Editors' Suggestion

Neural-network force field backed nested sampling: Study of the silicon *p-T* phase diagram

Nico Unglert [1], Jesús Carrete [1,2], Livia B. Pártay [3], and Georg K. H. Madsen [1,*]

[1] *Institute of Materials Chemistry, TU Wien, 1060 Vienna, Austria*

[2] *Instituto de Nanociencia y Materiales de Aragón, CSIC-Universidad de Zaragoza, 50009 Zaragoza, Spain*

[3] *Department of Chemistry, University of Warwick, Coventry CV4 7AL, United Kingdom*

(Received 24 August 2023; accepted 17 November 2023; published 20 December 2023)

Nested sampling is a promising method for calculating phase diagrams of materials. However, if accuracy at the level of *ab initio* calculations is required, the computational cost limits its applicability. In the present work, we report on the efficient use of a neural-network force field in conjunction with the nested-sampling algorithm. We train our force fields on a recently reported database of silicon structures evaluated at the level of density functional theory and demonstrate our approach on the low-pressure region of the silicon pressure-temperature phase diagram between 0 and 16 GPa. The simulated phase diagram shows good agreement with experimental results, closely reproducing the melting line. Furthermore, all of the experimentally stable structures within the investigated pressure range are also observed in our simulations. We point out the importance of the choice of exchange-correlation functional for the training data and show how the r2SCAN meta-generalized gradient approximation plays a pivotal role in achieving accurate thermodynamic behavior. We furthermore perform a detailed analysis of the potential energy surface exploration and highlight the critical role of a diverse and representative training data set.

DOI: 10.1103/PhysRevMaterials.7.123804

I. INTRODUCTION

Nested sampling (NS) is a powerful Bayesian method that can efficiently sample high-dimensional parameter spaces [1,2]. The applications of NS in materials science have progressed steadily in the past decade. While early investigations mainly focused on simple model systems such as Lennard-Jones [3] and hard-sphere models [4], more recent work has used embedded-atom potentials to study a variety of metallic systems, including elemental metals such as Fe, Zr, and Li [5–7], as well as alloys like CuAu [8,9], AgPd [10], and CuPt nanoparticles [11].

With the emergence of efficient machine-learned force fields (MLFFs), the sampling of potential energy surfaces (PESs) at a level of accuracy similar to the underlying *ab initio* method becomes affordable for NS. In this context, it has been applied in conjunction with Gaussian approximation potentials and moment tensor potentials to predict the thermodynamic behavior of carbon [12], platinum [13], and AgPd [14] alloys.

MLFFs use statistical learning techniques to approximate the PES of a material [15]. Unlike classical interatomic potentials, MLFFs do not require extensive parametrization. Instead, they provide a highly flexible functional form that has the ability to generalize across different chemical environments. Trained on datasets obtained from *ab initio* calculations, MLFFs can thus capture the physics of the system on par with the underlying method. However, two critical factors are paramount for their successful application: first, the diversity and representativeness of the training dataset [16] and, second, the quality of the selected *ab initio* method.

Concerning the *ab initio* method, Kohn-Sham density functional theory (DFT) has been the method of choice for calculating the properties of solid-state materials. A crucial aspect of DFT is the exchange-correlation (xc) functional, which incorporates electron-electron interactions within the system [17]. As an exact formulation of this functional is not available, various approximations are employed, and the choice of approximation significantly impacts the accuracy of DFT for a given problem. Traditionally, the suitability of xc functionals has been assessed based on ground-state properties, such as lattice parameters, or cohesion energies [18–20]. Despite their importance for practically relevant predictions, finite-temperature properties, such as the melting point, have been less explored as target properties for evaluating the suitability of functionals due to the computational complexity of obtaining them [21,22]. Using NS together with MLFF-based models to conduct an exhaustive exploration of the PES and give access to finite-temperature thermodynamic behavior can bridge this gap and thus open the door for a much more comprehensive evaluation of functional performance in a broader range of conditions.

Here we demonstrate this aspect in a NS study of the low-pressure silicon *p-T* phase diagram. We show how the choice of a suitable xc functional crucially influences the predicted melting temperature of Si over a large pressure range. In contrast to simple metallic systems, silicon stands out because of its variability in chemical bonding. In its low-pressure allotrope, strong directional bonds lead to the characteristic tetrahedral coordination of the semiconducting cubic diamond phase. At higher pressures the system transitions

*georg.madsen@tuwien.ac.at

to more closely packed structures like the well-known β-Sn phase. These circumstances complicate the use of classical interatomic force fields. Due to their rigid functional form these models are usually very poorly transferable and thus work only for the specific phases and properties they were designed for [23].

This diversity of chemical bonding requires an equally diverse set of training data to be representative of the rich phase behavior. To address this, we perform a detailed analysis of the configurations explored by NS, revealing a wide range of attraction basins and regions of the PES explored during the simulation. For the training data we reevaluate the database of Bartók *et al.* [24], which contains around 2475 manually curated silicon structures. We show how the database, a result of the continuous efforts to create general-purpose MLFFs, possesses the diversity and representativeness necessary to deliver accurate thermodynamic predictions from a MLFF-backed NS simulation.

II. METHODOLOGY

A. DFT

To assess the effect of the xc functional, we recomputed the energies and forces of the database provided by Bartók *et al.* [24]. For the DFT calculations the Perdew-Burke-Ernzerhof (PBE) [25] and r2SCAN [26] functionals as implemented in VASP [27,28] were used. The cutoff energy for the plane wave basis was chosen as 300 eV. The partial occupancies for the orbitals were determined employing Fermi smearing with a smearing parameter of 0.025 eV. The reciprocal space sampling was performed on a Monkhorst-Pack grid with a k spacing of 0.3 Å^{-1}, and the energy convergence criterion was set to 10^{-5} eV. For the evaluation of energy-volume curves we used a denser sampling with a k spacing of 0.2 Å^{-1} and a tighter convergence criterion of 10^{-8} eV. We removed one configuration from the database, where a single atom is placed in a large vacuum.

B. Neural-network force field

The simulations in this work use our recently developed neural-network force field (NNFF) architecture, NEURALIL [29,30]. Atomic coordinates are encoded into atom-centered descriptors that are invariant with respect to global rotations and translations, with relative positions of neighbors transformed into second-generation spherical Bessel descriptors [31]. The descriptors are fed into a dense $64\times32\times16$ feature extractor inspired by deep residual network (ResNet) architectures [30,32] using a Swish-1 differentiable activation function [33]. We include a repulsive Morse contribution to avoid unphysical behavior for short interatomic distances [34]. The implementation uses JAX [35] for just-in-time compilation and automatic differentiation and FLAX [36] for simplified model construction and parameter bookkeeping.

In order to compute the descriptors for our training configurations, we rely on the minimum image convention. This means that we require the training dataset to have cells large enough to fit a sphere with the corresponding cutoff radius. However, since training datasets often contain structures with a variety of different cell sizes, we need to apply a special procedure to handle this. We use an iterative process to generate diagonal supercells of increasing size and perform a Minkowski reduction [37] to make the cell as compact as possible. This process continues until the desired cutoff fits into the cell, at which point the cycle is stopped. The result is a set of supercell structures that conform to the cutoff parameter for the descriptor generation. However, this set may still contain configurations with significantly varying numbers of atoms. To ensure that our JAX-based approach is efficient, it is important to have static array sizes. Therefore, we perform a padding procedure with ghost atoms to fill up all supercells until the number of atoms is constant. This allows us to generate second-generation spherical Bessel descriptors efficiently and accurately for all configurations in the training dataset, regardless of their cell size and number of atoms.

For the descriptor generation we consider atomic environments within a cutoff of $r_{\text{cut}} = 4$ Å and choose $n_{\text{max}} = 6$ [29]. For each training we randomly split the database and use 3/4 of the data for training and the rest for validation.

C. Nested sampling

NS partitions the configuration space into a nested sequence of phase space volumes confined by surfaces of isolikelihood. Moving from the outer shells to the higher-likelihood inner shells in this nested sequence corresponds to the transition from a high-entropy fluid phase to more ordered crystalline states. Each iteration of the NS algorithm peels off a layer of the nested sequence, resulting in a corresponding sample. This approach to sampling ensures that only thermodynamically relevant structures are sampled, ultimately enabling the calculation of the thermodynamic partition function [10].

During the NS process, a group of K walkers is continuously updated by replacing the highest-energy walker at each iteration. Its energy defines a bounded region of configuration space from which a new walker has to be uniformly sampled. This sampling process is carried out by performing a Markov chain Monte Carlo (MCMC) random walk with a cloned version of one of the remaining walkers, taking a total of L steps. In the case of a constant pressure simulation, the MCMC steps involve modifying the simulation cell, through isotropic volume changes, shear transformations, or stretching operations and atom steps modifying the positions of individual atoms. In this study, the Galilean Monte Carlo (GMC) technique [38] is applied to decorrelate the positional degrees of freedom. Every GMC trajectory is initiated by sampling a random vector from the surface of the $3N$-dimensional unit sphere. This establishes N independent three-dimensional directions along which the atoms are propagated with a given step size. In cases where a step fails to meet the acceptance criterion, the direction vector is reflected against the normalized force vector [38]. This assists in navigating the trajectory back within the allowed energy bound. Our NS calculations were performed with a modified version of the PYMATNEST code [8,10]. We retained most of the logic of PYMATNEST but adapted the parallel workflow to be managed by a scheduler provided in the PYTHON library DASK. In the original message passing interface (MPI)-based parallelization scheme [8], instead of the clone taking all the L steps required to decorrelate

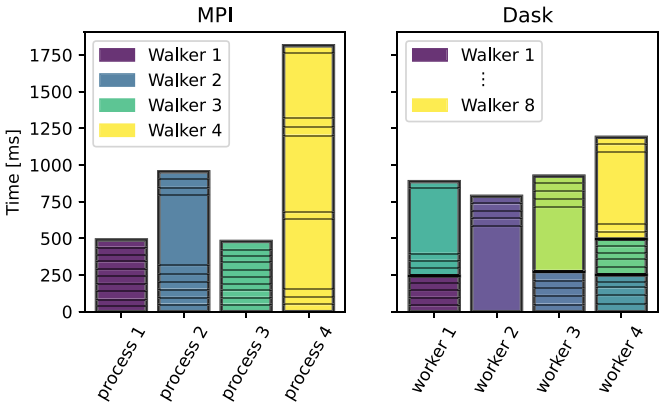

FIG. 1. Synthetic scenario illustrating the advantage of the DASK parallelization scheme. Boxes delimited by thin lines indicate individual cell steps or GMC atom trajectories. Note that the larger boxes correspond to the more time-consuming atom move trajectories. Boxes delimited by thick lines indicate whole random walks of particular walkers. For better visibility, the latter are also colored.

the configuration in one iteration, each of the available n_p processes is used to have n_p different walkers perform L/n_p steps each. Since the atom moves involve more expensive force evaluations, they typically take significantly longer than single cell steps, which require only energy evaluations. As a result, depending on the random choice of step types the computational work allocated to each processor can vary substantially. In contrast, our DASK implementation uses a pool of n_w workers to which n_t individual walks are dynamically assigned. The idea of the DASK-based parallel workflow is schematically depicted in Fig. 1, showing two artificial scenarios for the original and DASK parallelizations. In both cases, the workload is handled by $n_{p/w} = 4$ processing units, and the total walk length is $L = 40$ steps. In the original parallelization this corresponds to four walkers being walked for $L/n_p = 10$ steps. In this scenario, the MPI code would be required to wait for all processes to complete their random walk, resulting in significant computational overhead. However, for the DASK example the random walk is split into $n_t = 8$ walkers of length $L/n_t = 10$. These are dynamically allocated to the four workers to obtain an improved load balance (see Fig. 1). In the limiting case of $n_t = n_w$, the DASK implementation becomes equivalent to the MPI scheme.

After conducting a series of convergence tests on the pristine silicon system, we chose $K = 600$ walkers and a walk length of $L = 1000$ steps for our simulations. We realize the cell steps by isotropic volume moves, which sample the cell volume uniformly. To alter the cell shape, volume-preserving shear and stretch step moves are proposed with a uniform distribution in strain. The step probability ratio for volume, stretch, shear, and atom steps was set to 2:1:1:1, respectively. We restrict our simulation cell to a minimum aspect ratio of 0.8 to avoid pathologically thin cells forming at the early stages of the sampling [10]. Steps violating this constraint are discarded.

To initiate the sampling process, a set of $K = 600$ uniformly random distributed configurations has to be generated. We achieve this by replicating a cubic diamond cell with a density of 2.31 g/cm^3 a number of K times. To randomize these configurations, we perform a two-step procedure. In the first step, we diversify the cell shapes of these structures by an initial isotropic volume scaling and a subsequent series of 1000 cell shape modifying steps. In each of these steps one shear move and one stretch move are performed in random order. In a second step, to further decorrelate the walkers and, especially, the positional degrees of freedom, a series of 10 NS random walks each with walk lengths of 100 steps is performed on the generated structures. The energy threshold for these walks was chosen to be $U_{\text{initial}} = U_{\max} + N \times 1\,\text{eV}$, where $U_{\max}$ is the energy of the highest energy walker and N is the number of atoms. We find this procedure to be more numerically stable than sampling positions from a uniform distribution, which tends to result in artificially clumped structures. We stop our simulations once the current estimated finite-difference temperature [8] is below 200 K. This corresponds to approximately 5×10^8 energy evaluations for each individual simulation, underscoring the need for efficient surrogate models to estimate the potential energy.

From the converged NS runs we compute the isobaric heat capacities according to

$$C_p = \frac{3Nk_B}{2} + k_B\beta^2\{\langle Y^2\rangle - \langle Y\rangle^2\}, \tag{1}$$

where N is the number of atoms, k_B is the Boltzmann constant $\beta = (k_B T)^{-1}$, and Y is the microscopic enthalpy. The latter is given by $Y = U + PV$, with the potential energy U and the pressure-volume term PV. The thermodynamic expectation values in Eq. (1) are evaluated using the NS partition function

$$\langle O\rangle = \frac{\sum_i w_i O(R_i)\exp(-\beta Y_i)}{\sum_i w_i \exp(-\beta Y_i)}, \tag{2}$$

where the sums run over all acquired samples. Y_i and R_i are the microscopic enthalpy and configuration of a given sample i, and O is an arbitrary observable depending on the configuration R_i. The weights $w_i = (\chi_{i-1} - \chi_i)V_i^N$ contain the NS configuration space volumes χ_i and the cell volumes V_i. We use the isobaric heat capacity C_p to locate first-order phase transitions.

D. Structure representation in two dimensions

Visualizing the structural variety occurring in high-dimensional spaces, such as the $3N_{\text{atoms}}$-dimensional space of potential configurations for a system with N_{atoms}, requires a projection into a lower-dimensional space. For that purpose, we utilize the same spherical Bessel descriptors used for encoding atomic environments for the NNFF. For a structure composed of N_{atoms}, this results in a matrix with shape (N_{atoms}, n_{features}), which describes the complete structure. To make this representation invariant with respect to atom permutations within the structure, we compute the distributions of each of the n_{features} features as a histogram and divide it by the number of atoms N_{atoms}. After flattening, this yields a vector of length $n_{\text{features}} n_{\text{bins}}$ which is permutation invariant and independent of the system size. For this work, the histograms are calculated using $n_{\text{bins}} = 128$ and in a range from 0 to 4. To visualize the permutation-invariant structure descriptors, we use dimensionality reduction techniques, which have become

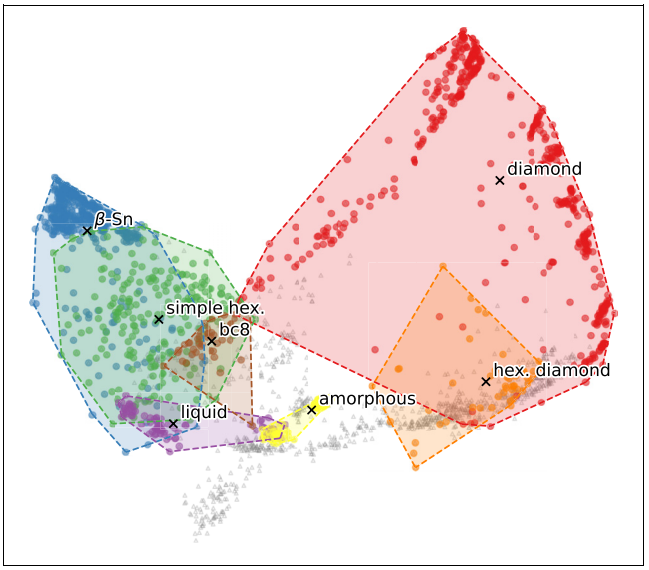

FIG. 2. The first two principal components of spherical Bessel descriptors highlighting different silicon phases present in the structural database. Each gray triangle represents one structure; colored areas show the convex hull around sets of structures corresponding to a certain phase. Only a few example phases are highlighted.

a widespread tool, e.g., in the analysis of molecular dynamics trajectories [39,40]. Following a comprehensive evaluation of various methodologies, we opted to utilize the principal component analysis (PCA) technique as implemented in the SCIKIT-LEARN library [41]. We show a comparison of the PCA, t-distributed stochastic neighbor embedding [42], and uniform manifold approximation and projection [39] techniques applied to our dataset in the Supplemental Material [43].

E. Optimization and symmetry determination

For the analysis of the walker population we perform rough relaxations of the atomic positions using our NNFF model. For that purpose, we use the Broyden–Fletcher–Goldfarb–Shanno implementation contained in the JAX [35] library with a loose force convergence criterion of 0.01 eV/Å.

For symmetry determination we employ SPGLIB [44] as implemented in the PYMATGEN package [45]. Since the finite-temperature structures may have slightly distorted cells, we employ a very loose symmetry precision parameter of 0.3 Å to determine the space group.

For runs that converge into strongly disordered or amorphous metastable minima, we determine the nature of the minimum by eye and by looking at the radial distribution function.

III. RESULTS

A. Neural-network force field

Based on the spherical Bessel descriptors, the configuration space spanned by the structures in the training dataset [24] is illustrated in Fig. 2. The map is divided into several distinct regions, each representing the most prominent phases of silicon. One of the most striking features of the map is the energetically lowest phase, cubic diamond, which occupies a significant area towards the northeast. Moving towards the west, we can see the most relevant phases at intermediate pressures, such as β-Sn and simple hexagonal. The liquid configurations of silicon are represented by only two patches located in the bottom left part of the map. Figures 3(a) and 3(b) show detailed parity plots as well as averaged statistics for the models trained on the r2SCAN and PBE datasets. The energy and force errors are on the order of 10 meV/*atom* and 100 meV/Å, respectively, with a slightly better result for the PBE database. The errors are similar to what we obtain by training on the original CASTEP data [24]. The differences in error statistics between the training and validation sets are given in Fig. 3 and indicate that no significant overfitting occurred in the training process.

To further test the transferability and accuracy of our trained models, we created a test set of structures that are not included in the training dataset. For this we extracted crystal structures of the most prominent silicon phases in the investigated pressure range from the Materials Project database. In addition, we added the body-centered orthorhombic *Imma* phase [46] since it was not present in the Materials Project database. For each of these six phases, we created a set of isotropically scaled cells around the equilibrium volume and evaluated the energies by DFT and the corresponding NNFFs. The resulting energy-volume curves are shown in Fig. 3 for both functionals. The NNFFs reproduced the respective curves with very similar performance. Compared to the PBE energies, the r2SCAN energies exhibit a significant increase in energy differences. For example, the energy difference between the cubic diamond and β-Sn minimum is increased by almost 50% in the case of r2SCAN.

B. Phase diagram

We conducted NS calculations using our NNFF models across a range of eight different pressures, spanning from 0 to 16 GPa. To account for finite-size effects on the calculated quantities, we performed simulations on systems consisting of 16 and 32 silicon atoms. Additionally, for a few specific data points, we extended the simulations to include also 64-atom systems. The resulting constant pressure heat capacities are shown in Fig. 4. An overview of all calculations and the corresponding identified most stable phases is summarized in Table I.

Theoretically, the heat capacity diverges at first-order phase transitions in the thermodynamic limit. This behavior is consistent with our findings, as the heat capacity peaks become more pronounced as the system size increases from 16 to 64 atoms. Moreover, we observe a slight shift towards higher temperatures in the smaller systems. Interestingly, this finite-size effect appears to be pressure dependent. Comparing the 16- and 32-atom simulations we observe that at 0 GPa, the deviation is more pronounced, and as the pressure increases, the deviation gradually decreases. For the three pressures we ran using 64 atoms, at 4 and 10 GPa the melting temperature decreases by around 100 K; at 16 GPa almost no shift appears. A similar trend was recently observed in a NS study of carbon, where the finite-size effect almost diminished above a pressure of 100 GPa [12].

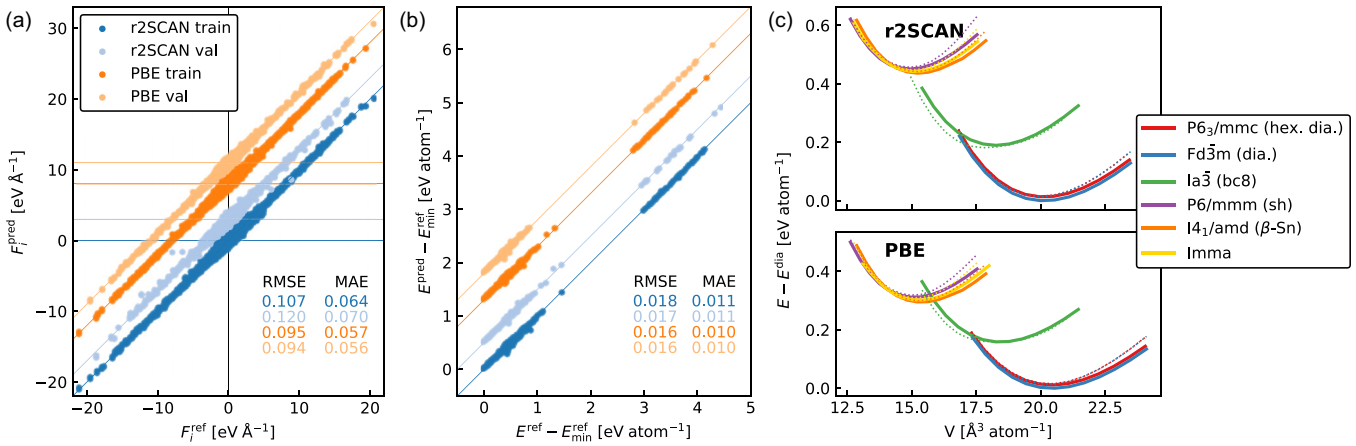

FIG. 3. (a) Force and (b) energy training and validation set parity plots for the NNFF. Energies per atom are given relative to the minimum energy occurring in the respective dataset. (c) Energy-volume curves for several crystalline phases of silicon evaluated using DFT with a certain functional (solid lines) and corresponding NNFF models (dotted lines) that were trained on a database evaluated using the same functional. Top: r2SCAN. Bottom: PBE. Energies per atom are given relative to the minimum energy of the cubic diamond phase for both functionals.

Based on the calculated melting temperatures from our simulations we can construct a $p-T$ phase diagram. It is depicted in Fig. 5 together with the experimental phase diagram reported by Voronin *et al.* [47], which we briefly describe below. In the low-pressure regime up to approximately 10 GPa, the predominant phase is the cubic diamond phase, and the melting line exhibits a consistent negative slope of $-60\,\mathrm{K}/GPa$. Increasing the pressure above 10 GPa, the β-Sn becomes stable. The cubic diamond-β-Sn-liquid triple point is found at 10.5 GPa and 1003 K. Within a relatively narrow region of around 2 GPa, the β-Sn phase occupies a distinct range and is separated from the orthorhombic *Imma* phase by a phase boundary characterized by a negative slope at approximately 13 GPa. As pressure increases further, the equilibrium structure transitions to a simple hexagonal phase. The β-Sn, *Imma*, and simple hexagonal phases are separated by a gentle positively sloped melting line from the liquid phase.

The simulated r2SCAN melting temperatures (see Fig. 5, black dashed and dotted lines) reproduce the experimentally observed trends. The 32-atom simulations show a negatively sloped melting line until 10 GPa, close to the experimental cubic diamond-β-Sn-liquid triple point. For the higher pressures, the melting line follows the experiment with a slightly decreased slope and a small constant shift to lower temperatures. Our simulations of different system sizes indicate that a small finite-size effect even for the 64-atom runs remains for the lower pressures, while it seems to be almost absent for the higher-pressure domain. We discuss the regions of stability for the r2SCAN calculations in the following section, where we perform a detailed analysis of these NS runs.

The 16-atom PBE simulations (see Fig. 5, gray dashed line) correctly predict the slope of the low-pressure melting line. However, they fail in the prediction of the absolute values, which are shifted to lower temperatures by approximately 300 K. This is in line with previous *ab initio* molecular dynamics simulations that predict the 0 GPa melting temperatures to be 1687 and 1450 K for the SCAN and PBE functionals, respectively [22]. For higher pressures, PBE captures neither the correct trend nor the correct magnitude of the experimental melting line, with a continuing negative slope down to a transition temperature of 706 K at 16 GPa. We note that these differences in melting temperatures do not arise from the algorithm finding different phases. A comparison of the data in Table I reveals that overall similar phases are found for PBE and r2SCAN. Instead, we relate this observation to a

TABLE I. Space groups of final structures the nested sampling converges to for different runs (I: cubic diamond $Fd\bar{3}m$, II: β-Sn $I4_1/amd$, III: bc8 $Ia\bar{3}$, IV: hexagonal diamond $P6_3/mmc$, V: simple hexagonal $P6/mmm$, I*: disordered cubic diamond $P\bar{1}$, A: amorphous).

Functional	Size	Seed	0 GPa		4 GPa		9 GPa		10 GPa		11 GPa		12 GPa		13 GPa		16 GPa	
PBE	16	0	IV	1446	I	1216	II	911	II	951	II	901	II	951	V	856	V	706
r2SCAN	16	0	IV	1791	I	1676	I	1346	IV	1431	II	976	II	981	V	1001	V	1031
		1	I	1821	I	1696	IV	1231	III	1011	II	951	II	996	V	1021	V	1056
		2	I	1801	I	1686	I	1251	IV	1361	II	941	II	981	V	1031	V	1046
	32	0	I	1696	I	1636	I	1231	II	886	II	951	II	976	V	991	V	1036
		1	I	1566	I	1616	I	1316	I*	986	II	961	II	956	V	966	V	1006
		2	IV	1696	I	1601	I	1231	II	936	II	926	II	936	V	966	V	1001
	64	0			I	1586			A	901							V	1001
		1			IV	1466			A	836							V	1001
		2			I*	1451			I*	891							V	996

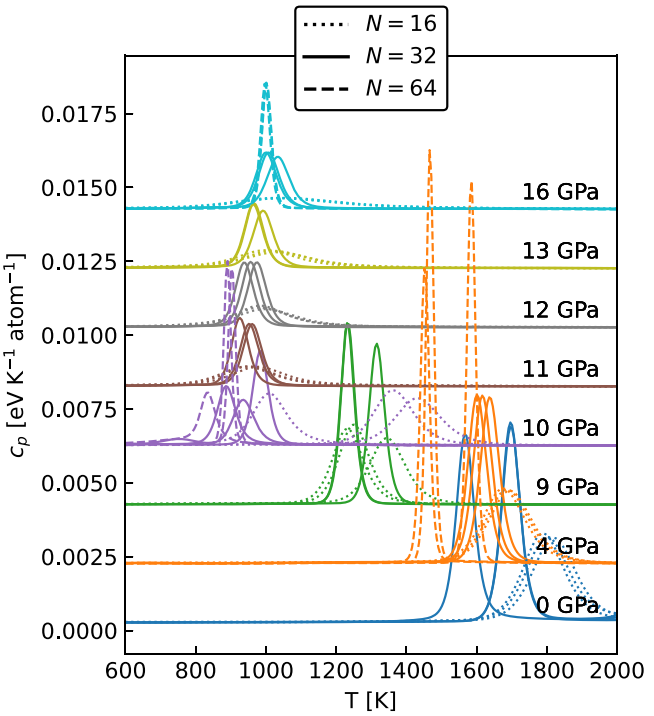

FIG. 4. Heat capacities for a series of pressures in the range of 0 to 16 GPa as calculated for system sizes of 16 (dotted lines), 32 (solid lines), and 64 atoms (dashed lines) using the r2SCAN-based model (see Table I). For better visibility, 16- and 64-atom heat capacities are scaled by factors of 3 and 0.3, respectively.

misprediction of the relative energies of the different silicon phases shown in Fig. 3(c). Due to the smaller energetic differences, the observed phase transitions can occur already at lower temperatures. In the following, we restrict our analysis to the more accurate r2SCAN results.

C. Analysis of NS runs

Figure 6 shows the evolution of the walker live set for the 32-atom NS run (seed = 0) at 10 GPa in the two-dimensional (2D) structure representation map (see Fig. 2). In order to assign each of the walkers to a certain basin of the PES, we relaxed their atomic positions and determined the corresponding space group. Initially, all walkers reside in the liquid configuration area of the training database. At this point, the walkers are in highly disordered liquid or even gaslike states characterized by large cells and low coordination numbers. Therefore, even after ionic relaxation, the system remains in a low-symmetry crystalline configuration, and all walkers are assigned to space group $P1$ at the start of the sampling. As the iteration progresses, the cloud of walkers leaves the liquid area and enters the domain spanned by the β-Sn and simple hexagonal $P6/mmm$ phase structures of the training database. Although the majority of walkers remain in the $P1$ space group, we observe a diverse population of other space groups, notably $Imma$, $I4_1/amd$, and $C2/m$. In subsequent snapshots, the presence of strongly disordered, liquidlike walkers diminishes. The simulation predominantly focuses on the $Imma$ phase, with the $I4_1/amd$ and $C2/m$ phases being weakly represented. Towards the end of the simulation, sampling

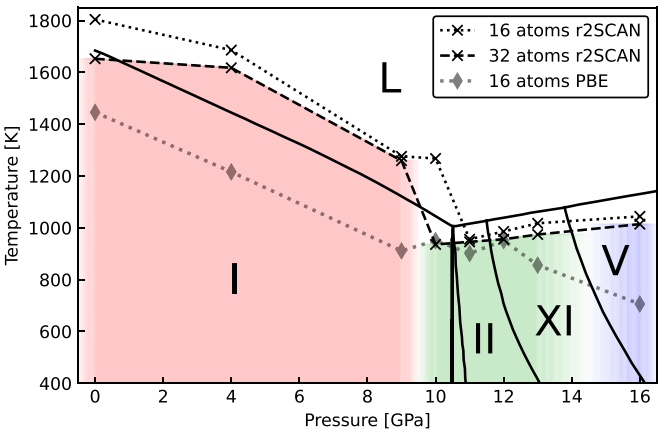

FIG. 5. Pressure-temperature phase diagram of silicon. Solid black lines show the experimental phase equilibrium lines according to Ref. [47] (I: cubic diamond $Fd\bar{3}m$, II: β-Sn, XI: $Imma$, V: simple hexagonal $P6/mmm$, L: liquid). Dotted and dashed black lines show 16- and 32-atom simulations using the NNFF model trained on r2SCAN data. Data points for r2SCAN simulations represent the average of three independent simulations at each pressure (see Table I). The gray dotted line shows the melting line of a series of 16-atom simulations using a model trained on PBE data. Colored areas show regions of stability we deduce from our runs (Red: $Fd\bar{3}m$, blue: $P6/mmm$, green: $I4_1/amd$-$Imma$-$P6/mmm$; see Sec. III D for a detailed explanation).

lower enthalpy levels, a shift occurs in the population towards the $I4_1/amd$ phase, which turns out to be the most stable phase at this pressure. This contradicts the depiction in the 2D configuration space map, where the trajectory ends at the tip of the cubic diamond region, not falling into the actual $I4_1/amd$ region. We interpret this behavior as an artifact of the PCA, which cannot always preserve the full information from the high-dimensional space upon dimensionality reduction. Nevertheless, the visualization in Fig. 6 provides insight into how the NS algorithm explores the configuration space during the simulation. Throughout the process, the walker set encompasses a wide region in configuration space until eventually converging into the most stable basin.

A summary of the walker populations for all investigated pressures in the 32-atom, seed = 0 calculation series is presented in Fig. 7(a). Additional analyses for the other runs can be found in the Supplemental Material [43]. For the three lowest pressures, a similar pattern emerges with a predominant population of the cubic diamond $Fd\bar{3}m$ phase. Although the NS algorithm visits alternative basins such as $Imma$ and $I4_1/amd$, these are quickly disregarded due to the exceptional stability of the cubic diamond phase under those conditions. In the intermediate pressure range of 10 to 12 GPa, two competing phases are observed. The $Imma$ phase experiences a substantial initial increase in population alongside a gradual representation of the $I4_1/amd$ phase. The $Imma$ phase later becomes depopulated towards the end of the simulation, with the $I4_1/amd$ phase emerging as the most stable. The presence of the $P6/mmm$ phase is also noted, gaining significance between 10 and 12 GPa. Beyond 12 GPa, the walker population is dominated by the $P6/mmmm$ phase, which becomes the ground state. The exploration of the $I4_1/amd$ phase

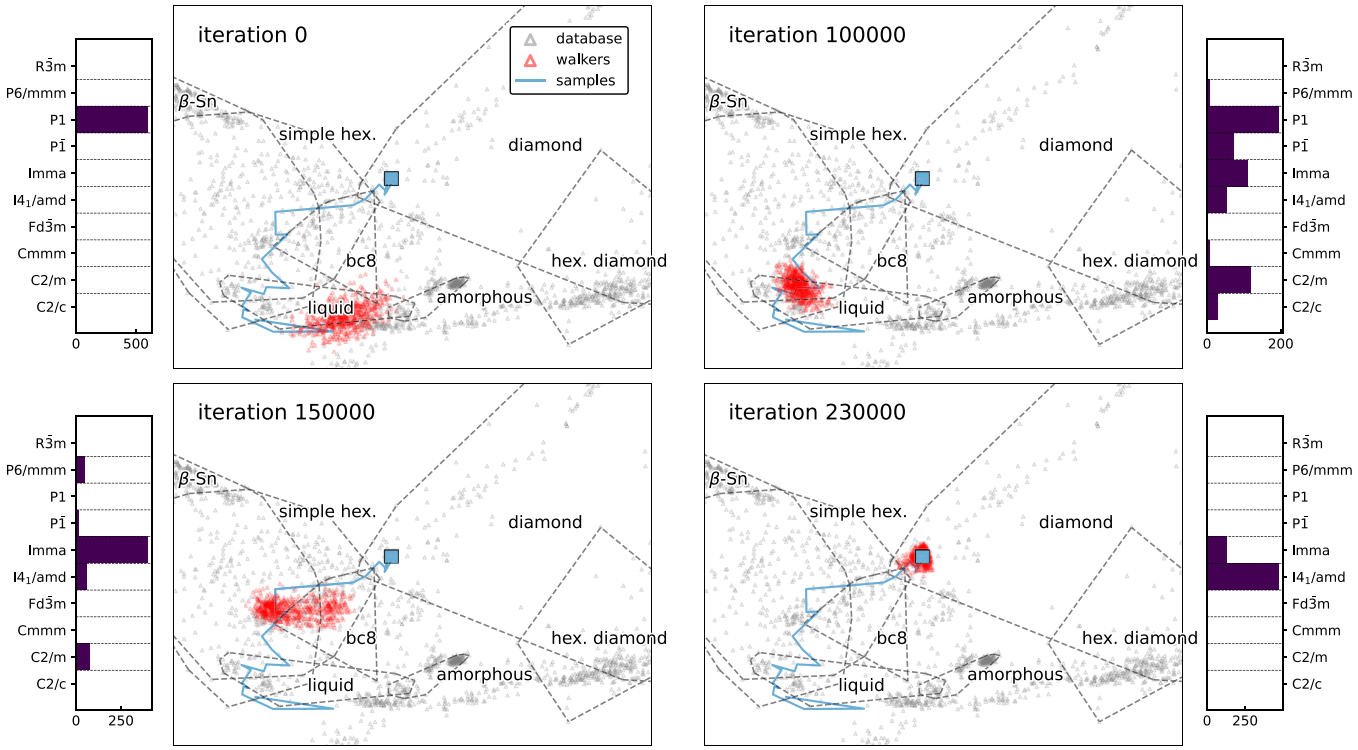

FIG. 6. Evolution of the walker population of the 32-atom, seed = 0, simulation at 10 GPa in configuration space over time illustrated using the 2D configuration space map from Fig. 2. Gray points correspond to configurations in the training database; dashed lines indicate the convex hull of configurations belonging to a certain phase. Blue lines show the trajectory of NS samples (plotting only every 10 000th sample). Red points show the walker population at the given iteration. Histograms to the side show the population of space groups determined for optimized walkers.

diminishes in importance, while the *Imma* phase maintains a degree of population throughout the simulation.

For the pressures above 10 GPa all r2SCAN simulations converge to the same phases consistently (see Table I). However, discrepancies arise among the runs at lower pressures. Regardless of the system size, between 0 and 9 GPa multiple runs converge into the hexagonal diamond $P6_3/mmc$ phase, which has been observed experimentally as a minor phase in indented cubic silicon [48]. We attribute this to the small energetic difference between the actual ground state $Fd\bar{3}m$ and the $P6/mmm$ phase [see Fig. 3(c)]. Furthermore, we observe a disordered cubic diamond phase for the 4 GPa

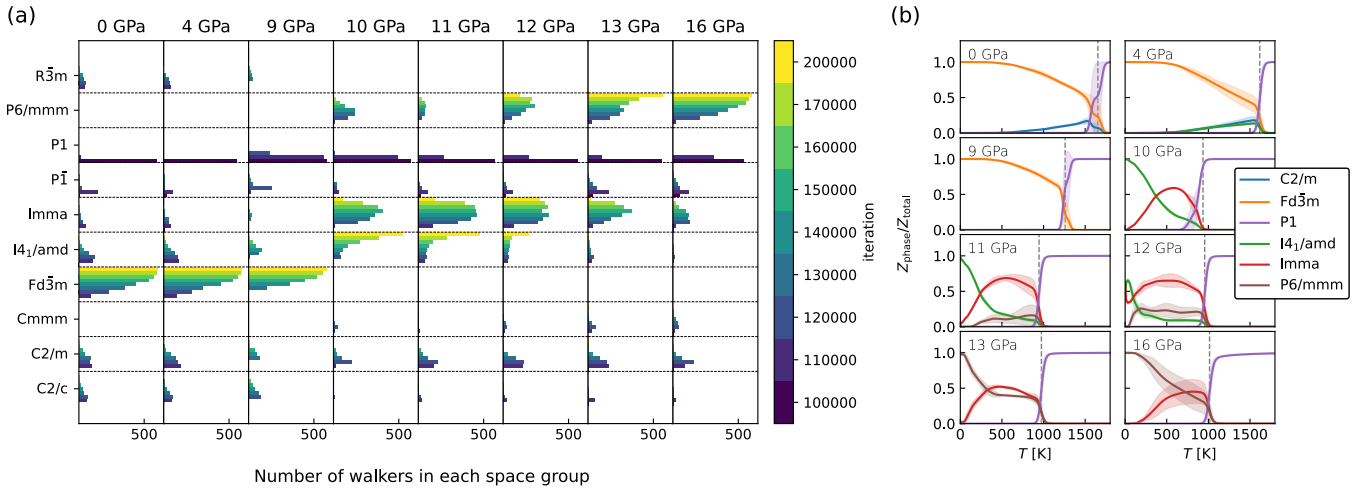

FIG. 7. Analysis of the basins that are explored during the nested sampling for the 32-atom, seed = 0, simulations. (a) Population of the most prominent space groups over time for the walkers at each simulated pressure. Colors indicate the iteration. (b) Partition function ratios of different occurring phases for all simulated pressure values, averaged over all three independent 32-atom runs, excluding two outliers discussed in the text. Colored areas show the standard deviation. Dashed lines show averaged heat capacity c_p peak positions corresponding to melting (see Table I).

64-atom simulation. At 10 GPa, two runs of the 16-atom simulations converge to the $P6_3/mmc$ phase, while the third run results in the $Ia\bar{3}$ space group, known as the cubic body-centered BC8 phase of silicon. The BC8 phase is metastable at ambient pressure [47] and can be obtained by slowly decompressing the metallic β-Sn phase and remains metastable unless heated above 200 °C [49]. In the 32- and 64-atom case, we observe one simulation collapsing into a disordered cubic diamond minimum. The remaining 64-atom simulations converge to amorphous structures. We connect these discrepancies to the experimental findings of multiple metastable phases in this pressure range [47]. Thus, the problem becomes strongly multimodal under these conditions, hampering the sampling.

Figure 6 furthermore gives a visual impression of how the number of walkers determines the granularity of the PES sampling. Smaller numbers increase the likelihood of the walker cloud missing the entry point to a particular basin funnel. As a result, the set of walkers can become trapped in a metastable minimum since the Galilean Monte Carlo walk cannot traverse large energy barriers. This interpretation is supported by our comprehensive analysis of the walker populations (see the Supplemental Material [43]). In cases where a run converges to a metastable phase, the actual most stable phase is never populated, indicating that its entry point has not been found due to its small phase space volume. The more frequent occurrence of the $P6_3/mmc$ phase in the 16-atom simulations may also be influenced by the minimum aspect ratio constraint imposed on the cell shape. We speculate that in the 16-atom case this constraint may favor the formation of the $P6_3/mmc$ phase compared to the competing $Fd\bar{3}m$ phase.

D. Solid-solid phase transition

To facilitate the identification of solid-solid phase transitions, we calculate the ratios of the contributions to the partition function by the competing phases. To achieve this, we perform optimizations on every 10th sample obtained during the NS process and determine the corresponding space group. This enables us to assign specific samples to particular basins of the PES and separate the overall partition function into individual contributions from different phases:

$$Z(\beta) = \sum_i w_i e^{-\beta E_i} = \sum_{i\in\text{dia}} w_i e^{-\beta E_i} + \sum_{i\in\beta-\text{Sn}} w_i e^{-\beta E_i} + \cdots . \tag{3}$$

The result is shown as an average over the independent runs for the 32-atom simulations (excluding the two outliers at 0 and 10 GPa discussed above) in Fig. 7(b). In all cases the melting transition is clearly visible in the form of a sharp step in the $P1$ partition function contribution. For pressures above 9 GPa the competition between different phases becomes apparent. The intersections of the $I4_1/amd$ and the $Imma$ contributions at 10, 11, and 12 GPa indicate a solid-solid phase transition occurring at 372, 247, and 96 K, respectively. Although the $Imma$ phase is significant at 13 GPa, determining a clear phase transition point is challenging in this case. To summarize, in the pressure range from 10 to 13 GPa, three distinct phases, namely, $I4_1/amd$, $Imma$, and $P6/mmm$, interact in a complex manner (see the green shaded area in Fig. 5). At 16 GPa we observe an unambiguous dominance of the $P6/mmm$ phase.

IV. CONCLUSION

In the current study, we successfully combined the nested sampling method with a fully automatically differentiable neural-network force field. By employing this powerful methodology, we can achieve *ab initio* precision in our predictions, which we demonstrated by accurately simulating the pressure-temperature phase diagram of silicon. Through a comparison of the predicted melting lines from two common exchange-correlation functionals, we have demonstrated that the performance of a machine-learning model is limited by the quality of its corresponding ground truth data.

The successful use of neural-network force fields together with nested sampling adds to the growing field of calculating phase diagrams using machine-learned force fields [12,50,51]. Despite their success, machine-learned force fields still heavily rely on the quality, size, and diversity of the training datasets to deliver accurate and reliable results. This requirement can be demanding and hinder their transferability and widespread applicability. In this study, we utilized a meticulously curated configuration database. However, adapting our method to a different system would necessitate the creation of an entirely new database. In this respect, the inherent capability of NNFFs to handle large amounts of data facilitates the adoption of active learning methods. By developing efficient NNFF-backed nested sampling active learning approaches, we may mitigate the necessity for intricate manually curated training databases. This opens up new possibilities for purely data-driven configuration space exploration, enhancing our understanding of complex systems.

A compatible version of NEURALIL, including example scripts for training and evaluation, is available on GitHub [52]. The PYMATNEST code on which our implementation is based is available on GitHub [53].

A dataset containing the energy and sample trajectories of all presented nested sampling calculations as well as the DFT-evaluated training databases is available on Zenodo [54].

ACKNOWLEDGMENT

L.B.P. acknowledges support from EPSRC through the individual Early Career Fellowship (EP/T000163/1).

[1] J. Skilling, Nested sampling, AIP Conf. Proc. **735**, 395 (2004).
[2] J. Skilling, Nested sampling for general Bayesian computation, Bayesian Anal. **1**, 833 (2006).
[3] L. B. Pártay, A. P. Bartók, and G. Csányi, Efficient sampling of atomic configurational spaces, J. Phys. Chem. B **114**, 10502 (2010).

[4] L. B. Pártay, A. P. Bartók, and G. Csányi, Nested sampling for materials: The case of hard spheres, Phys. Rev. E **89**, 022302 (2014).

[5] L. B. Pártay, On the performance of interatomic potential models of iron: Comparison of the phase diagrams, Comput. Mater. Sci. **149**, 153 (2018).

[6] G. A. Marchant and L. B. Pártay, Nested sampling of materials potential energy surfaces: Case study of zirconium, Phys. Sci. Forum. **5**, 5 (2022).

[7] J. Dorrell and L. B. Pártay, Pressure temperature phase diagram of lithium, predicted by embedded atom model potentials, J. Phys. Chem. B **124**, 6015 (2020).

[8] R. J. N. Baldock, N. Bernstein, K. M. Salerno, L. B. Pártay, and G. Csányi, Constant-pressure nested sampling with atomistic dynamics, Phys. Rev. E **96**, 043311 (2017).

[9] A. Gola and L. Pastewka, Embedded atom method potential for studying mechanical properties of binary CuAu alloys, Modell. Simul. Mater. Sci. Eng. **26**, 055006 (2018).

[10] L. B. Pártay, G. Csányi, and N. Bernstein, Nested sampling for materials, Eur. Phys. J. B **94**, 159 (2021).

[11] K. Rossi, L. B. Pártay, G. Csányi, and F. Baletto, Thermodynamics of CuPt nanoalloys, Sci. Rep. **8**, 9150 (2018).

[12] G. A. Marchant, M. A. Caro, B. Karasulu, and L. B. Pártay, Exploring the configuration space of elemental carbon with empirical and machine learned interatomic potentials, npj Comput Mater **9**, 131 (2023).

[13] J. Kloppenburg, L. B. Pártay, H. Jónsson, and M. A. Caro, A general-purpose machine learning Pt interatomic potential for an accurate description of bulk, surfaces, and nanoparticles, J. Chem. Phys. **158**, 134704 (2023).

[14] C. W. Rosenbrock, K. Gubaev, A. V. Shapeev, L. B. Pártay, N. Bernstein, G. Csányi, and G. L. W. Hart, Machine-learned interatomic potentials for alloys and alloy phase diagrams, npj Comput Mater **7**, 24 (2021).

[15] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Machine learning force fields, Chem. Rev. **121**, 10142 (2021).

[16] V. Zaverkin, D. Holzmüller, I. Steinwart, and J. Kästner, Exploring chemical and conformational spaces by batch mode deep active learning, Digital Discovery **1**, 605 (2022).

[17] A. D. Becke, Perspective: Fifty years of density-functional theory in chemical physics, J. Chem. Phys. **140**, 18A301 (2014).

[18] F. Tran, J. Stelzl, and P. Blaha, Rungs 1 to 4 of DFT Jacob's ladder: Extensive test on the lattice constant, bulk modulus, and cohesive energy of solids, J. Chem. Phys. **144**, 204120 (2016).

[19] P. Borlido, J. Schmidt, A. W. Huran, F. Tran, M. A. L. Marques, and S. Botti, Exchange-correlation functionals for band gaps of solids: Benchmark, reparametrization and machine learning, npj Comput. Mater. **6**, 96 (2020).

[20] P. Kovács, F. Tran, P. Blaha, and G. K. H. Madsen, What is the optimal mGGA exchange functional for solids?, J. Chem. Phys. **157**, 094110 (2022).

[21] L.-F. Zhu, B. Grabowski, and J. Neugebauer, Efficient approach to compute melting properties fully from *ab initio* with application to Cu, Phys. Rev. B **96**, 224202 (2017).

[22] F. Dorner, Z. Sukurma, C. Dellago, and G. Kresse, Melting Si: Beyond density functional theory, Phys. Rev. Lett. **121**, 195701 (2018).

[23] J. V. Michelin, L. G. Gonçalves, and J. P. Rino, On the transferability of interaction potentials for condensed phases of silicon, J. Mol. Liq. **285**, 488 (2019).

[24] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, Machine learning a general-purpose interatomic potential for silicon, Phys. Rev. X **8**, 041048 (2018).

[25] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. **77**, 3865 (1996).

[26] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun, Accurate and numerically efficient r2SCAN meta-generalized gradient approximation, J. Phys. Chem. Lett. **11**, 8208 (2020).

[27] G. Kresse and J. Furthmüller, Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, Phys. Rev. B **54**, 11169 (1996).

[28] G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B **59**, 1758 (1999).

[29] H. Montes-Campos, J. Carrete, S. Bichelmaier, L. M. Varela, and G. K. H. Madsen, A differentiable neural-network force field for ionic liquids, J. Chem. Inf. Model. **62**, 88 (2022).

[30] J. Carrete, H. Montes-Campos, R. Wanzenböck, E. Heid, and G. K. H. Madsen, Deep ensembles vs committees for uncertainty estimation in neural-network force fields: Comparison and application to active learning, J. Chem. Phys. **158**, 204801 (2023).

[31] E. Kocer, J. K. Mason, and H. Erturk, Continuous and optimally complete description of chemical environments using spherical Bessel descriptors, AIP Adv. **10**, 015021 (2020).

[32] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2016), pp. 770–778.

[33] P. Ramachandran, B. Zoph, and Q. V. Le, Searching for activation functions, arXiv:1710.05941.

[34] S. Bichelmaier, J. Carrete, R. Wanzenböck, F. Buchner, and G. K. H. Madsen, Neural-network-backed effective harmonic potential study of the ambient pressure phases of hafnia, Phys. Rev. B **107**, 184111 (2023).

[35] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python+NumPy programs (2018), http://github.com/google/jax.

[36] J. Heek, A. Levskaya, A. Oliver, M. Ritter, B. Rondepierre, A. Steiner, and M. van Zee, Flax: A neural network library and ecosystem for JAX (2020), http://github.com/google/flax.

[37] H. Minkowski, U eher die positiven quadratischen Formen und über kettenbruchähnliche Algorithmen, J. Reine Angew. Math. **1891**, 278 (1891).

[38] J. Skilling, Galilean and Hamiltonian Monte Carlo, Proceedings **33**, 19 (2019).

[39] F. Trozzi, X. Wang, and P. Tao, UMAP as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: A comparison study, J. Phys. Chem. B **125**, 5022 (2021).

[40] R. Tamura, M. Matsuda, J. Lin, Y. Futamura, T. Sakurai, and T. Miyazaki, Structural analysis based on unsupervised learning: Search for a characteristic low-dimensional space by local

structures in atomistic simulations, Phys. Rev. B **105**, 075107 (2022).

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. **12**, 2825 (2011).

[42] L. v. d. Maaten and G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. **9**, 2579 (2008).

[43] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevMaterials.7.123804 for a comparison of the PCA, t-SNE, and UMAP dimensionality reductions, the different walker populations, and nested sampling trajectories.

[44] A. Togo and I. Tanaka, `Spglib`: A software library for crystal symmetry search, arXiv:1808.01590.

[45] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis, Comput. Mater. Sci. **68**, 314 (2013).

[46] M. I. McMahon and R. J. Nelmes, New high-pressure phase of Si, Phys. Rev. B **47**, 8337 (1993).

[47] G. A. Voronin, C. Pantea, T. W. Zerda, L. Wang, and Y. Zhao, *In situ* x-ray diffraction study of silicon at pressures up to 15.5 GPa and temperatures up to 1073 K, Phys. Rev. B **68**, 020102(R) (2003).

[48] V. G. Eremenko and V. I. Nikitenko, Electron microscope investigation of the microplastic deformation mechanisms of silicon by indentation, Phys. Status Solidi A **14**, 317 (1972).

[49] H. Zhang, H. Liu, K. Wei, O. O. Kurakevych, Y. Le Godec, Z. Liu, J. Martin, M. Guerrette, G. S. Nolas, and T. A. Strobel, BC8 silicon (Si-III) is a narrow-gap semiconductor, Phys. Rev. Lett. **118**, 146601 (2017).

[50] T. Chen, F. Yuan, J. Liu, H. Geng, L. Zhang, H. Wang, and M. Chen, Modeling the high-pressure solid and liquid phases of tin from deep potentials with *ab initio* accuracy, Phys. Rev. Mater. **7**, 053603 (2023).

[51] N. M. Chtchelkatchev, R. E. Ryltsev, M. V. Magnitskaya, S. M. Gorbunov, K. A. Cherednichenko, V. L. Solozhenko, and V. V. Brazhkin, Local structure, thermodynamics, and melting of boron phosphide at high pressures by deep learning-driven *ab initio* simulations, J. Chem. Phys. **159**, 064507 (2023).

[52] H. Montes-Campos, J. Carrete, S. Bichelmaier, L. M. Varela, and G. K. H. Madsen, NEURALIL, https://github.com/Madsen-s-research-group/neuralil-public-releases.

[53] N. Bernstein, R. J. N. Baldock, L. B. Pártay, J. R. Kermode, T. D. Daff, A. P. Bartók, and G. Csányi, PYMATNEST, https://github.com/libAtoms/pymatnest.

[54] N. Unglert, J. Carrete, L. B. Pártay, and G. K. H. Madsen, Neural-network force field backed nested sampling: Study of the silicon p-T phase diagram, 2023, doi: 10.5281/zenodo.8143152.