Meta-Learning of Interatomic Potential Models for Accelerated Materials Simulations

Joe D. Morrow and Volker L. Deringer

1 Department of Chemistry, Inorganic Chemistry Laboratory, University of Oxford, Oxford OX1 3QR, United Kingdom

Machine learning (ML) based interatomic potentials are emerging tools for materials simulations but require a trade-off between accuracy and speed. Here we show how one can use one ML potential model to train another: we use an existing, accurate, but more computationally expensive model to generate reference data (locations and labels) for a series of much faster “meta-learned” potentials. Without the need for quantum-mechanical reference computations at the meta-learning stage, extensive reference datasets can be easily generated, and we find that this improves the quality of fast potentials with less flexible functional forms. We apply the technique to disordered silicon, including a simulation of vitrification and polycrystalline grain formation under pressure with a system size of a million atoms. Our work provides conceptual insight into the machine learning of interatomic potential models, and it suggests a route toward accelerated simulations of condensed-phase systems and nanostructured materials.

The properties of materials are governed by the atomic structure and the by forces acting on atoms. Creating accurate computational models for interatomic forces and potential-energy surfaces is therefore a central task in the physical sciences. Today, machine learning (ML) methods are increasingly used to represent quantum-mechanical potential-energy surfaces, typically based on density-functional theory (DFT) ground-truth data and with prediction accuracy to within a few meV per atom [1]. This way, quantum-mechanical quality simulations become accessible at a cost that is orders of magnitude cheaper than DFT [2,8]. ML-based potential models have begun to be applied to challenging problems in physics and related fields: the structures and properties of amorphous materials [9–11], the phase transformations of matter including behavior under extreme conditions [12–14], or surface science and catalysis [15–16]. Recent overviews summarize the progress in the field [17–20].

However, ML potentials still face challenges. A key one is that they incur much larger computational cost than efficient empirical counterparts. It has been argued that various types of ML potentials exist on a “Pareto front” with a given trade-off between accuracy and speed [1]. In addition to developing accurate and computationally efficient potentials, there is a similarly important need for validating an ML potential model once it has been fitted. This is challenging because there is less physical information, or even none, encoded into the model a priori, and because there is not always an unambiguous correlation between the numerical error of a potential (on a fixed test set) and its behavior in simulations [21]. The importance of physically-motivated tests has been demonstrated, e.g., in a comparative study of carbon interatomic potentials [22]. A recent benchmark study of ML potentials compared a range of different fitting frameworks and their numerical performance for multiple systems [1]. Specifically for elemental silicon (Si), various tests including random search have been devised to assess the quality of ML potentials [21,23–24].

In the present work, we outline a computational approach for generating ML interatomic potential models that relies on the idea of “meta-learning” (or “meta-labeling”): we use one ML model to generate reference data for another. The former is accurate and flexible, but more computationally expensive than the latter. We demonstrate that meta-learning a potential for disordered Si largely retains the quality of prediction of the initial model (both in terms of numerical errors and physically motivated tests), whilst speeding up the simulation substantially. This way, much larger-scale simulations become accessible—or conversely, the cost of a given simulation is reduced to a few percent.

Before describing the approach, we briefly discuss terminology. The term “meta-learning” is used in different ways by ML researchers to describe the idea of using knowledge about learning to increase the efficiency of the learning process itself [25]. Related concepts include boosting [26,27], where a chain of models is set up, each trained (except the base model) to correct the error of the preceding one, and stacking, where an ensemble of models is fitted and their predictions combined to yield a more accurate one [28]. A meta-learning algorithm might adjust the way stacked predictions are combined, or automatically optimize model hyperparameters. We use the term in a slightly different way: to refer to the indirect learning of the ground truth through predictions of an intermediate model, which is primarily motivated by the computational cost advantage in applying (“querying”) the meta-learned model once already trained.

Methodology.—Machine-learned potentials approximate a potential-energy surface (PES) based on reference data, typically from small-scale DFT computations [29]:

\[ \{x_i, y_i\}_{i=1...N} \xrightarrow{M_1} \hat{y}(x), \]

where \(x_i\) is a descriptor for the atomic environment [30], \(y_i\) is one of \(N\) observations of the reference PES, and \(\hat{y}\) denotes the ML model fitted with a given method [31]. We call this method “M1” in the following.
Firstly, we can build an almost arbitrarily large database produced computational cost of GAP compared to DFT. Secondly, we can evaluate the fidelity of the fit via numerical errors on a wide-ranging set of out-of-sample testing data, and also test behavior in large-scale molecular dynamics (MD) simulations. The practical advantages of the present application of meta-learning rely on the observation that the GAP fitting approach may be able to more accurately interpolate across a sparsely sampled PES compared to MTPs, which is not yet fully investigated [1, 41]. We expect, however, that the same concept could be applied to any pair of fitting methodologies where a cost–accuracy disparity exists. The insights into database development and physically-guided validation of ML potentials are independent of the particular M1 and M2 methods chosen.

For the present study, we aim to replicate the behavior of GAP-18 under pressure, which we find to be an instructive test case, and which relates to questions of experimental interest [42–44]. To improve interpolation for high-p structures, we use a large training database (D1 in Fig. 1) containing 250 structures with 1,000 atoms each, corresponding to compression to 20 GPa (as in Ref. [1]). These structures are supplemented with the original GAP-18 database (Fig. 2), which we re-label with GAP-18 energies and forces to make the input data consistent.
Numerical errors.—The accuracy of an ML potential is generally characterized by an error versus its reference method [1], typically evaluated on a test set not included during training. In this vein, Fig. 3 shows the mean absolute error on different test sets for a series of M′ and M″ models of increasing complexity and cost. On the one hand, we test on random-structure-searching (RSS) configurations which are different from the fitting database; on the other hand, we test on snapshots from small-scale compression MD simulations, directly mirroring the physical situation that is to be “meta-learned”.

For both test sets, Fig. 3 provides evidence of efficient meta-learning of the M0 PES: comparing their predictions to M0 (DFT), meta-learned MTP models tend to at least match, if not improve upon, the corresponding directly-learned (D0-trained) MTPs. Inspecting the force-component errors (lower panels in Fig. 3) indicates that our meta-learned MTP models faithfully recover the predictions of M1—suggesting that the existing M1 PES is smooth enough to be learned by a different fitting method. More generally, the error analysis suggests that improvements to the M1 reference can be expected to transfer to subsequent meta-learned models.

We emphasize that the tests in Fig. 3 are not designed, nor able, to assess the quality of MTP models in absolute terms. In previous studies, reference databases for MTP fitting have been built using active learning [45, 46], and it would now be interesting to include such a gradually improved reference database in the construction of MTP models for high-pressure disordered Si.

Physically-guided validation.—While numerical errors provide some indication of performance, the correlation between both is unreliable enough to warrant the extensive validation of ML potentials based on physical behavior—for example, assessing whether structural features of a reference simulation are reproduced. Here we use the Smooth Overlap of Atomic Positions (SOAP) kernel [34] which has proven useful for analyzing local structure in amorphous Si [11, 47]. The computational speed of ML potentials, and of MTPs in particular, allows us to compare several different meta-learned potentials in large-scale (100,000-atom) simulations. For Si, this provides a more stringent test than small-scale structures, allowing us to directly assess the description of the nucleation of Shi crystallites and their intervening grain boundaries, and to quantify this description using SOAP.

Figure 4 benchmarks a series of MTPs, directly-learned (blue; M′L) or meta-learned (red; M″L), with increasing quality and cost. All compression simulations start from the same low-density amorphous (LDA) structure from...
FIG. 4. Performance of meta-learned potential models of increasing flexibility in 100,000-atom compression simulations of amorphous Si. We report the average SOAP similarity to pressure-adjusted sh Si during compression MD (cf. Fig. S1), benchmarking MTPs (solid) versus GAP-18 (dashed). The difference for each MTP to GAP-18 is plotted underneath the absolute values, with $M''_L$ offset for clarity. Representative snapshots during the run are displayed, colored according to atomistic SOAP similarity to sh (A–E). Points at which MD simulations fail (for $M'_{14}$ and $M'_{20}$) are marked with asterisks.

TABLE I. Quality metrics for 100,000-atom LDA Si models produced by $10^{11}$ Ks$^{-1}$ quench simulations: the proportion of 3- and 5-fold connected atoms ($N_3$ and $N_5$), the inverse height of the first sharp diffraction peak ($H^{-1}$) [53], and the mean ($\bar{\theta}$) and width ($\Delta \theta$) of the bond-angle distribution.

| Model  | $N_3$ (%) | $N_5$ (%) | $H^{-1}$ | $\bar{\theta}$ (deg) | $\Delta \theta$ (deg) | $\Delta E$ (eV) |
|--------|-----------|-----------|----------|-----------------------|-----------------------|----------------|
| $M'_{16}$ | 0.57 | 2.37 | 0.603 | 108.99 | 10.85 | 0.156 |
| $M'_{16}$ | 0.68 | 1.64 | 0.560 | 109.15 | 10.23 | 0.152 |
| $M''_{20}$ | 0.47 | 3.65 | 0.498 | 108.93 | 10.91 | 0.146 |
| $M''_{20}$ | 0.69 | 1.33 | 0.504 | 109.16 | 9.93 | 0.144 |
| GAP-18 | 0.70 | 0.95 | 0.564 | 109.17 | 9.99 | 0.145 |

Ref. [11] and are run using LAMMPS [49]. In all cases, we use the similarity to sh Si as quality measure, aiming to quantify how well a given model reproduces the behavior of GAP-18 ($M_1$). Overall, meta-learned MTPs perform much more reliably at the compression test than those directly fitted (to $D_0$), with every $M''_L$ model producing the expected sequence of phases [11], viz: LDA $\rightarrow$ very-high-density amorphous (VHDA) $\rightarrow$ polycrystalline (pc) sh. In contrast, only 2 out of 5 the directly-learned models perform acceptably. Comparisons of this type are only feasible because $M_1$ is efficient enough to produce a reference simulation (albeit with considerable effort), and additionally because $M_2$ is efficient enough to readily produce many of such simulations.

We finally test our meta-learned potentials for a task for which they were not specifically trained (in D1): the production of LDA structural models via simulated quenching from the melt. A number of quantitative measures are available [50–52] to measure how well our meta-learned models reproduce the behavior of GAP-18 in this well-established test for interatomic potentials (Table I). We run quenching simulations with 100,000 atoms, as in Ref. [11] now comparing two representative levels of MTP fitting. In both cases, the meta-learned model outperforms a directly learned one at the same level. Both $M''_L$ models come close to reproducing the defect count, with that of 3-fold connected atoms being almost identical (0.68%, 0.69%, and 0.70% with $M''_{16}$, $M''_{20}$, and GAP-18, respectively), and that of 5-fold connected ones being somewhat more varied. There is not, however, a unambiguous advantage of $M''_{20}$ over $M''_{16}$: specifically, the inverse height of the first sharp diffraction peak, $H^{-1}$, is much better described by the latter. This underlines the subtleties in assessing the behavior of ML potentials.

Application.—To demonstrate its usefulness in practice, we employ our approach to generate million-atom structural models of LDA and pc-sh Si. The nanoscale grain structure of pc-sh Si invites the use of larger-scale
models for more extended studies of the typical structure in the grain-boundary region and the size and orientation of crystallites. The efficiency of MTPs permits fast simulations on long timescales without having to compromise on system size. Figure 5 shows results of a melt-quench (Fig. 5a) and compression MD (Fig. 5b) simulation of a system of 1M Si atoms, using the M''\textsubscript{16} model [53].

The LDA structure, visualized in Fig. 5a, has a bond-angle distribution closely similar to that for the 100k-atom GAP-18 structure (Fig. 5c). However, a more detailed analysis of the environments of the N = 5 atoms ("coordination defects") is possible with larger system size, as evident from the much lower scatter in the results shown in Fig. 5c. The histogram bin widths are 0.25° for both system sizes for fair comparison, which is sufficiently narrow to permit analysis of small deviations away from the ideal tetrahedral angle. The wide distribution of bond angles in N = 5 atoms is consistent with the relatively wide spread in local structure, as characterized in Ref. 47 for much smaller simulation cells.

Upon compression, the expected pressure-induced collapse of the largely fourfold-connected structure into a VHDA phase [11, 51] is reproduced by the meta-learned potential, the onset occurring at only slightly lower pressure compared to the reference GAP-18 simulation. There are no discernible differences between the volume-pressure curves of 100k-atom and 1M-atom simulations carried out using the same meta-learned model (Fig. 5d), making system-size effects for VHDA formation unlikely. However, system-size effects in the nucleation of sh crystallites are clearly still important even in 100k-atom systems. On the top face of the GAP-18 structure in Fig. 5b, a crystallite extends all across the cell and hence is infinite in that direction under periodic boundary conditions. In the 1M-atom simulation, the structure has similar grain sizes, but the larger cell dimensions ensure that no grain approaches the length of the cell itself.

The differences in the volume-pressure curves between GAP-18 and M''\textsubscript{16} (Fig. 5d) are of a similar magnitude to those between GAPs trained using different reference methods in Ref. 11. The crystallization and volume of the pc-sh phase are well reproduced, as is the behavior as characterized using SOAP (Fig. 5e). Figure 5 therefore suggests that our meta-learned M''\textsubscript{16} model can, to a large extent, match predictions of GAP-18. This way, more detailed analyses become possible—for example, of the grain structure in pc-sh Si, as well as a closer study of the LDA phase which is the subject of ongoing work.

In future studies, we envisage using more diverse combinations of fitting frameworks in a similar vein, e.g.,
neural networks \[2\] and SNAP \[1\], and automating this model selection. In addition, we propose to incorporate ensembles of base models, including using different levels of theory, to further augment the capabilities of a meta-learned model. We hope that such developments will help to make quantum-mechanically accurate materials simulations with millions of atoms more commonplace in the years ahead.

J.D.M. acknowledges funding from the EPSRC Centre for Doctoral Training in Inorganic Chemistry for Future Manufacturing (OxICFM), EP/S023828/1. The authors acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work (http://dx.doi.org/10.5281/zenodo.22558), as well as support from the John Fell Oxford University Press (OUP) Research Fund. We thank A. Shapeev for helpful comments on the manuscript.

* volker.deringer@chem.ox.ac.uk

[1] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, and S. P. Ong, J. Phys. Chem. A 124, 731 (2020)

[2] J. Behler and M. Parrinello, Phys. Rev. Lett. 98, 146401 (2007)

[3] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. 104, 136403 (2010)

[4] A. Thompson, L. Swiler, C. Trott, S. Foisie, and G. Tucker, J. Comput. Phys. 285, 316 (2015)

[5] Z. Li, J. R. Kermode, and A. De Vita, Phys. Rev. Lett. 114, 096405 (2015)

[6] A. V. Shapeev, Multiscale Model. Simul. 14, 1153 (2016)

[7] L. Zhang, J. Han, H. Wang, R. Car, and W. E, Phys. Rev. Lett. 120, 143001 (2018)

[8] R. Jimmouchi, J. Lahsteiner, F. Karsai, G. Kresse, and M. Bokdam, Phys. Rev. Lett. 122, 225701 (2019)

[9] G. C. Sosso, G. Miceli, S. Caravati, F. Giberti, J. Behler, and M. Bernasconi, J. Phys. Chem. Lett. 4, 4241 (2013)

[10] M. A. Caro, V. L. Deringer, J. Koskinen, T. Laurila, and G. Csányi, Phys. Rev. Lett. 120, 166101 (2018)

[11] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, and S. R. Elliott, Nature 589, 59 (2021)

[12] B. Cheng, G. Mazzola, C. J. Pickard, and M. Ceriotti, Nature 585, 217 (2020)

[13] J. S. Smith, B. Nebgen, N. Mathew, J. Chen, N. Lubbens, L. Burakovsky, S. Tretiak, H. A. Nam, T. Gerhard, S. Fensin, and K. Barros, Nat. Commun. 12, 1257 (2021)

[14] H. Zong, V. N. Robinson, A. Hermann, L. Zhao, S. Scandolo, X. Ding, and G. J. Ackland, Nat. Phys. 17, 955 (2021)

[15] N. Artrith and A. M. Kolpak, Nano Lett. 14, 2670 (2014)

[16] J. Timmermann, F. Kraushofer, N. Resch, P. Li, Y. Wang, Z. Mao, M. Riva, Y. Lee, C. Staaeck, M. Schmid, C. Scheurer, G. S. Parkinson, U. Diebold, and K. Reuter, Phys. Rev. Lett. 125, 206101 (2020)

[17] J. Behler, Angew. Chem. Int. Ed. 56, 12828 (2017)

[18] V. L. Deringer, M. A. Caro, and G. Csányi, Adv. Mater. 31, 1902765 (2019)

[19] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, Annu. Rev. Phys. Chem. 71, 361 (2020)

[20] P. Friederich, F. Häse, J. Proppe, and A. Aspuru-Guzik, Nat. Mater. 20, 750 (2021)

[21] J. George, G. Hautier, A. P. Bartók, G. Csányi, and V. L. Deringer, J. Chem. Phys. 153, 044104 (2020)

[22] C. de Tomas, A. Aghajamali, J. L. Jones, D. J. Lim, M. J. López, I. Suarez-Martinez, and N. A. Marks, Carbon 155, 624 (2019)

[23] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, Phys. Rev. X 8, 041048 (2018)

[24] Y. Lysoyngorskii, C. van der Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner, and R. Drautz, npj Comput. Mater. 7, 97 (2021)

[25] R. Villalba and Y. Drissi, Artif. Intell. Rev. 18, 77 (2002)

[26] Y. Freund and R. E. Schapire, J. Comput. Syst. Sci. 55, 119 (1997)

[27] J. Friedman, T. Hastie, and R. Tibshirani, Ann. Stat. 28, 337 (2000)

[28] T. G. Dietterich, in Multiple Classifier Systems (Springer, Berlin, Heidelberg, 2000) pp. 1–15.

[29] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, Chem. Rev. 121, 10073 (2021)

[30] F. Musil, A. Grisafi, A. P. Bartók, C. Ortner, G. Csányi, and M. Ceriotti, Chem. Rev. 121, 9759 (2021)

[31] In practice, the matter is more complex: typically, atomic energies are machine-learned from total energies (and, often, forces) \[2, 3\].

[32] B. Cheng, R.-R. Griffiths, S. Wengert, C. Kunkel, T. Stenczel, B. Zhu, V. L. Deringer, N. Bernstein, J. T. Margraf, K. Reuter, and G. Csányi, Acc. Chem. Res. 53, 1981 (2020)

[33] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, Phys. Rev. X 8, 041048 (2018)

[34] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B 87, 184115 (2013)

[35] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, Mach. Learn.: Sci. Technol. 2, 025002 (2021)

[36] We note that the idea of mapping a kernel-based ML potential to a fast approximator has been previously explored: via cubic spline interpolation across a pre-calculated grid of points \[55, 56\].

[37] C. J. Pickard and R. J. Needs, Phys. Rev. Lett. 97, 045504 (2006)

[38] C. J. Pickard and R. J. Needs, J. Phys.: Condens. Matter 23, 53201 (2011)

[39] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B 46, 6671 (1992)

[40] S. J. Clark, M. D. Segall, C. J. Pickard, P. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, Z. Kristallogr. 220, 567 (2005)

[41] C. W. Rosenbrock, K. Gubaev, A. V. Shapeev, L. B. Pártay, N. Bernstein, G. Csányi, and G. L. W. Hart, npj Comput. Mater. 7, 1 (2021)

[42] S. K. Deb, M. Wilding, M. Somayazulu, and P. F. McMillan, Nature 414, 528 (2001)

[43] M. C. Wilding, M. Wilson, and P. F. McMillan, Chem. Soc. Rev. 35, 964 (2006)

[44] P. F. McMillan, Nature 589, 22 (2021)

[45] E. V. Podryabinkin and A. V. Shapeev, Comput. Mater.
[46] E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev, and A. R. Oganov, Phys. Rev. B 99, 064114 (2019).
[47] N. Bernstein, B. Bhattarai, G. Csányi, D. A. Drabold, S. R. Elliott, and V. L. Deringer, Angew. Chem. Int. Ed. 58, 7057 (2019).
[48] R. Xie, G. G. Long, S. J. Weigand, S. C. Moss, T. Carvalho, S. Roorda, M. Hejna, S. Torquato, and P. J. Steinhardt, Proc. Natl. Acad. Sci. U. S. A. 110, 13250 (2013).
[49] S. Plimpton, J. Comput. Phys. 117, 1 (1995).
[50] R. Atta-Fynn and P. Biswas, J. Chem. Phys. 148, 204503 (2018).
[51] V. L. Deringer, N. Bernstein, A. P. Bartók, M. J. Cliffe, R. N. Kerber, L. E. Marbella, C. P. Grey, S. R. Elliott, and G. Csányi, J. Phys. Chem. Lett. 9, 2879 (2018).
[52] D. K. Limbu, S. R. Elliott, R. Atta-Fynn, and P. Biswas, Sci. Rep. 10, 7742 (2020).
[53] For recent, large-scale atomistic simulations using ML potentials, see, e.g., Ref. 13 or: W. Jia, H. Wang, M. Chen, D. Lu, L. Lin, R. Car, W. E, and L. Zhang, in SC ’20: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 5 (2020).
[54] M. Durandurdu and D. A. Drabold, Phys. Rev. B 64, 014101 (2001).
[55] A. Glielmo, C. Zeni, and A. De Vita, Phys. Rev. B 97, 184307 (2018).
[56] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, and B. Kozinsky, npj Comput. Mater. 6, 120 (2020).
SOAP SIMILARITY ANALYSIS FOR PHYSICALLY-GUIDED VALIDATION

The Smooth Overlap of Atomic Positions (SOAP) descriptor is a fingerprint originally developed in the context of ML potentials [34] and widely employed in GAP fitting [29]; however, SOAP has also been used for classifying and understanding complex structures across a wide range of chemical systems [30, S1–S2]. Briefly, the SOAP formalism encodes the local environment of atom $i$ as a neighbor density $\rho^{(i)}(r)$ by placing a 3D Gaussian function of width $\sigma_{at}$ on $i$ and each of its neighbors up to a cutoff. A similarity measure $k_{SOAP}(i,j)$ between two such environments is given by evaluating the overlap integral of their respective $\rho(r)$ and averaging over all possible rotations $\hat{R}$ of one of the local environments.

Here we use the configuration-averaged SOAP kernel (cf. Ref. S2) to measure the similarity of MD frames to crystal structures adopted by Si, which provides a physical interpretation of the average structure in a large simulation system. We are interested in the structural changes that occur with increasing pressure beyond the trivial isotropic contraction of bond lengths. To subtract this effect from that of the SOAP similarity, we compare our simulation trajectory not only to ambient-pressure crystalline phases ("fixed" in Fig. S1), but also to crystals with lattice parameters optimized as a function of pressure. More precisely, we optimize the lattice parameters at 1 GPa intervals with DFT and interpolate between them when calculating the similarity so that the reference crystal is under the same pressure as is measured over each ps of MD simulation time ("scaled" in Fig. S1).

Figure S1 shows the evolution of the SOAP similarity with pressure during the compression simulation reported in Ref. [11] (carried out for a 100,000-atom system using GAP-18) which we here use as reference, assessing the per-atom similarity to crystalline phases in two different ways: either fixing them to their ambient-pressure structures (dashed lines), or allowing for freely varying lattice parameters with pressure (solid lines). The atomic environments in the quenched LDA phase at 0 GPa have a rather close resemblance to those in the cubic diamond (dia) structure, with both featuring tetrahedral Si environments. Up to about 12 GPa, the similarity to pressure-adjusted dia remains practically constant.

FIG. S1. A structural benchmark for high-pressure disordered Si. We evaluate the average SOAP similarity of the compression trajectory of Ref. [11] to selected crystalline structures. Dashed lines are for reference crystals with a fixed lattice at 0 GPa with DFT. Solid lines are similarities to crystals with their lattice parameters optimised as a function of pressure. The structure images show representative snapshots, color-coded according to SOAP similarity to the sh crystalline phase, with lighter colors indicating higher similarity (drawn with data from Ref. [11]).
TABLE S1. General hyperparameters used for SOAP-based structural analysis, carried out using the QUIP/GAP code as interfaced to the quippy Python package (see https://github.com/libAtoms/QUIP). SOAP vectors were configuration-averaged (average = $T$ in quippy) for each MD frame before their transformation to the power spectrum and then kernel via a dot product with the reference crystal. The dot product was raised to a power of $\zeta = 4$.

- $\sigma_{at}$: 0.5 Å
- $R_{cut}$: 5.0 Å
- $n_{max}$: 6
- $l_{max}$: 10
- $\text{cutoff\_transition\_width}$: 2.0 Å
- $\zeta$: 4

More detailed inspection of the MD trajectory confirms the interpretation that the LDA phase does not change substantially beyond an isotropic contraction until it collapses to a very-high-density amorphous phase (VHDA) around 12 GPa [11], seen as a sudden drop in the dia similarity. The VHDA phase has the closest similarity among tested crystals to the high-pressure Si-VI phase with space group $Cmca$, which appears experimentally upon compression to about 40 GPa, after the formation of sh and before hexagonal close-packed (hcp) Si. At about 14 to 16 GPa, the nucleation of sh crystallites begins [11], so the similarity to $Cmca$ decays slightly and similarity to sh increases. At this point, characterizing inhomogeneous polycrystalline sh (pc-sh) with any single variable becomes fraught, which also explains why the $Cmca$ similarity remains so high (along with $\beta$-Sn; see Fig. S2). The interfaces between sh grains, which are more similar to $Cmca$ and $\beta$-Sn than they are to sh, occupy a reasonable fraction of the simulation-cell volume, hence contribute appreciably to the average similarity. Nevertheless, the analysis in Fig. S1 summarizes the structural changes during compression in a comprehensive way.

FIG. S2. Extension of Fig. S1 to now include further crystal structures: comparing the average atomic environment during compression MD to reference structures. Dashed lines are for reference crystals with fixed structural parameters as optimized using DFT at 0 GPa. Solid lines indicate SOAP similarities to crystals with their lattice parameters optimized using DFT as a function of pressure. Rhombohedral phase R8 results from decompressing $\beta$-Sn, containing 5 and 6-membered rings of tetrahedrally-coordinated Si (see Ref. S3). bct is a hypothetical four-fold coordinated structure predicted to be stable under tensile stress [S4]. The metallic phases commonly labeled with their orthorhombic spacegroups, $Imma$ and $Cmca$, appear with increasing pressure respectively between $\beta$-Sn and sh and between sh and hexagonally close-packed (hcp) phases. Anzellini et al. [S5] provide an introduction to the numerous Si phases and their nomenclature.
MELT–QUENCH SIMULATIONS

The static structure factor, \( S(Q) \), is a particularly useful quantity because it can be experimentally observed, and high-quality reference data for amorphous Si are available (e.g., Refs. 48 and S6). Figure S3 shows the structure factor of an LDA model produced using the variable quench-rate protocol of Ref. 51. There is almost perfect agreement between \( M'_16 \) and GAP-18, which also matches the experimental data from Ref. 48 closely. Smaller-scale simulations suggest that levels 16 and 20 span most of the variation in \( H^{-1} \) among the MTPs studied here (Fig. S4), but system sizes of the order of 100k atoms are required for repeatable results due to the stochastic nature of the simulations.

We limit ourselves to two levels because of the expense associated with the long timescales of \( 10^{11} \text{ K s}^{-1} \) melt-quenching, but note that other meta-learned MTPs tested tend to overestimate the FSDP height, with \( M'_12 \) and \( M'_14 \) producing regions of diamond-like crystallinity.

FIG. S3. Variable-rate melt quenches as a benchmark for the behavior of ML potentials for Si. The figure compares D0-trained and meta-learned MTPs to M1 (upper panels) and experiment (lower panel); the simulations follow the protocol of Ref. 51. Although the structure factor, \( S(Q) \), as predicted by \( M'_16 \), agrees almost perfectly with the GAP-18 result, the much larger FSDP predicted by \( M'_20 \) suggests this may in fact be serendipitous. It is nevertheless encouraging that the other peaks are also well-matched by the \( M'_16 \) curve. The increased ordering implied by the higher FSDP of the \( M'_20 \) curve does not obviously manifest as crystalline grains, nor is any peak splitting characteristic of paracrystalline ordering clear in the other diffraction peaks.

FIG. S4. Static structure factors of 4096-atom low-density amorphous Si structures, produced via the same variable-temperature melt-quench protocol as for Figure S3, and using directly and indirectly fitted MTPs covering a broader range of \( \text{lev}_{\text{max}} \) values. Some of the variability in the height of the FSDP can be explained by the stochastic nature of grain formation during cooling in combination with the relatively small system size (the height varies from simulation to simulation for the same MTP and same protocol: e.g., compare \( M'_20 \) in this figure and Fig. S3). However, the more flexible meta-learned MTPs do seem to induce some over-ordering. More detailed studies using meta-learned MTPs specifically trained for melt-quench simulations will be the subject of future work.

(as do \( M'_12 \) and \( M'_14 \)); this requires further work in a future study, and we focus on the level 16 and 20 models herein. The strong performance of \( M'_16 \) supports our choice to use it for a large-scale proof-of-concept (Fig. 5 in the main text).
### COMPUTATIONAL DETAILS

**Moment tensor potential fitting**

MTP models have been applied to a range of research questions, including phase diagrams [11], reaction dynamics of small molecules [S7], and lithium ion conduction [S8]. Full details of their construction are given in the original work by Shapeev and co-workers [6,35], which we summarize briefly here. The descriptor of a local environment is defined as the moment tensor,

$$M_{\mu,\nu}(n_i) = \sum_j f_\mu(r_{ij}) \bigotimes_{\nu \text{ times}} r_{ij},$$  

(S1)

where $n_i$ denotes the positions of the $i$-th atom and all neighbors $j$ up to a cutoff, $R_{\text{cut}}$, which we set to 5 Å consistent with that of the descriptor used for GAP-18 [24]. The radial part $f_\mu$ depends only on the position vectors $r_{ij}$ of each $j$-th atom from the $i$-th one, and the angular information is encoded by tensors of rank $\nu$ produced from the repeated outer product of $r_{ij}$ vectors. These $M_{\mu,\nu}$ are contracted to give the basis functions $B_\alpha$. Training involves finding the regression coefficients $\xi_\alpha$ that minimize a loss function based on the errors in predicted energies,

$$E_{\text{MTP}} = \sum_i \varepsilon_i \equiv \sum_i \sum_\alpha \xi_\alpha B_\alpha(n_i),$$  

(S2)

as well as forces and stresses, which can be expressed in terms of the first and second derivatives of $E_{\text{MTP}}$ with respect to $r_{ij}$. The number of basis functions defining the particular functional form of an MTP (and hence its flexibility and computational efficiency) grows exponentially with the important hyperparameter $\text{lev}_{\text{max}}$, where the level (“lev”) is defined as

$$\text{lev}(M_{\mu,\nu}) = 2 + 4\mu + \nu$$  

(S3a)

$$\text{lev}(M_{\mu,\nu}; n_\alpha) = \text{lev}(M_{\mu,\nu}) + \text{lev}(n_\alpha)$$  

(S3b)

and “⊗” represents an appropriate tensor contraction operation. All possible $M_{\mu,\nu}$, and contractions of one or more $M_{\mu,\nu}$, are included as degrees of freedom in the functional form of an MTP, provided the resulting basis function satisfies $\text{lev}(B_\alpha) \leq \text{lev}_{\text{max}}$. In the present work, we investigated the range $\text{lev}_{\text{max}} = 12–24$. For brevity, we denote $\text{lev}_{\text{max}}$ as “L” in the main text.

We found that the performance of MTPs was quite sensitive to the relative weights of energies, forces, and stresses in the fit. After some testing, we settled on 10 : 1 : 0.01 respectively. Our chosen force weight is a factor of 10 larger than the commonly-used default value, as we found both direct and meta-learned MTPs suffered from unstable MD and poor predictions of the physical behavior under pressure using the default force weight. Full details of all remaining MTP hyperparameter choices are given in Table S2.

\[\begin{array}{ll}
\text{min_dist} & 1.5 \text{ Å} \\
\text{max_dist} & 5.0 \text{ Å} \\
\text{radial_basis_type} & \text{RBChebyshev} \\
\text{energy-weight} & 1.0 \\
\text{force-weight} & 0.1 \\
\text{stress-weight} & 0.001 \\
\text{MTP version} & 1.1.0 \\
\end{array}\]

**TABLE S2.** General hyperparameters used for all MTPs trained in this work. The default radial\_basis\_size for each value of $\text{lev}_{\text{max}}$ was used throughout, as was the case for any parameters not listed below.

---

**FIG. S5.** Mean absolute error of predicted energies versus quantity of compression MD training data of two types: starting separately from LDA structures (dashed lines) and randomly placed hard spheres (solid lines). 3 MTPs with the same settings were trained for each database and we report their average errors in energy prediction vs. GAP-18 on two independent validation sets: structures derived from melt-quench MD (blue) and compression MD (yellow). The median standard deviation in error across identically-trained MTPs, $\sigma_{\Delta E}/|\Delta E|$, is in the region of 10–45%. MAEs show improvement with the addition of similar data to the training set, which is converged with respect to database size by $240 \times 10^3$ atomic environments. The errors for melt-quench MD begin to degrade with the increased proportion of compression MD included during training between $120 \times 10^3$--$240 \times 10^3$ environments. Including even more environments leads to an unstable potential for MD simulations.
D1 database

The reference configurations for the meta-learning step were obtained with an MD simulation protocol of 50 ps thermalization of a random initial structure at 500 K, followed by a constant pressure ramp to 20 GPa over 200 ps (as used in Ref. [11]). In Fig. S5, we study errors for compression MD with respect to quantity of compression training data, which we find to converge by around $2 \times 10^5$ atomic environments. Meanwhile, the errors for melt-quench MD start to diverge at this point (becoming worse with larger D1), suggesting that over-weighting environments encountered during compression MD comes at the expense of performance in other regions of configuration space. Based on these results, and our observations of unstable MD for compression data sets much greater than $2 \times 10^5$ environments, we found potentials trained on $2.4 \times 10^5$ environments from random starting points to be suitable and therefore use a database of this size for validation.

Test configurations

The “RSS” (random structure searching) test set, for which results are shown in Fig. 3a in the main text, consists of 100 structures produced using the buildcell code of the Ab Initio Random Structure Searching (AIRSS) software [37, 38] and optimized with M1 at pressures in the range 0–32 GPa. The initial geometries of the random structures were constrained to have a minimum interatomic separation of 2 Å, a volume of $20 \pm 5$ Å$^3$ per atom, 1–4 random symmetry operations, and 1–64 atoms in the unit cell. The corresponding buildcell settings were NATOM=1-16, VARVOL=20, MINSEP=2.0, SYMMOPS=1-4. Each structure was optimized at a random pressure taken from the (positive) exponential distribution centered at 0 GPa with width 5 GPa. See Ref. [38] for details of the AIRSS framework.

The “compression MD” set (Fig. 3b) consists of structures from 64-atom compression runs of LDA Si (snapshots taken between 5 and 20 GPa). Both test sets were evaluated in single-point DFT computations with settings corresponding to Ref. [23] (cf. https://doi.org/10.5281/zenodo.1250555), using the PW91 functional [39], a 250 eV energy cut-off (with automatic finite basis set correction), 0.05 eV Fermi–Dirac smearing, and an on-the-fly pseudopotential, as implemented in CASTEP [40].

Computational cost

Timing data (used to construct the horizontal axis in Fig. 3 in the main text) were acquired by averaging across 9 MD runs of 5,000 timesteps with 10,000 atoms, parallelized across 48 CPU cores using LAMMPS. The protocol, initial structure, and random seed used were the same for each potential.

SUPPLEMENTARY REFERENCES

S1. S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Phys. Chem. Chem. Phys. 18, 13754 (2016).
S2. J. Mavračič, F. C. Mocanu, V. L. Deringer, G. Csányi, and S. R. Elliott, J. Phys. Chem. Lett. 9, 2985 (2018).
S3. J. Crain, G. J. Ackland, J. R. Maclean, R. O. Piltz, P. D. Hatton, and G. S. Pawley, Phys. Rev. B 50, 13043(R) (1994).
S4. Y. Fujimoto, T. Koresu, S. Saito, T. Miyake, and A. Oshiyama, New J. Phys. 10, 083001 (2008).
S5. A. Anzellini, M. T. Wharmby, F. Miozzi, A. Kleppe, D. Daisenberger, and H. Wilhelm, Sci. Rep. 9, 15537 (2019).
S6. K. Lazaìri, S. Kycia, S. Roorda, M. Chicoine, J. L. Robertson, J. Wang, and S. C. Moss, Phys. Rev. B 60, 13520 (1999).
S7. I. S. Novikov, Y. V. Suleimanov, and A. V. Shapeev, Phys. Chem. Chem. Phys. 20, 29503 (2018).
S8. C. Wang, K. Aoyagi, P. Wisesa, and T. Mueller, Chem. Mater. 32, 3741 (2020).