

<https://doi.org/10.1038/s41524-025-01550-4>

# Data-efficient construction of high-fidelity graph deep learning interatomic potentials

**Tsz Wai Ko**✉ & **Shyue Ping Ong**✉

Machine learning potentials (MLPs) have become an indispensable tool in large-scale atomistic simulations. However, most MLPs today are trained on data computed using relatively cheap density functional theory (DFT) methods such as the Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation (GGA) functional. While meta-GGAs such as the strongly constrained and appropriately normed (SCAN) functional have been shown to yield significantly improved descriptions of atomic interactions for diversely bonded systems, their higher computational cost remains an impediment to their use in MLP development. In this work, we outline a data-efficient multi-fidelity approach to constructing Materials 3-body Graph Network (M3GNet) interatomic potentials that integrate different levels of theory within a single model. Using silicon and water as examples, we show that a multi-fidelity M3GNet model trained on a combined dataset of low-fidelity GGA calculations with 10% of high-fidelity SCAN calculations can achieve accuracies comparable to a single-fidelity M3GNet model trained on a dataset comprising  $8 \times$  the number of SCAN calculations. This work provides a pathway to the development of high-fidelity MLPs in a cost-effective manner by leveraging existing low-fidelity datasets.

Atomistic simulations are an essential tool in materials science. A key input to atomistic simulations is an interatomic potential (IP) or force field that describes the potential energy surface (PES). Over the past decade, machine learning (ML) has emerged as a powerful approach to constructing IPs<sup>1–3</sup> by learning the PES from a dataset of reference quantum mechanics (QM) calculations<sup>4–10</sup>. Such machine learning potentials (MLPs) typically achieve remarkably improved accuracies in terms of energy and force errors over traditional empirical IPs<sup>11</sup> and can be fitted more readily to more complex chemistries<sup>12,13</sup>. MLPs have enabled studies inaccessible to either QM or empirical IPs, including the study of short-range chemical order and dislocation motion in high-entropy alloys<sup>14–16</sup>, long-time scale diffusion in lithium superionic conductors<sup>17–19</sup>, etc. More recently, a new class of MLPs based on graph deep learning (GLPs) has been developed<sup>20–23</sup>. GLPs utilize a natural graph representation for a collection of atoms, where the atoms are the nodes and the bonds between them are the edges. The information then passes through the graph via a series of message-passing steps, typically modeled using neural networks. The primary advantage of GLPs is their ability to handle arbitrary complex chemistries, without combinatorial explosion in the feature space associated with local environment-based MLPs<sup>24</sup>. Indeed, universal MLPs with coverage of the entire periodic table,

such as the Materials 3-body Graph Network (M3GNet)<sup>20</sup>, Crystal Hamiltonian Graph Network (CHGNet)<sup>21</sup> and Message-passing Atomic Cluster Expansion (MACE)<sup>22</sup>, have been trained using large public datasets of QM structural relaxations from the Materials Project<sup>25</sup> and have broad applications in materials discovery and dynamic simulations.

The biggest bottleneck in MLP construction is the computation of the reference QM data set. For this reason, most MLPs today are fitted using data computed using relatively low-cost semi-local density functional theory (DFT) methods at the generalized gradient approximation (GGA) level, such as the Perdew-Burke-Ernzerhof (PBE) functional<sup>26</sup>. More accurate functionals exist. For instance, the strongly constrained and appropriately normed (SCAN) functional designed to recover all 17 exact constraints presently known for meta-GGA functionals has been demonstrated to yield significantly improved geometries and energies for diverse materials<sup>27</sup>. In quantum chemistry, post-Hartree Fock (post-HF) methods such as CCSD(T) can achieve chemical accuracy in energies (within 1 kcal mol<sup>−1</sup> of experiments). However, such improved accuracies generally come at a much higher computational expense and poorer scaling<sup>28</sup>.

Various approaches have been developed to minimize the computational expense involved in generating training data for high fidelity (hifi)

Aiiso Yufeng Li Family Department of Chemical and Nano Engineering, University of California San Diego, 9500 Gilman Dr, Mail Code 0448, La Jolla, CA, 92093-0448, USA. ✉e-mail: [t1ko@ucsd.edu](mailto:t1ko@ucsd.edu); [ongsp@ucsd.edu](mailto:ongsp@ucsd.edu)

MLPs. One widely used approach is transfer learning (TL), in which a pre-trained MLP constructed using low fidelity (lofi) data, such as DFT and HF, is fine-tuned using a reduced amount of hifi data<sup>29–31</sup>. TL has also been demonstrated in property prediction models<sup>32–35</sup>. A common TL approach is to fix the embedding layers in a GNN, which is expected to improve the transferability to unseen structures and accelerate the convergence of model training. However, this can potentially reduce model accuracy due to a decrease in the number of learnable parameters. Recently, Shoghi et al.<sup>36</sup> proposed a supervised joint multi-domain pretraining scheme for transfer learning in combination with data normalization and temperature sampling to address the imbalance of dataset sizes and range of energies and forces. Such a scheme enables model training across multiple datasets spanning diverse chemical domains within a multi-task framework. Another popular approach is  $\Delta$ —learning, which trains a model to reproduce the difference between the baseline and target methods<sup>37–41</sup>. However,  $\Delta$ —learning requires the same data to be present in both the baseline and target methods, which is a severe constraint for MLP development.

Uniquely among the GLP architectures developed so far, the M3GNet architecture incorporates a global state feature (Fig. 1). Previously, Chen et al.<sup>42</sup> have demonstrated the use of the global state feature for fidelity embedding to train a Materials Graph Network (MEGNet) model to predict band gaps using data from multiple fidelity (mfi) sources (different functionals, experiments). The inclusion of a large quantity of low-fidelity (lofi) PBE band gaps was found to greatly enhance the resolution of latent structural features in MEGNet, which can markedly decrease the mean absolute errors (MAEs) of hifi band gap predictions by 22–45%. Most critically, this improvement in accuracy was achieved without an increase in the amount of hifi training data. Similar works based on mfi band gap prediction for metal-organic framework<sup>43</sup> and optical crystals<sup>44</sup>, formation enthalpy of materials<sup>45</sup> and optical molecular peak<sup>46</sup> have also been reported. Compared to transfer learning or  $\Delta$  — learning, the multi-fidelity approach is more flexible as it does not require a pre-trained baseline model or the same data in different fidelities.

In this work, we demonstrate the application of the mfi approach to construct highly accurate M3GNet GLPs for two model systems—silicon and water. Both silicon and water have been extensively investigated with DFT at different levels of approximations<sup>47–53</sup>, and the reported properties can be qualitatively different depending on the reference methods. For silicon, the radial distribution function (RDF) of liquid silicon computed with the SCAN functional is in better agreement with the experimental RDF due to its clearer discrimination of metallic and covalent bonds<sup>49,54</sup>. Furthermore, recent studies show that an MLP trained on SCAN data outperforms that trained on PBE data in terms of predictions of structural transitions from amorphous to polycrystalline phases at different external pressures<sup>55</sup>, with a  $\Delta$  — learning MLP trained on more accurate random phase approximation (RPA) calculations as a reference<sup>56</sup>. For water, meta-GGA and hybrid functionals such as SCAN<sup>57,58</sup> and SCANO<sup>59</sup> provide more accurate descriptions of hydrogen bonding and van der Waals interactions than GGA functionals such as revised-PBE (RPBE) and BLYP<sup>60–62</sup>. We will demonstrate that with appropriate sampling, an mfi-M3GNet can be extremely efficient in terms of hifi data requirements, requiring 10% coverage of SCAN data to achieve similar accuracies as an M3GNet GLP trained on an extensive SCAN dataset. We also provide comprehensive benchmarks reproducing the structural properties of liquid water and amorphous silicon, the Murnaghan equation of state, and the relaxed geometries of crystals. These results illustrate a data-efficient pathway to the construction of hifi GLPs.

## Results

### Multi-fidelity M3GNet architecture

The M3GNet architecture has already been extensively covered in our previous work, and interested readers are referred to ref. 20 for details. Here, we will discuss only the modifications for the treatment of data of multiple fidelities. For brevity, we will discuss the modifications in the context of datasets comprising only two fidelities, which is the most common scenario

for PES datasets; extension beyond two fidelities is trivial<sup>42</sup>. The fidelity information (lofi RPBE/PBE and hifi SCAN) is encoded as integers (lofi: 0 and hifi: 1) and embedded as a vector in the global state feature input to the M3GNet model (Fig. 1). This fidelity embedding encodes the complex functional relationship between different fidelities and their associated PESs and is automatically learned from the training dataset. The edges, nodes, and global state features are repeatedly updated through a series of interaction blocks that involve sequential three-body interactions and graph convolutions. The resulting atomic feature vectors, which represent the combined information of local chemical environments and fidelity, are fed into gated multi-layer perceptrons for the calculation of atomic energies.

In the following sections, we will present the benchmarks for the performance of mfi M3GNet models for silicon and water against DFT and 1fi M3GNet models. Figure 2 illustrates the selection of lofi PBE/RPBE and hifi SCAN data for training, validation and test sets for constructing 10%-mfi SCAN and 80%-1fi SCAN models. As there is large overlap in structures between the lofi and hifi SCAN datasets<sup>55</sup>, only 80% of the lofi data combined with 10% of the hifi data selected from structures within the 80% lofi data were used to train the mfi models. The 20% of SCAN data points, where the structures did not appear in both the lofi and hifi training sets, were divided into equal validation and test sets. This process avoids the possibility of including the same structures in the training and the validation/test sets, which allows for a robust comparison of different sampling sizes and techniques.

### Silicon

In this section, we develop mfi and 1fi M3GNet IPs for silicon and benchmark their performance in reproducing not only basic energy and forces, but also structural properties of crystalline polymorphs and amorphous silicon as well as derived bulk properties such as the bulk modulus.

**Convergence with percentage of hifi data points.** Figure 3 shows the convergence of the energy and force errors of the mfi M3GNet models for silicon with respect to the percentage of hifi (SCAN) data. Here, the Dimensionality-Reduced Encoded Cluster with Stratified (DIRECT) sampling approach developed by the current authors was used to ensure robust coverage of the configuration space regardless of the size of the SCAN dataset. With only 10% of SCAN data, the mfi M3GNet model already achieves comparable energy and force MAEs as the 1fi M3GNet potential trained on the 80% of SCAN data. Further decreases in MAEs with the increasing amount of SCAN training data are relatively small. Furthermore, we note that the MAEs of the 1fi M3GNet models are significantly higher than that of the mfi M3GNet models with comparable SCAN training data sizes. For example, the 10%-1fi M3GNet model has energy and force MAEs of 0.062 eV atom<sup>-1</sup> and 0.127 eV Å<sup>-1</sup>, respectively, compared to the 10%-mfi M3GNet model energy and force MAEs of 0.032 eV atom<sup>-1</sup> and 0.100 eV Å<sup>-1</sup>, respectively. Even for 50% SCAN data, the mfi model significantly outperforms the 1fi model. Hence, the inclusion of the large quantity of lofi PBE data improves the quality of hifi SCAN predictions.

**Effect of sampling.** To explore the effect of sampling of hifi data on the performance of mfi M3GNet models, we trained 10%-mfi M3GNet models, with the 10% SCAN data sampled (a) from a narrow region (denoted as “10% mfi-narrow”), (b) in random manner (denoted as “10% mfi-random”), and (3) DIRECT sampling. The “narrow” sampling only includes primarily diamond crystal structures with distortions, strains and defects and excludes amorphous and other structures with more diverse local environments. Figure 4 illustrates the coverage of the three different sampling techniques using the first two principal components (PCs) of the encoded structure features, which were obtained using the pretrained M3GNet formation energy model as a structure featurizer.

From Fig. 5, it can be observed that the 10%-mfi-DIRECT and 10%-mfi-random models exhibit similar test MAEs in energies and forces, but the

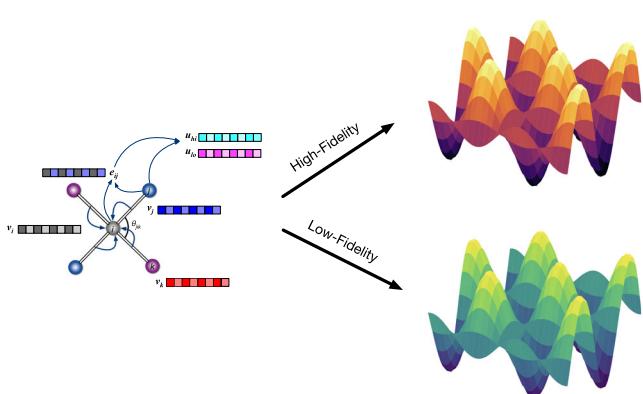
10%-mfi-narrow model exhibits extremely high test MAEs. These results illustrate that the 10%-mfi-narrow model extrapolates poorly beyond the training domain for the majority of the validation and test structures. This can be supported by the plot of the first two principal components of the training, validation, and test structures, as provided in Fig. S1. The validation errors, which share the same observations as the test errors, are plotted in Fig. S2. To further evaluate the accuracy of 10%-mfi-random and 10%-mfi-DIRECT, we selected 297 bulk structures with great structural diversity from GAP-18 dataset<sup>56</sup>, which is used for constructing general-purpose IPs of silicon, to compare the error of energies and forces with respect to DFT-SCAN. The MAEs in energies and forces of 10%-mfi-random are 0.114 eV atom<sup>-1</sup> and 0.088 eV Å<sup>-1</sup>, respectively, while those of the 10%-mfi-DIRECT are 0.040 eV atom<sup>-1</sup> and 0.091 eV Å<sup>-1</sup>, respectively. This clearly shows that the GLP achieves better accuracy with respect to unseen structures with DIRECT sampling.

**Equation of state of silicon polymorphs.** Figure 6 a shows the calculated Murnaghan equation of state (EOS) for four silicon polymorphs. The EOS curves obtained from the 10%-mfi-random and 10%-mfi-DIRECT M3GNet models generally match DFT and 80%-1fi%-SCAN results in terms of the curvature and the energy minimum of crystals, even though the majority of the scaled crystals are outside the training domain (see Fig. 6b). We further note that they are qualitatively different from the 80%-1fi-PBE results in terms of the energetics of crystals. For

instance, 80%-1fi-PBE underestimates the energy difference between the  $\beta$ -Sn and diamond phases compared to 80%-1fi-SCAN, 10%-mfi-random and 10%-mfi-DIRECT, which is consistent with previous DFT results<sup>54,63</sup>. Table 1 summarizes the equilibrium normalized volumes, minimum energies and bulk moduli of crystals extracted from the fitted Murnaghan equation of state. Interestingly, the 10%-mfi-narrow M3GNet has the smallest error in minimum energy and bulk modulus for the common diamond structure among all the models, but has much larger errors for the other silicon polymorphs. This is attributed to the comprehensive coverage of diamond structures and very limited coverage of other polymorphs by the 10%-mfi-narrow GLP. In contrast, the 10%-mfi-random and 10%-mfi-DIRECT M3GNet GLPs provide considerably more diverse coverage in configuration space, leading to markedly improved prediction errors (especially for the bulk modulus) for the non-diamond structures, but with higher prediction errors for the diamond structure itself.

**Pressure-induced phase transition of amorphous silicon.** Amorphous silicon is known to undergo pressure-induced phase transitions. Here, isothermal compression MD simulations were performed on amorphous silicon cells containing 1728 atoms (see Methods section). Figure 7a shows the volume change of amorphous silicon under a uniform pressurization rate of 0.05 GPa ps<sup>-1</sup> with the various M3GNet GLPs. It can be observed that MD simulations using both 10%-mfi-random and 10%-mfi-DIRECT successfully reproduce the phase transitions observed in 80%-1fi-SCAN and previous work using MD simulations with a Gaussian Approximation Potential (GAP) on 100,000 amorphous silicon atoms<sup>55</sup>. For instance, a volume collapse to very-high-density amorphous (VHDA) states is observed at ~ 15 GPa, and recrystallization is observed with further compression. MD simulations with the 10%-mfi-narrow M3GNet GLP, on the other hand, predict the structural collapse to occur at a much lower pressure (~ 7 GPa), and no recrystallization was observed with further compression. From Fig. 7b, we can observe a sharp increase in potential energy at the phase transition pressure. This is followed by a small drop in potential energy with recrystallization in the 10%-mfi-random and 10%-mfi-DIRECT simulations, while no such drop was observed in the 10%-mfi-narrow simulations. Additionally, we note that the 80%-1fi-PBE model predicts the phase transition around 8 GPa, which is noticeably lower than those predicted by other SCAN GLPs.

Figure 8 shows that the local environments of structures extracted from simulations under 20 GPa at 500 K using all three GLPs. From Fig. 8b, it can be clearly observed that the structures extracted from the 10%-mfi-random and 10%-mfi-DIRECT simulations exhibit a clearly crystalline nature with lower coordination numbers ( $7 \leq N_c \leq 8$ ), while the structure extracted from the 10%-mfi-narrow simulation appears to remain amorphous with a greater percentage of highly coordinated Si atoms ( $N_c > 8$ ). This poor performance of the 10%-mfi-narrow GLP can be attributed to the underestimation of short-range repulsive forces as can be seen in the relatively flat curvature of EOS curves (Fig. 6b) and the underestimation of the energy of high energy structures (Fig. 5a). MD simulations with the 80%-1fi-PBE also do not result



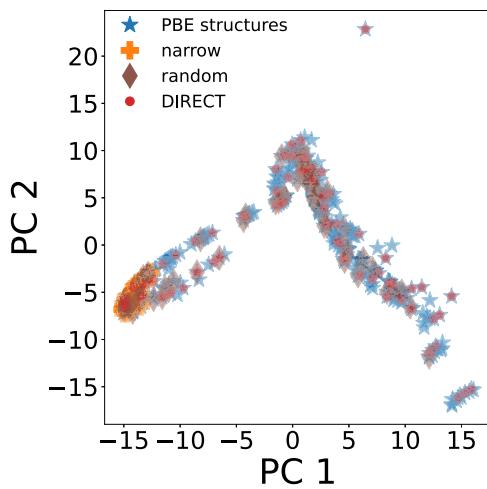
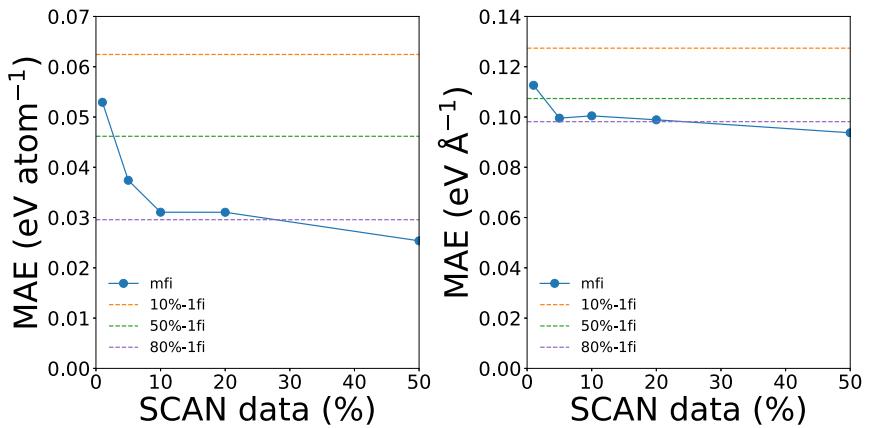
**Fig. 1 | Efficient construction of high-fidelity M3GNet interatomic potential.** Every structure is represented as a graph, considering atoms as nodes, bonds as edges, and the fidelity of reference methods as global state attributes. Atomic numbers and radial basis functions are embedded in latent node features and edge features, respectively. The fidelity of each data point is encoded as an integer (e.g. 0: low-fidelity, 1: high-fidelity in this work) to distinguish different levels of reference electronic structure methods. These integers are then embedded into the corresponding learnable state features. The information between nodes, edges and global state features are iteratively exchanged via sequential three-body interactions and graph convolutions. The resulting atomic features are fed into the readout layer yielding atomic energies. The summation of atomic contributions is equal to the potential energy of the system computed with the given fidelity method.



**Fig. 2 | Training data selection protocol for 10%-mfi-SCAN and 80%-1fi-SCAN models.** In the pie charts, similar positions correspond to similar structures. **a** The training set (dark blue regions) for the 10%-mfi SCAN model comprises 10% of the hifi SCAN data and 80% of the lofi PBE/RPBE data. The validation (orange region)

and test (green regions) sets are constructed by splitting equally the 20% of the hifi SCAN data that do not appear in either the PBE/RPBE or SCAN training data. **b** For the 80%-1fi SCAN models, 80% of SCAN data were used for training. The same structures were used for validation and test sets as the mfi SCAN model.

**Fig. 3 | Performance of multi-fidelity M3GNet potentials for silicon.** The test mean absolute errors (MAEs) of mfi M3GNet GLPs trained with different percentages of SCAN data for silicon. The dashed lines represent the MAEs of the 1fi M3GNet potential trained on 10%, 50% and 80% of SCAN data for reference.



**Fig. 4 | Coverage of the first two principal components (PCs) of the M3GNet-encoded latent structure features for the three different sampling techniques.** The narrow sampling approach only samples structures from the far left of the PC space. The random sampling approach significantly improves the coverage of the entire space, but structures in the extrema of the latent space are missed. Finally, DIRECT sampling ensures coverage of the entire latent space, including extreme regions such as an isolated atom and highly distorted structures.

in the formation of crystalline structures, highlighting the qualitative differences between the GLPs trained on PBE and SCAN functionals.

Interestingly, the 10%-mfi-random structure contains about 40% of atoms in a sixfold-coordinated  $\beta$ -Sn-like phase and 30% as sevenfold coordinated and eightfold-coordinated simple hexagonal (sh)-like phase. The 10%-mfi-DIRECT structure contains around 65% of atoms as eightfold-coordinated sh crystallites, 25% in sevenfold coordination, and the remaining atoms in a sixfold-coordinated  $\beta$ -Sn-like phase. The 80%-1fi-SCAN structures contain more than 75% of the atoms in the  $\beta$ -Sn-like crystal. The larger variations in the coordination distribution between the different SCAN GLPs can be attributed to the very small energy difference between the  $\beta$ -Sn and sh crystals, as seen in the EOS results, as well as sensitivity in the coordination analysis. Nevertheless, the main observation is that the mfi M3GNet models successfully reproduce the pressure-induced phase transition of disordered silicon observed in the 80%-fi-SCAN and DFT-SCAN simulations, unlike the 80%-1fi-PBE and 10%-mfi-narrow models.

## Water

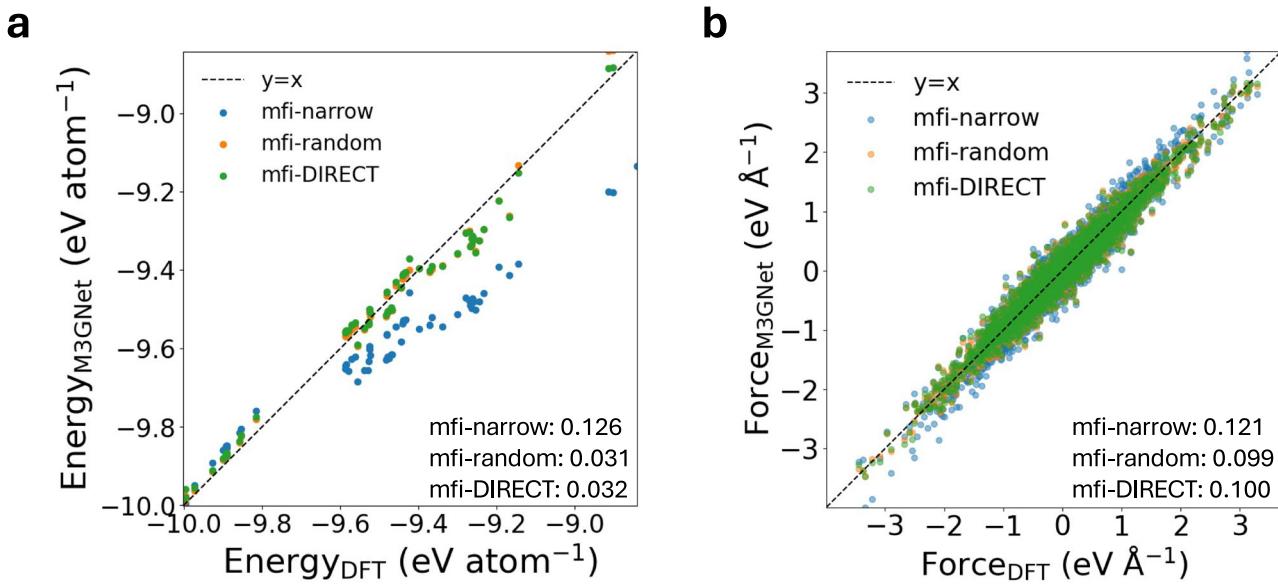
In this section, we develop mfi and 1fi M3GNet GLPs for water. Given the relatively similar performances of DIRECT and random sampling for the

mfi M3GNet GLPs for silicon, only random sampling was used to select hifi training data for water.

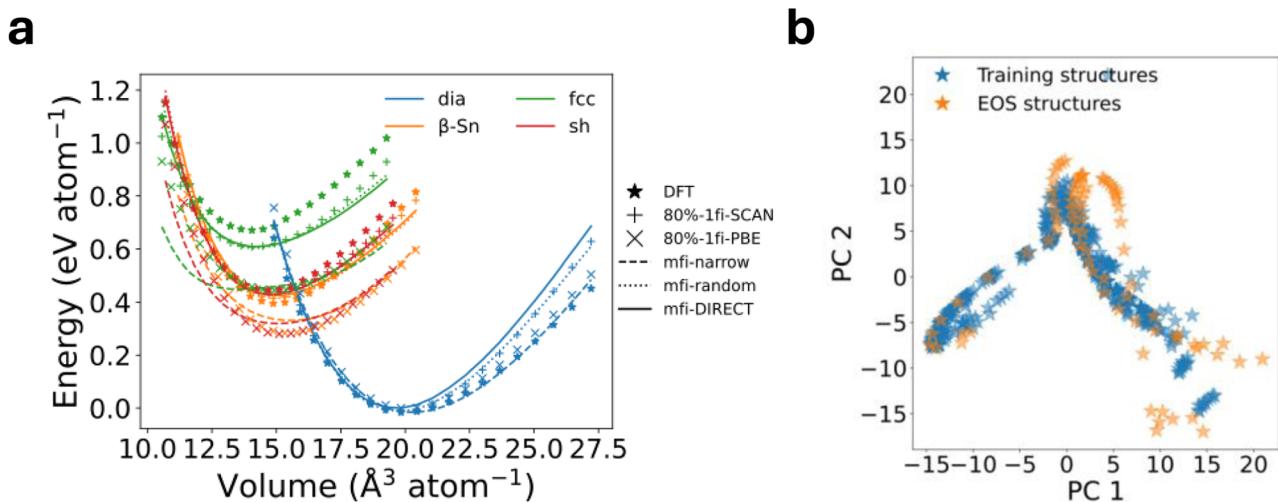
**Energy and force errors.** Figure 9 shows the convergence of the energy and force MAEs of mfi M3GNet models for water with respect to the percentage of hifi (SCAN) data. The parity plots of energies and forces for all GLPs are provided in Figs. S4, S5. The 10%-mfi M3GNet has energy and force MAEs of  $0.2 \text{ meV atom}^{-1}$  and  $0.014 \text{ eV } \text{\AA}^{-1}$ , respectively, which is less than half those of the 10%-1fi model and outperforms even the 80%-1fi M3GNet GLP.

**Structure of liquid water.** NPT MD simulations were performed on a model containing 64 formula units of  $\text{H}_2\text{O}$  using the 10%-mfi and 1fi M3GNet IPs. Figure 10 compares the radial distribution functions (RDFs) of liquid water at 300K obtained using the mfi M3GNet GLPs and prior SCAN AIMD results. The 10%-mfi SCAN model is in agreement with 80%-1fi SCAN in terms of the position and magnitude of peaks. Notably, 10%-mfi SCAN predicted a lower first peak in the RDF of oxygen-oxygen, which better aligns with SCAN-AIMD results<sup>64</sup> compared to 80%-1fi SCAN. Similarly, the 10%-mfi SCAN generally matches the shape of the angular distribution function (ADF) of O-O-O with 80%-1fi SCAN as shown in Figure 11. Notably, the magnitude of peaks in the ADF predicted by 10%-mfi SCAN is even in better agreement with SCAN-AIMD<sup>64</sup>. All these analyses show that the mfi M3GNet model enables the efficient construction of hifi MLPs with state-of-the-art accuracy using a fraction of data points, where the lofi dataset provides additional coverage in configuration space as shown in the first two PCs of reduced structural latent features from Fig. S6. The RDFs and ADFs at other temperatures can be also found in Figs. S7, S8, respectively.

**Structure of ice polymorphs.** We also compared the performance of 10%-1fi-SCAN, 80%-1fi-SCAN and 10%-mfi-SCAN M3GNet with DFT-SCAN by conducting geometry relaxation on various small ice crystals ( $N_{\text{atoms}} < 100$ ) reported in ref. 65, which covers the majority of experimentally known phases. To measure the similarity between two crystal structures, we employed CrystalNN algorithm<sup>66</sup> to compute the structural fingerprint of periodic structures based on the Voronoi algorithm combined with the solid angle weights to determine the probability of various coordination environments. The smaller fingerprint distance indicates the higher similarity between the two crystal structures. Figure 12 shows the distribution of fingerprint distance among three different M3GNet models. The 10%-mfi SCAN generally achieves the lowest fingerprint distance between DFT-relaxed and M3GNet-relaxed structures among the three models. The main reason is that the lofi training set covers several crystal structures, which allows the mfi model to provide a more informative representation of local environments for other crystals resulting in better agreement with DFT-SCAN. In contrast, the other two



**Fig. 5 | Test mean absolute errors (MAEs) of energies and forces for M3GNet GLPs for silicon.** Parity plots of (a) energies and (b) forces predicted by 10%-mfi-narrow, 10%-mfi-random and 10%-mfi-DIRECT. The numbers indicate the MAEs of energies in  $\text{eV atom}^{-1}$  and forces in  $\text{eV \AA}^{-1}$ .



**Fig. 6 | Equation of state of silicon polymorphs from 1fi and mfi M3GNet models.** a Calculated DFT and 1fi/mfi M3GNet energies of different silicon crystals as a function of the normalized volume. All predicted SCAN energies are relative to the lowest DFT energy of the diamond structure, while the zero of predicted PBE

energies is shifted to the lowest energy prediction. b Visualization of the first two principal components of the encoded features using the M3GNet formation energy model for isotropically deformed crystals ("EOS structures") and training structures.

models only contain the liquid water structures for training and therefore the extrapolation limits the accuracy of both models. The main reason is that the lofi training set covers various crystal structures. This enables the mfi model to offer a more informative representation of local environments for other crystals, resulting in better agreement with DFT-SCAN. In contrast, the training set for 1fi models only includes liquid water structures for training, which provides less accurate predictions for crystals.

## Discussion

In the preceding sections, we have demonstrated that a multi-fidelity graph network architecture provides a data-efficient pathway to construct GLPs for higher levels of theory. From Figs. 3 and 9, we estimate that an mfi M3GNet GLP trained with a mfi dataset containing just 5–10% of SCAN data can achieve similar energy and force accuracies as a 1fi M3GNet GLP

trained on a much larger 80% hifi SCAN dataset. This represents a savings of ~ 87.5 to ~ 93.75% in the amount of hifi training data that needs to be computed. We note that lofi GGA-based data are often already available and the computational cost of a hifi SCAN calculation is several times higher than that of a lofi GGA calculation. Hence, the computational cost of constructing a hifi GLP is dominated by the cost of generating the hifi dataset.

Figure 13 compares the t-distributed stochastic neighbor embedding (T-SNE) analysis of the atomic features extracted from the silicon test set from a mfi and 1fi M3GNet trained on a 10% DIRECT SCAN data. Consistent with prior analyses, the mfi M3GNet model exhibits better separation between structures of high and low atomic energies in the latent space compared to the 1fi M3GNet model. We surmise that it is this improvement in latent representation with the inclusion of a large lofi PBE dataset that led to the improvement in the performance of the mfi M3GNet model.

**Table 1 | Equations of state for silicon polymorphs**

Crystals	$E_0$ (eV atom $^{-1}$ )	$V_0$ (Å $^3$ atom $^{-1}$ )	$B$ (GPa)
<b>Diamond</b>			
1fi-80%-SCAN	-10.003 (0.003)	19.798 (0.99%)	118.450 (24.21%)
1fi-80%-PBE	-5.399	20.073 (0.38%)	105.64 (10.77%)
mfi-narrow	-10.008 (0.002)	20.286 (1.45%)	102.092 (7.05%)
mfi-random	-10.000 (0.006)	19.751 (1.23%)	118.101 (23.84%)
mfi-DIRECT	-9.989 (0.017)	19.629 (1.84%)	125.178 (31.26%)
<b>beta-Sn</b>			
1fi-80%-SCAN	-9.577 (0.018)	14.993 (0.20%)	106.952 (7.29%)
1fi-80%-PBE	-5.117	15.484 (3.48%)	99.797 (13.49%)
mfi-narrow	-9.669 (0.074)	15.403 (2.94%)	79.528 (31.06%)
mfi-random	-9.577 (0.018)	15.052 (0.59%)	98.365 (14.73%)
mfi-DIRECT	-9.568 (0.027)	15.074 (0.74%)	97.788 (15.23%)
<b>FCC</b>			
1fi-80%-SCAN	-9.387 (0.069)	14.091 (0.85%)	89.401 (7.01%)
1fi-80%-PBE	-4.957	14.879 (6.48%)	79.908 (16.89%)
mfi-narrow	-9.553 (0.235)	13.794 (1.28%)	48.805 (49.24%)
mfi-random	-9.390 (0.072)	14.194 (1.58%)	91.839 (4.48%)
mfi-DIRECT	-9.386 (0.068)	14.234 (1.87%)	82.718 (13.97%)
<b>sh</b>			
1fi-80%-SCAN	-9.554 (0.008)	14.838 (0.82%)	99.725 (9.20%)
1fi-80%-PBE	-5.120	15.246 (3.59%)	97.331 (11.38%)
mfi-narrow	-9.679 (0.133)	14.897 (1.22%)	72.821 (33.69%)
mfi-random	-9.563 (0.016)	14.823 (0.72%)	94.846 (13.64%)
mfi-DIRECT	-9.563 (0.017)	14.853 (0.93%)	97.944 (10.82%)

The numbers reported are fitted values of the minimum energy ( $E_0$ ), equilibrium volume ( $V_0$ ) and bulk modulus ( $B$ ) with Murnaghan equation of state. The numbers inside the parentheses indicate the mean absolute error for  $E_0$  and absolute percentage error for  $V_0$  and  $B$  with respect to DFT.

As noted in the introduction, the mfi approach offers several advantages over transfer learning (TL) and  $\Delta$ -learning. The mfi approach does not require an initial baseline model. It retains full model flexibility unlike the TL approach. And it does not require the same data to be generated in different fidelities like  $\Delta$ -learning. Most importantly, a single mfi GLP can potentially be constructed for >2 fidelities, extracting the maximum value from scarce materials PES data. For comparison, we also trained two models using TL with the same 10% and 50% SCAN data (10%-TL and 50%-TL SCAN models, respectively). The same PBE training data was used for constructing the pretrained MLP, and the model weights in the embedding layer were fixed during the fine-tuning with SCAN data. The energy and force test MAEs of the 10%-TL SCAN model are 0.035 eV atom $^{-1}$  and 0.101 eV Å $^{-1}$ , respectively, and those of the 50%-TL SCAN model are slightly lower at 0.031 eV atom $^{-1}$  and 0.097 eV Å $^{-1}$ , respectively. In both cases, the corresponding mfi M3GNet models achieve slightly lower test MAEs in both energies and forces with the same amount of lofi and hifi data. This is likely due to the greater flexibility in the mfi model to learn the complex relationships between the lofi and hifi PESs, compared to the TL approach.

Though we have demonstrated the multi-fidelity concept using two relatively simple model systems (silicon and water), the impact of this work goes beyond custom MLPs for specific chemistries. For instance, a major area of active research is in the development of universal GLPs (also so-called “foundational” MLPs) with broad coverage of the periodic table of the elements. Generating a robust training dataset, even with standard GGA methods, for such universal GLPs is a major challenge. Using the multi-fidelity approach outlined in this work, we anticipate that training universal GLPs based on state-of-the-art meta-GGAs such as the SCAN functional can be made noticeably more data-efficient. This approach can also be extended

to other GLP architectures through appropriate modification to include a global fidelity embedding feature and additional message-passing operations.

## Methods

### M3GNet architecture

The Materials 3-body Graph Network (M3GNet) model architecture has been extensively covered previously, and interested readers are referred to ref. 20 for details. Only the major parameter settings used in this work will be discussed here. The distance cut-off that defines the edges  $e_{ij}$  and the angles between two edges  $e_{ij}$  and  $e_{ik}$  was chosen as 5 Å. The lengths of the node, edge, and global state feature vectors were set at 64, 64, and 16, respectively. The node feature vector is a learned embedding based on the atomic number of each atom, while the global state feature vector is a learned embedding based on the fidelity of the data. The bond angles were expanded using a spherical harmonic basis with  $m = 3$  and  $l = 0$ . The number of message-passing blocks (comprising sequential three-body interactions, bond, atom and global state updates) is set to 3. The numbers of training parameters in the 1fi and mfi M3GNet models for silicon are 221,597 and 325,757, respectively, while those for water are 221,661 and 325,821, respectively. Each dense layer contains 64 output neurons that are fully connected from the input neurons. The update functions are modeled using gated multi-layer perceptrons (MLPs) that contain two layers with 64 neurons for each layer. Each updated atomic feature vector is fed into a 64-64-1 gated MLP to predict the corresponding atomic energy. The total energy is expressed as the sum of the atomic energies, while the forces and stresses are calculated by taking the partial derivatives of the total energy with respect to the atomic positions and lattice vectors.

### Potential training

All model parameters were optimized using the Adam algorithm<sup>67</sup>. The initial learning rate was set at 0.001. The cosine scheduler was used to gradually reduce the learning rate to 0.01 of the original value in 100 epochs. Early stopping of the model training was triggered when the validation loss did not reduce for 200 epochs. A batch size of 16 and 8 was used in the model training for silicon and water, respectively. However, a smaller batch size of 4 was used for mfi M3GNet training for water to avoid training crashes caused by excessive memory assumption. The loss function for the potential training was given by

$$L = \text{MSE}\left(\frac{E_{\text{total}}^{\text{M3GNet}}}{N_{\text{atoms}}}, \frac{E_{\text{total}}^{\text{DFT}}}{N_{\text{atoms}}}\right) + w_f \text{MSE}\left(F_{i,\alpha}^{\text{M3GNet}}, F_{i,\alpha}^{\text{DFT}}\right), \quad (1)$$

where MSE represents the mean squared error loss function,  $E_{\text{total}}$  denotes the total energy of the system,  $N_{\text{atoms}}$  denotes the number of atoms, and  $F_{i,\alpha}$  denotes the force component acting on atom  $i$  along the  $\alpha = x, y, z$  axis.

$E_f$  is obtained by subtracting the elemental reference energies  $E_{\text{elem}}$  from the total energy, via the following expression:

$$E_i = E_{\text{total}} - \mathbf{n}_{\text{elem}} \cdot \mathbf{E}_{\text{elem}} \quad (2)$$

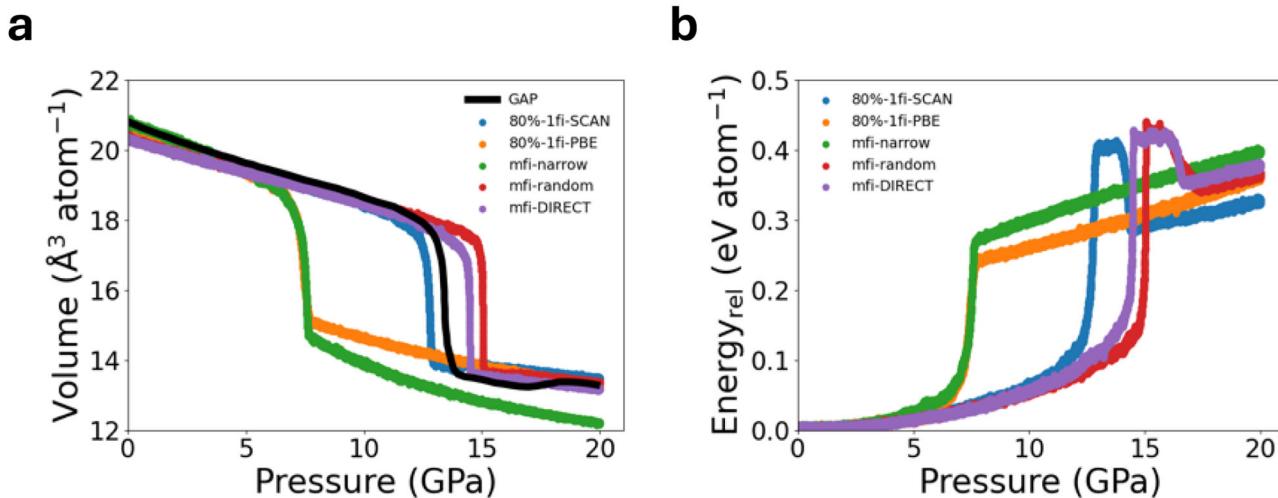
where  $\mathbf{n}_{\text{elem}}$  is a row vector representing the number of each element in the structure. The elemental reference energies  $E_{\text{elem}}$  were fitted using linear regression of the target total energies, using the following expression:

$$\mathbf{E}_{\text{elem}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{E}_{\text{total}}, \quad (3)$$

where  $\mathbf{E}_{\text{total}}$  is the matrix of total energies with dimensions  $N_{\text{struct}} \times N_{\text{elem}}$ ,  $N_{\text{struct}}$  is the total number of structures,  $N_{\text{elem}}$  is the total number of elements, and  $\mathbf{A}$  is the composition matrix obtained by stacking  $\mathbf{n}_{\text{elem}}$  for all structures.

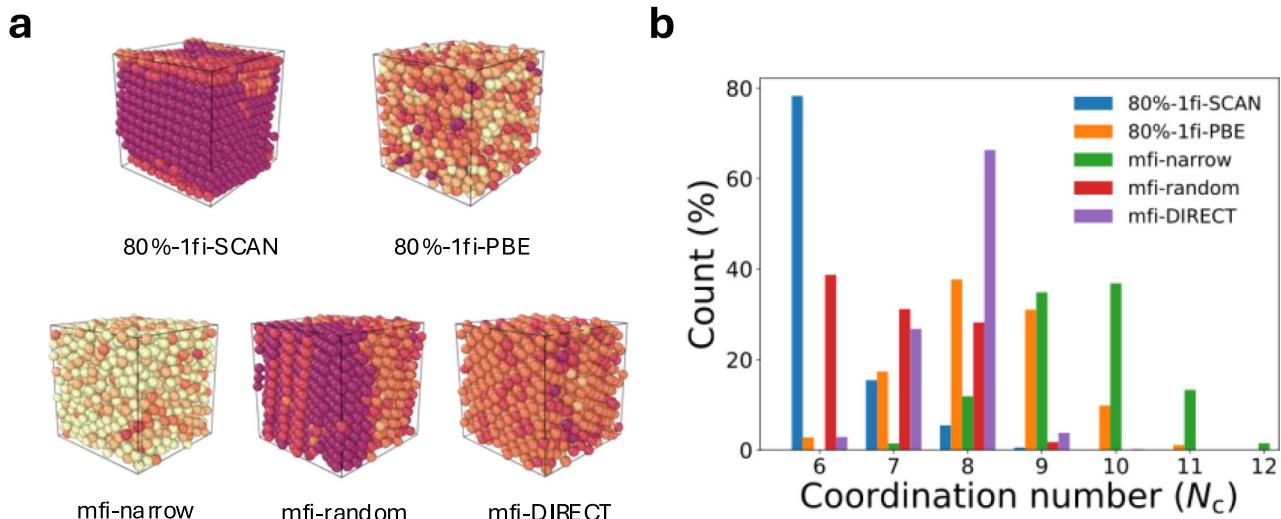
To recover the total energy from the atomic energies, the following expression is used:

$$E_{\text{total}} = \sum_i^{N_{\text{atom}}} \sigma E_i + \mathbf{n}_{\text{elem}} \cdot \mathbf{E}_{\text{elem}} \quad (4)$$



**Fig. 7 | Structural transition of amorphous silicon induced by external pressurization.** The evolution of (a) normalized volume and (b) potential energy of amorphous silicon with 1728 atoms is simulated by increasing the pressurization

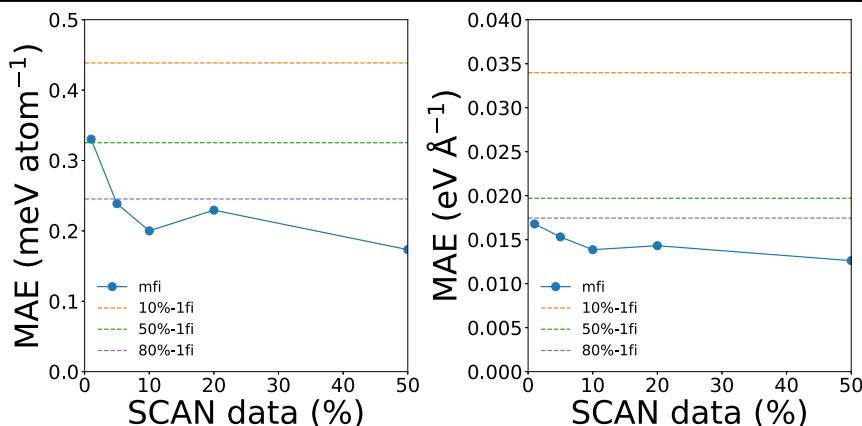
rate = 0.05 GPa ps $^{-1}$  from 0 GPa to 20 GPa. The energies are relative to the lowest energy of snapshots during pressurization. The simulated volume using Gaussian approximation potential (GAP) is reproduced from ref. 55.

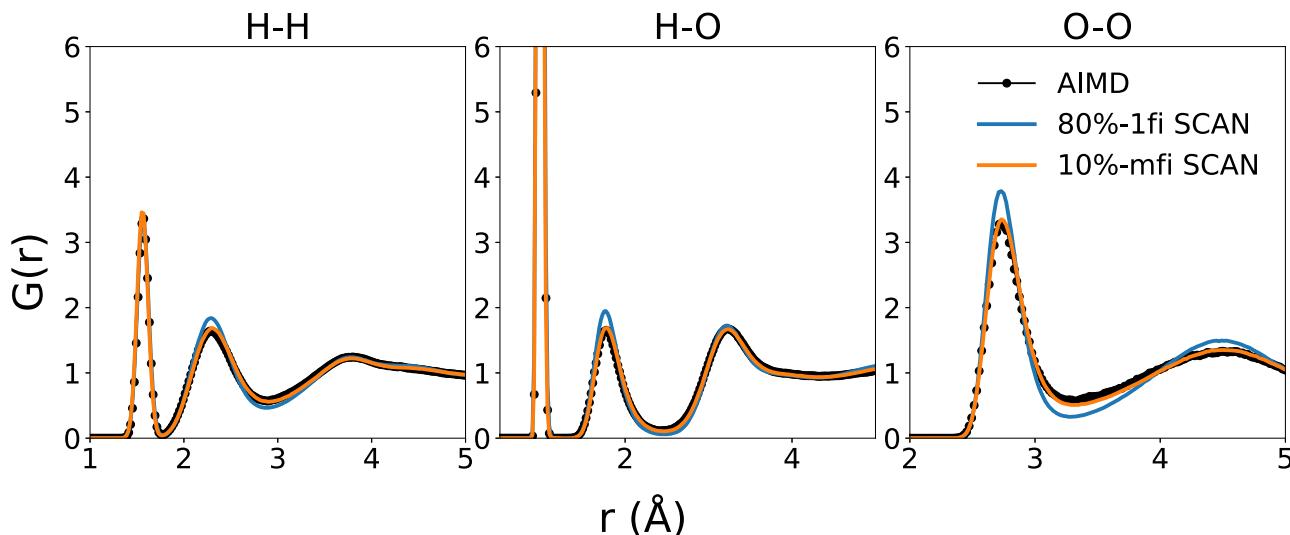


**Fig. 8 | Coordination analysis of disordered silicon at 20 GPa and 500 K.** a The snapshot of isothermal compressed silicon structures. The heat map indicates the coordination number with a spatial cutoff = 2.85 Å. b The statistic of coordination

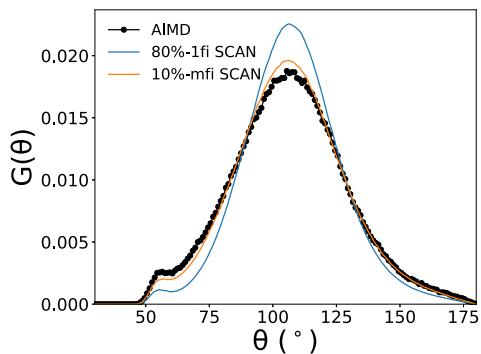
number is obtained from 100 ps NPT simulations. The structures with color coding were visualized using Ovito<sup>78</sup>.

**Fig. 9 | Performance of mfi M3GNet potentials for water.** The test mean absolute errors (MAEs) of mfi M3GNet GLPs trained with different percentages of SCAN data for silicon. The dashed lines represent the MAEs of the 1fi M3GNet potential trained on 10%, 50% and 80% of SCAN data for reference.





**Fig. 10 | Radial distribution function (RDF) of liquid water.** The RDFs were extracted from NPT simulations with the 80%-1fi and 10%-mfi M3GNet GLPs at 300K and 1 atm. The SCAN-AIMD results were obtained from ref. 64.



**Fig. 11 | Angular distribution function of liquid water.** The O-O-O triplet angular distribution function of liquid water was computed with 80%-1fi SCAN and 10%-mfi SCAN at 300K and 1 atm. The AIMD results were obtained from ref. 64.

where the scaling factor  $\sigma$  is calculated by taking the inverse of the root mean square of all atomic force components from the training set.

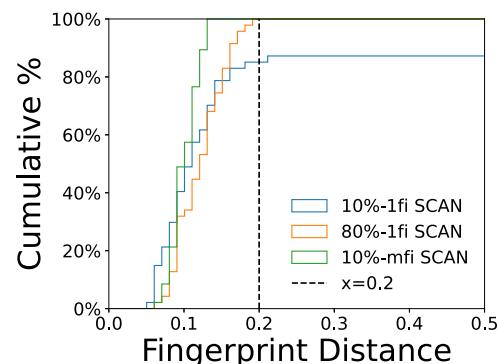
#### DFT static calculations

All DFT calculations were performed using VASP (version: 6.3.2) with spin polarization. The projector augmented wave method<sup>68</sup> was employed to describe core and valence electrons with pseudopotentials. The electronic convergence criterion was set at  $10^{-5}$  eV. For the Murnaghan equation of state for silicon crystals, the energy cut-off and k-spacing were chosen to be 800 eV and  $0.2 \text{ \AA}^{-1}$ .

**Geometry relaxation.** For the high-throughput geometry relaxation of ice crystals with DFT, the ionic and electronic convergence were set to  $0.05 \text{ eV \AA}^{-1}$  and  $10^{-4}$  eV, respectively. The conjugate gradient algorithm<sup>69</sup> was employed to update the ionic positions. Moreover, the energy cutoff and k-spacing were chosen to be  $0.35 \text{ \AA}^{-1}$ , respectively. The same force threshold was applied to the M3GNet geometry relaxation via ASE interface<sup>70</sup> using the LBFGS<sup>71</sup> algorithm.

#### MD simulations

All MD simulations were performed with LAMMPS<sup>72,73</sup> (version: 2Jun2022). A Nose-Hoover thermostat was employed to control temperature and pressure for NPT simulations. The time step was chosen to be 1 fs

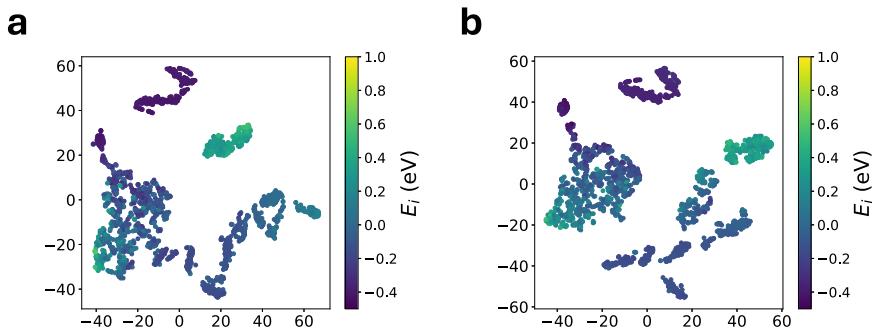


**Fig. 12 | Distribution of structural fingerprint distances for ice crystals relaxed with M3GNet and SCAN.** Cumulative histogram of calculated CrystalNN fingerprint distance between different M3GNet- and SCAN-relaxed ice structures. All these crystals are obtained from ref. 65. The legend indicates the fraction of data points used for training. A smaller fingerprint distance means greater similarity between two relaxed structures and the vertical dashed line is used to visualize the fraction of structures that fall below 0.2 in terms of fingerprint distance.

and the damping constants for temperature and pressure were set to 0.1 ps and 1.0 ps, respectively.

Amorphous silicon structures were generated by the melt-and-quench approach. The initial structure was obtained by random displacement of 1728 silicon atoms in the cubic cell with a density of  $2.56 \text{ g/cm}^3$  and then relaxed under a force threshold of  $10^{-3} \text{ eV \AA}^{-1}$ . The relaxed structure was equilibrated at 1500 K for 110 ps and then cooled down to 500 K. A fast quenching rate of  $10^{13} \text{ K s}^{-1}$  was used for quenching from 1500 K to 1250 K and from 1050 K to 500 K, while a slower quenching rate of  $10^{11} \text{ K s}^{-1}$  was used for quenching from 1250 K and 1050 K. The simulated temperature and the corresponding potential energy of silicon during the melt-and-quench are provided in Fig. S3. After an additional 50 ps equilibration at 500 K, the amorphous silicon was isotropically compressed at a rate of  $0.05 \text{ GPa (ps)}^{-1}$  from 0 GPa to 20 GPa. The snapshot of silicon structures at 20 GPa was further equilibrated for 20 ps at 500 K. Finally, their structural properties were obtained from 100-ps NVE simulations. For the NPT simulations of liquid water, an initial cubic cell containing 64  $\text{H}_2\text{O}$  molecules with  $\rho = 1.0 \text{ g cm}^{-3}$  was equilibrated for 50 ps under 1 atm. The density and structural properties were extracted from a production run of 0.5 ns.

**Fig. 13 | Two-dimensional t-distributed Stochastic neighbor embedding of atomic features for silicon.** The atomic features for the test set of silicon were extracted from (a) 10%-1fi-DIRECT, and (b) 10%-mfi-DIRECT.



### DIRECT sampling

The DIimensionally-Reduced Encoded Clusters with sTratified (DIRECT) sampling workflow has already been detailed in previous work<sup>74</sup>. First, a pre-trained M3GNet Materials Project formation energy model was used to encode structures into a 128-element vector. Next, principal components analysis (PCA) was performed on the encoded structure features, and the first six PCs with eigenvalues greater than 1 (Kaiser's rule) were selected as a reduced-dimensionality representation. Finally, the balanced iterative reducing and clustering using hierarchies (BIRCH) algorithm<sup>75</sup> was used to divide all encoded structures into  $n$  clusters and  $k$  structures are sampled from each cluster for static DFT calculations. In this work, the values of  $n$ ,  $k$ , and the threshold used to construct the 1fi and mfi M3GNet GLPs of silicon are given in Table S1.

### Training data selection

For silicon, 80% of the training structures were selected from a total of 608 structures, with 70% randomly sampled and 10% DIRECT sampled to ensure the training set covered mostly diverse structures while keeping sufficient structural diversity for meaningful validation and testing. The hifi SCAN data points for training the 10%-mfi-narrow M3GNet model were selected based on PCA, where their first and second PC values were lower than -12.5 and -2.5, respectively. As for water, all training structures were randomly sampled for constructing 1fi and mfi M3GNet GLPs.

### T-SNE visualization

The t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis was performed using the OpenTSNE library<sup>76,77</sup>. The cosine distance was used as the distance metric. All other settings were kept at the default values.

### Data availability

The RPBE water dataset is available at <https://zenodo.org/records/2634098>, while the SCAN water dataset is available in ref. 64. Both PBE and SCAN silicon datasets are given in <https://doi.org/10.5281/zenodo.4174139>. All data necessary to replicate the training results will be publicly available upon acceptance of the manuscript.

### Code availability

All codes used in this work are available in public Github repositories. The M3GNet implementation is available at <https://github.com/materialsvirtuallab/m3gnet>. The DIRECT sampling is implemented in the MAterials Machine Learning (maml) (<https://github.com/materialsvirtuallab/maml>) repository.

Received: 5 December 2024; Accepted: 24 February 2025;

Published online: 07 March 2025

### References

- Unke, O. T. et al. Machine learning force fields. *Chem. Rev.* **121**, 10142–10186 (2021).
- Ko, T. W., Finkler, J. A., Goedecker, S. & Behler, J. General-purpose machine learning potentials capturing nonlocal charge transfer. *Acc. Chem. Res.* **54**, 808–817 (2021).
- Kocer, E., Ko, T. W. & Behler, J. Neural network potentials: A concise overview of methods. *Ann. Rev. Phys. Chem.* **73**, 163–186 (2022).
- Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
- Ko, T. W., Finkler, J. A., Goedecker, S. & Behler, J. A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer. *Nat. Commun.* **12**, 398 (2021).
- Ko, T. W., Finkler, J. A., Goedecker, S. & Behler, J. Accurate fourth-generation machine learning potentials by electrostatic embedding. *J. Chem. Theory Comput.* **19**, 3567–3579 (2023).
- Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
- Thompson, A. P., Swiler, L. P., Trott, C. R., Foiles, S. M. & Tucker, G. J. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *J. Comput. Phys.* **285**, 316–330 (2015).
- Shapeev, A. V. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Model. Simul.* **14**, 1153–1173 (2016).
- Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **99**, 014104 (2019).
- Zuo, Y. et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **124**, 731–745 (2020).
- Li, X.-G., Chen, C., Zheng, H., Zuo, Y. & Ong, S. P. Complex strengthening mechanisms in the NbMoTaW multi-principal element alloy. *npj Comput. Mater.* **6**, 70 (2020).
- Lee, T. et al. Atomic-scale origin of the low grain-boundary resistance in perovskite solid electrolyte Li0.375Sr0.4375Ta0.75Zr0.25O3. *Nat. Commun.* **14**, 1940 (2023).
- Kostiuchenko, T., Körmann, F., Neugebauer, J. & Shapeev, A. Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials. *npj Comput. Mater.* **5**, 55 (2019).
- Santos-Florez, P. A. et al. Short-range order and its impacts on the BCC MoNbTaW multi-principal element alloy by the machine-learning potential. *Acta Mater.* **255**, 119041 (2023).
- Byggmästar, J., Nordlund, K. & Djurabekova, F. Modeling refractory high-entropy alloys with efficient machine-learned interatomic potentials: Defects and segregation. *Phys. Rev. B* **104**, 104101 (2021).
- Qi, J. et al. Bridging the gap between simulated and experimental ionic conductivities in lithium superionic conductors. *Mater. Today Phys.* **21**, 100463 (2021).
- Krenzer, G. et al. Nature of the superionic phase transition of lithium nitride from machine learning force fields. *Chem. Mater.* **35**, 6133–6140 (2023).

19. Lacivita, V., Artrith, N. & Ceder, G. Structural and compositional factors that control the Li-ion conductivity in LiPON electrolytes. *Chem. Mater.* **30**, 7077–7090 (2018).
20. Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
21. Deng, B. et al. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
22. Batatia, I., Kovacs, D. P., Simm, G., Ortner, C. & Csányi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Adv. Neural Inf. Process. Syst.* **35**, 11423–11436 (2022).
23. Batzner, S. et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453 (2022).
24. Ko, T. W. & Ong, S. P. Recent advances and outstanding challenges for machine learning interatomic potentials. *Nat. Comput. Sci.* **3**, 998–1000 (2023).
25. Jain, A. et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
26. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865 (1996).
27. Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402 (2015).
28. Furness, J. W., Kaplan, A. D., Ning, J., Perdew, J. P. & Sun, J. Accurate and numerically efficient r2SCAN meta-generalized gradient approximation. *J. Phys. Chem. Lett.* **11**, 8208–8215 (2020).
29. Zaverkin, V., Holzmüller, D., Bonfirraro, L. & Kästner, J. Transfer learning for chemically accurate interatomic neural network potentials. *Phys. Chem. Chem. Phys.* **25**, 5383–5396 (2023).
30. Smith, J. S. et al. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **10**, 2903 (2019).
31. Chen, M. S. et al. Data-efficient machine learning potentials from transfer learning of periodic correlated electronic structure methods: Liquid water at AFQMC, CCSD, and CCSD(T) accuracy. *J. Chem. Theory Comput.* **19**, 4510–4519 (2023).
32. Buterez, D., Janet, J. P., Kiddle, S. J., Oglic, D. & Lió, P. Transfer learning with graph neural networks for improved molecular property prediction in the multi-fidelity setting. *Nat. Commun.* **15**, 1517 (2024).
33. Hoffmann, N., Schmidt, J., Botti, S. & Marques, M. A. Transfer learning on large datasets for the accurate prediction of material properties. *Digital Discov.* **2**, 1368–1379 (2023).
34. Jha, D. et al. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nat. Commun.* **10**, 5316 (2019).
35. Lansford, J. L., Barnes, B. C., Rice, B. M. & Jensen, K. F. Building chemical property models for energetic materials from small datasets using a transfer learning approach. *J. Chem. Inf. Model.* **62**, 5397–5410 (2022).
36. Shoghi, N. et al. From molecules to materials: Pre-training Large Generalizable Models for Atomic Property Prediction. The Twelfth International Conference on Learning Representations. (2024)
37. Batra, R., Pilania, G., Uberuaga, B. P. & Ramprasad, R. Multifidelity information fusion with machine learning: A case study of dopant formation energies in Hafnia. *ACS Appl. Mater. Interfaces* **11**, 24906–24918 (2019).
38. Zaspel, P., Huang, B., Harbrecht, H. & Lilienfeld, O. A. Boosting quantum machine learning models with a multilevel combination technique: Pople diagrams revisited. *J. Chem. Theory Comput.* **15**, 1546–1559 (2019).
39. Dral, P. O., Owens, A., Dral, A. & Csányi, G. Hierarchical machine learning of potential energy surfaces. *J. Chem. Phys.* **152**, 204110 (2020).
40. Zheng, P., Zubatyuk, R., Wu, W., Isayev, O. & Dral, P. O. Artificial intelligence-enhanced quantum chemical method with broad applicability. *Nat. Commun.* **12**, 7022 (2021).
41. Yan, Z., Wei, D., Li, X. & Chung, L. W. Accelerating reliable multiscale quantum refinement of protein–drug systems enabled by machine learning. *Nat. Commun.* **15**, 4181 (2024).
42. Chen, C., Zuo, Y., Ye, W., Li, X. & Ong, S. P. Learning properties of ordered and disordered materials from multi-fidelity data. *Nat. Comput. Sci.* **1**, 46–53 (2021).
43. Rosen, A. S. et al. High-throughput predictions of metal–organic framework electronic properties: Theoretical challenges, graph neural networks, and data exploration. *npj Comput. Mater.* **8**, 1–10 (2022).
44. Yu, Z., Xue, P., Xie, B.-B., Shen, L. & Fang, W.-H. Multi-fidelity machine learning for predicting bandgaps of nonlinear optical crystals. *Phys. Chem. Chem. Phys.* **26**, 16378–16387 (2024).
45. Gong, S. et al. Calibrating DFT formation enthalpy calculations by multifidelity machine learning. *JACS Au* **2**, 1964–1977 (2022).
46. Greenman, K. P., Green, W. H. & Gómez-Bombarelli, R. Multi-fidelity prediction of molecular optical peaks with deep learning. *Chem. Sci.* **13**, 1152–1162 (2022).
47. Cooper, N., Goringe, C. & McKenzie, D. Density functional theory modelling of amorphous silicon. *Comput. Mater. Sci.* **17**, 1–6 (2000).
48. Pedersen, A., Pizzagalli, L. & Jónsson, H. Optimal atomic structure of amorphous silicon obtained from density functional theory calculations. *N. J. Phys.* **19**, 063018 (2017).
49. Remsing, R. C., Klein, M. L. & Sun, J. Dependence of the structure and dynamics of liquid silicon on the choice of density functional approximation. *Phys. Rev. B* **96**, 024203 (2017).
50. Remsing, R. C., Klein, M. L. & Sun, J. Refined description of liquid and supercooled silicon from ab initio simulations. *Phys. Rev. B* **97**, 140103 (2018).
51. Cheng, B., Engel, E. A., Behler, J., Dellago, C. & Ceriotti, M. Ab initio thermodynamics of liquid and solid water. *Proc. Natl Acad. Sci.* **116**, 1110–1115 (2019).
52. Ruiz Pestana, L., Marsalek, O., Markland, T. E. & Head-Gordon, T. The quest for accurate liquid water properties from first principles. *J. Phys. Chem. Lett.* **9**, 5009–5016 (2018).
53. Forster-Tonigold, K. & Groß, A. Dispersion corrected RPBE studies of liquid water. *J. Chem. Phys.* **141**, 064501 (2014).
54. Sun, J. et al. Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. *Nat. Chem.* **8**, 831–836 (2016).
55. Deringer, V. L. et al. Origins of structural and electronic transitions in disordered silicon. *Nature* **589**, 59–64 (2021).
56. Bartók, A. P., Kermode, J., Bernstein, N. & Csányi, G. Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **8**, 041048 (2018).
57. Zheng, L. et al. Structural, electronic, and dynamical properties of liquid water by ab initio molecular dynamics based on SCAN functional within the canonical ensemble. *J. Chem. Phys.* **148**, 164505 (2018).
58. Chen, M. et al. Ab initio theory and modeling of water. *Proc. Natl. Acad. Sci.* **114**, 10846–10851 (2017).
59. Zhang, C. et al. Modeling liquid water by climbing up Jacob’s ladder in density functional theory facilitated by using deep neural network potentials. *J. Phys. Chem. B* **125**, 11444–11456 (2021).
60. Morawietz, T., Singraber, A., Dellago, C. & Behler, J. How van der Waals interactions determine the unique properties of water. *Proc. Natl. Acad. Sci.* **113**, 8368–8373 (2016).
61. Sprik, M., Hutter, J. & Parrinello, M. Ab initio molecular dynamics simulation of liquid water: Comparison of three gradient-corrected density functionals. *J. Chem. Phys.* **105**, 1142–1152 (1996).
62. Fernández-Serra, M., Ferlat, G. & Artacho, E. Two exchange-correlation functionals compared for first-principles liquid water. *Mol. Simul.* **31**, 361–366 (2005).

63. Dorner, F., Sukurma, Z., Dellago, C. & Kresse, G. Melting Si: Beyond density functional theory. *Phys. Rev. Lett.* **121**, 195701 (2018).
64. Yao, Y., Kanai, Y. Temperature dependence of nuclear quantum effects on liquid water via artificial neural network model based on SCAN meta-GGA functional. *J. Chem. Phys.* **153**, (2020).
65. Monserrat, B., Brandenburg, J. G., Engel, E. A. & Cheng, B. Liquid water contains the building blocks of diverse ice phases. *Nat. Commun.* **11**, 5757 (2020).
66. Pan, H. et al. Benchmarking coordination number prediction algorithms on inorganic crystal structures. *Inorg. Chem.* **60**, 1590–1603 (2021).
67. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. In International Conference on Learning Representations, (OpenReview.net, 2019).
68. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953 (1994).
69. Fletcher, R. & Reeves, C. M. Function minimization by conjugate gradients. *Comput. J.* **7**, 149–154 (1964).
70. Larsen, A. H. et al. The atomic simulation environment—a Python library for working with atoms. *J. Condens. Matter Phys.* **29**, 273002 (2017).
71. Nocedal, J. Updating quasi-Newton matrices with limited storage. *Math. Comput.* **35**, 773–782 (1980).
72. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
73. Thompson, A. P. et al. LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **271**, 108171 (2022).
74. Qi, J., Ko, T. W., Wood, B. C., Pham, T. A. & Ong, S. P. Robust training of machine learning interatomic potentials with dimensionality reduction and stratified sampling. *npj Comput. Mater.* **10**, 43 (2024).
75. Zhang, T., Ramakrishnan, R. & Livny, M. BIRCH: an efficient data clustering method for very large databases. *SIGMOD Rec.* **25**, 103–114 (1996).
76. Poličar, P. G., Stražar, M. & Zupan, B. openTSNE: a modular Python library for t-SNE dimensionality reduction and embedding. *J. Stat. Softw.* **109**, 1–30 (2024).
77. Linderman, G. C., Rachh, M., Hoskins, J. G., Steinerberger, S. & Kluger, Y. Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nat. Methods* **16**, 243–245 (2019).
78. Stukowski, A. Visualization and analysis of atomistic simulation data with OVITO—the Open Visualization Tool. *Model. Simul. Mat. Sci. Eng.* **18**, 015012 (2009).

## Acknowledgements

This work was intellectually led by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and

Engineering Division under contract No. DE-AC02-05-CH11231 (Materials Project program KC23MP). This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy Office of Science User Facility using NERSC award DOE-ERCAP0026371. T. W. Ko also acknowledges the support of the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program.

## Author contributions

T. W. Ko and S. P. Ong conceived the idea and designed the work. T. W. Ko modified the original M3GNet implementations and performed the analysis. T. W. Ko and S. P. Ong wrote the manuscript and contributed to the discussion and revision.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41524-025-01550-4>.

**Correspondence** and requests for materials should be addressed to Tsz Wai Ko or Shyue Ping Ong.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025