JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC

Article

Hyperparameter Optimization for Atomic Cluster Expansion Potentials

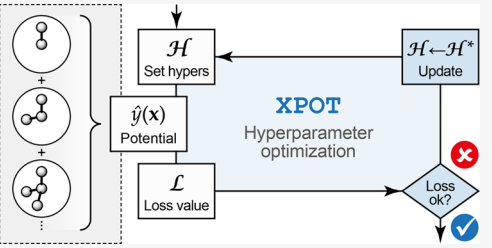
Daniel F. Thomas du Toit, Yuxing Zhou, and Volker L. Deringer*

Cite This: J. Chem. Theory Comput. 2024, 20, 10103−10113

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Machine learning-based interatomic potentials enable accurate materials simulations on extended time- and length scales. ML potentials based on the atomic cluster expansion (ACE) framework have recently shown promising performance for this purpose. Here, we describe a largely automated computational approach to optimizing hyperparameters for ACE potential models. We extend our openly available Python package, XPOT, to include an interface for ACE fitting, and discuss the optimization of the functional form and complexity of these models based on systematic sweeps across relevant hyperparameters. We showcase the usefulness of the approach for two example systems: the covalent network of silicon and the phase-change material Sb_2Te_3. More generally, our work emphasizes the importance of hyperparameter selection in the development of advanced ML potential models.



INTRODUCTION

Machine learning (ML) approaches to interatomic potential fitting are now widely used in computational chemistry.[1−5] ML potentials provide surrogate models for quantum-mechanical (QM) potential-energy surfaces at a fraction of the cost, providing longer-time scale and larger-length scale simulations than would be accessible with direct QM methods, while maintaining comparable accuracy. As a result, they are now becoming widespread tools to simulate materials and molecules, with applications ranging from the structure of disordered solids[6,7] to modeling nuanced effects such as anharmonic phonons[8] or noncollinear magnetism in iron,[9] as well as materials not yet synthesized.[10]

Many methods for fitting ML potentials have now been developed, from the early Behler−Parrinello neural-network[11,12] and Gaussian Approximation Potential (GAP)[13−15] models to more recent atomic cluster expansion (ACE) potentials,[16−18] as well as graph-neural-network architectures.[19−22] Each of these methods has different characteristics: for example, graph-based architectures currently define the state-of-the-art in terms of accuracy, but inference (prediction) is still relatively expensive for large systems.[23,24] Our focus herein is on ML potentials that are based on ACE descriptors and fitted using the `PACEmaker` software by Drautz and coworkers,[17,18] which have shown to provide efficient and accurate predictions, e.g., for data from ref 25.[17]

Despite their popularity, ML potentials pose challenges because they largely lack physically motivated functional forms. First, they require high-quality reference data:[26] recent handcrafted ML potentials for carbon[6] and Ge−Sb−Te[7] each used more than 300,000 atomic environments for training; for potentials encompassing many different elements, the data set sizes can easily range in the millions.[27,28] Second, ML potentials require well-chosen hyperparameters—those parameters which must be chosen before fitting starts, and thus determine the nature of the fitting. For mathematically complex ML models in particular, the effects of changing hyperparameters are not always immediately clear. For example, in the ACE approach we use here, the prefactors and exponents of atomic property terms which contribute to the energy (see Methods section below) can be difficult to tune manually. It would therefore be desirable to explore the space of hyperparameters automatically and to optimize them with minimal user input.

We previously showed how hyperparameter optimization for spectral neighbor analysis potential (SNAP; ref 29.) models can improve their prediction accuracy without increased cost at runtime.[30] Conversely, with poorly chosen hyperparameters, accuracy and robustness may suffer. We note that more widely, key advances have been made in recent years in terms of choosing hyperparameters for ML potentials.[30−34]

In this work, we study the effect of hyperparameter optimization for ACE ML potentials, fitted to existing high-quality data sets for silicon[35] and Sb_2Te_3.[7] We investigate both numerical and structural predictions and benchmark the performance of the resulting ACE models. For this purpose, we extended our automated hyperparameter optimization package, "Cross-platform optimizer for machine learning

Received: August 2, 2024
Revised: September 24, 2024
Accepted: October 16, 2024
Published: November 6, 2024

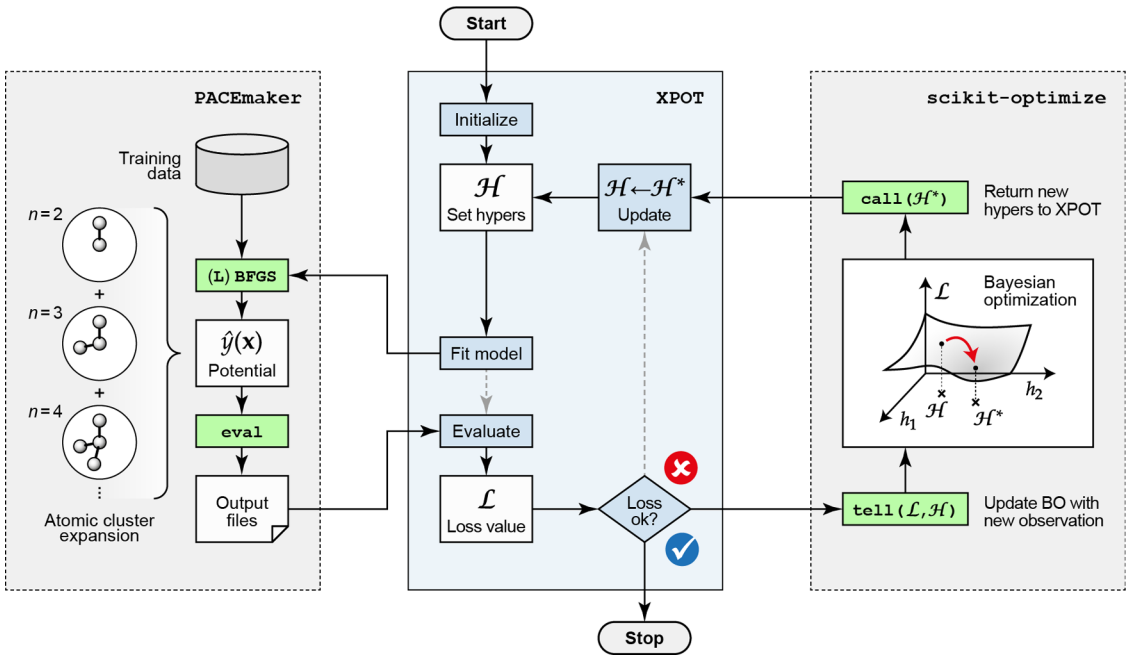


Figure 1. Overview of the methodology for automated optimization of ACE potentials, showing a simplified flowchart of the computational tasks involved. The core functionality of `XPOT` is highlighted in the central box in blue, and external calls to `PACEmaker` (left) and `scikit-optimize` (right) are indicated.

interatomic potentials" (XPOT),[30] to include support for ACE potentials fitted with `PACEmaker`[18] (Figure 1). In doing so, we emphasize that XPOT enables efficient transfer between fitting frameworks, here from GAP to ACE—giving access to potentials which are cheaper to run while maintaining comparable accuracy for the same training data.

METHODS

Atomic Cluster Expansion. The atomic cluster expansion (ACE) produces a complete set of basis functions which span the space of local atomic environments, as described in refs 16−18. In recent years, the ACE framework has shifted the Pareto front for cost and accuracy of ML potentials[17] and has underpinned the development of state-of-the-art interatomic potential models.[6,36−38] For example, a general-purpose ACE potential for carbon[6] outperforms previously reported models[39] in terms of both accuracy and speed, and an ACE model for Si−O was shown to accurately describe complex nanoscale structural features in this key binary system.[37]

ACE-based ML models are built from general atomic "properties", φ_i, expanded across body-ordered functions from the set of neighbors of each atom. Any given property of atom i (index p) can therefore be expanded as a function of this atom's local environment:

$$\varphi_i^{(p)} = \sum_{\mathbf{v}} c_{\mathbf{v}}^{(p)} \mathbf{B}_{i\mathbf{v}} \tag{1}$$

where the expansion coefficients $c_{\mathbf{v}}^{(p)}$ and basis functions ($\mathbf{B}_{i\mathbf{v}}$) share multi-indices $\mathbf{v}$ which describe the list of basis functions in a cluster. The potential energy of the ith atom—typically the prediction target for ML potential models—can be described as a combination of several atomic properties. In the simplest case, the atomic energy, ε_i, depends on only a single value:

$$\varepsilon_i = \varphi_i^{(1)} \tag{2}$$

which leads to a linear model. However, ACE potentials are not limited to using a single property: introducing nonlinear behavior through a Finnis−Sinclair-like embedding, for example, results in

$$\varepsilon_i = \varphi_i^{(1)} + \sqrt{\varphi_i^{(2)}} \tag{3}$$

where the two properties are generalizations of the pairwise repulsion and density of the widely used Finnis−Sinclair potential.[40] This approach can be further extended to an arbitrary number of P atomic properties (p = 1, ..., P) which all enter as arguments into a nonlinear function, denoted F:

$$\varepsilon_i = F(\varphi_i^{(1)}, ..., \varphi_i^{(P)}) \tag{4}$$

where i again is the atomic index. For example, in ref 37., it was shown that a complex functional form (P = 8) outperformed linear and Finnis−Sinclair type embeddings for ACE models for the Si−O system while only leading to a ≈15% increase in computational requirements. A full systematic study of the number, exponents, etc. of the terms that enter F is yet to be reported, and it is likely that there is some scope for optimizing nonlinear ACE models in this regard.

XPOT. We use our Python package, XPOT,[30] to fit optimized ACE models. XPOT automates the optimization of hyperparameters within ranges specified by the user, by iteratively minimizing a combined energy and force loss function which is defined as

$$\mathcal{L} = \frac{\alpha}{[E]} \times \mathcal{L}_{\mathrm{E}} + \frac{(1-\alpha)}{[\mathbf{F}]} \times \mathcal{L}_{\mathrm{F}} \tag{5}$$

In this expression, square brackets indicate the units (in the present work, we use energies in eV $n_{\mathrm{at}}^{-1/2}$ and forces in eV Å^{-1} respectively), emphasizing that the loss function, $\mathcal{L}$, itself is dimensionless. By adjusting the relative energy and force loss weighting, controlled by the single parameter α, the potential

https://doi.org/10.1021/acs.jctc.4c01012
J. Chem. Theory Comput. 2024, 20, 10103−10113

fit can be optimized toward the desired characteristics. XPOT evaluates $\mathcal{L}$ over a validation data set, which is not used in training the model.[30]

The energy contribution to the loss function, $\mathcal{L}_E$, is given by

$$\mathcal{L}_E = \sqrt{\frac{1}{N_{\text{cells}}} \sum_{i \in \mathbf{D}_{\text{val}}} \left(\frac{\hat{E}_i - E_i}{\sqrt{n_{\text{at},i}}} \right)^2} \tag{6}$$

where N_{cells} is the number of structures in the validation set, $\mathbf{D}_{\text{val}}$, and $n_{\text{at},i}$ is the number of atoms in the ith structure from that set. The energy prediction of the ACE model is $\hat{E}_i$, whereas the "correct" reference energy in the validation set is E_i. The difference between the two is divided by $\sqrt{n_{\text{at},i}}$—not $n_{\text{at},i}$—in order to weight the contribution of energy errors from each structure equally, regardless of the number of atoms in any given validation-set structure.[41]

The force contribution to the loss function,

$$\mathcal{L}_F = \sqrt{\frac{1}{N_{\text{at}}} \sum_{j \in \mathbf{D}_{\text{val}}} |\hat{\mathbf{F}}_j - \mathbf{F}_j|^2} \tag{7}$$

is determined by measuring the magnitude of the error between the predicted and reference force vectors. The force error is averaged over all N_{at} atoms in the data set.

Implementation of ACE Support in XPOT. Figure 1 lays out the workflow of XPOT, outlining the distribution and ordering of tasks within the hyperparameter optimization loop. This figure specifically relates to `PACEmaker` and ACE potentials, but the XPOT workflow is similar for other fitting methods. XPOT interfaces both to potential fitting software (left-hand side in Figure 1) and to the Bayesian optimization (BO) interface of `scikit-optimize` (right-hand side). The latter supports discrete and continuous variable optimization, as discussed in ref 30. The user-defined hyperparameter ranges are parsed, before potentials are fitted iteratively, and new hyperparameters are chosen through BO predictions until the loss reaches the desired value, or the maximum number of iterations is reached. A full description of the XPOT code can be found in ref 30.

To begin the process, the user initializes the optimization in Python, specifying an input script containing the hyperparameters to be optimized. XPOT parses this input and uses `scikit-optimize` to set values for each iteration. Then, `PACEmaker` is invoked, fitting the ACE potential model, and calculating the errors on both the training and validation data sets. XPOT parses the output files and converts the prediction errors into a single loss function, $\mathcal{L}$ (see eq 5). At this stage, if optimization is determined to be finished, XPOT exits, otherwise calling another estimation of the loss surface and updating the hyperparameters to continue the process.

While the optimization process in XPOT is agnostic to the specific fitting methodology, file parsing between XPOT and the desired fitting software must be implemented. Here, to reduce compute-time and installation requirements, while evaluation of the loss function is done by XPOT, we use the detailed per-structure predictions on the validation data set (performed by `PACEmaker` after fitting) as the errors to calculate the loss. As a result, unlike XPOT's interface to `fitsnap`, `LAMMPS` is not required for optimization of ACE potentials. GPU fitting with `PACEmaker` is supported in XPOT, and can be controlled by specifying the desired CUDA device at runtime using environment variables.

Data Sets. We briefly describe the data sets that we use in the present study for fitting and validating potentials. For details, we refer to the cited original publications; for an overview of validation techniques for ML potentials more generally, we refer to ref 41.

In the case of silicon, we use three different data sets for specific purposes. First, we take the training data set for the Si-GAP-18 general-purpose potential by Bartók et al.[35] as an example of a highly developed and largely handcrafted data set. We fit ACE models to this data set using XPOT, and we use the corresponding test set for validation (both taken from ref 35.). Second, we use DFT-labeled snapshots from a GAP-MD simulation described in ref 42. (referred to as "MQ-MD" in the following). This data set contains diamond-type supercells with vacancy defects, liquid, and amorphous structures, as well as transitions between these phases. The structures were generated with Si-GAP-18-driven MD and subsequently labeled with single-point DFT computations.[42] Finally, we test our XPOT-fitted models on random structure search (RSS) configurations from ref 43. These allow us to test the potentials' ability to describe structures different from those on which they were trained.

In Figure 2a, we characterize the various training and testing sets used by plotting the energy against the mass density for each structure. The bulk structures mainly have densities relevant to amorphous, crystalline, and liquid silicon, with sp-like chain structures constituting the data set entries at very low density and high energy. Despite the difference in the

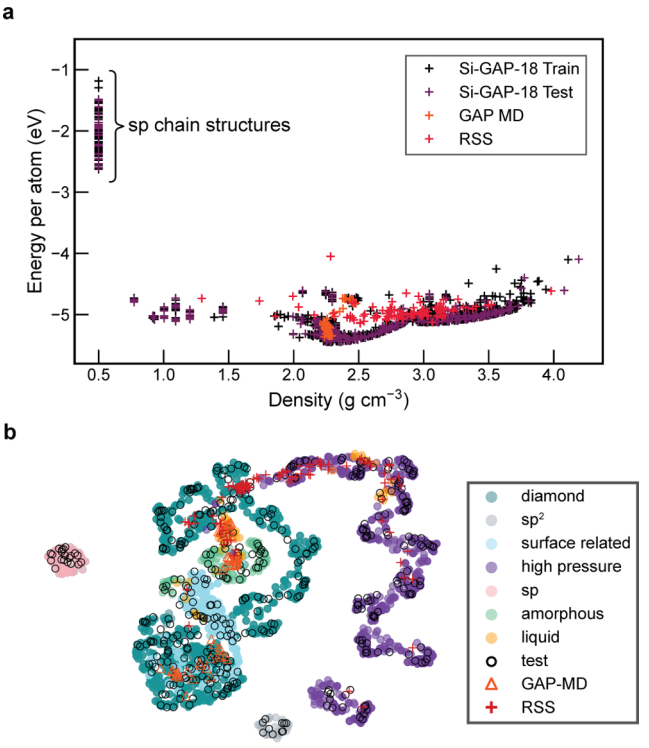


Figure 2. Visualization of the data for Si used in this work. (a) An energy−density plot showing the heterogeneous nature of the Si-GAP-18 training data set (ref 35.), with the "sp" chain structures highlighted. (b) UMAP embedding of the averaged ACE vectors for structures in the silicon data sets, highlighting the overlap in local atomic environments between structures. The categories are partly simplified compared to ref 35.: here, for example, "diamond" also includes defective diamond-type Si structures.

https://doi.org/10.1021/acs.jctc.4c01012
J. Chem. Theory Comput. 2024, 20, 10103−10113

composition of the RSS, MQ-MD, and Si-GAP-18 test data sets, the structures seem broadly related when compared in this plot.

In Figure 2b, we show a similarity map, obtained by UMAP dimensionality reduction[44] on the ACE vectors of the atomic environments averaged for each structure. This map indicates distinct regions, mirroring the varied structure types included in the data set. By color-coding the points in the map according to the configuration types defined in the Si-GAP-18 training data set,[35] we can visualize the isolated nature of the low-coordinate ("sp" and "sp^2") and high-pressure structures (β-Sn-type and simple hexagonal), and the relation between the various phases included. Additionally, we can visualize the coverage of the various test sets and their similarity to the training data. The RSS data set includes entries that (although clearly higher in energy; cf. Figure 2a) resemble diamond-like, high-pressure, and liquid-like structures. Finally, we note that all surface-related structures (slab models) in the Si-GAP-18 training set are based on the diamond-type form, and so the similarity of these structures to bulk diamond-type Si is expected.

The data set used in our study of Sb_2Te_3 is taken from the GST-GAP-22 data set.[7] In this case, we remove all structures which include Ge, yielding a data set for the binary Sb−Te system. We randomly split this data set into train and test data (80:20), for each configuration type as defined in the data set, such as "`crystalline`", "`aimd`", and "`liquid`". This way, we can ensure that the validation data span a wide range of structure types, and reduce the possibility of optimizing overly toward a single subgroup of the data set.

RESULTS AND DISCUSSION

Silicon(I): Numerical Performance. We fitted ACE potentials to the general-purpose Si-GAP-18 training data set from ref 35. Previously, Lysogorskiy et al. fitted a linear potential to the same data,[17] and several nonlinear models ($P > 1$) have been fitted recently for other materials.[6,37] Here, we create ACE models over the range of $1 \leq P \leq 4$, using fewer basis functions compared to ref 17., in order to improve efficiency at runtime. Our XPOT-ACE-2 model achieves accuracy comparable to the potential fitted in ref 17. ("Ref-ACE", Figure 3), taking advantage of optimized hyperparameters and nonlinear ACE models for improved efficiency.

XPOT optimizes hyperparameters by minimizing the combined loss, $\mathcal{L}$ (eq 5). As such, XPOT is guided by changes in the errors on the specific validation set (here, the Si-GAP-18 test set[35]). As the MQ-MD and RSS data sets are *not* used as targets in XPOT optimization, we leverage them as distinct benchmarks for accuracy and robustness. The MQ-MD data set from ref 42. is similar in character to the parts of the Si-GAP-18 test set, allowing further accuracy tests on larger structures. In contrast, the RSS data set contains randomly generated structures which are typically higher in energy (Figure 2a). No explicit RSS structures are present in either the Si-GAP-18 test or train sets. Using these two data sets, we can validate the performance of the Si-GAP-18 test set as a general-purpose optimization target.

For each value of P, we optimized up to 6 hyperparameters at once, and performed 32 fitting iterations. The first four of these were initialization fits (shaded areas in Figure 3), where a pseudorandom sampling method is used to determine hyperparameters, before Bayesian Optimization (BO) is applied for the remaining iterations. For all models, we

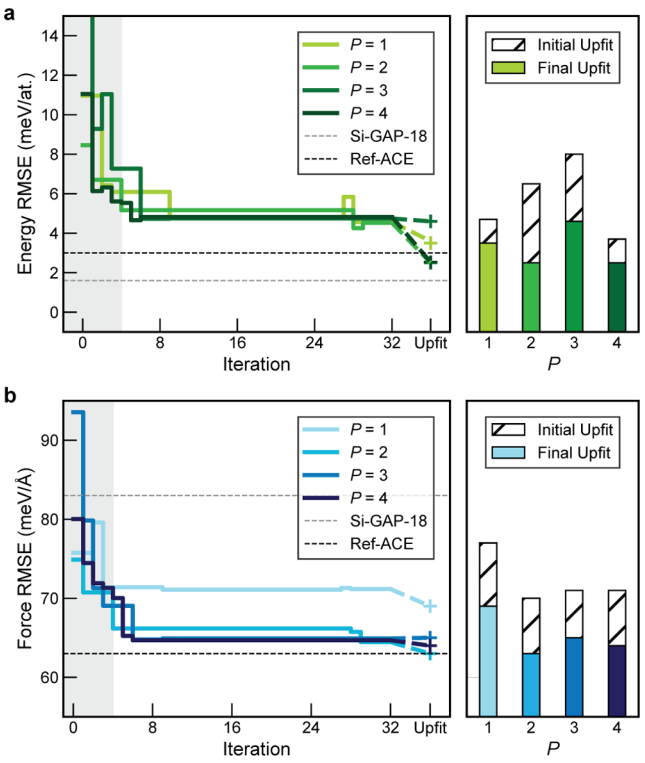


Figure 3. Evolution of energy and force errors for silicon ML potentials through iterative optimization using XPOT. We show (a) the per-atom energy RMSE and (b) the force component RMSE across iterations. Errors are evaluated on the Si-GAP-18 test set from ref 35. On the left, the gray region indicates the initial sampling stage of optimization (see ref 30.). The bar charts show the errors on the test set for "upfitted" potentials, as described in the text: both for the best potentials from the initialization protocol (hatched), and the best potentials after Bayesian optimization (solid). P refers to the number of atomic properties, as defined in the Methods section (cf. eq 4). The data for $P = 3$ and $P = 4$ refer to the "-3F" and "-4F" potentials from Table 1, respectively.

optimized φ_i exponents, cutoff, radial basis function (including `radbaseparameter`), and `dcut` (the smoothing distance at the outer limit of the cutoff). We used universal structure weighting—unlike REF-ACE, which includes weightings based on the type of structure in the training set.[17] This was done to show that even without a "bespoke" approach to weighting data set entries, hyperparameter optimization could improve ML potentials. In future work, we hope to study the effect of weighting techniques to further improve fitting for smaller data sets.

After the optimization, the best potential for each optimization sweep was "upfitted". By this term, we mean the approach of continuing the fitting process from an existing potential. PACEmaker uses a ratio value, κ, to determine the relative weighting of energy and force errors in the fitting process. A higher value of κ to prioritizes forces when fitting, and a lower κ value prioritizes energies. Upfitting potentials (by varying κ) can potentially improve the accuracy of ACE models.[45] This process is analogous to the stage two (`--swa`) setting used in MACE fitting, whereby energy weighting is increased for the final 20% of the fitting process to improve energy errors.[46] Herein, we first fitted using $\kappa = 0.8$, before upfitting the same potential with $\kappa = 0.02$. This process was the same for all XPOT-ACE potentials in the present work. After

https://doi.org/10.1021/acs.jctc.4c01012
J. Chem. Theory Comput. 2024, 20, 10103−10113

Table 1. Energy and Force RMSE Values of Silicon Potentials, Evaluated on Three Different Test Sets[a]

			energy RMSE (meV at.$^{-1}$)			force RMSE (meV Å^{-1})			
	P	# func.	Si-GAP-18-test[35]	MQ-MD[42]	RSS[43]	Si-GAP-18-test	MQ-MD	RSS	MD speed
XPOT-ACE-1	1	3000	3.5	5.1	27.1	69	105	158	47
XPOT-ACE-2	2	3000	2.5	5.0	23.1	63	97	150	46
XPOT-ACE-3	3	3000	318	5.2	>10^6	300	98	>10^8	45
XPOT-ACE-4	4	3000	4.8	5.5	62.6	63	99	274	43
XPOT-ACE-3F	3	2000	4.6	4.1	20.5	65	97	139	65
XPOT-ACE-4F	4	1625	2.5	5.4	72.5	64	100	187	80
XPOT-ACE-6827	1	6827	3.0	4.4	34.5	63	104	179	16
REF-ACE[17]	1	6827	3.2	4.3	42.1	77	124	175	16
Si-GAP-18[35]	N/A	N/A	1.6	8.5	34.9	83	139	177	1

[a]For the XPOT-ACE models, the number of atomic properties, P, as defined in eq 4, ranges from 1 to 4. An "F" denotes a potential where the number of functions was optimized by XPOT. XPOT-ACE-6827 is an optimized model using the same number of radial basis functions as the linear ACE potential fitted by Lysogorskiy et al.[17] (denoted as "REF-ACE" here). The MD speed is given relative to that of Si-GAP-18.

the final potentials were upfitted, we performed numerical validation on two external data sets (see Methods section).

Initially, we fitted potentials with 3000 functions each for all values of P, but found an increased likelihood of overfitting for $P \geq 3$. Table 1 includes results from optimizing potentials for $P = 3$ and $P = 4$ with 3000 functions, highlighting the reduced accuracy and robustness compared to the $P \leq 2$ potentials. Specifically, the XPOT-ACE-3 potential has become much worse during the upfitting procedure, producing highly unreasonable energy and force errors on the RSS test set. When analyzing the errors per structure for the upfitted XPOT-ACE-3 potential further, it was found that the increased error on the Si-GAP-18 test set comes from two individual structures. We emphasize that these issues would not have become apparent from testing only on the MQ-MD data set, and instead a more wide-ranging analysis including RSS test data reveals that the potential is not sufficiently robust. The nonupfitted potential result from XPOT has a testing error of 7.9 meV at.$^{-1}$—note that while this seems to be an acceptable result, it is already 4× the training error, and thus this potential as well appears to be overfitted.

As such, we used XPOT to optimize the number of functions for the $P \geq 3$ potentials as an additional variable hyperparameter, from 500 up to an upper limit of 2000. By reducing the maximum number of functions, we restricted the flexibility of these models, to optimize toward more robust behavior. We note that XPOT-ACE-3F uses the maximum number of functions (2000), while XPOT-ACE-4F is optimized to 1625 functions. These optimized potentials are denoted with an "F" suffix and the results are included in Table 1. In addition to having improved numerical accuracy and efficiency, these potentials were able to complete melt−quench simulations for 4096-atom cells (Table 2), which the XPOT-ACE-3 and XPOT-ACE-4 potentials were not. However, as the optimization still occurs for the Si-GAP-18 test set, a good performance on RSS data is still not guaranteed, as shown by the high RSS error for XPOT-ACE-4F.

The evolution of energy and force errors is visualized in Figure 3, including the upfitting procedure which is undertaken at the end of the optimization process, as described above. For comparison, we also upfitted the best potentials from the initialization phase, to assess whether the BO part of the process brings additional benefit. We see that across all values of P, the final optimized potentials are more accurate on the Si-GAP-18 test set used to evaluate $\mathcal{L}$.

Table 2. Energy Per-Atom Differences of Quenched a-Si Structures Relative to (Diamond-Type) Crystalline Silicon[a]

quench rate (K s^{-1})	ΔE (meV at.$^{-1}$)			
	XPOT-ACE quenched			
	XPOT	REF	GAP	DFT
10^{14}	205(6)	209(6)	215(7)	210(7)
10^{13}	183(4)	187(4)	192(4)	185(5)
10^{12}	158(3)	161(3)	166(4)	156(4)
10^{11}	140(4)	142(5)	146(5)	137(5)
	REF-ACE quenched			
	XPOT	REF	GAP	DFT
10^{14}	211(4)	208(4)	218(3)	217(6)
10^{13}	190(4)	187(3)	197(4)	185(5)
10^{12}	160(4)	158(3)	166(4)	155(5)
10^{11}	147(4)	145(4)	153(4)	144(5)

[a]Structures were quenched and relaxed using the ML potentials in the top row, comparing the differences between XPOT-ACE and REF-ACE.[17] The standard deviation across the sampled structures is reported using parentheses.

The upfitting process for the initial potentials does not always improve testing error, despite the training error reducing throughout the fitting process. This suggests that the initial hyperparameters selected are not providing a generalizable description of silicon, indicated by relatively poor predictive accuracy on unseen environments. Additionally, not all upfitted initial potentials were found to be stable in MD simulations up to 1800 K. We further quantify the accuracy of these potentials in the Supporting Information.

For the nonlinear XPOT-ACE-4F model, we observe that the robustness and accuracy on the test set are sufficient, but it is prone to large errors on other data sets—that is, the model appears to be overfitted. This finding again demonstrates that, as we discussed in ref 30., the makeup of the validation set is paramount in optimizing a potential toward the desired characteristics. Additionally, setting bounds for the flexibility of the potential reduces the likelihood of overfitting, but the composition of the validation set still directly affects the loss function, and thus the minimum in hyperparameter space toward which the potentials are optimized.

Taking into account the results above, XPOT-ACE-2 seems to offer the best combination of accuracy and robustness among the entries of Table 1. We therefore take this potential

https://doi.org/10.1021/acs.jctc.4c01012
J. Chem. Theory Comput. 2024, 20, 10103−10113

forward for further testing, and from here onward we refer to it simply as "XPOT-ACE".

Silicon(II): Physics-Guided Validation. We plot the errors for each structural snapshot from the MQ-MD test set in Figure 4. Doing so provides a more nuanced view of the types of configurations for which the potential shows higher or lower errors.[42] We show that XPOT-ACE improves the force predictions compared to REF-ACE, while having very slightly higher errors (to within 1 meV at.$^{-1}$). Our potential is comparatively most improved for the higher-energy liquid structures, while improvements for crystalline and amorphous structures are reduced, which we presume to be due to the uniform structure weighting. Notably, the higher weighting of the crystalline phases for existing potentials[17,35] offers improved accuracy for crystalline configurations, whereas our potential still improves the force accuracy for those.

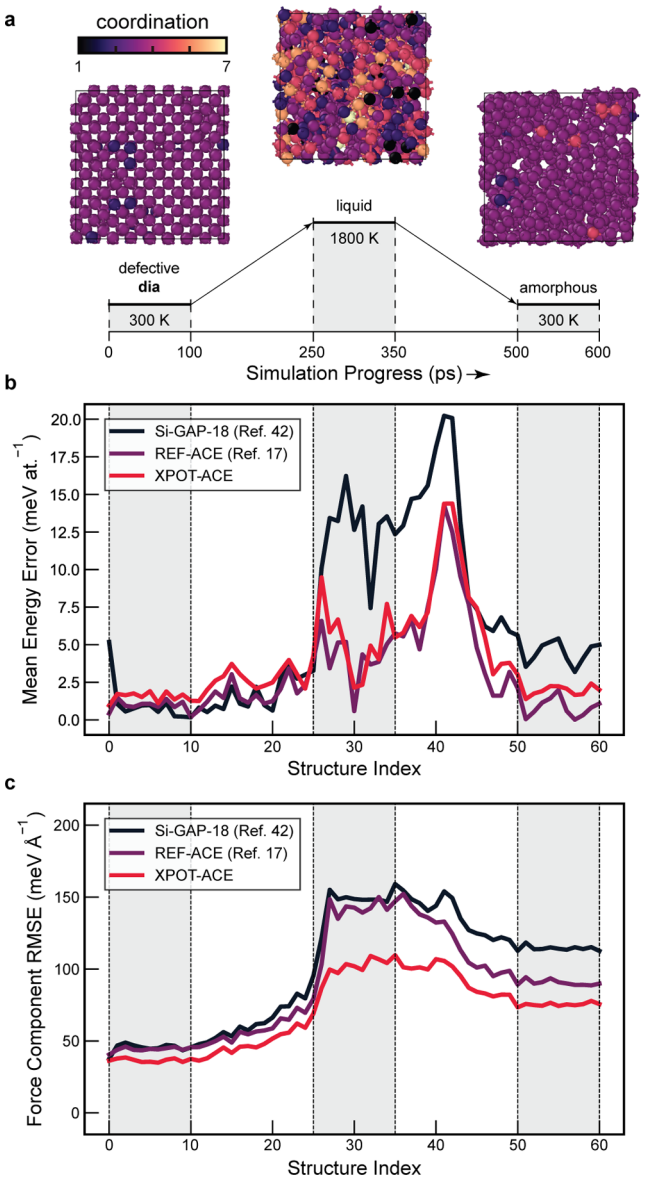


Figure 4. Accuracy of silicon ML potentials evaluated on DFT-labeled snapshots from a Si-GAP-18-driven melt−quench simulation reported in ref 42., and visualized in the style of that prior work. (a) An overview of the constant-pressure simulation protocol, adapted from ref 42. The images show the three classes of structure seen at each stage of the simulation, color-coded according to coordination number. DFT snapshots were computed every 10 ps throughout the simulation. (b) Energy errors compared to DFT snapshots along the simulation trajectory. (c) Force errors for the same structures. XPOT-ACE outperforms all other potentials studied here in terms of force errors, but trails in energy errors to the ACE from ref 17. The schematic in panel (a) and the overall style are adapted from ref 42., which is published under a CC BY license (https://creativecommons.org/licenses/by/4.0/).

Across the simulation snapshots, the same structures are resulting in "spiking" (high error) energies across all three models. This suggests some underlying characteristic of the Si-GAP-18 training data set results in these less accurate predictions across models, especially on freezing of the liquid state (which is not strongly represented in the training data). For the latter, all three potentials show increased energy errors. The fluctuations are much reduced in force predictions, but there is still a visible "bump" in these predictions for the same structures.

Next, we quenched 500-atom randomized structures at fixed rates from 10^{14} to 10^{11} K s^{-1}. Both XPOT-ACE and REF-ACE potentials were used and we compared the energies of the structures produced. On top of labeling these structures with XPOT-ACE, REF-ACE, and Si-GAP-18, we compute DFT energies for the structures, to quantify the predictive accuracy of all three potentials (akin to our studies of quenched SiO_2 in ref 49). For each rate, 5 random structures were quenched, with both mean and standard deviation reported in Table 2. We tested faster quench rates, but found that rates of 10^{14} and 10^{15} K s^{-1} led to structures within 2 meV at.$^{-1}$ of each other for all models. We therefore do not include results for quench rates of >10^{14} K s^{-1} in Table 2.

Although the energies of structures for 10^{12} and 10^{13} K s^{-1} are very similar, XPOT-ACE predicts structures with slightly lower energies for 10^{11} K s^{-1}, suggesting a more relaxed and therefore more stable a-Si structure. All potentials provide a good match to DFT. When compared to our findings in Table 1 and Figure 4, this is in line with our expectations.

Finally, we performed structural validation tests for the compression of silicon, as described in ref 47. This test was comprised of compressing a 100,000 atom low density amorphous silicon model up to 20 GPa. The pressurization rate is 0.1 GPa ps^{-1}, and the temperature is held at 500 K. We observed similar behavior to what was seen in simulations using Si-GAP-18,[47] whereby low-density amorphous (LDA) silicon upon compression collapses into a very-high-density amorphous (VHDA) phase at ≈12 GPa, and subsequently simple-hexagonal (sh) crystallites nucleate and grow. This result demonstrates that our XPOT-ACE model has learned much of the same behavior as Si-GAP-18 from the training set.

The formation of the VHDA phase in the compression simulation the using XPOT-ACE potential occurred at a pressure 0.5 GPa higher than that observed for Si-GAP-18, as shown by the relative lack of highly coordinated silicon atoms in Figure 5b—an effect which is corroborated by Figure 5e, where the percentage of N = 8 atoms does not rise substantially until 12.5 GPa, in contrast to the Si-GAP-18 simulation. Aside from this slightly delayed formation of VHDA, the potentials predicted almost identical densities throughout the simulation, and show behavior that is consistent with experiments for both VHDA formation[50] and crystallization.[51]

Antimony Telluride. To test our approach for a more complex material system, we fitted a potential for Sb_2Te_3,

https://doi.org/10.1021/acs.jctc.4c01012
J. Chem. Theory Comput. 2024, 20, 10103−10113

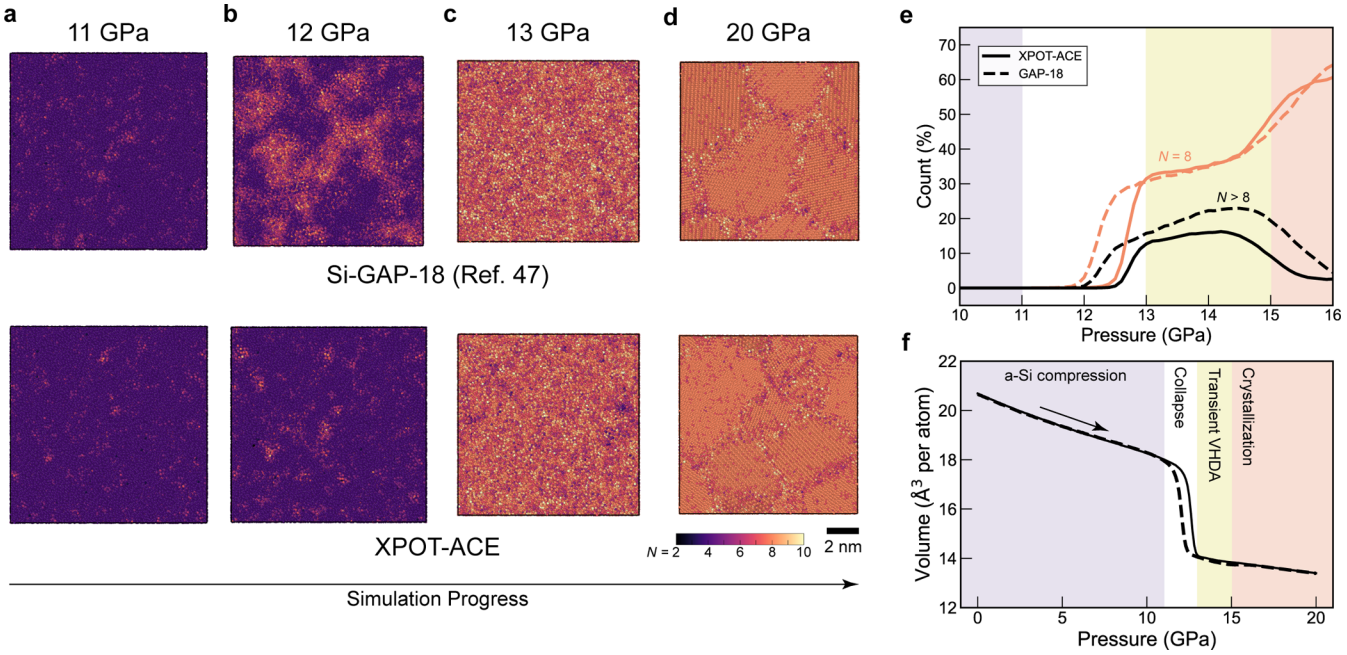


Figure 5. (a−d) Simulations of amorphous silicon under isothermal compression with both XPOT-ACE-2 ("XPOT-ACE" for brevity) and Si-GAP-18,[35] similar to the simulations reported in ref 47. from which data for the GAP simulation are taken. Both potentials predict a collapse into VHDA occurring between 12−13 GPa from which simple hexagonal crystallites then form. (e) Coordination numbers as a function of pressure in the trajectories (determined by counting neighbors up to 2.85 Å). The initial increase in $N > 8$ atoms corresponds to the formation of the VHDA phase before crystallization occurs. (f) Volume against pressure during the simulations. Simulations were carried out using LAMMPS.[48] The results are visualized in a similar way as in ref 47.

which is an important chalcogenide material used in various phase-change materials (PCM)-based devices for ultrafast data storage[54] and high-performance neuromorphic computing tasks.[55,56] PCMs have long served as key application cases for ML potentials, including early work on the binary material GeTe[57] and the ternary $Ge_2Sb_2Te_5$.[58] The former potential has been used for studies of crystallization[59] and thermal properties of GeTe;[60] the latter has been applied to study structure and bonding in $Ge_2Sb_2Te_5$[61] and tested for the binary Sb_2Te_3.[62]

Our previous work has introduced an ML potential based on the GAP framework for Ge−Sb−Te (GST) alloys located along the compositional tie-line between GeTe and Sb_2Te_3.[7] This GAP model, which we call "GST-GAP-22", can accurately describe disordered structures of GST alloys and complex phase transition processes under practical programming conditions (e.g., nonisothermal heating) on the length scale of real-world devices.[7] We took a subset of the GST-GAP-22 data set, which only contains elemental crystal structures of Sb and Te as well as binary bulk structures (including crystalline, amorphous, and intermediate crystallization configurations) found in the Sb−Te system.

A validation set was created using the protocol described in the Methods section, providing a representative sample of structures, and allowing us to consistently quantify the performance and robustness of our potentials. This is important for defining the loss function in XPOT (cf. eq 5), and so the validation set is fixed for all potentials fitted during optimization.

Additionally, to confirm that the reduction in scope of the training set (from the full Ge−Sb−Te system to only the Sb−Te system) was not unfairly advantaging our own optimized potentials, we fitted a GAP with the same hyperparameters as for the GST-GAP-22 potential,[7] but using only the Sb−Te subset of the data for training. We use this potential ("SbTe-GAP" in the following) as a benchmark as it provides very similar force errors to GST-GAP-22, but resulted in improved accuracy for energy prediction.

In this case, we cannot directly compare to GST-GAP-22 numerically, as the validation data are taken from the GST-GAP-22 training data set. Therefore, to quantify the accuracy of the potentials, we created a new benchmark data set similar to that built for silicon in ref 42. (cf. Figure 4). Specifically, we performed a GAP-MD simulation in which Sb_2Te_3 was melted starting from a defective rocksalt-like crystal with an anion (Te) vacancy, before being quenched at 10^{13} K s^{-1} to form the amorphous phase. This trajectory was then labeled with DFT to produce a benchmark set representing crystalline, liquid, and amorphous Sb_2Te_3. Snapshots of these three phases are shown in Figure 6a. The atoms are color-coded by crystallinity as defined by the Smooth Overlap of Atomic Positions (SOAP)-based similarity,[52] with respect to rocksalt-like Sb_2Te_3, as discussed in ref 53.

In Figure 6, we show that our XPOT-optimized ACE potential has improved not only upon the GST-GAP-22 predictions, but also the SbTe-GAP fitted on the Sb−Te subset of the full GST-GAP-22 database. The numerical accuracy is improved, and predictions are over 400× faster than for the GAP potentials. Figure 6b shows that the energy errors on the defective crystalline structures are consistently higher than for the liquid or amorphous phases across all potentials, with the same structural snapshots showing higher errors in this region. This occurs due to the relatively small number of crystalline structures with vacancy defects in the training database, and while our potentials all predict the energy to within 15 meV at.$^{-1}$, we see that these structures present more of a challenge

https://doi.org/10.1021/acs.jctc.4c01012
J. Chem. Theory Comput. 2024, 20, 10103−10113

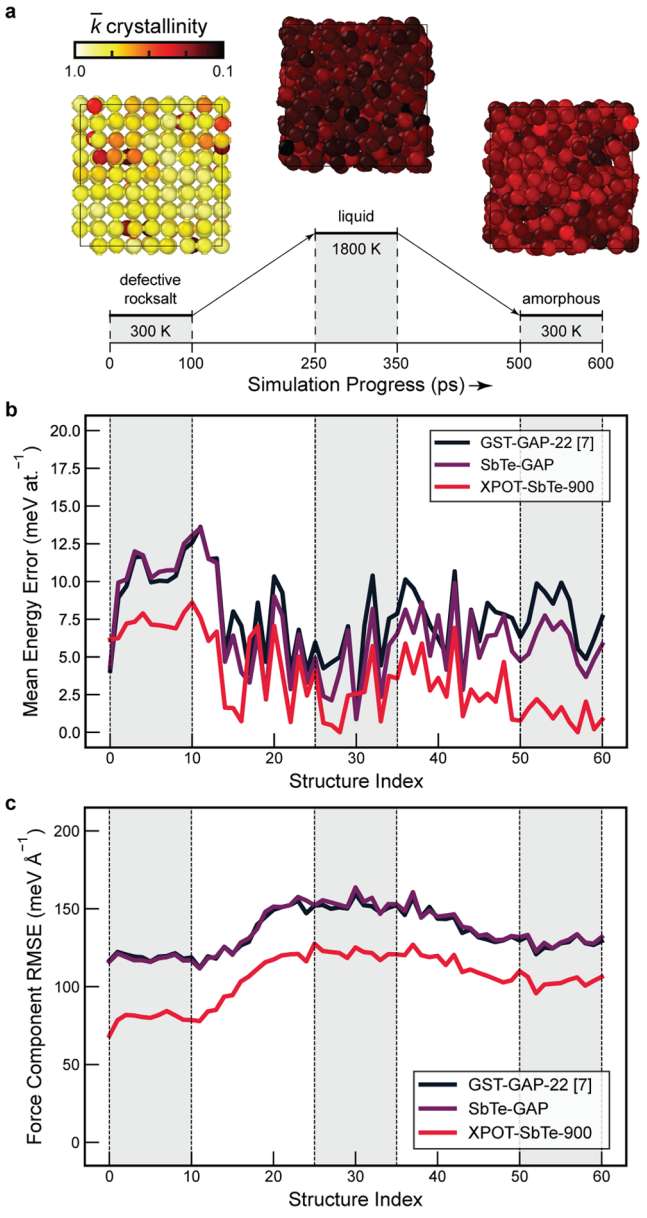


Figure 6. Prediction accuracy of Sb_2Te_3 ML potentials across DFT-labeled snapshots from a GST-GAP-22 melt-quench simulation. (a) An overview of the MD-based benchmark protocol created in a way similar to ref 42., now for Sb_2Te_3. Structures show the three classes of structures seen at each stage of the simulation, color-coded according to a per-atom crystallinity measure.[52,53] Snapshots were labeled with DFT every 10 ps throughout the simulation. (b) Energy errors compared to DFT snapshots across the simulation trajectory. (c) Force errors across the same structures. XPOT-ACE outperforms both GAP potentials on both energy and force errors.

to these ML potentials than the liquid or amorphous structures do. The SbTe-XPOT-ACE potential is significantly more accurate than both GAPs in predicting energies and forces for amorphous structures, and the force predictions are uniformly over 10% closer to DFT than those of by either GAP potential.

We carried out further physically guided validation, testing the similarity of structural predictions between AIMD, GST-GAP-22, and our XPOT-ACE potential, aiming to verify that numerical errors correspond to well-described physical properties and processes in simulation. We first characterize the structure of amorphous Sb_2Te_3. We ran a melt-quench simulation on a crystalline Sb_2Te_3 model containing 360 atoms. In Figure 7a–b, we show that our liquid and amorphous structures generated using SbTe-XPOT-ACE closely match the RDF and ADF of structures generated using AIMD and GST-GAP-22, including the shoulder in the ADF of the liquid phase. Both GST-GAP-22 and the XPOT-ACE potential marginally overorder the amorphous phase of Sb_2Te_3, evidenced by a slightly larger peak at 90° in the ADF and a larger second peak of the RDF, as compared to the AIMD reference (Figure 7b).

To move beyond structural validation, we produced a 9600-atom structural model of amorphous Sb_2Te_3 (in a box of $4.3 \times 9.0 \times 8.5$ nm^3) with a prefixed crystalline template to simulate the crystal-growth process. Figure 7c shows a crystallization simulation for this templated structural model. As in Figure 6, we used SOAP-based similarity[52] with respect to rocksalt-like Sb_2Te_3 to quantify the per-atom crystallinity during the crystallization process.[53] Our structural model was annealed at 600 K for 4 ns, and the growth proceeded quickly at the rough crystalline–amorphous interface. Upon nanosecond crystallization, we found many defects (e.g., point defects and layer stacking faults) in the recrystallized model (cf. the dark red atoms in Figure 5c), indicating competing growth of different crystalline regions with different crystal orientations. We note that such local disorder is challenging to fully characterize due to the short time scales on which they happen (e.g., in the programming operations of real-world devices), and high-temperature annealing can help to eliminate the local defects, e.g., via vacancy ordering,[63] resulting in an energetically more favorable crystalline phase with fewer defects.[64]

CONCLUSIONS

Hyperparameter optimization can improve the accuracy and efficiency of ML potentials. Alongside data set construction and fitting architectures, the choice of hyperparameters remains an important aspect of fitting performant ML models. Well-chosen hyperparameters can lead to more accurate, and more efficient, ML potentials for accelerated materials modeling.

In the present work, we have described an extension of our openly available XPOT code to ACE model fitting via `PACEMaker`.[17,18] We have shown example applications for Si and Sb_2Te_3, two systems with diverse chemistry, across a wide range of configurational space (including liquid, amorphous, and crystalline phases). We thoroughly validated these potentials using a number of numerical and physical tests, and we believe that those types of benchmarks can be useful for other systems as well (see also ref 41). In particular, expanding on ref 30., we studied how suitable different types of test sets are to assess the quality of ML potentials across a range of structures—including "external" test sets that are different from the validation set used during optimization.

As ML potentials become larger and more complex, selecting suitable hyperparameters for fitting has become increasingly important to maximize performance and accuracy for a given training data set. Furthermore, our results in Table 1 suggest that series of potential fits with systematically varied hyperparameters could help to diagnose possible areas of failures for candidate ML potentials: specifically, the XPOT-ACE-3 model has low numerical errors on static MQ-MD snapshots but fails when used to drive MD itself, and this may be correlated with an extremely high prediction error on the RSS test set, which is generated distinctly from both the

https://doi.org/10.1021/acs.jctc.4c01012
J. Chem. Theory Comput. 2024, 20, 10103–10113

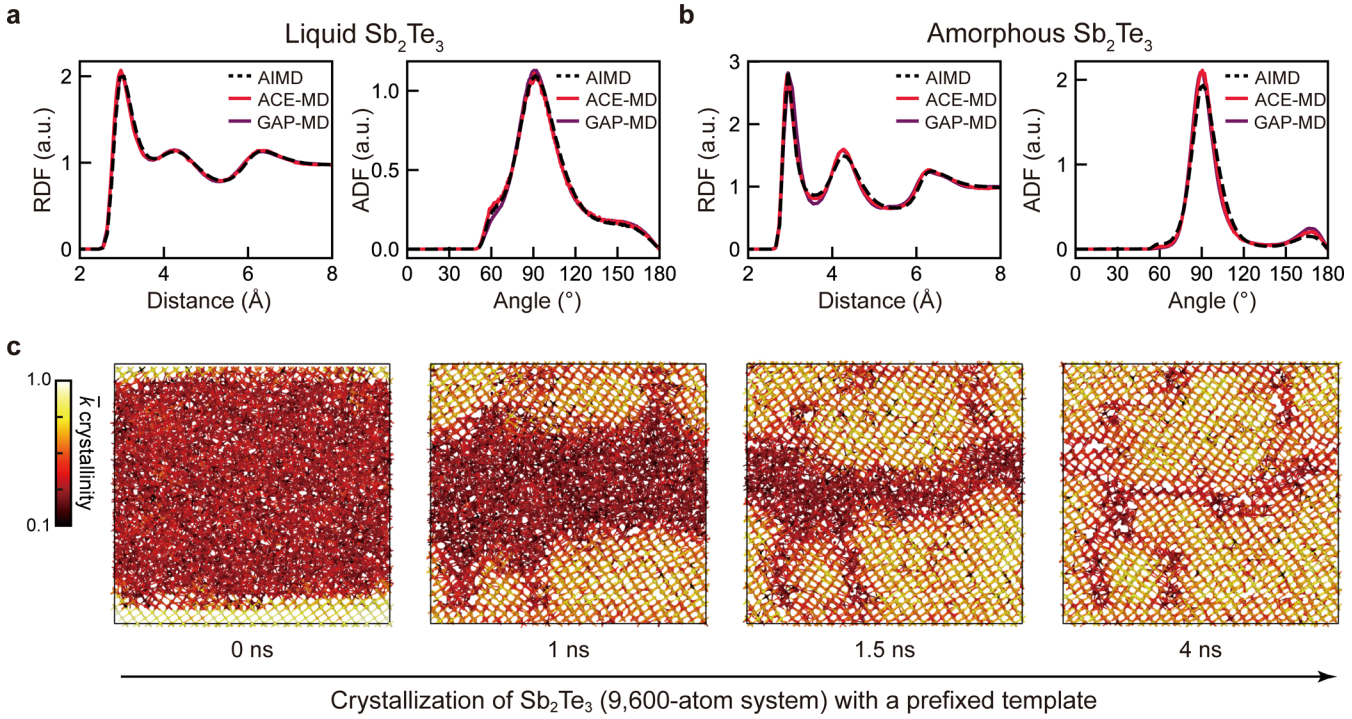


Figure 7. Structure and dynamics of Sb_2Te_3. (a–b) RDF and ADF plots for liquid and amorphous structures simulated by the SbTe-XPOT-ACE (red) and GST-GAP-22 (purple) models as well as AIMD (black). The data for the latter two are taken from ref 7. (c) Snapshots of an MD crystal-growth simulation. SOAP similarity is used to highlight the growth of crystalline Sb_2Te_3 (yellow) across snapshots. Several crystal grains and grain boundaries are visible. Simulations were carried out using LAMMPS.[48]

physically motivated MQ-MD and test data sets. In this way, series of XPOT runs might help users to judge whether more data are required, or whether the model hyperparameters require improvement. We hope for this work to contribute to the wider uptake of ML potentials and their application to increasingly challenging research questions in chemistry and materials science.

■ ASSOCIATED CONTENT

Data Availability Statement

The XPOT Python code is openly available via GitHub at https://github.com/dft-dutoit/xpot. XPOT includes interfaces to third-party software, including `PACEmaker`[17] which is freely available for academic noncommercial use. Data and code to reproduce the results from the present paper will be provided openly upon journal publication.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.4c01012.

Additional numerical results and technical information as a supplement to Table 1, and hyperparameter ranges used for optimization (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Volker L. Deringer − *Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford OX1 3QR, U.K.*; orcid.org/0000-0001-6873-0278; Email: volker.deringer@chem.ox.ac.uk

Authors

Daniel F. Thomas du Toit − *Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford OX1 3QR, U.K.*; orcid.org/0000-0002-4952-6076

Yuxing Zhou − *Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford OX1 3QR, U.K.*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.4c01012

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank John Gardner, Minaam Qamar, Dr. Anton Bochkarev, Dr. Yury Lysogorskiy, and Prof. Ralf Drautz for helpful discussions and Thomas Nicholas for testing early versions of the code. This work was supported through a U.K. Research and Innovation Frontier Research grant [grant number EP/X016188/1]. Y.Z. acknowledges a China Scholarship Council-University of Oxford scholarship. We are grateful for computational support from the U.K. national high performance computing service, ARCHER2, for which access was obtained via the UKCP consortium and funded by EPSRC grant ref EP/X035891/1.

■ REFERENCES

(1) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828–12840.

(2) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.* **2019**, *31*, No. 1902765.

(3) Noé, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annu. Rev. Phys. Chem.* **2020**, *71*, 361–390.

(4) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142−10186.
(5) Friederich, P.; Häse, F.; Proppe, J.; Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **2021**, *20*, 750−761.
(6) Qamar, M.; Mrovec, M.; Lysogorskiy, Y.; Bochkarev, A.; Drautz, R. Atomic Cluster Expansion for Quantum-Accurate Large-Scale Simulations of Carbon. *J. Chem. Theory Comput.* **2023**, *19*, 5151−5167.
(7) Zhou, Y.; Zhang, W.; Ma, E.; Deringer, V. L. Device-scale atomistic modelling of phase-change memory materials. *Nat. Electron.* **2023**, *6*, 746−754.
(8) Ouyang, Y.; Yu, C.; He, J.; Jiang, P.; Ren, W.; Chen, J. Accurate description of high-order phonon anharmonicity and lattice thermal conductivity from molecular dynamics simulations with machine learning potential. *Phys. Rev. B* **2022**, *105*, No. 115202.
(9) Rinaldi, M.; Mrovec, M.; Bochkarev, A.; Lysogorskiy, Y.; Drautz, R. Non-collinear magnetic atomic cluster expansion for iron. *npj Comput. Mater.* **2024**, *10*, No. 12.
(10) Li, F.; Cheng, X.; Lu, L.-L.; Yin, Y.-C.; Luo, J.-D.; Lu, G.; Meng, Y.-F.; Mo, H.; Tian, T.; Yang, J.-T.; Wen, W.; Liu, Z.-P.; Zhang, G.; Shang, C.; Yao, H.-B. Stable All-Solid-State Lithium Metal Batteries Enabled by Machine Learning Simulation Designed Halide Electrolytes. *Nano Lett.* **2022**, *22*, 2461−2469.
(11) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, No. 146401.
(12) Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121*, 10037−10072.
(13) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, No. 136403.
(14) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian Process Regression for Materials and Molecules. *Chem. Rev.* **2021**, *121*, 10073−10141.
(15) Klawohn, S.; Darby, J. P.; Kermode, J. R.; Csányi, G.; Caro, M. A.; Bartók, A. P. Gaussian approximation potentials: Theory, software implementation and application examples. *J. Chem. Phys.* **2023**, *159*, No. 174108.
(16) Drautz, R. Atomic Cluster Expansion for Accurate and Transferable Interatomic Potentials. *Phys. Rev. B* **2019**, *99*, No. 014104.
(17) Lysogorskiy, Y.; van der Oord, C.; Bochkarev, A.; Menon, S.; Rinaldi, M.; Hammerschmidt, T.; Mrovec, M.; Thompson, A.; Csányi, G.; Ortner, C.; Drautz, R. Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon. *npj Comput. Mater.* **2021**, *7*, No. 97.
(18) Bochkarev, A.; Lysogorskiy, Y.; Menon, S.; Qamar, M.; Mrovec, M.; Drautz, R. Efficient parametrization of the atomic cluster expansion. *Phys. Rev. Mater.* **2022**, *6*, No. 013804.
(19) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, No. 2453.
(20) Batatia, I.; Batzner, S.; Kovács, D. P.; Musaelian, A.; Simm, G. N. C.; Drautz, R.; Ortner, C.; Kozinsky, B.; Csányi, G. The Design Space of E(3)-Equivariant Atom-Centered Interatomic Potentials, arXiv:2205.06643. arXiv.org e-Print archive, 2022. https://arxiv.org/abs/2205.06643.
(21) Chen, C.; Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2022**, *2*, 718−728.
(22) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling. *Nat. Mach. Intell.* **2023**, *5*, 1031−1041.
(23) Stocker, S.; Gasteiger, J.; Becker, F.; Günnemann, S.; Margraf, J. T. How robust are modern graph neural network potentials in long and hot molecular dynamics simulations? *Mach. Learn.: Sci. Technol.* **2022**, *3*, No. 045010.
(24) Gardner, J. L. A.; Baker, K. T.; Deringer, V. L. Synthetic pre-training for neural-network interatomic potentials. *Mach. Learn.: Sci. Technol.* **2024**, *5*, No. 015003.
(25) Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P. Performance and Cost Assessment of Machine Learning Interatomic Potentials. *J. Phys. Chem. A* **2020**, *124*, 731−745.
(26) Ben Mahmoud, C.; Gardner, J. L. A.; Deringer, V. L. Data as the next challenge in atomistic machine learning. *Nat. Comput. Sci.* **2024**, *4*, 384−387.
(27) Batatia, I.; Benner, P.; Chiang, Y.; Elena, A. M.; Kovács, D. P.; Riebesell, J.; Advincula, X. R.; Asta, M.; Avaylon, M.; Baldwin, W. J.; Berger, F.; Bernstein, N.; Bhowmik, A.; Blau, S. M.; Cărare, V.; Darby, J. P.; De, S.; Della Pia, F.; Deringer, V. L.; Elijošius, R.; El-Machachi, Z.; Falcioni, F.; Fako, E.; Ferrari, A. C.; Genreith-Schriever, A.; George, J.; Goodall, R. E. A.; Grey, C. P.; Grigorev, P.; Han, S.; Handley, W.; Heenen, H. H.; Hermansson, K.; Holm, C.; Jaafar, J.; Hofmann, S.; Jakob, K. S.; Jung, H.; Kapil, V.; Kaplan, A. D.; Karimitari, N.; Kermode, J. R.; Kroupa, N.; Kullgren, J.; Kuner, M. C.; Kuryla, D.; Liepuoniute, G.; Margraf, J. T.; Magdău, I.-B.; Michaelides, A.; Moore, J. H.; Naik, A. A.; Niblett, S. P.; Norwood, S. W.; O'Neill, N.; Ortner, C.; Persson, K. A.; Reuter, K.; Rosen, A. S.; Schaaf, L. L.; Schran, C.; Shi, B. X.; Sivonxay, E.; Stenczel, T. K.; Svahn, V.; Sutton, C.; Swinburne, T. D.; Tilly, J.; van der Oord, C.; Varga-Umbrich, E.; Vegge, T.; Vondrák, M.; Wang, Y.; Witt, W. C.; Zills, F.; Csányi, G. A foundation model for atomistic materials chemistry, arXiv:2401.00096. arXiv.org e-Print archive, 2024. https://arxiv.org/abs/2401.00096.
(28) Yang, H.; Hu, C.; Zhou, Y.; Liu, X.; Shi, Y.; Li, J.; Li, G.; Chen, Z.; Chen, S.; Zeni, C.; Horton, M.; Pinsler, R.; Fowler, A.; Zügner, D.; Xie, T.; Smith, J.; Sun, L.; Wang, Q.; Kong, L.; Liu, C.; Hao, H.; Lu, Z. MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures, arXiv:2405.04967. arXiv.org e-Print archive, 2024. https://doi.org/10.48550/arXiv.2405.04967.
(29) 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.* **2015**, *285*, 316−330.
(30) Thomas du Toit, D. F.; Deringer, V. L. Cross-platform hyperparameter optimization for machine learning interatomic potentials. *J. Chem. Phys.* **2023**, *159*, No. 024803.
(31) Natarajan, S. K.; Caro, M. A. Particle Swarm Based Hyper-Parameter Optimization for Machine Learned Interatomic Potentials, arXiv:2101.00049. arXiv.org e-Print archive, 2021 https://doi.org/10.48550/arXiv.2101.00049.
(32) Adams, B. M.; Bohnhoff, W. J.; Dalbey, K. R.; Ebeida, M. S.; Eddy, J. P.; Eldred, M. S.; Hooper, R. W.; Hough, P. D.; Hu, K. T.; Jakeman, J. D.; Khalil, M.; Maupin, K. A.; Monschke, J. A.; Ridgway, E. M.; Rushdi, A. A.; Seidl, D. T.; Stephens, J. A.; Swiler, L. P.; Winokur, J. G. *Dakota A Multilevel Parallel Object-Oriented Framework for Design Optimization Parameter Estimation Uncertainty Quantification and Sensitivity Analysis: Version 6.15 User's Manual*, Sandia Technical Report SAND2020-12495; OSTI: 2021.
(33) Poelking, C.; Faber, F. A.; Cheng, B. BenchML: an extensible pipelining framework for benchmarking representations of materials and molecules at scale. *Mach. Learn.: Sci. Technol.* **2022**, *3*, No. 040501.
(34) Park, H.; Zhu, R.; Huerta, E. A.; Chaudhuri, S.; Tajkhorshid, E.; Cooper, D. End-to-end AI framework for interpretable prediction of molecular and crystal properties. *Mach. Learn.: Sci. Technol.* **2023**, *4*, No. 025036.
(35) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine Learning a General-Purpose Interatomic Potential for Silicon. *Phys. Rev. X* **2018**, *8*, No. 041048.
(36) Liang, Y.; Mrovec, M.; Lysogorskiy, Y.; Vega-Paredes, M.; Scheu, C.; Drautz, R. Atomic cluster expansion for Pt−Rh catalysts:

From ab initio to the simulation of nanoclusters in few steps. *J. Mater. Res.* **2023**, *38*, 5125−5135.
(37) Erhard, L. C.; Rohrer, J.; Albe, K.; Deringer, V. L. Modelling atomic and nanoscale structure in the silicon−oxygen system through active machine learning. *Nat. Commun.* **2024**, *15*, No. 1927.
(38) Ibrahim, E.; Lysogorskiy, Y.; Mrovec, M.; Drautz, R. Atomic cluster expansion for a general-purpose interatomic potential of magnesium. *Phys. Rev. Mater.* **2023**, *7*, No. 113801.
(39) Deringer, V. L.; Csányi, G. Machine Learning Based Interatomic Potential for Amorphous Carbon. *Phys. Rev. B* **2017**, *95*, No. 094203.
(40) Finnis, M. W.; Sinclair, J. E. A simple empirical N-body potential for transition metals. *Philos. Mag. A* **1984**, *50*, 45−55.
(41) Morrow, J. D.; Gardner, J. L. A.; Deringer, V. L. How to validate machine-learned interatomic potentials. *J. Chem. Phys.* **2023**, *158*, No. 121501.
(42) George, J.; Hautier, G.; Bartók, A. P.; Csányi, G.; Deringer, V. L. Combining Phonon Accuracy with High Transferability in Gaussian Approximation Potential Models. *J. Chem. Phys.* **2020**, *153*, No. 044104.
(43) Morrow, J. D.; Deringer, V. L. Indirect learning and physically guided validation of interatomic potential models. *J. Chem. Phys.* **2022**, *157*, No. 104105.
(44) McInnes, L.; Healy, J.; Saul, N.; Grossberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Software* **2018**, *3*, No. 861.
(45) Bochkarev, A. Personal communication. 2023.
(46) MACE 0.1.0 Documentation, 2024. https://mace-docs.readthedocs.io/en/latest/guide/training.html#swa-and-ema. accessed July 30, 2024.
(47) Deringer, V. L.; Bernstein, N.; Csányi, G.; Ben Mahmoud, C.; Ceriotti, M.; Wilson, M.; Drabold, D. A.; Elliott, S. R. Origins of structural and electronic transitions in disordered silicon. *Nature* **2021**, *589*, 59−64.
(48) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, No. 108171.
(49) Erhard, L. C.; Rohrer, J.; Albe, K.; Deringer, V. L. A machine-learned interatomic potential for silica and its relation to empirical models. *npj Comput. Mater.* **2022**, *8*, No. 90.
(50) McMillan, P. F.; Wilson, M.; Daisenberger, D.; Machon, D. A density-driven phase transition between semiconducting and metallic polyamorphs of silicon. *Nat. Mater.* **2005**, *4*, 680−684.
(51) Pandey, K. K.; Garg, N.; Shanavas, K. V.; Sharma, S. M.; Sikka, S. K. Pressure induced crystallization in amorphous silicon. *J. Appl. Phys.* **2011**, *109*, No. 113511.
(52) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87*, No. 184115.
(53) Xu, Y.; Zhou, Y.; Wang, X.-D.; Zhang, W.; Ma, E.; Deringer, V. L.; Mazzarello, R. Unraveling Crystallization Mechanisms and Electronic Structure of Phase-Change Materials by Large-Scale Ab Initio Simulations. *Adv. Mater.* **2022**, *34*, No. 2109139.
(54) Rao, F.; Ding, K.; Zhou, Y.; Zheng, Y.; Xia, M.; Lv, S.; Song, Z.; Feng, S.; Ronneberger, I.; Mazzarello, R.; Zhang, W.; Ma, E. Reducing the stochasticity of crystal nucleation to enable subnanosecond memory writing. *Science* **2017**, *358*, 1423−1427.
(55) Ding, K.; Wang, J.; Zhou, Y.; Tian, H.; Lu, L.; Mazzarello, R.; Jia, C.; Zhang, W.; Rao, F.; Ma, E. Phase-change heterostructure enables ultralow noise and drift for memory operation. *Science* **2019**, *366*, 210−215.
(56) Zhang, W.; Mazzarello, R.; Wuttig, M.; Ma, E. Designing Crystallization in Phase-Change Materials for Universal Memory and Neuro-Inspired Computing. *Nat. Rev. Mater.* **2019**, *4*, 150−168.
(57) Sosso, G. C.; Miceli, G.; Caravati, S.; Behler, J.; Bernasconi, M. Neural Network Interatomic Potential for the Phase Change Material GeTe. *Phys. Rev. B* **2012**, *85*, No. 174103.
(58) Mocanu, F. C.; Konstantinou, K.; Lee, T. H.; Bernstein, N.; Deringer, V. L.; Csányi, G.; Elliott, S. R. Modeling the Phase-Change Memory Material, $Ge_2Sb_2Te_5$, with a Machine-Learned Interatomic Potential. *J. Phys. Chem. B* **2018**, *122*, 8998−9006.
(59) Sosso, G. C.; Miceli, G.; Caravati, S.; Giberti, F.; Behler, J.; Bernasconi, M. Fast Crystallization of the Phase Change Compound GeTe by Large-Scale Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2013**, *4*, 4241−4246.
(60) Sosso, G. C.; Donadio, D.; Caravati, S.; Behler, J.; Bernasconi, M. Thermal transport in phase-change materials from atomistic simulations. *Phys. Rev. B* **2012**, *86*, No. 104301.
(61) Konstantinou, K.; Mocanu, F. C.; Lee, T.-H.; Elliott, S. R. Revealing the intrinsic nature of the mid-gap defects in amorphous $Ge_2Sb_2Te_5$. *Nat. Commun.* **2019**, *10*, No. 3065.
(62) Konstantinou, K.; Mavračić, J.; Mocanu, F. C.; Elliott, S. R. Simulation of Phase-Change-Memory and Thermoelectric Materials using Machine-Learned Interatomic Potentials: Sb_2Te_3. *Phys. Status Solidi (B)* **2020**, *258*, No. 2000416.
(63) Zhang, B.; Zhang, W.; Shen, Z.; Chen, Y.; Li, J.; Zhang, S.; Zhang, Z.; Wuttig, M.; Mazzarello, R.; Ma, E.; Han, X. Element-resolved atomic structure imaging of rocksalt $Ge_2Sb_2Te_5$ phase-change material. *Appl. Phys. Lett.* **2016**, *108*, No. 191902.
(64) Wang, J.; Zhou, C.; Yu, Y.; Zhou, Y.; Lu, L.; Ge, B.; Cheng, Y.; Jia, C.-L.; Mazzarello, R.; Shi, Z.; Wuttig, M.; Zhang, W. Enhancing thermoelectric performance of Sb_2Te_3 through swapped bilayer defects. *Nano Energy* **2021**, *79*, No. 105484.

https://doi.org/10.1021/acs.jctc.4c01012
J. Chem. Theory Comput. 2024, 20, 10103−10113