty$	7.31	7.02(LDA)[36], 6.94 (PBE)[39]	6.52[42]
$\omega_{\text{LO}}(\Gamma)$ (THz)	28.54	28.56(LDA)[36]	29.16[37]
$\omega_{\text{TO}}(\Gamma)$ (THz)	23.28	23.36(LDA)[36]	23.87[37]
LO–TO splitting (THz)	5.26	5.20 (LDA)[36]	5.29[37]

constant matrices:[33–35]

$${}^{\text{NA}}C_{I\alpha,J\beta} = \frac{4\pi}{\Omega} \frac{(\mathbf{q} \cdot Z_I^*)_\alpha (\mathbf{q} \cdot Z_J^*)_\beta}{\mathbf{q} \cdot \varepsilon_\infty \cdot \mathbf{q}}, \quad (17)$$

where  $Z_I^*$  is the Born effective charge on the  $I$ -th atom,  $\varepsilon_\infty$  is the dielectric matrix,  $\alpha, \beta, \gamma$  are indices of three dimensions, and  $\mathbf{q}$  is the wavevector of the phonon.

In Table S7, we list the lattice constant  $a$ , Born effective charge on Silicon  $Z_{\text{Si}}^*$ , dielectric constant  $\varepsilon_\infty$ , longitudinal optical phonon frequency at  $\Gamma$  point  $\omega_{\text{LO}}(\Gamma)$ , transverse optical phonon frequency at  $\Gamma$  point  $\omega_{\text{TO}}(\Gamma)$  and the LO–TO splitting predicted with MatterSim of 3c-SiC at zero pressure. MatterSim predicts the LO–TO splitting with an error of only 0.06 THz with respect to previous literature employing *ab initio* computational methods[36], and 0.03 THz when compared to experimental measurements.[37]

### S5.3 Phase boundary of B1–B2 MgO

To construct an accurate phase boundary of the B1–B2 phases of MgO, we calculate the free energies of the two phases using the QHA methodology as implemented in Phonopy,[30]. The computational setup we employ is the same as in Section S4.3.1.

### S5.4 Hysteresis curve of BaTiO<sub>3</sub>

The simulation of BaTiO<sub>3</sub> under an applied electric field is based on the electric enthalpy,

$$U = U_0 - \mathbf{E} \cdot \mathbf{P}, \quad (18)$$

where  $U_0$  is the potential energy in the absence of an electric field,  $\mathbf{E}$  is the applied electric field, and  $\mathbf{P}$  is the polarization of the material. Using the Born effective charges predicted by MatterSim-MT, we evaluate the polarization of unit cell  $u$  as

$$\mathbf{P}^{(u)} = \frac{1}{2} \sum_{O \in u} \mathbf{Z}_O^* \cdot \Delta r_{O-\text{Ti}}^{(u)} + \frac{1}{8} \sum_{\text{Ba} \in u} \mathbf{Z}_{\text{Ba}}^* \cdot \Delta r_{\text{Ba}-\text{Ti}}^{(u)}, \quad (19)$$

where  $\Delta r$  denotes the relative displacement between atoms, and  $\mathbf{Z}^*$  is the Born effective charge tensor. The applied electric field also introduces an additional force on each atom during the MD simulation:

$$F_i = F_i^0 + e\mathbf{Z}_i^* \cdot \mathbf{E}, \quad (20)$$

where  $F_i^0$  is the force in the absence of the electric field.

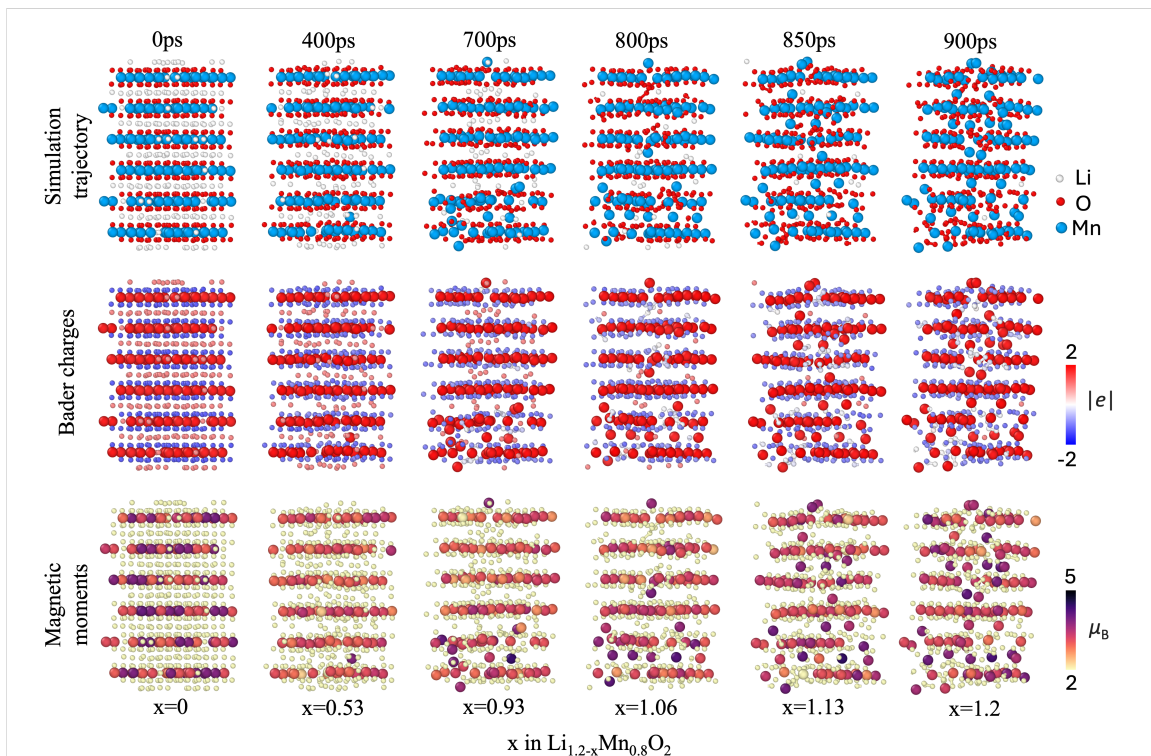
With this setup, we performed MD simulations at target temperatures of 100, 200, and 300 K using a Nosé–Hoover thermostat with time step 2 fs under a sinusoidal electric field,

$$\mathbf{E}(t) = E_{\max} \cos\left(\frac{2\pi t}{T}\right), \quad (21)$$

where the maximum electric field is  $E_{\max} = 4 \text{ MV/cm}$  and the period is  $T = 200 \text{ ps}$ . The hysteresis curve shown in Fig. 3(c) is averaged over 5 cycles.

### S5.5 Delithiation of Li<sub>1.2-x</sub>Mn<sub>0.8</sub>O<sub>2</sub>

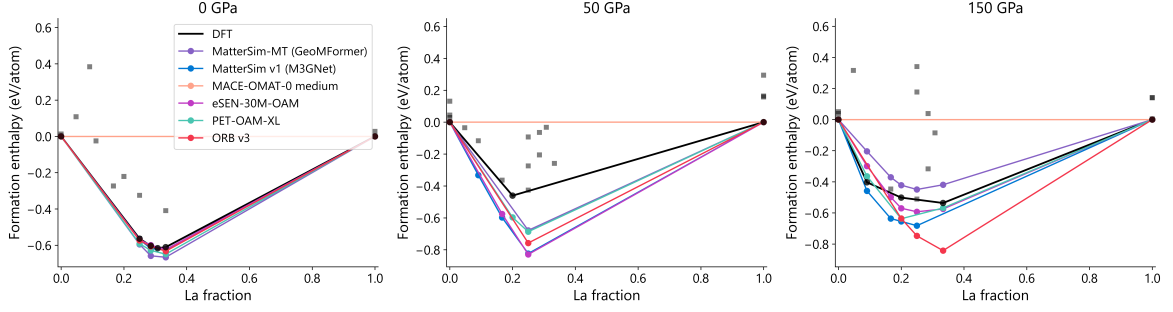
The molecular dynamics simulations of Li<sub>1.2-x</sub>Mn<sub>0.8</sub>O<sub>2</sub> are conducted using the MatterSim-MT-10M model, following protocols similar to those in Ref. 43. For the simulation, we construct an initial structure of Li<sub>1.2-x</sub>Mn<sub>0.8</sub>O<sub>2</sub> by generating a  $5 \times 4 \times 2$  supercell from the conventional cell of LiCoO<sub>2</sub>



**Fig. S9:** Snapshots of multi-task molecular dynamics simulation on  $\text{Li}_{1.2-x}\text{Mn}_{0.8}\text{O}_2$

(mp-22526). Subsequently, all Co atoms are replaced with Mn, and 30 Mn atoms are randomly selected and substituted with Li. During the simulation, the structure undergoes initial relaxation on both ionic and lattice degrees of freedom. Molecular dynamics simulations are then performed under the NVT ensemble at 1000 K with a 1 fs timestep. One lithium atom is removed every 5 ps until complete delithiation, resulting in a total simulation time of approximately 900 ps. In addition to forces, energies, and stress tensors, Bader charges and atomic magnetic moments are computed using MatterSim for real-time analysis.

Fig. S9 illustrates the structural evolution at various lithium concentrations. Initially, Mn acts as the primary redox center, characterized by a change in magnetic moment from approximately  $3\text{--}4\mu_{\text{B}}$ . When lithium depletion reaches  $x \approx 0.5$  in  $\text{Li}_{1.2-x}\text{Mn}_{0.8}\text{O}_2$ , Mn ions begin migrating from octahedral to tetrahedral sites in the lithium layers, accompanied by oxygen sublattice distortions. At this stage, most oxygen atoms maintain their anionic state. Further delithiation to  $x \approx 0.9$  triggers oxygen dimer formation, typically following Mn migration to the Li layer. The evolution of these unstable dimers leads to molecular oxygen formation within the lattice, signifying irreversible anionic redox. This phenomenon is corroborated by Bader charge analysis, where the  $\text{O}_2$  molecules exhibit near-zero charges. Continued delithiation destabilizes the lattice structure, resulting in increased molecular oxygen formation and eventual transition to nanofluidic states.



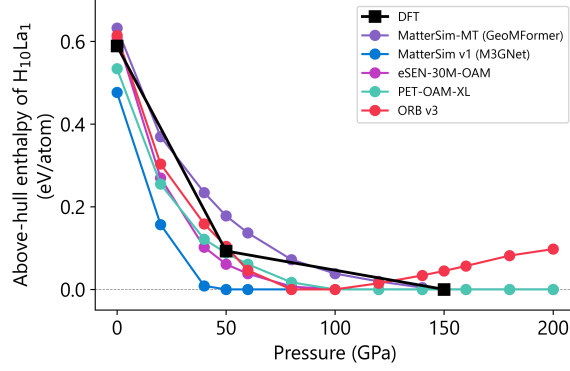
**Fig. S10:** Convex hull of formation enthalpies in the La–H system at 0, 50, and 150 GPa. For clarity, we only show for DFT the phases above the hull (black squares). Coloured lines show the convex hulls predicted by ML interatomic potentials.

## S5.6 Comparison to other foundational interatomic potentials

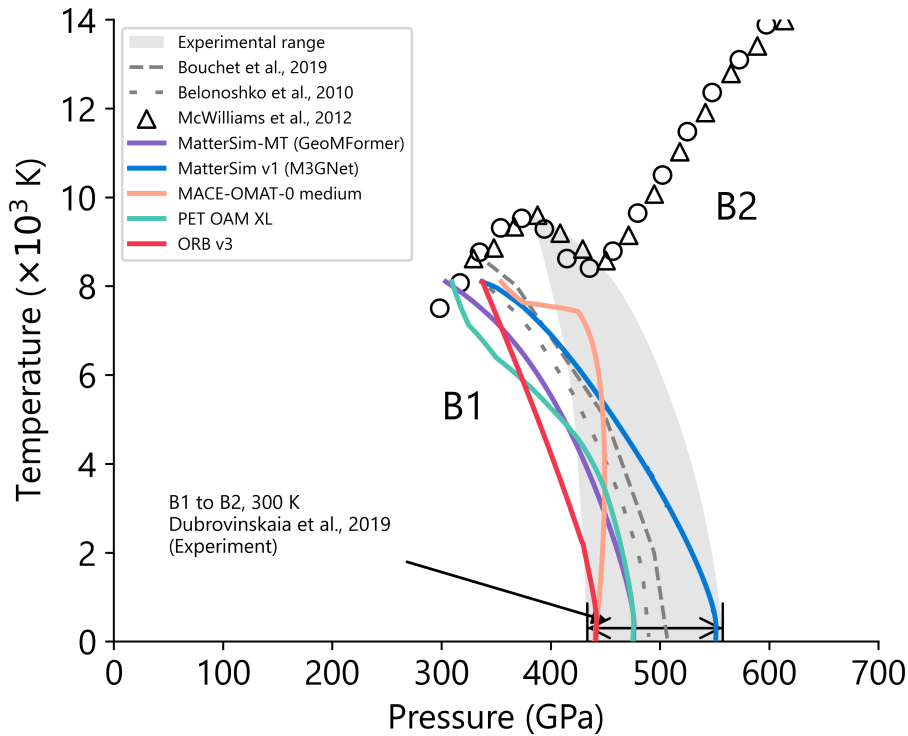
To put the performance of **MatterSim-MT** into context, we assess how well popular and established ML interatomic potentials capture the energetics relevant to high-pressure phase stability. Specifically, we benchmark several publicly available models including MACE-OMAT-0 [23], PET OAM XL [], ORB v3 [], eSEN-30M-OAM [], and MatterSim v1 [2] based on the M3GNet architecture [16], on two tasks from the main text: the La–H convex hull under pressure and the MgO B1–B2 phase boundary.

We begin with the La–H system. For each model, we follow the workflow described in section S5.1 and relax all structures at pressures between 200 and 0 GPa with a force tolerance of 0.02 eV/Å. In Fig. S10, we construct the convex hull of formation enthalpies for each model at 0, 50, and 150 GPa and compare the results to the DFT calculations, reported in Fig. S8. At ambient pressure, all models apart from MACE-OMAT-0 medium reproduce the DFT hull within  $\approx 50$  meV/atom, correctly identifying LaH<sub>2</sub>, LaH<sub>3</sub>, and La<sub>2</sub>H<sub>5</sub> as stable phases. MACE-OMAT-0 medium produces unphysical relaxed geometries at elevated pressures that propagate to the ambient-pressure hull through the chained relaxation protocol. A reverse sweep from 0 to 200 GPa confirms that the model reproduces the DFT hull at 0 GPa but diverges at pressures as low as 50 GPa. At 50 GPa, the hull topology is sensitive to small energy differences and varies across models. DFT places only LaH<sub>4</sub> on the hull, whereas most models also stabilise LaH<sub>3</sub> and some LaH<sub>5</sub>, which lie only  $\approx 6$  and  $\approx 21$  meV/atom above the DFT hull, respectively. At 150 GPa, all models except MACE-OMAT-0 medium and ORB v3 reproduce the DFT hull, including the stabilisation of LaH<sub>10</sub>, LaH<sub>4</sub>, and LaH<sub>2</sub>. Similar to 50 GPa, some models place additional phases (LaH<sub>3</sub>, LaH<sub>5</sub>) on the hull which are marginally above the DFT hull.

Fig. S8 tracks the above-hull enthalpy of LaH<sub>10</sub> as a function of pressure. This phase is experimentally observed to stabilize between 137 and 218 GPa as a high-temperature superconductor [32]. In DFT, LaH<sub>10</sub> lies on the hull at 150 GPa. GeoMFormer reproduces this stabilisation pressure, while M3GNet, PET-OAM-XL, and eSEN-30M-OAM stabilise LaH<sub>10</sub> somewhat earlier



**Fig. S11:** Above-hull enthalpy of  $\text{LaH}_{10}$  as a function of pressure predicted by DFT and a range of ML interatomic potentials.



**Fig. S12:** B1–B2 phase boundary of  $\text{MgO}$  computed by DFT and different ML interatomic potentials compared to experimental measurements.

(50–100,GPa). ORB v3 shows qualitatively different behaviour where  $\text{LaH}_{10}$  briefly reaches the hull around 80–100 GPa but destabilises again at higher pressures.

For  $\text{MgO}$ , we compute the B1–B2 phase boundary using the QHA workflow described in Section S5.3 and compare to experiments [44, 45] and *ab-initio* calculations [46, 47] in Fig.S12. The shape of the boundary is governed by the Clausius–Clapeyron relation and reflects how the two phases differ in volume and vibrational entropy. Reproducing this shape therefore tests whether a model captures these thermodynamic quantities correctly. We note that eSEN-30M-OAM is excluded from this benchmark as the  $4 \times 4 \times 4$  supercells required for phonon calculations exceeded the memory capacity

on a NVIDIA A100 80GB GPU. MatterSim-MT and MatterSim v1 both reproduce the shape predicted by theory [46, 47]. While MatterSim v1 overestimates the transition pressure by  $\approx 60$  GPa, MatterSim-MT agrees within  $\approx 15$  GPa at 300 K. PET-OAM-XL reproduces the boundary at low temperatures but develops an unphysically steep slope above  $\approx 5000$  K. ORB v3 produces a nearly linear boundary, failing to capture the curvature. MACE-OMAT-0 medium yields an essentially flat boundary with no meaningful temperature dependence, indicating that the model does not resolve the free energy difference between the two phases.

## S6 MatterSim as a continual active learner

As a universal predictive model for material properties, MatterSim may not yield satisfactory accuracy for highly complex systems that have not been previously encountered in its training dataset. Under such circumstances, an active learning approach can be employed to selectively filter data, followed by finetuning of the MatterSim model. This procedure facilitates the rapid development of a sufficiently accurate and operational model. In this work, we present an example of the complex ionic superconductor  $\text{Li}_2\text{B}_{12}\text{H}_{12}$ , as shown in the inset of Fig. 2(a) in the main text. To generate the training dataset, we performed NVT simulations on  $\text{Li}_2\text{B}_{12}\text{H}_{12}$  using VASP[12, 13] with a time step of 0.5 fs for a total duration of 5.0 ps at a simulation temperature of 2000 K. The test dataset belongs to trajectories that are not included in the training set.

The active learning procedure can significantly improve the accuracy of MLFFs by augmenting the dataset with data points that exhibit high uncertainty based on the ensemble models. The uncertainty arises from the variability inherent in different pre-trained models, each initialized with a distinct random number seed. This variability is quantified as follows:

$$\text{unc} = \max \left[ \frac{1}{N} \sum_{i=1}^N (|F_{ia}| - |\bar{F}_a|)^2 \right] \quad (22)$$

where  $F_{ia}$  denotes the predicted atomic force on the  $a$ -th atom by the  $i$ -th model and  $N$  represents the total number of pretrained models utilized.  $\bar{F}_a$  signifies the average of the atomic forces predicted by the ensemble models.

## S7 Finetuning and molecular dynamics on liquid water

This section details the application of finetuning to simulate properties of liquid water. We first outline the parameter settings for the finetuning process and MD simulations. Next, we describe the post-processing of the oxygen-oxygen-oxygen angular distribution function (ADF), present the oxygen-hydrogen and hydrogen-hydrogen radial distribution functions (RDF), assess data efficiency, and

detail the calculation of diffusion coefficients. The three trained models are referred to as MatterSim-MT, MatterSim scratch-900, and finetune-60. The MatterSim-MT model, trained with PBE-level theory, is discussed in the main text without finetuning and lacks liquid water configurations in its training data. The scratch-900 model is trained from scratch using 900 bulk liquid water configurations from Ref. [48, 49] with rev-PBE0-D3 level theory. The finetune-60 model is fine-tuned using only 60 of these 900 configurations.

### S7.1 Parameter settings for finetuning

To address the limitations imposed by the level of theory of the training data, we implemented finetuning on the MatterSim model. For finetuning, we discard the pretrained head, reinitialize parameters, and normalize the data. The parameters for finetuning and training from scratch are shown in Table S8.

We randomly sampled 100 liquid water configurations, which were divided into validation and test datasets, with the remaining 900 configurations reserved as potential training candidates. Among the 900 available candidates, 60 configurations were selected at random to create various training sets through the alteration of random seeds, which were subsequently employed to fine-tune the MatterSim model.

**Table S8:** Water data training hyperparameters

Training parameters	finetune	from scratch
optimizer	AdamW[6]	AdamW[6]
head max learning rate	2e-4	2e-4
backbone max learning rate	5e-5	2e-4
weight decay	0.0	0.0
$\beta_1$	0.9	0.9
$\beta_2$	0.999	0.999
$\epsilon$	1e-7	1e-7
training epochs	200	200
learning rate schedule	stepLR	stepLR
step size epochs	8	8
$\omega_f$	10.0	10.0

### S7.2 Simulation settings for molecular dynamics

In this work, we evaluated the performance of the finetuning scheme by probing the structural and dynamical properties of liquid water. MD simulations are conducted up to nanoseconds using the ASE package. The initial structure used for these simulations is composed of a cubic liquid water box containing 512 water molecules with a box length of 24.68. Production runs of the MD simulations are carried out in the NVT ensemble at 298 K and we regulate the temperature using the Nosé-Hoover thermostat[50–54] every 100 steps. The timestep used to propagate the dynamics is chosen to be

0.5 fs. Lastly, the first 200 ps out of the nanosecond trajectories are discarded for pre-equilibration. ADF and RDF, along with diffusion coefficients, are analyzed using the General Purpose Trajectory Analyser (GPTA) software tool.[55]

### S7.3 ADF of Liquid Water

As illustrated in Fig. 2(d), the ADF of the oxygen species ( $P_{OOO}(\theta)$ ), is multiplied by the sine of the angle  $\sin(\theta)$ , corresponding to the triples of oxygen atoms. The product  $P_{OOO}(\theta) \sin(\theta)$  allows for a direct comparison with angular distribution extracted from empirical potential structural refinement (EPSR) based on joint X-ray/neutron scattering measurements,[56] thus enabling better understanding of local arrangements of the water molecules in condensed phase. Following the methodology outlined in previous studies[57–59],  $P_{OOO}(\theta)$  is normalized such that  $\int_0^\pi P_{OOO}(\theta) \sin(\theta) d\theta$  goes to unity. Similarly, the cutoff value for identifying oxygen triples is chosen so that the oxygen-oxygen coordination number averages around 4.[57]

### S7.4 Deriving Diffusion Coefficients from Mean Squared Displacements

Diffusion coefficients ( $\mathcal{D}$ ) of liquid water at 300 K are determined via the Einstein relation[60, 61]:

$$\mathcal{D} = \frac{1}{6} \lim_{\tau \rightarrow \infty} \frac{d\lambda}{d\tau}, \quad (23)$$

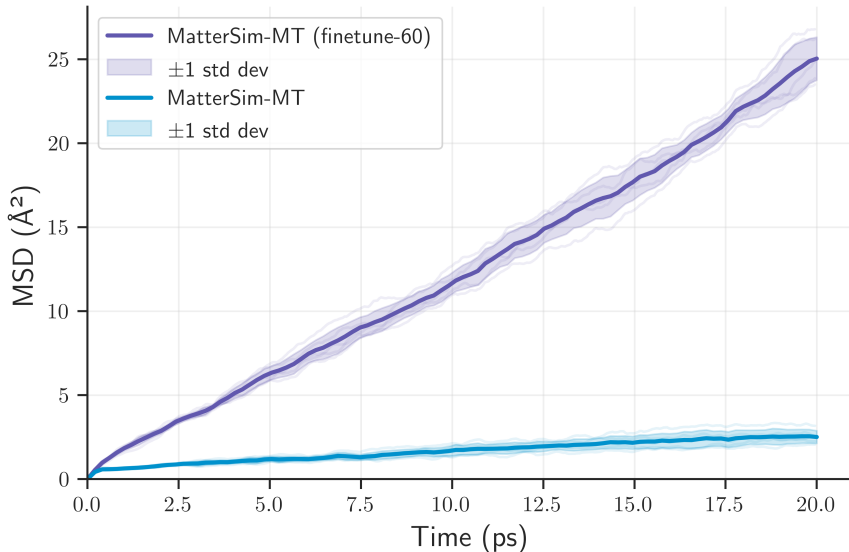
where  $\lambda$  stands for the mean squared displacements along a MD trajectory, and  $\tau$  represents the correlation time chosen. Eq. 23 suggests that  $\mathcal{D}$  can be obtained by extracting the slope from a linear fit between  $\lambda$  and  $\tau$ . In this work,  $\lambda$  is computed along the same trajectory used to compute the RDFs and ADFs. The MSDs computed with the MatterSim-MT and finetuned MatterSim-MT models are reported in Fig. S13, with the shaded regions indicating the standard deviation among 5 segments extracted from the trajectory. As listed in Table S9, MatterSim-MT predicts the diffusion coefficient to be  $0.018(4) \text{ \AA}^2/\text{ps}$ , which is compatible with simulation results reported in literature, for example  $0.018(2) \text{ \AA}^2/\text{ps}$  in Ref. 62. Moreover, after finetuning with only 60 snapshots of water at the level of rev-PBE0-D3, MatterSim is able to correctly predict the diffusion coefficient to be  $0.193(11) \text{ \AA}^2/\text{ps}$ , which is also close to the experimental measurements  $0.187 \text{ \AA}^2/\text{ps}$ ,[63] and  $0.230\text{--}0.240 \text{ \AA}^2/\text{ps}$ . [64, 65]

### S7.5 Data efficiency of finetuning

Fig. S14 illustrates the model’s data efficiency during finetuning, showing that only a minimal amount of data is required to achieve the same performance as training from scratch.

Type	$\mathcal{D}$ ( $\text{\AA}^2/\text{ps}$ )	Source
MatterSim-MT finetune	$0.193\pm 0.011$	<b>This work</b>
MatterSim-MT	$0.018\pm 0.004$	<b>This work</b>
Simulation with PBE functional	$0.018\pm 0.002$	[62]
Simulation with SCAN functional	$0.190\pm 0.025$	[62]
Experiment	$0.230\sim 0.240$	[64, 65]
Experiment	0.187	[63]

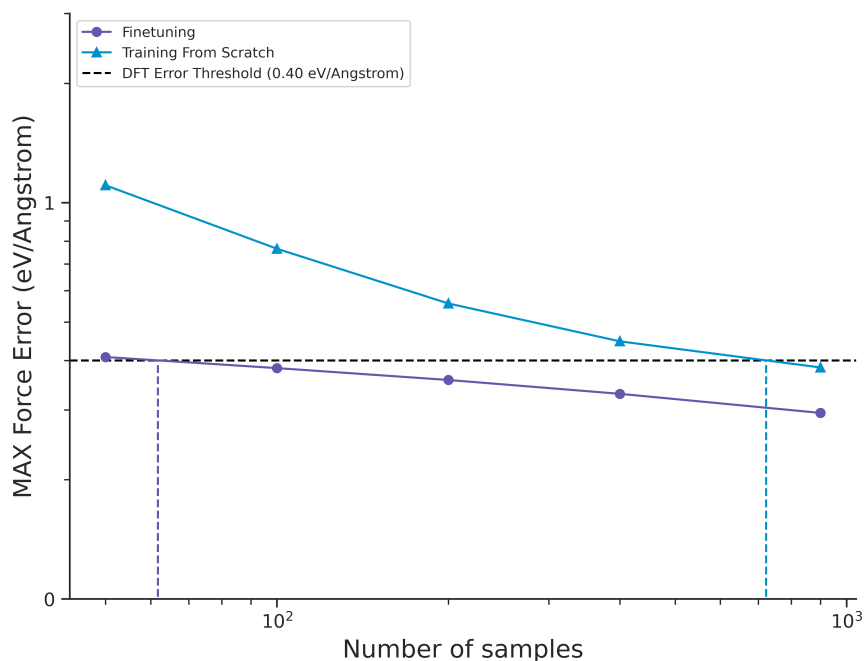
**Table S9:** The diffusion coefficient of water at 300 K predicted with MatterSim-MT and finetuned models of MatterSim-MT, compared with first-principles simulations reported in the literature and experimental measurements.



**Fig. S13:** Mean squared displacement ( $\text{\AA}$ ) analysis of MatterSim-MT and finetuned models. Shaded regions indicate one standard deviation of five 20-ps segments from the same trajectory.

## References

- [1] T. Chen, S. Luo, D. He, S. Zheng, T.Y. Liu, L. Wang, Geomformer: A general architecture for geometric molecular representation learning (2023)
- [2] H. Yang, C. Hu, Y. Zhou, X. Liu, Y. Shi, J. Li, G. Li, Z. Chen, S. Chen, C. Zeni, et al., Mattersim: A deep learning atomistic model across elements, temperatures and pressures. arXiv preprint arXiv:2405.04967 (2024)
- [3] A. Vaswani, Attention is all you need. Advances in Neural Information Processing Systems (2017)
- [4] T. Xie, J.C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. Physical review letters **120**(14), 145301 (2018)
- [5] Y. Zhong, H. Yu, X. Gong, H. Xiang, Edge-based tensor prediction via graph neural networks. arXiv preprint arXiv:2201.05770 (2022)



**Fig. S14:** MAX force error against first-principles results in the test set from the finetuned model and the model trained from scratch

- [6] I. Loshchilov, F. Hutter, Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
- [7] A.R. Tan, J.C. Dietschreit, R. Gomez-Bombarelli, Enhanced sampling of robust molecular datasets with uncertainty-based collective variables. arXiv preprint arXiv:2402.03753 (2024)
- [8] A.M. Krajewski, J.W. Siegel, J. Xu, Z.K. Liu, Extensible structure-informed prediction of formation energy with improved accuracy and usability employing neural networks. *Computational Materials Science* **208**, 111254 (2022)
- [9] A. Thomas-Mitchell, G. Hawe, P.L. Popelier, Calibration of uncertainty in the active learning of machine learning force fields. *Machine Learning: Science and Technology* **4**(4), 045034 (2023)
- [10] S. Thaler, G. Doehner, J. Zavadlav, Scalable bayesian uncertainty quantification for neural network potentials: promise and pitfalls. *Journal of Chemical Theory and Computation* **19**(14), 4520–4532 (2023)
- [11] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V.L. Chevrier, K.A. Persson, G. Ceder, Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **68**, 314–319 (2013)

- [12] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science* **6**(1), 15–50 (1996)
- [13] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B* **54**(16), 11169 (1996)
- [14] P.E. Blöchl, Projector augmented-wave method. *Physical review B* **50**(24), 17953 (1994)
- [15] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple. *Physical review letters* **77**(18), 3865 (1996)
- [16] C. Chen, S.P. Ong, A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science* **2**(11), 718–728 (2022)
- [17] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C.J. Bartel, G. Ceder, Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nature Machine Intelligence* **5**(9), 1031–1041 (2023)
- [18] J. Schmidt, N. Hoffmann, H.C. Wang, P. Borlido, P.J. Carriço, T.F. Cerqueira, S. Botti, M.A. Marques, Machine-learning-assisted determination of the global zero-temperature phase diagram of materials. *Advanced Materials* **35**(22), 2210788 (2023)
- [19] L. Barroso-Luque, M. Shuaibi, X. Fu, B.M. Wood, M. Dzamba, M. Gao, A. Rizvi, C.L. Zitnick, Z.W. Ulissi, Open materials 2024 (omat24) inorganic materials dataset and models. *arXiv preprint arXiv:2410.12771* (2024)
- [20] S. Curtarolo, G.L. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design. *Nature materials* **12**(3), 191–201 (2013)
- [21] A. Togo. Atsushi togo. URL <https://doi.org/10.48505/nims.4197>
- [22] F. Giannessi, S. Di Cataldo, S. Saha, L. Boeri, A database of high-pressure crystal structures from hydrogen to lanthanum. *Scientific Data* **11**(1), 766 (2024)
- [23] I. Batatia, P. Benner, Y. Chiang, A.M. Elena, D.P. Kovács, J. Riebesell, X.R. Advincula, M. Asta, W.J. Baldwin, N. Bernstein, et al., A foundation model for atomistic materials chemistry. *arXiv preprint arXiv:2401.00096* (2023)
- [24] Y. Park, J. Kim, S. Hwang, S. Han, Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations. *Journal of Chemical Theory and*

Computation (2024)

- [25] M. Neumann, J. Gin, B. Rhodes, S. Bennett, Z. Li, H. Choubisa, A. Hussey, J. Godwin, Orb: A fast, scalable neural network potential. arXiv preprint arXiv:2410.22570 (2024)
- [26] A. Togo, L. Chaput, T. Tadano, I. Tanaka, Implementation strategies in phonopy and phono3py. *J. Phys. Condens. Matter* **35**(35), 353001 (2023). <https://doi.org/10.1088/1361-648X/acd831>
- [27] A. Togo, First-principles phonon calculations with phonopy and phono3py. *J. Phys. Soc. Jpn.* **92**(1), 012001 (2023). <https://doi.org/10.7566/JPSJ.92.012001>
- [28] J.P. Perdew, A. Ruzsinszky, G.I. Csonka, O.A. Vydrov, G.E. Scuseria, L.A. Constantin, X. Zhou, K. Burke, Restoring the density-gradient expansion for exchange in solids and surfaces. *Physical review letters* **100**(13), 136406 (2008)
- [29] A. Togo. Private communication (2024)
- [30] A. Togo, L. Chaput, T. Tadano, I. Tanaka, Implementation strategies in phonopy and phono3py. *Journal of Physics: Condensed Matter* **35**(35), 353001 (2023)
- [31] I.A. Kruglov, D.V. Semenok, H. Song, R. Szcześniak, I.A. Wrona, R. Akashi, M.M. Davari Esfahani, D. Duan, T. Cui, A.G. Kvashnin, et al., Superconductivity of LaH<sub>10</sub> and LaH<sub>16</sub> polyhydrides. *Physical Review B* **101**(2), 024508 (2020)
- [32] A. Drozdov, P. Kong, V. Minkov, S. Besedin, M. Kuzovnikov, S. Mozaffari, L. Balicas, F. Balakirev, D. Graf, V. Prakapenka, et al., Superconductivity at 250 K in lanthanum hydride under high pressures. *Nature* **569**(7757), 528–531 (2019)
- [33] W. Cochran, R. Cowley, Dielectric constants and lattice vibrations. *Journal of Physics and Chemistry of Solids* **23**(5), 447–450 (1962)
- [34] P. Giannozzi, S. De Gironcoli, P. Pavone, S. Baroni, Ab initio calculation of phonon dispersions in semiconductors. *Physical Review B* **43**(9), 7231 (1991)
- [35] S. Baroni, S. De Gironcoli, A. Dal Corso, P. Giannozzi, Phonons and related crystal properties from density-functional perturbation theory. *Reviews of modern Physics* **73**(2), 515 (2001)
- [36] K. Karch, F. Bechstedt, P. Pavone, D. Strauch, Pressure-dependent dynamical and dielectric properties of cubic SiC. *Journal of Physics: Condensed Matter* **8**(17), 2945 (1996)

- [37] D. Olego, M. Cardona, Pressure dependence of raman phonons of ge and 3 c-sic. *Physical Review B* **25**(2), 1151 (1982)
- [38] J. Wang, L. Zhang, Q. Zeng, G.L. Vignoles, L. Cheng, Surface relaxation and oxygen adsorption behavior of different sic polytypes: a first-principles study. *Journal of Physics: Condensed Matter* **22**(26), 265003 (2010)
- [39] J.H. Skone, M. Govoni, G. Galli, Self-consistent hybrid functional for condensed systems. *Physical Review B* **89**(19), 195112 (2014)
- [40] O. Madelung, *Physics of group iv elements and iii-v compounds*. Landolt-Bornstein, Numerical Data and Functional Relationships in Science and Technology, New Series **17** (1982)
- [41] D. Olego, M. Cardona, P. Vogl, Pressure dependence of the optical phonons and transverse effective charge in 3 c-sic. *Physical Review B* **25**(6), 3878 (1982)
- [42] P. Yu, Fundamentals of semiconductors (Springer, 2005)
- [43] K. McColl, S.W. Coles, P. Zarabadi-Poor, B.J. Morgan, M.S. Islam, Phase segregation and nanoconfined fluid o<sub>2</sub> in a lithium-rich oxide cathode. *Nature Materials* pp. 1–8 (2024)
- [44] R.S. McWilliams, D.K. Spaulding, J.H. Eggert, P.M. Celliers, D.G. Hicks, R.F. Smith, G.W. Collins, R. Jeanloz, Phase transformations and metallization of magnesium oxide at high pressure and temperature. *Science* **338**(6112), 1330–1333 (2012)
- [45] N. Dubrovinskaia, S. Petitgirard, S. Chariton, R. Tucoulou, J. Garrevoet, K. Glazyrin, H.P. Liermann, V.B. Prakapenka, L. Dubrovinsky, B1-b2 phase transition in mgo at ultra-high static pressure. arXiv preprint arXiv:1904.00476 (2019)
- [46] A.B. Belonoshko, S. Arapan, R. Martonak, A. Rosengren, Mgo phase diagram from first principles in a wide pressure-temperature range. *Physical Review B* **81**(5), 054110 (2010)
- [47] J. Bouchet, F. Bottin, V. Recoules, F. Remus, G. Morard, R. Bolis, A. Benuzzi-Mounaix, Ab initio calculations of the b1-b2 phase transition in mgo. *Physical Review B* **99**(9), 094113 (2019)
- [48] B. Cheng, E.A. Engel, J. Behler, C. Dellago, M. Ceriotti, Ab initio thermodynamics of liquid and solid water. *Proceedings of the National Academy of Sciences* **116**(4), 1110–1115 (2019)
- [49] B. Monserrat, J.G. Brandenburg, E.A. Engel, B. Cheng, Liquid water contains the building blocks of diverse ice phases. *Nature communications* **11**(1), 5757 (2020)

- [50] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **52**(12), 7182–7190 (1981)
- [51] S. Nosé, A unified formulation of the constant temperature molecular dynamics methods. *The Journal of chemical physics* **81**(1), 511–519 (1984)
- [52] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms. *The Journal of chemical physics* **101**(5), 4177–4189 (1994)
- [53] W. Capinski, H. Maris, E. Bauser, I. Silier, M. Asen-Palmer, T. Ruf, M. Cardona, E. Gmelin, Thermal conductivity of isotopically enriched si. *Applied physics letters* **71**(15), 2109–2111 (1997)
- [54] W. Shinoda, M. Shiga, M. Mikami, Rapid estimation of elastic constants by molecular dynamics simulation under constant stress. *Physical Review B* **69**(13), 134103 (2004)
- [55] M. Praiteri. GPTA: Github repository for gpt assisted tasks. <https://github.com/praiteri/GPTA> (2023). Accessed: 2024-03-18
- [56] A. Soper, C. Benmore, Quantum differences between heavy and light water. *Physical review letters* **101**(6), 065502 (2008)
- [57] R.A. DiStasio, B. Santra, Z. Li, X. Wu, R. Car, The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water. *The Journal of chemical physics* **141**(8) (2014)
- [58] A.P. Gaiduk, J. Gustafson, F. Gygi, G. Galli, First-principles simulations of liquid water using a dielectric-dependent hybrid functional. *The journal of physical chemistry letters* **9**(11), 3068–3073 (2018)
- [59] L. Zheng, M. Chen, Z. Sun, H.Y. Ko, B. Santra, P. Dhuvad, X. Wu, Structural, electronic, and dynamical properties of liquid water by ab initio molecular dynamics based on scan functional within the canonical ensemble. *The Journal of Chemical Physics* **148**(16) (2018)
- [60] A. Einstein, *Annalen der physik*. Nr **10**, 891ff (1905)
- [61] E.J. Maginn, R.A. Messerly, D.J. Carlson, D.R. Roe, J.R. Elliot, Best practices for computing transport properties 1. self-diffusivity and viscosity from equilibrium molecular dynamics [article v1. 0]. *Living Journal of Computational Molecular Science* **1**(1), 6324–6324 (2019)

- [62] M. Chen, H.Y. Ko, R.C. Remsing, M.F. Calegari Andrade, B. Santra, Z. Sun, A. Selloni, R. Car, M.L. Klein, J.P. Perdew, et al., Ab initio theory and modeling of water. *Proceedings of the National Academy of Sciences* **114**(41), 10846–10851 (2017)
- [63] R. Mills, Self-diffusion in normal and heavy water in the range 1-45. deg. *The Journal of Physical Chemistry* **77**(5), 685–688 (1973)
- [64] K. Krynicky, C.D. Green, D.W. Sawyer, Pressure and temperature dependence of self-diffusion in water. *Faraday Discussions of the Chemical Society* **66**, 199–208 (1978)
- [65] E.H. Hardy, A. Zygar, M.D. Zeidler, M. Holz, F.D. Sacher, Isotope effect on the translational and rotational motion in liquid water and ammonia. *The Journal of Chemical Physics* **114**(7), 3174–3181 (2001)