

Atomic cluster expansion potential for the Si-H system

Louise A. M. Rosset¹ and Volker L. Deringer¹**Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford OX1 3QR, United Kingdom*

(Received 23 October 2025; accepted 24 February 2026; published 21 April 2026)

The silicon-hydrogen system is of key interest for solar-cell devices, including both crystalline and amorphous modifications. Elemental amorphous Si is now well understood, but the atomic-scale effects of hydrogenating the silicon matrix remain to be fully explored. Here, we present a machine-learned interatomic potential model based on the atomic cluster expansion (ACE) framework that can describe a wide range of Si-H phases, from crystalline and amorphous bulk structures to surfaces and molecules. We perform numerical and physical validation across a range of hydrogen concentrations and compare our results to experimental findings. Our work constitutes an advancement toward the exploration of large structural models of *a*-Si:H at realistic device scales.

DOI: [10.1103/891y-wm1f](https://doi.org/10.1103/891y-wm1f)

I. INTRODUCTION

Amorphous silicon (*a*-Si) is a prototypical disordered material which continues to be of interest for a wide range of applications, from interferometer mirror coatings [1] to novel solar-cell heterojunction technologies [2,3]. In these devices, the amorphous matrix often contains a varied concentration of hydrogen, from 5 to 20 at.-% H [4], and the resulting materials are denoted as *a*-Si:H.

The properties of *a*-Si:H have long been rationalized in terms of the underlying structure [5]. Adding hydrogen formally disrupts the (idealized) fourfold-connected continuous random network of Si atoms [6], and leads to the passivation of defect states such that *a*-Si:H structure models show a clean gap in the electronic density of states (DOS) [7,8]. Hydrogen enhances the stability and carrier mobility in *a*-Si:H [4], as well as reduces electron-hole recombinations at defect sites, providing higher conversion efficiency in p-n solar cells [9]. A challenge in exploring the effect of hydrogenating silicon has been the cost of first-principles (*ab initio*) simulation methods and the need for very short molecular-dynamics (MD) timesteps to describe the motion of the light hydrogen atoms. Thus, many open questions still remain around *a*-Si:H, such as its degradation under light, known as the Staebler-Wronski effect [10], or the nature of “protocrystalline” solar cells [11].

Atomistic modeling of *a*-Si and, to a lesser extent, *a*-Si:H has been accelerated by machine-learned interatomic potential (MLIP) models [12]. An important milestone in this regard has been a general-purpose Gaussian approximation potential (GAP) for elemental silicon [13], subsequently denoted “Si-GAP-18”: this model has been applied to the study

of ambient- and high-pressure disordered forms of silicon [14–16]. It was later shown how Si-GAP-18 could be distilled into a faster “student” model [17], and the latter has been applied to study defects in a million-atom simulation of *a*-Si [18] as well as the presence of local structural order in the amorphous phase [19]. There was also a report of a GAP model for *a*-Si:H [20], to which we refer as “SiH-GAP-22” in the following, which was applied to modeling crystalline/*a*-Si:H interfaces [21]. However, the latter model was deliberately focused on the disordered phases, and a full MLIP-driven exploration of the binary Si-H system has not yet been reported to the best of our knowledge.

A unified description of the Si-O system was achieved in a previous study [22] with the atomic cluster expansion (ACE) framework [23–25]. We now report a description of the Si-H system using similar methodology. Indeed, bespoke ACE-based MLIPs have reached highly competitive simulation length and time scales and have provided insights into complex structural transitions [26,27]; a recent benchmark across a range of MLIP fitting frameworks showed that non-linear ACE models enable fast inference while maintaining good accuracy [28].

In the present work, we introduce an MLIP model for the binary Si-H system based on the ACE framework. We describe the development of our training dataset for Si-H through an active-learning cycle, the careful validation of the model against DFT reference data, and comparison with experimental observables where applicable. We evaluate the potential’s performance across bulk and surface benchmarks and illustrate its usefulness by studying the mechanical properties of *a*-Si:H.

II. METHODOLOGY

A. A training dataset for the Si-H system

Figure 1 presents the overall approach for the generation of the reference dataset and the iterative training of the potential using an active-learning (AL) workflow. AL aims to iteratively

*Contact author: volker.deringer@chem.ox.ac.uk

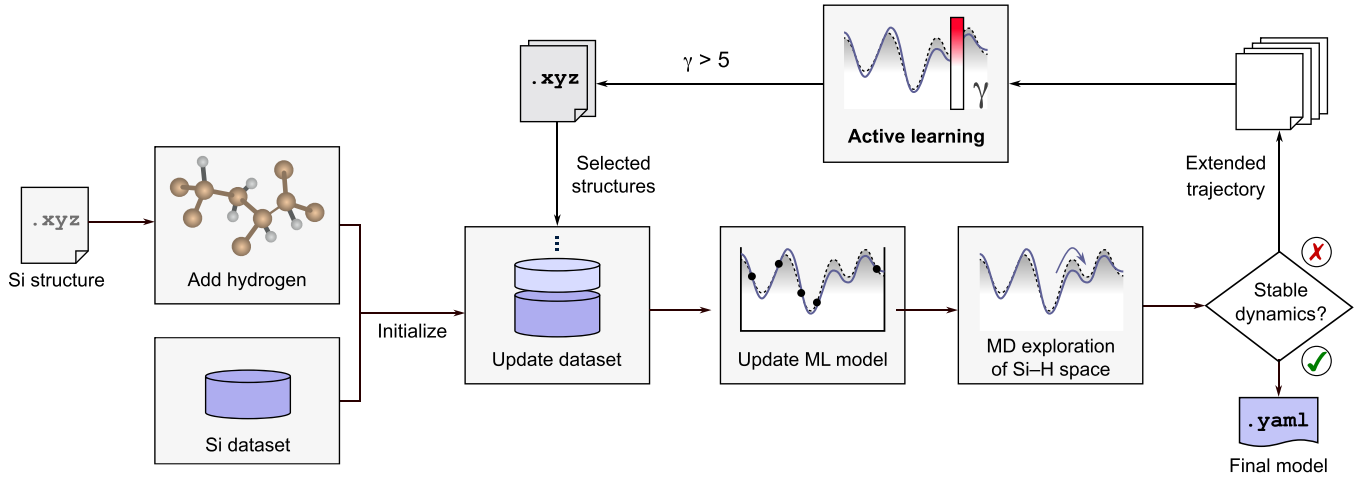


FIG. 1. Simplified overview of the approach for MLIP training dataset construction, including iterative exploration and ACE model fitting, adopted in the present study.

expand a potential’s training dataset in regions of the configurational space where its predictions are uncertain. A range of AL strategies have been proposed [29–34], typically relying on uncertainty metrics that assess the potential’s confidence in its own predictions [35,36].

As an initial dataset, we took large α -Si structures generated in a previous study [19], to which we added crystalline Si configurations obtained by systematically scaling and distorting relevant unit cells, and randomly displacing atoms from their relaxed positions. We also added liquid Si structures of varying densities from separate MD runs at 5000 and 9000 K, respectively.

To supplement this dataset with Si-H configurations, we added SiH_4 , Si_2H_6 , and Si_3H_8 molecules as well as SiH_2 and SiH_3 radicals. Using a simple script for adding hydrogen, we generated additional diamond-type, liquid, and amorphous structures, each hydrogenated with up to 20 at.-% H. The hydrogenation script used rules informed by literature [37–41]: adding H to dangling Si bonds, creating voids by removing Si atoms and decorating the new coordination defects, or placing H_2 molecules in voids. Some H atoms were also introduced as interstitial or substitutional defects. We note that hydrogen can adopt a wealth of conformations in the silicon matrix [42], and that the script only provided a starting point for hydrogenation.

Prior to fitting, we applied filters to our initial dataset and to all subsequent additions made via AL. We removed all structures with positive DFT energies relative to the energy of isolated Si and H atoms, with a maximum force magnitude of >50 eV/Å, with Si-Si distances of <1.6 Å, or with Si-H distances of <1 Å. Details of the DFT computations are described in Appendix A. We then fit an initial MLIP model to this “Iter-0” dataset, and used it to extend the breadth of the configurational space covered by the dataset.

To choose which structures to add to the dataset with AL, we used the extrapolation grade, γ [31], based on the D-optimality criterion [46] as a measure of the interpolation ($\gamma < 1$) or extrapolation ($\gamma > 1$) behavior of the model on an unseen configuration, relative to the subset of its training dataset that spans the widest space. By evaluating γ over MD

simulations, we selected all structures encountered during the simulation that had $\gamma > 5$, i.e., a value placing the structure firmly outside the interpolation regime. We performed Farthest Point Sampling [47] on this set of structures, which is a well-established sampling strategy for building MLIP datasets [48–50]; we note in passing that other sampling algorithms have been recently explored [51]. We use SOAP descriptors [44] as our distance metric, to determine the subset of structures that would provide the widest sampling of the relevant configurational space.

Once the potential no longer produced any $\gamma > 5$ structures for simulations of amorphous and crystalline Si-H, we simulated surface structures by cleaving and relaxing bulk amorphous structures. We also decorated crystalline (diamond-type) Si surfaces with hydrogen, specifically the (100), (111), (221), and (311) surfaces. We added small random clusters by random structure search and relaxation, following the ideas of Refs. [52] and [53], as these have been shown to substantially improve model stability in MD [27]. The full AL protocol is detailed in Appendix B. We stopped AL when we no longer encountered $\gamma > 5$ configurations when performing MD in relevant phase space. The potential performed well across the Si-H space and would no longer be expected to benefit substantially from additional training structures.

The complete dataset composition, by configuration type and AL iteration, is detailed in Table I and visualized in Fig. 2. As a metric of comparison, the Si-GAP-18 training dataset contains 2475 structures and 171 815 atomic environments [13], while the SiH-GAP-22 training dataset contains 390 structures with 60 770 Si and 5137 H atoms for a total of 65 907 atomic environments.

We set aside 20% of the entire set of structures generated by the protocol as a validation set, splitting by configuration type to enforce proportional representation. We also defined two further test sets: 10% of the training set of Si-GAP-18, to which we refer as “Si-GAP-18-set”, and the totality of the training set of SiH-GAP-22 (“SiH-GAP-22-set”). Both datasets were relabeled with the same DFT settings as were used for dataset generation (Appendix A). We evaluate

TABLE I. Composition of the training dataset for the SiH-ACE-25 potential, detailing the number of structures and number of atomic environments per element for each configuration type.

Type	Structures	Atomic environments		
		Si	H	Total
Amorphous	2030	265 453	14 233	279 686
Liquid	697	56 515	12 450	68 965
Crystalline	1289	51 296	2995	54 291
High pressure	717	16 776	—	16 776
Molecules	438	997	4487	5484
Surfaces	54	2812	2550	5362
Random clusters	382	808	1650	2458
Total	5607	394 657	38 365	433 022

our model on these training sets as they are more diverse and comprehensive than the test sets defined in the original publications [13,20]: these test sets only contain selected configurations, rather than splits from the respective full training sets.

B. ACE model parametrization

With the training dataset in place, we turn to the parametrization of the potential. A particularity of our training dataset is that it contains $10\times$ more Si environments than H environments (Table I), and that these Si environments are very diverse, as illustrated in Fig. 2. This focus on Si is a reflection of our interest in hydrogenated silicon phases, with hydrogen content below 30 at.-% H. To best describe the

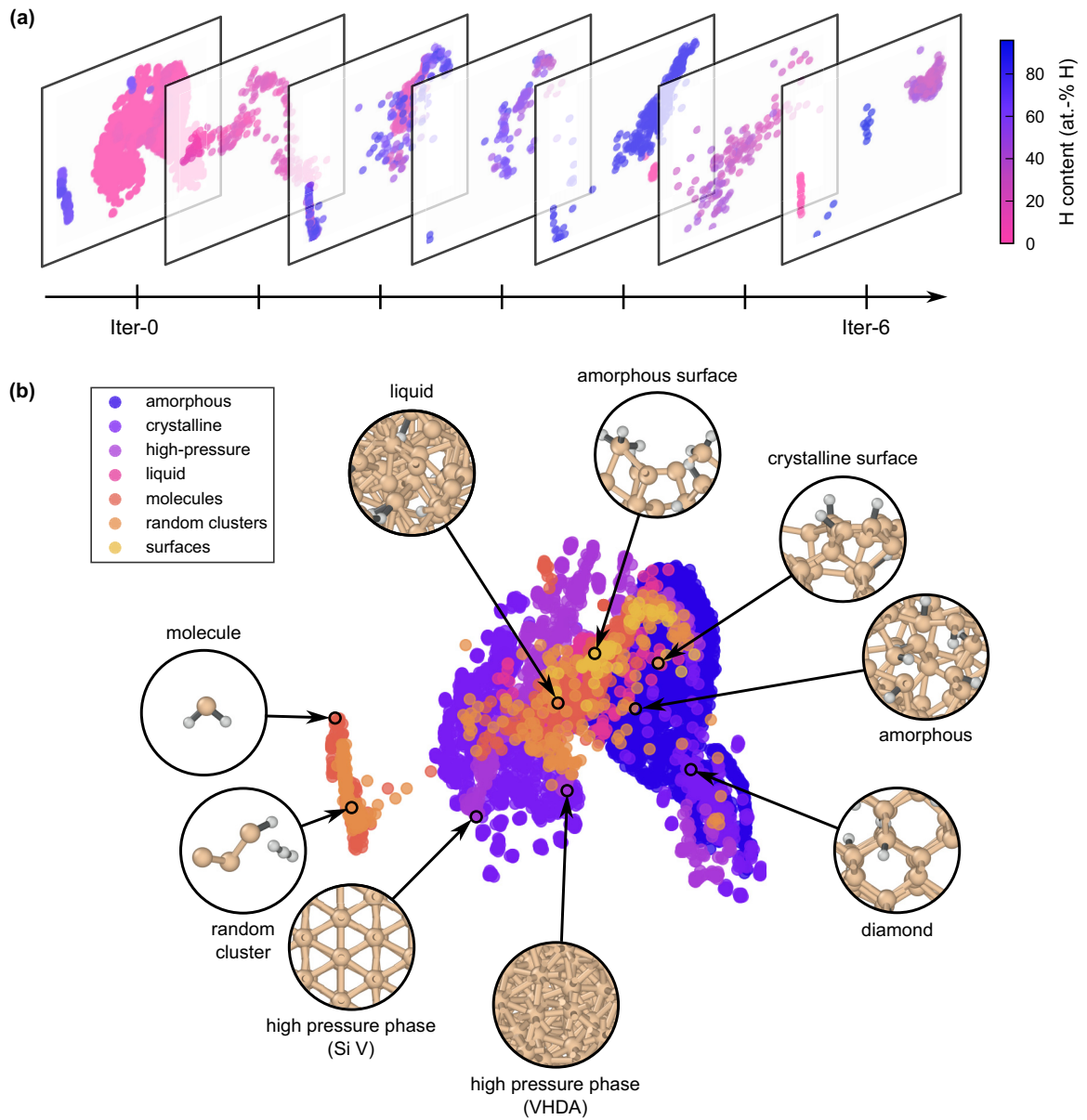


FIG. 2. The SiH-25 dataset. (a) Training dataset evolution over active learning iterations, color-coded by hydrogen content, visualized using UMAP [43] and SOAP descriptors [44]. (b) Projected map of the full dataset, where each point represents a structure in the training set and is color-coded by structure type. Representative structures of each configurational type, visualized using OVITO [45], are shown as insets.

Si-H system, we opt for a custom potential shape that contains different numbers of basis functions for each element, unlike some other multielement ACE potentials reported earlier [27,54]. We compare the numerical predictions of these custom, “asymmetric” models to “symmetric” models of the same total size in Appendix C, and show that the asymmetric models perform marginally better.

To inform our choice of model parametrization, we systematically vary the number of basis functions per block¹, while keeping the other three blocks fixed at 1600, 800, and 600 for Si, H, and SiH respectively, chosen as “medium” parametrization for each. We train five randomly initialized models for each parametrization. As our interest lies in the overall trends, we only use a 20% split of our total training dataset to fit these models. For each basis size, we evaluate a combined energy and force loss function on the SiH-GAP-22-set [Fig. 3(a)].

For both the Si and SiH elemental blocks, increasing the number of functions and hence the complexity of the model does not offer a substantial decrease in the loss. Conversely, increasing the number of H basis functions first improves, then notably *worsens* the loss—this could reflect an overfitting scenario related to the dataset’s heterogeneity. Figure 3(a) warns against too many H basis functions, but does not otherwise indicate a preferential basis size for the Si and SiH blocks.

To further guide our choice, we fit three potentials of increasing computational cost: a “cheap” model with 2300 total functions, a “medium” model with 3600 total functions, and an “expensive” model with 4600 total functions, where the elemental basis block sizes are shown as vertical lines in Fig. 3(a). As a metric of cost comparison, the “cheap”, “medium”, and “expensive” potentials can run 1.825, 1.284, and 1.084 ns/day, respectively, for a system of 1000 atoms and a timestep of 0.1 fs on an NVIDIA A100 GPU card.

We then compare the individual energy and force predictions from these three models on the SiH-GAP-22-set, in Fig. 3(b), and ensure that the results are converged for the size of the training dataset, such that increasing the dataset size does not substantially improve the energy or force RMSE. All three models perform comparably in terms of energy and force metrics, achieving energy errors under 10 meV/at. and force errors around 150 meV/Å.

We move beyond numerical testing with a domain-specific test on a difficult benchmark: a 10 ps simulation in the NpT ensemble of a liquid Si-H system of 1000 atoms with 50 at.-% H at 2000 K. This constitutes a challenging test as hydrogen is highly mobile and diffuses quickly at this elevated temperature, leading to short interatomic distances and high forces. We assess the stability of the MD simulations by plotting the average box size as a function of simulation time, with five repeats of the protocol for each model, comparing our “asymmetric” models to their “symmetric” counterparts [Fig. 3(c)].

Despite their good numerical accuracy (Appendix C), all three symmetric models are unstable in MD: the simulation box quickly expands unreasonably (“explodes”) for the

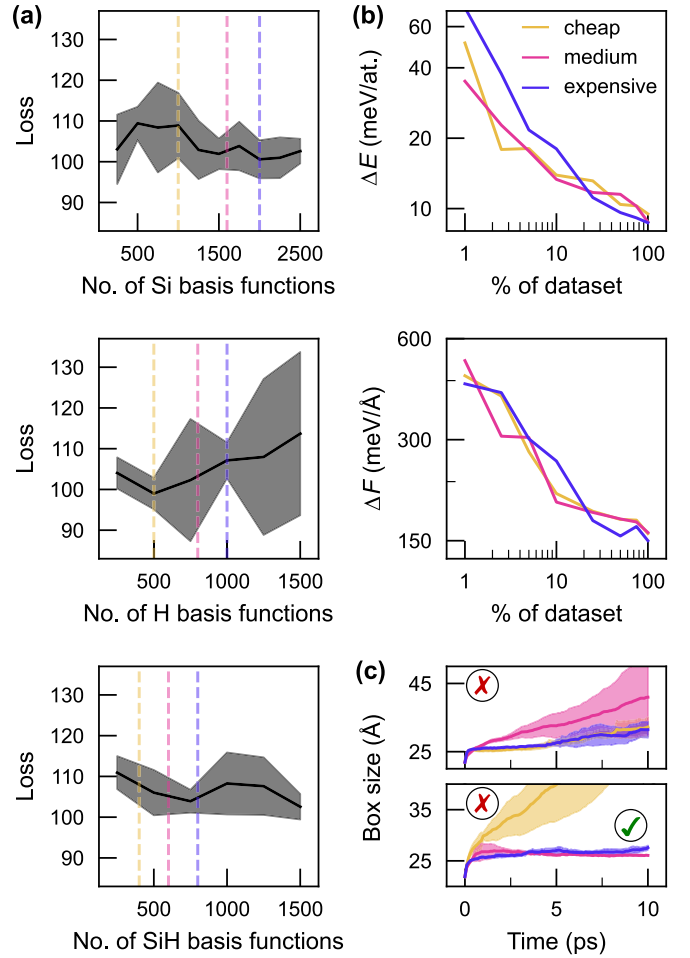


FIG. 3. (a) Combined energy and force loss against number of elemental basis functions for Si (top), H (middle) and SiH (bottom), where we plot the average of five randomly initiated models and the standard deviation in gray. Colored vertical dashed line indicate the basis sizes for three models of increasing cost. (b) Energy and force RMSE as a function of dataset size for three models of increasing cost, where we systematically pick the best of five randomly initiated models, presented as log-log plots. (c) Evolution of box size against simulation time for three models of increasing cost, plotting the average and the standard deviation of five repeats for the “symmetric” models (top) and “asymmetric” models (bottom).

“medium” model, while the box expansion has a delayed onset of around 5 ps for the “cheap” and “expensive” models. Similarly, the “cheap” asymmetric potential is too simple to stabilize the liquid Si-H mixture at 2000 K, and the simulation box explodes. However, the “medium” and “expensive” asymmetric potentials do capture the correct dynamics of this challenging test scenario. This is likely due to their greater number of Si basis functions relative to their symmetric counterparts of the same size, which results in twice as many four-body and five-body Si terms, for example, in the asymmetric “medium” model. These asymmetric models therefore provide a more complete parametrization of the higher body order interactions, which are important in complex phases like liquids [55], although our understanding of the behavior of a model as a function of its “body-orderedness” is still primitive

¹Basis functions are organized by species blocks. An elemental model will have a singular block, while an A-B binary model will have four blocks: A-A, B-B, A-B, and B-A.

TABLE II. Predictions of the SiH-ACE-25 potential across three datasets: a validation set obtained as 20% of our AL-generated dataset; a 10% split of the Si-GAP-18 training set; and the SiH-GAP-22 training set. We detail validation-set composition by structure and atomic environments, and provide energy and force-component RMSE values.

Type	ΔE (meV/at.)	ΔF (eV/Å)		
		Si	H	All
Amorphous	11.5	0.12	0.23	0.13
Liquid	19.8	0.22	0.44	0.28
Crystalline	11.9	0.11	0.21	0.12
High pressure ^a	14.1	0.22	–	0.22
Molecules	79.4	0.82	0.44	0.53
Surfaces	24.4	0.24	0.18	0.22
Random clusters	147.0	0.80	0.61	0.67
Total	46.3	0.15	0.36	0.18
Si-GAP-18-set	24.5	0.13	–	0.13
SiH-GAP-22-set	8.7	0.15	0.30	0.16

^aSi structures only

[56]. This justifies the choice of a custom potential shape: for the same overall model size, the “medium” and “expensive” asymmetric models are stable while their symmetric counterparts are not.

As the “medium” and “expensive” potentials perform comparably well both numerically and physically, we select the “medium” potential shape for its reduced complexity and cost. With this custom “asymmetric” shape, a final model was fitted with optimized hyperparameters (Appendix A), hereafter denoted as “SiH-ACE-25”.

C. The SiH-ACE-25 model

Table II details the SiH-ACE-25 model’s root-mean-square error (RMSE) values for energy and force predictions across all three validation sets considered. The errors on our validation set have been separated out by structure type to highlight which of the configuration types are more challenging to describe.

From our validation dataset, the more disordered configurations generally incur higher errors, with the random clusters (along with molecules) having the largest errors likely as they contain larger forces on average than other configurations. Furthermore, these configurations account for less than 2% of the total atomic environments seen in training (cf. Table I); more training data might reduce these errors. The hydrogen environments are more challenging for the potential, as seen by the higher force errors on H than Si, by about a factor of 2 for bulk configurations. This stems from the unbalanced dataset that contains far fewer H atomic environments. Overall, the model performs well across the dataset, particularly on bulk amorphous and crystalline configurations.

We further test the model’s performance on the Si-GAP-18-set and SiH-GAP-22-set validation sets. The Si-GAP-18-set contains 229 structures and 52 630 atomic environments, while the SiH-GAP-22-set contains 358 structures and 52 630 atomic environments, of which only 3698 are H environments.

TABLE III. Properties of diamond-type silicon, comparing SiH-ACE-25 predictions to SCAN and literature values for the lattice constant (a_0), the atomization energy (E_0), and the bulk modulus (B) of diamond-type silicon, and the formation energies of a silicon vacancy (V_{Si}), a hydrogen interstitial (H_i), and a H_2 interstitial (H_{2i}).

	This work		Expt. (Lit.)
	SCAN	SiH-ACE-25	
a_0 (Å)	5.43	5.42	5.43 [58]
E_0 (eV/at.)	–5.86	–5.86	–
B (GPa)	94.17	98.08	99.0 [59]
	This work		DFT (Lit.)
	SCAN	SiH-ACE-25	
V_{Si} (eV)	3.94	3.41	3.17 [60]
H_i (eV)	1.56	0.55	1.04 [61]
H_{2i} (eV/at.)	0.66	0.23	0.80 [62]

The potential shows very good performance on both sets, with much lower average errors than on our validation dataset. This is likely due to the fact that the configurational space sampled in these datasets is more limited, such that our validation set is a more challenging numerical benchmark for the potential.

III. VALIDATION

A. Crystalline-phase properties

We first evaluate the performance of our model on a range of properties of the crystalline diamond-type modification of silicon, **dia**-Si (Table III). We include formation energies for hydrogen defects in the crystalline network, reporting values for the two most common defect types: interstitial H atoms and H_2 molecules [57]. We note that these configurations have not been explicitly included in the training set.

The SiH-ACE-25 model provides excellent agreement with experiment for the lattice parameter, a_0 , of **dia**-Si. It also provides excellent agreement with DFT for the atomization energy (relative to an isolated Si atom), E_0 . The potential’s prediction of the bulk modulus, B , closely matches experiment. We note that a benchmark study of estimated bulk moduli of **dia**-Si found a variation of about 10 GPa across different DFT functionals [63].

We next turn to the energies of various defects in **dia**-Si. We focus on the relaxed structures and the relative ordering of defect energies, rather than on the absolute values as these vary strongly with DFT functionals and parametrizations [64]. The relaxation around silicon vacancy (V_{Si}) defects is a well-documented, challenging test that requires large structural models and well-converged parameters, or will otherwise lead to incorrect outward relaxation behavior [60]. SiH-ACE-25 correctly predicts the Si atoms to relax inward around the V_{Si} site, with an inward movement of magnitude 0.28 Å, compared to a SCAN-DFT prediction of a magnitude of 0.16 Å.

Interstitial H atoms adopt a bond-centered configuration, wherein the H_i defect sits at the center of a nearest-neighbor Si-Si bond, causing a substantial outward relaxation [65]. Our model reproduces the correct relaxed symmetry, stretching the

bond from 2.35 to 3.19 Å. This compares to a relaxed bond length of 3.20 Å from DFT and 3.25 Å from literature [65]. However, the energy prediction from SiH-ACE-25 is substantially lower than those from both our SCAN computations and literature.

Finally, we assess the model's predictions of an interstitial H_2 molecule in the **dia**-Si structure. The H_{2i} energies quoted in Table III are given per H atom in an interstitial H_2 molecule in the bulk, relative to that atom's energy in a H_2 molecule in a vacuum. The SiH-ACE-25 potential correctly predicts that an H atom in an H_2 interstitial defect is more stable than it would be as a singular H_i defect [57,66]. The potential relaxes the H_2 molecule bond length to 0.774 Å, compared to 0.766 Å for DFT.

B. Amorphous structures

We use simulated melt-quenching to generate a range of α -Si:H configurations with a total of 1000 atoms each, and hydrogen content ranging from 0 to 50 at.-% H in increments of 5 at.-% H, with ten repeats at each concentration for better statistics. We note that the relatively small model size does not appear to induce size effects,² and that further details of the protocol are outlined in Appendix A. We then study the structural properties of these configurations.

We first plot the partial structure factors in Fig. 4, alongside experimental neutron-scattering data measured for α -Si:H samples of 16 at.-% H made by sputtering (taken from Ref. [68]). Hydrogen immediately affects the connectivity of the silicon backbone network, even in the medium regime of 10 at.-% H: it disrupts the network by bonding to Si atoms that would otherwise be bonded to other Si atoms, reflected by a broadening of the partial structure factor and decrease of all peak heights. Further increase of the amount of hydrogen in the amorphous matrix further diffuses these peaks. Both the partial structure factors for Si-H and H-H sharpen as the hydrogen content in the amorphous matrix increases. Absent at low hydrogen content, a clear signature of the H-H partial structure factor emerges at high hydrogen content, which reflects local ordering of the H atoms.

Comparing to the experimental sample with ~ 16 at.-% H, the melt-quenched models match the experimental data for Si-Si and Si-H structure factors well, reflecting the quality in prediction of the underlying amorphous matrix. Aside from the first sharp peak, the experimental H-H signal is quite noisy from strong incoherent scattering [5,68,69]. While Ref. [69] constitutes a more recent study of the neutron-scattering data for hydrogenated and deuterated α -Si, it only provides a noise- and background-corrected pattern for the deuterated sample, so we opt to compare to Ref. [68] instead. Further experimental measurements of the structure factor of α -Si:H across preparation methods and a range of hydrogen content would substantially strengthen the comparison to experiment and validation of computational models.

We found that the real-space structural characterization of these α -Si:H samples, including silicon coordination, bond

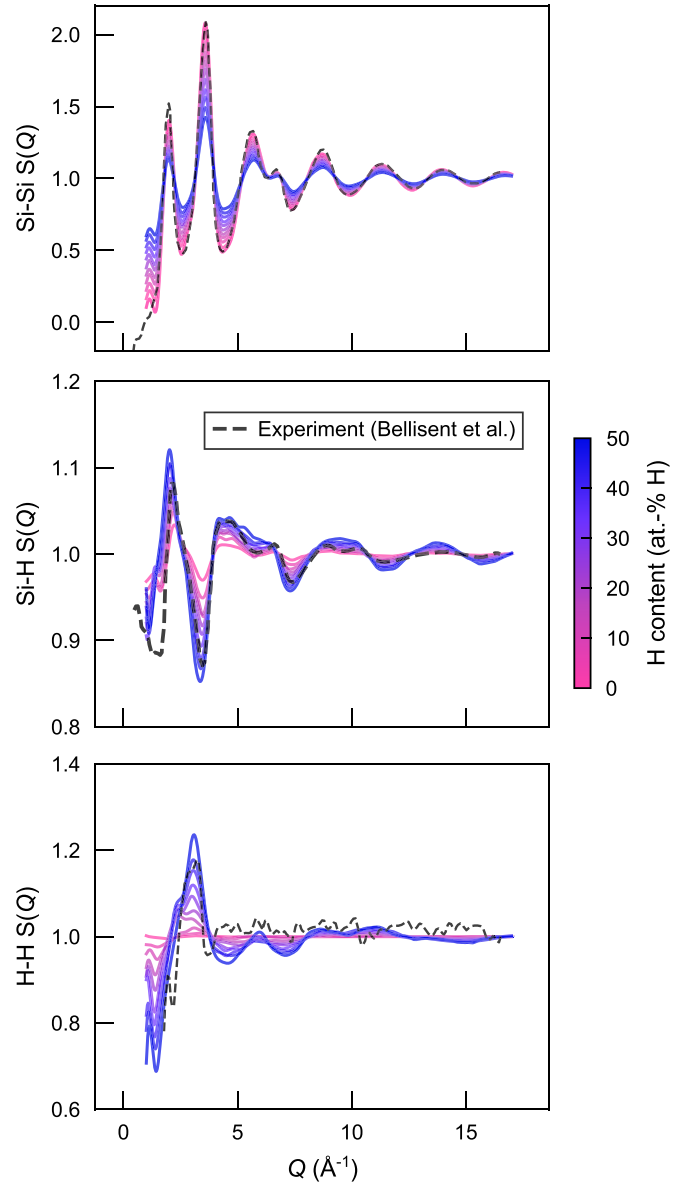


FIG. 4. Average partial structure factors of annealed α -Si:H structures, showing Si-Si (top), Si-H (middle), and H-H (bottom) contributions. Structure factors are averaged over ten repeats at each concentration. Experimental measurements from Ref. [68] for a sample with ~ 16 at.-% H are plotted as dashed black lines, normalized for direct comparison.

angle and ring distributions, is largely unchanged by increasing hydrogen content³—instead, we turn to a detailed picture of the evolution of atomic environments in α -Si:H with increasing hydrogen content in Fig. 5. Pores in pure α -Si are artefacts of the deposition method [71], and hence are not expected in a simulated quenched structure [72]. In contrast, for α -Si:H, as the hydrogen content increases, pores do appear and grow to account for a substantial fractional volume of the box. Here, we calculate the pore volume by placing a fine 3D mesh on the structure and evaluating the number of empty grid

²We show that models of 10 000 atoms reproduce these findings at [67].

³Structural analysis supporting this statement is available at [67].

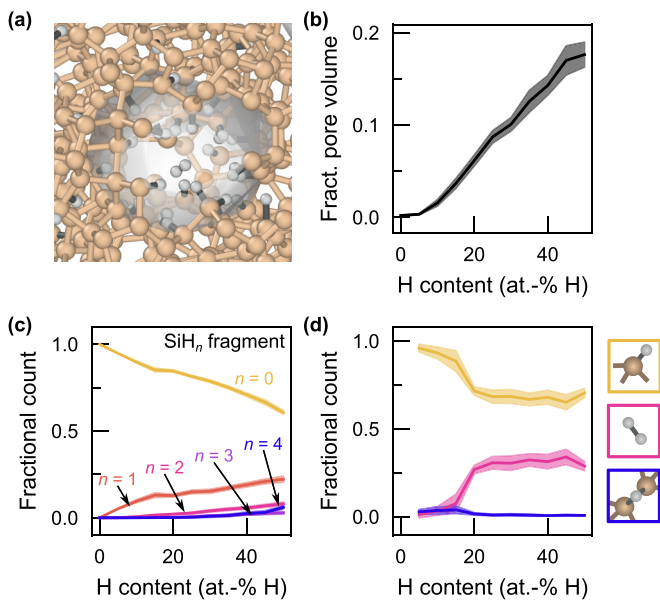


FIG. 5. (a) Structure visualization of a pore in an a -Si:H sample with 10 at.% H, showing the pore edges shaded in gray using the OVITO alpha-shape mesh modifier [45,70]. (b) Fractional pore volume in the a -Si:H structures as a function of hydrogen content. (c) Count of the occurrence of SiH_n fragments in the a -Si:H structures relative to the total number of Si atoms, for n taking values from 0 to 4, as a function of hydrogen content in the structures. (d) Count of the occurrence of three different H fragments relative to the total number of H atoms as a function of hydrogen content in the structures, showing the occurrence of H atoms bonded to one Si atom (top, yellow), H atoms bonded to one H atom to form a H_2 molecule (middle, pink), and bond-centered H atoms, bonded to two Si atoms (bottom, blue). Visualizations are added to illustrate the structural motifs. We report averaged results over ten repeats at each concentration; shading indicates the standard deviations across repeats.

elements, using a script developed in the context of Ref. [73], and present results in Fig. 5(a). Open and closed porosity are difficult to measure experimentally, but the fractional pore volume in the low-H regime is consistent with experimental observations of pore fractions around 1% to 3% for samples of H content around 3–8 at.% H [74]. As a metric of comparison, our 5 and 10 at.% H models display average pore fractions of 0.3% and 1.6%, respectively. Pore size and pore volume fraction vary with hydrogen content, but also with other parameters of film preparation, such as deposition rate or annealing time [74].

We analyze the local motifs present in the amorphous structures as the hydrogen content is varied, first looking at the occurrence of SiH_n fragments in Fig. 5(c). The preferred structural motif is the monohydride, SiH , fragment, in which a silicon atom is bonded to a singular hydrogen atom (in addition to other silicon atoms). This corresponds to a dangling bond in the Si network that has been passivated by a hydrogen atom. The abundance of this fragment type steadily grows as the hydrogen content increases. As hydrogen can be incorporated into amorphous silicon in excess of the dangling-bond population, the monohydride fragments likely occur in

a “broken-bond” model wherein a Si-Si bond is broken to accommodate for two monohydride environments [42].

The dihydride (SiH_2) fragment is also present, albeit in a much lower proportion, which is consistent with previous reports of $-\text{SiH}_2$ on the edges of large voids [75]. The fractional count of SiH_2 motifs grows steadily alongside the increase in pore volume. While very few trihydride (SiH_3) fragments appear even at very high hydrogen concentrations, a small amount of SiH_4 molecules ($\approx 6\%$ of H environments) appear at very high hydrogen concentrations, where they detach from the Si network to sit in large voids.

Turning to the hydrogen environments in the amorphous matrix in Fig. 5(d), the coordination of hydrogen to a single silicon atom is most prevalent (yellow), regardless of the SiH_n fragment. The “bridge” conformation [42], shown in blue, has been experimentally resolved as a secondary motif in a -Si:H by photoemission spectroscopy [76], which agrees with the low proportion of this conformation shown in our quenched structures.

As the hydrogen content increases, H_2 molecules form in large voids. Indeed, once an a -Si:H structure has been hydrogenated past its solubility limit, excess hydrogen will form H_2 molecules [37]. This has been observed experimentally across preparation methods [77–79]. While the relative fractional content is not straightforward to resolve experimentally, H_2 has been estimated to account for about 13% of the hydrogen content in films of 7.5 at.% H [80]. However, the solubility limit of hydrogen in the disordered silicon matrix varies greatly depending on the preparation method: Reference [37] predicts a solubility limit of 4 at.% H for ion-implanted samples, while samples prepared by chemical vapor deposition in Ref. [38] had a solubility of 16 at.% H.

C. Liquid structures

We study the liquid Si-H system by generating a range of liquid structures of 1000 atoms with increasing H content at a density of 2.58 g cm^{-3} , around the estimated density of liquid silicon [20,81]. We equilibrate these configurations at 2000 K in the NVT ensemble for 10 ps with a timestep of 0.1 fs, running ten repeats for each hydrogen concentration. We plot the results in Fig. 6, assessing real-space structural characteristics as a function of hydrogen content.

Figure 6(a) shows the coordination-number distributions for the Si atoms as a function of the hydrogen content in the melt. As the latter increases, the silicon coordination distribution shifts towards higher coordination numbers, from predominantly sixfold to sevenfold coordinated environments. The bond-angle distribution trend, shown in Fig. 6(b), reflects the increase in strained Si environments with an increase in angles of $\approx 60^\circ$ as the hydrogen content increases. The other dominant peak is at $\approx 95^\circ$, as determined by peak fitting, and weakens as more strained angles appear. The count of m -membered rings between Si atoms [Fig. 6(c)] shows the predominance of $m = 3$ rings regardless of hydrogen content, and indicates that increasing the hydrogen content reduces the count of intermediate rings ($4 \leq m \leq 7$) in favor of smaller $m = 3$ or larger $m > 7$ rings. This further confirms that increasing hydrogen content gives rise to more strained environments.

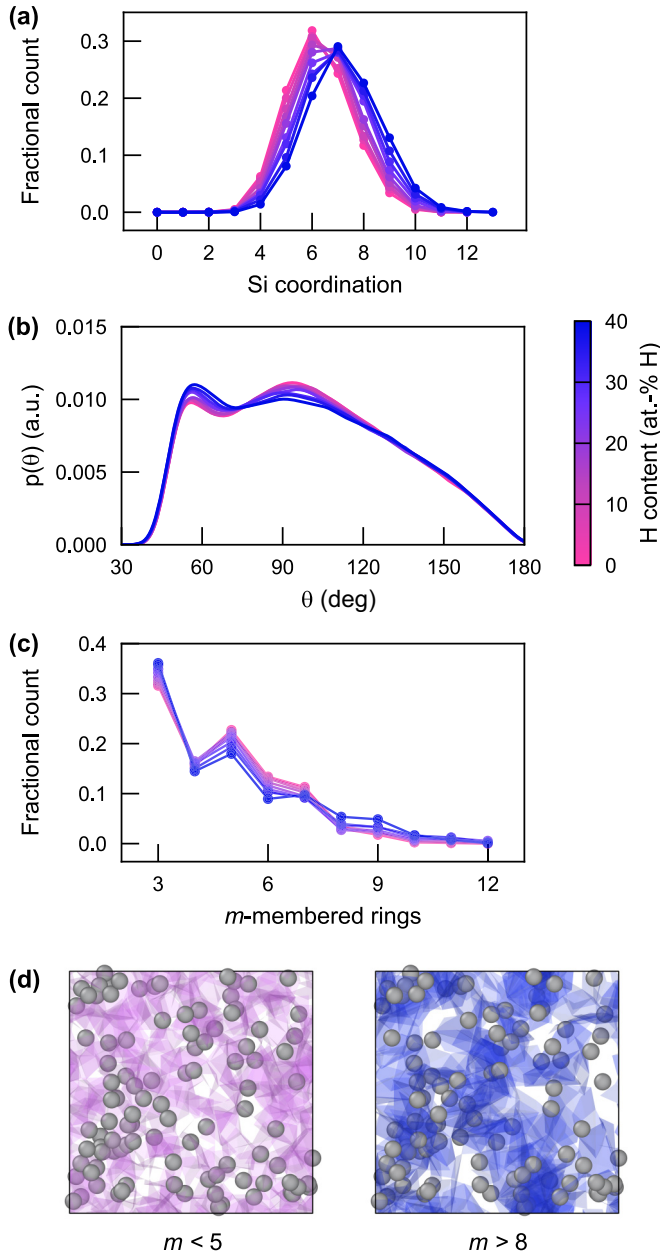


FIG. 6. (a) Fractional count of silicon coordination in liquid Si-H mixtures as a function of hydrogen content. (b) Si-Si-Si bond angle distribution. (c) Distribution of m -membered shortest-path rings. Distributions are averaged over ten repeats at each concentration. (d) Structure visualization of the H atoms in a Si-H melt with 10 at.-% H, highlighting m -membered rings with (left) $m < 5$ and (right) $m > 8$, respectively, visualized using OVITO [45].

This analysis offers insight into the effect of hydrogen on the structure of liquid silicon. One possible explanation, as explored in Fig. 6(d), is that concentration gradients of hydrogen appear in the melt, resulting in H-rich and H-poor areas. The H-rich areas show larger rings as hydrogen disrupts the Si network due to its low coordination count. The $m = 3$ rings would then originate from locally hydrogen-poor regions, where Si atoms have high coordination numbers. The more hydrogen, the greater the ring sizes and pores in the H-rich areas of the network, leading to the local densification

of the H-poor areas which see an increase in coordination and $m = 3$ ring count.

The Si-H liquid phase is not well characterized in the literature, as most preparation methods for α -Si:H rely on molecular precursors and bypass the liquid phase. Liquid silicon adopts complex conformations with several liquid-liquid phase transitions [81], but the effect of hydrogen on these transitions is unknown.

D. Crystalline surfaces

We assess the predictions of our model for clean and hydrogen-passivated crystalline (diamond-type) silicon surfaces in Fig. 7. We note that the training dataset includes hydrogenated crystalline surfaces, but no pristine Si ones, which therefore are “unseen” in training—although it is likely that these structures have some surface silicon atoms without any hydrogen in their environment. By comparison, the Si-GAP-18 training dataset does include a large number of diamond-type Si surface configurations [13]. Adding crystalline silicon surfaces to our training dataset would improve the SiH-ACE-25 model’s performance.

We first test our model’s performance on unseen clean Si surfaces, by evaluating the surface energies of the cleaved and relaxed Si(100) and Si(110) surfaces, as shown in Fig. 7(a). Our potential correctly orders the surfaces by relative stability, but slightly underestimates the Si(100) surface while overestimating Si(110) compared to DFT. The model relaxes the (100) surface to its (2×1) reconstruction [85], and simply relaxes the bond lengths of the Si(110) surface, for which there is no consensus on a reconstruction [86]. Overall, the potential provides a satisfying match to both DFT and experimental results.

A challenging benchmark for crystalline silicon is the energy of the Si(111) surface and its complex reconstructions [87], which we present in Fig. 7(b). The Si(111) reconstructions have been thoroughly studied and the (7×7) reconstruction has been identified as the most stable [88–90]. SiH-ACE-25 incorrectly predicts that the relaxed unreconstructed (1×1) surface is more stable than the reconstructed surfaces, likely as it has not seen any of these configurations during training. While Si-GAP-18 correctly predicts that the reconstructions are less energetic than the ideal surface, which it has seen in training, even this state-of-the-art potential incorrectly predicts the (5×5) reconstruction as the most stable. The correct ordering is shown as black squares from DFT calculations in Ref. [83]. Reproducing the correct ordering of the energies of the Si(111) surface reconstructions has been a longstanding challenge for both DFT and MLIPs [13]. The correct ordering was recently achieved with a bespoke neural-network potential targeting the Si(111) surface [91]. Despite this success, the Si(111) reconstructions remain highly non-trivial and complex configurations for any potential.

In Fig. 7(c), we show results for decorating Si(100) and (111) slabs with H atoms and investigating the relaxation behavior. Other, far more complex surface reconstructions with hydrogen have been investigated in the literature [92–95], but Si(100)-H and Si(111)-H are the most commonly occurring hydrogenated surfaces [86]. SiH-ACE-25 correctly predicts that the Si(100)-H surface is more stable than the

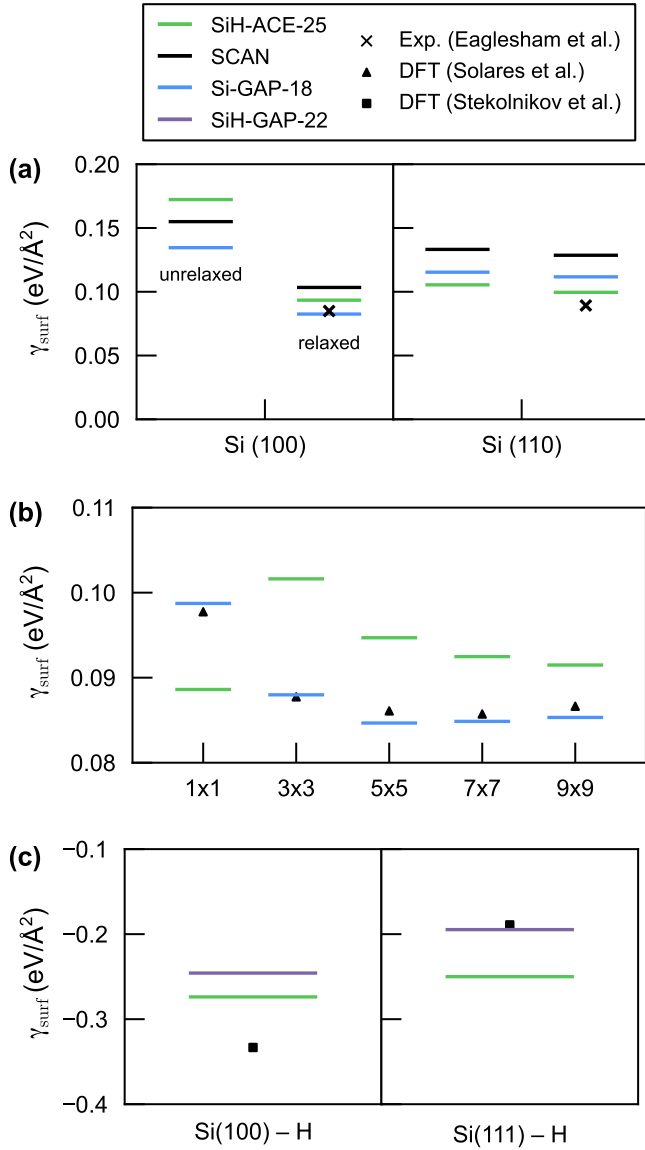


FIG. 7. (a) Surface energy, γ_{surf} , of the Si(100) and Si(110) surfaces in freshly cleaved and relaxed states, respectively, showing the results for DFT (black), SiH-ACE-25 (green), and Si-GAP-18 (blue) alongside experimental results (black crosses) [82]. (b) Surface energies for the ideal and reconstructed Si(111) surfaces, for SiH-ACE-25 (green), Si-GAP-18 (blue), and DFT (black triangles) [83]. These Si configurations are unseen for SiH-ACE-25 but seen for Si-GAP-18. (c) Surface energies of the passivated Si(100)-H and Si(111)-H surfaces after relaxation, for SiH-GAP-22 (purple), SiH-ACE-25 (green), and DFT (black squares) [84].

Si(110)-H surface. It adopts the correct dihydride and monohydride reconstructions for the Si(100)-H and Si(110)-H surfaces, respectively. It also provides satisfactory agreement with the DFT computations from Ref. [84].

E. Amorphous surfaces

Beyond crystalline silicon surfaces, we also assess the model's predictions on amorphous surfaces. We prepare 100 bulk *a*-Si:H structures of varying H content, and cleave them

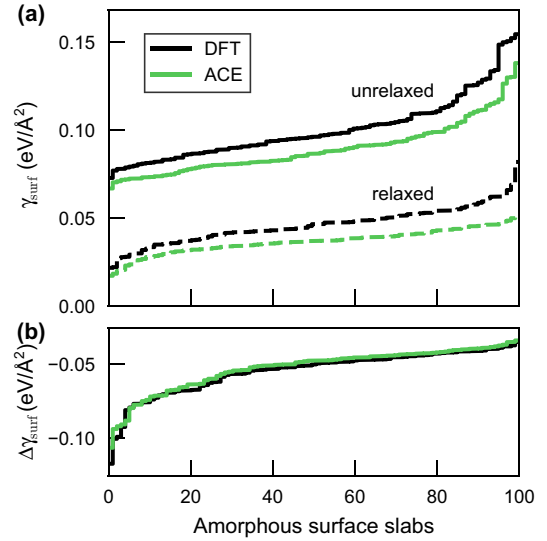


FIG. 8. (a) Surface energy, γ_{surf} , of 100 amorphous surfaces, freshly cleaved and after relaxation with the SiH-ACE-25 potential, comparing DFT and SiH-ACE-25 predictions. (b) Difference in predicted surface energy before and after relaxation for each model.

along the *xy* plane at random distances along the *z* axis. This results in a set of *a*-Si:H structures of varying size and H content, which we relax with the SiH-ACE-25 potential. We show the distributions of surface energies in Fig. 8, alongside the DFT predictions, ordered by stability similar to Ref. [22].

The distribution of both the unrelaxed and relaxed surface energies predicted by our model is in good agreement with its DFT reference. SiH-ACE-25 faithfully reproduces the distribution of energy changes on relaxation, without large outliers. There is a visible offset in the distribution of absolute surface energies between DFT and our model—this offset is constant, and we speculate that it could be due to the DFT and SiH-ACE-25 potential's differing treatment of the vacuum and periodic boundary conditions. Inspecting only the energetic effect of relaxation (that is, the surface-energy differences between unrelaxed and relaxed surfaces for individual slabs) confirms that the DFT and SiH-ACE-25 predictions are very similar in this regard [Fig. 8(b)].

In Fig. 9, we move beyond surface energetics and analyze the structure of a cleaved amorphous surface of 2500 atoms after an MD annealing treatment. The unrelaxed surface was obtained by cleaving a quenched bulk amorphous structure with 20 at.-% H along the *xy* plane. The cleaved structure was then annealed at 500 K for 10 ps in the *NVT* ensemble, with a timestep of 1 fs. We show structural fragments averaged over ten independent repeats of this protocol.

As is visible from the renders in Fig. 9(a), hydrogen atoms migrate to the surface during the annealing to passivate the dangling Si bonds. The slab also expands along the axis perpendicular to the cut plane, likely to relieve some of the internal stress caused by the surface cleavage.

Figure 9(b) exclusively considers the bonding environments of silicon atoms that are not bonded to hydrogen. It shows that the freshly cleaved surface contains a large pro-

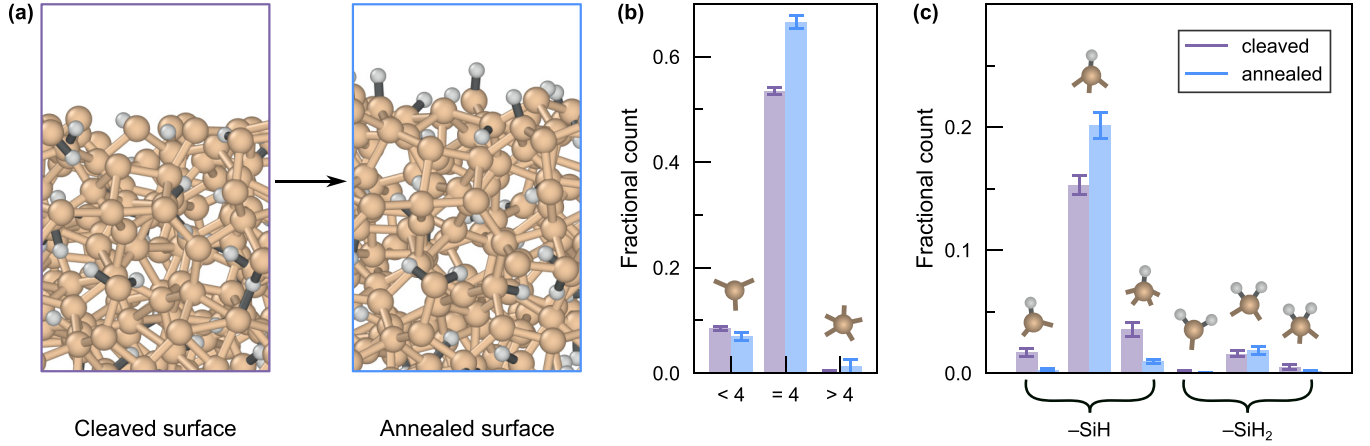


FIG. 9. (a) Visualizations of the freshly cleaved a -Si:H surface (left) and surface after the annealing treatment (right) with 20 at.-% H. (b) Histogram of the silicon coordination before and after annealing for different Si environments that are not bonded to any H atoms. (c) Histogram of the silicon coordination before and after annealing for different Si environments bonded to H atoms.

portion of undercoordinated silicon atoms, or dangling bonds, and far fewer overcoordinated silicon atoms relative to typical well-relaxed amorphous silicon films [18]. Upon annealing, the defect count decreases and the count of ideal, fourfold coordinated, silicon atoms increases to over 65% of the silicon environments.

Figure 9(c) provides a complementary understanding of the local structure after annealing. Within the monohydride fragments, the under- and overcoordinated conformations disappear upon annealing in favor of the typical environment of a silicon atom bonded to three other silicon atoms and a hydrogen atom. The other common fragment is the dihydride fragment, which anneals to a fourfold bonding environment where the silicon atom is bonded to two silicon atoms and two hydrogen atoms.

The low-temperature annealing treatment has lowered the surface energy by 0.072 ± 0.003 eV/Å², which is in line with the results from Fig. 8.

Concluding the validation thus far, we have shown that the SiH-ACE-25 potential provides good agreement with its DFT reference on amorphous surfaces, and that it can be used to simulate large-scale structural models of surfaces that are inaccessible to direct DFT simulations.

F. Molecules

We assess the predictions of the SiH-ACE-25 model on radicals and molecular species in Table IV, comparing to our DFT reference. The potential provides excellent agreement with SCAN on optimal bond length and bond angles of the SiH₄ molecule, as well as good agreement on these structural metrics for SiH₂, SiH₃, and Si₂H₆. This is expected as these configurations are covered by the training dataset. For comparison, SiH-GAP-22 that was not trained on molecules underpredicts the bond lengths across both molecules and radicals, with a relaxed bond length of 1.432 Å for SiH₄.

IV. APPLICATION TO A-SI:H

We now illustrate a practical use case for our potential by investigating the fundamental relationship between structure

and mechanical properties of a -Si:H, a topic that has gathered both experimental and computational interest [96–99]. Indeed, as a solar cell material, a -Si:H can face large mechanical loads under weathering conditions, and hence an understanding of the relationship between structure and mechanical properties could help to improve device lifetimes [100]. We note that no structures from explicit tensile deformation simulations were included in the training dataset, but that the amorphous samples in the dataset span a large configurational space as they were generated to cover a range of densities from 1.8 to 2.8 g cm⁻³, resulting in external pressures from about -4 to +6 GPa.

We prepare a set of five a -Si configurations of 10 000 atoms by simulated melt-quenching, using a quench rate of 10¹¹ K/s to obtain high-quality structural models: simulations with this quench rate were previously shown to provide excellent agreement with the structure factor and excess free energy from experiments [14–16,19]. We then hydrogenate these structures with increasing hydrogen content from 0 to 50 at.-% H, using the hydrogenation script developed for our dataset generation protocol. We anneal the hydrogenated structures for 10 ps at 500 K in the NpT ensemble. We then incrementally apply tensile strain along the z axis and perform fixed-volume geometry optimizations, uniaxially deforming the system to a total strain of 0.3 with a strain step of 0.0001. We present the results of our simulated tensile test in Fig. 10, where we

TABLE IV. Bond-length (d) and bond-angle ($\theta_{\text{H-Si-H}}$) predictions by SCAN and SiH-ACE-25 on molecules and radicals.

		d (Å)		$\theta_{\text{H-Si-H}}$ (°)	
		SCAN	SiH-ACE-25	SCAN	SiH-ACE-25
SiH ₂		1.520	1.505	90.8	95.3
SiH ₃		1.481	1.489	111.8	107.3
SiH ₄		1.479	1.481	109.5	109.5
Si ₂ H ₆	Si-Si	2.329	2.339		
	Si-H	1.482	1.480	108.5	109.1

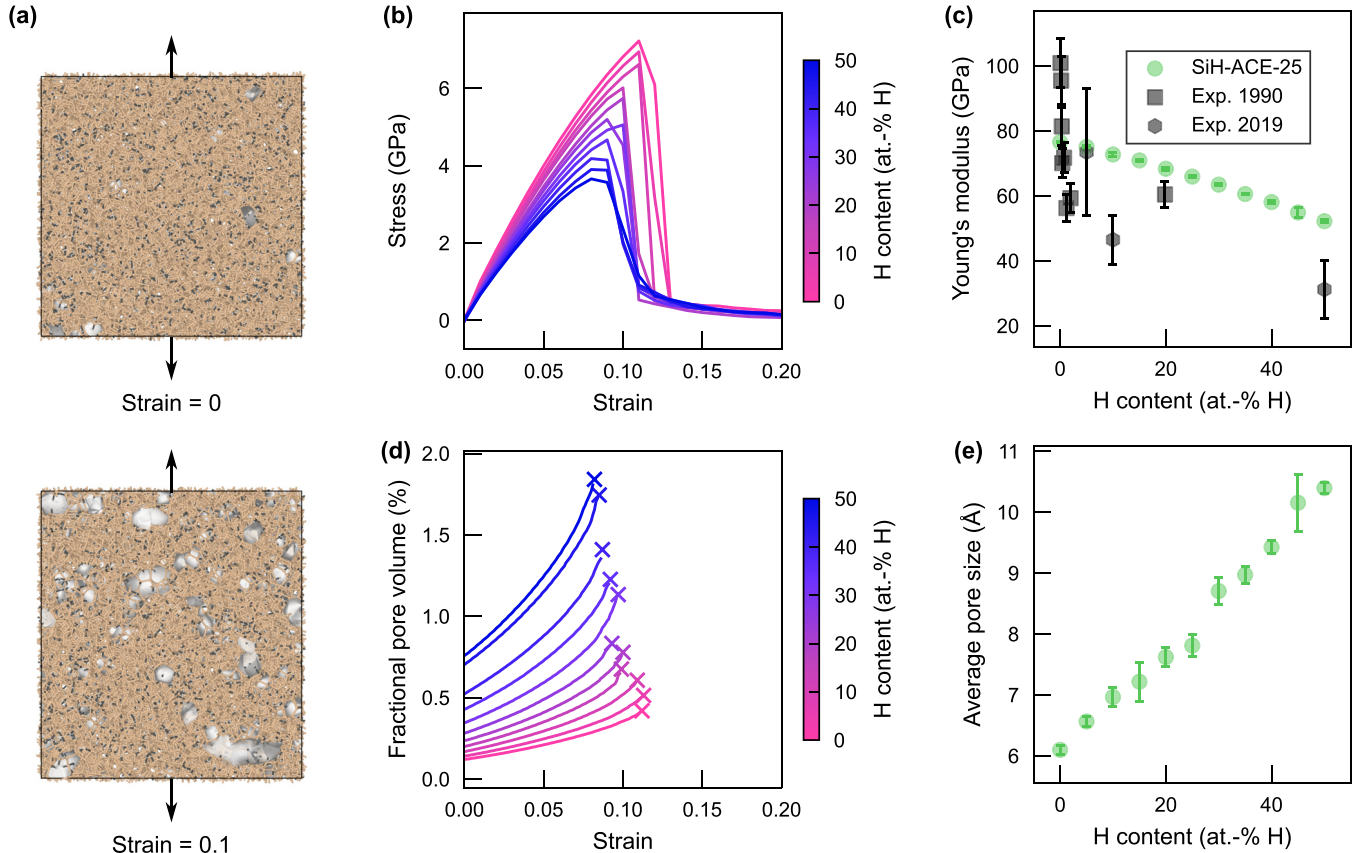


FIG. 10. Mechanical and structural properties of a -Si:H under uniaxial deformation as a function of hydrogen content. (a) Structure visualization of a 30 at.% H a -Si:H structure prior to deformation (top) and at a deformation of 10% strain (bottom), where gray regions show pores using the OVITO alpha-shape mesh modifier [45,70]. (b) Stress-strain profiles for a -Si:H models of increasing H content, averaged over five repeats at each concentration. (c) Average Young's modulus as a function of H content in the a -Si:H network, with error bars showing the standard deviation of each average measurement (green), alongside experimental data from Ref. [97] (hexagons) and Ref. [101] (squares). (d) Evolution of the average pore volume percentage against strain up to fracture, where fracture is represented as a cross. (e) Average pore size of the deformed material prior to fracture, with error bars showing the standard deviation over five repeats.

study the effect of H content on the mechanical and structural properties of a -Si:H.

Figure 10(a) depicts the uniaxial deformation protocol, showing an a -Si:H cell with 30 at.% H prior to and after straining to a strain of 0.1, where gray regions highlight the pores in the Si network.

In Figs. 10(b) and 10(c), we analyze the mechanical properties of our structures, plotting the average stress-strain curve and the Young's modulus over all five repeats. Introducing a small amount of hydrogen into the network immediately decreases the strength of the film, and increasing H content causes the strength to drop further. The deformation behavior at low H content shows a longer elastic than plastic deformation window, which agrees with experimental reports on a -Si:H samples with 5 at.% H [97]. The overall shape of the stress-strain curve is unaltered by H content, but with increasing hydrogenation, the film can accommodate less elastic deformation and thus its yield stress decreases considerably.

We estimate the Young's modulus, E , of each structure from the gradient of the stress-strain profile in the elastic deformation window, as reported in Fig. 10(c). We observe a monotonically decreasing trend in elastic strength as the

hydrogen content increases, which reproduces other computational results [96,97].

Mechanical properties of a -Si:H films vary greatly with preparation methods, as shown by the experimental data plotted in Fig. 10(c). The experimental samples in Ref. [97] were made by colloidal synthesis while the samples in Ref. [101] were prepared by radio frequency sputtering. Even for identical preparation conditions, mechanical properties are computed without accounting for closed porosity which varies from one sample to the next, artificially increasing the sample cross section and thus leading to large variances across sample repeats as reflected by the standard deviations of experimental measurements, shown as error bars in Fig. 10(c). As such, our results demonstrate very good agreement with the experimental ranges.

We assess the evolution of the pore-volume fraction as a function of strain in Fig. 10(d), plotting the fractional volume up to the point of failure. The pore volume shows constant growth during elastic deformation as the pores are slowly stretched, then a sharp increase during plastic deformation as bonds are broken and new voids appear.

Finally, in Fig. 10(e), we show the average pore size before rupture as a function of hydrogen content. Experimental find-

ings report a range of pore sizes and shapes, between 1 and 10 Å for Ref. [102], and between 10 to 20 Å for Ref. [103]. These wide fluctuations can be attributed to the contributions of both the volume and surface pores, the former usually smaller than the latter, and their relative proportions in the prepared sample [104].

Combining the mechanical and structural insights from Fig. 10, our simulations provide a clear nanoscopic picture of the effect of the hydrogen content in our films: increasing hydrogen content replaces Si-Si bonds with Si-H bonds, disrupting the silicon network and creating voids. Under deformation, a network with high Si-Si connectivity will require higher stress to rupture than a network with low Si-Si connectivity. At low H content, much more energy is required to break bonds and create pores, which will eventually connect and lead to failure. At high H content, pores are already present in the network, and must simply connect to cause failure, so the films yield at lower stress. Increasing H content leads to a porous material, weaker in both the elastic and the plastic deformation regimes. This interpretation is consistent with previous reports of hydrogen weakening the silicon tetrahedral network [96] and of hydrogen clusters appearing in the silicon matrix [97].

Beyond the pores and bonding connectivity, the homogeneity of the distribution of hydrogen within the Si network can also heavily affect the performance of the material under deformation and a heterogeneous distribution can lead to disparate deformation behavior [96]. A proposed model to understand this inhomogeneity is the formation of H-rich and H-poor domains, with corresponding low- and high-density amorphous networks, as described in Ref. [105]. This could explain why the mechanical strength of chemical-vapor-deposited *a*-Si:H films has been reported to peak around 10 at.-% H, which is also where the films were found to be most dense [98].

V. CONCLUSION

We have presented a machine-learned interatomic potential model for the Si-H system, using the nonlinear atomic cluster expansion (ACE) framework to reach inference speeds of ≈ 1 ns/day for a system of 1000 atoms and a timestep of 0.1 fs on an NVIDIA A100 GPU card. We validated this potential, referred to as SiH-ACE-25, using comprehensive and challenging tests spanning the elemental Si and binary Si-H configurational spaces. We showed that SiH-ACE-25 can be used to create a range of high-quality *a*-Si:H structures of varying hydrogen content, with porosity and coordination environments that coincide with experimental reports. We further assessed the predictions of our potential for bulk crystalline and liquid configurations, as well as crystalline and amorphous surfaces, and made comparisons to experimental data where relevant. We illustrated its use in elucidating structure-property relationships by probing the processes at play during mechanical testing of *a*-Si:H structures.

We expect that the new potential will be helpful for a range of modeling tasks, with particular focus on the *a*-Si:H system. One such task is the investigation of the diffusion kinetics of hydrogen in the amorphous matrix, where many open questions still prevail, like the nature of “floating bond centers”

[106], the possible trapping of H atoms in deep-state sites [107], or the effusion behavior of H atoms through surfaces [108]. The potential’s computational efficiency promises to unlock simulations at device-sizes and time scales that have never been explored for *a*-Si:H. Aside from the amorphous state, the Si-H liquid phase is not well understood, with very little available literature—in future work, SiH-ACE-25 could be used to simulate high hydrogen content in the silicon melt, and to study hydrogen dynamics and diffusion in liquid silicon.

Beyond the potential itself and its validation, the development of a large, high-quality dataset of structures across the Si-H configurational space also constitutes a key outcome of the present work. With the rise of “universal” [109,110] or “foundational” [111] MLIP models that are trained across a broad range of chemically diverse structures, our dataset could be repurposed (and, if required, relabeled) to create numerical benchmarks to assess the performance of such models specifically on amorphous structures.

ACKNOWLEDGMENTS

We thank D. A. Drabold, S. R. Elliott, and D. F. Thomas du Toit for useful discussions, and Y. Liu for access to his pore analysis code. This work was supported by UK Research and Innovation [Grant No. EP/X016188/1]. We are grateful for computational support from the UK national high performance computing service, ARCHER2, for which access was obtained via the UKCP consortium and funded by EPSRC Grant Ref. EP/X035891/1 (see also Ref. [112]). We are also grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/T022213/1, EP/W032260/1, and EP/P020194/1).

DATA AVAILABILITY

The data that support the findings of this article are openly available [113].

APPENDIX A: COMPUTATIONAL DETAILS

Reference energies and forces were calculated in VASP [114,115], using the projector augmented-wave formalism [116,117]. The SCAN exchange-correlation functional [118] was chosen as it has been extensively validated across the configurational space of silicon covering liquid, crystalline, and amorphous structures [16,81,119,120]. A plane-wave energy cutoff of 1200 eV and a k -point spacing of $0.23 \text{ \AA}^{-1}$ were chosen such that the calculated energies and forces were converged within 1 meV/atom and 10 meV/Å, respectively.

ACE potential fits were carried out using *pacemaker* [24]. The potential was trained on energies and forces, with a loss function of the form: $\mathcal{L} = (1 - \kappa)\Delta_E + \kappa\Delta_F + \Delta_{\text{coeff}} + \Delta_{\text{rad}}$, as described in Ref. [25], where $\kappa = 0.3315$, Δ_{coeff} is the regularization term, parametrized with $L_1 = 10^{-7}$, $L_2 = 10^{-8}$, and Δ_{rad} controls the smoothness of the radial basis set, with $w_1 = 10^{-7}$, $w_2 = 10^{-8}$, and $w_3 = 10^{-8}$. Hyperparameter optimization was performed using the Bayesian optimization routines implemented in *scikit-optimize* [121], through the *XPOT* interface [122]. Hyperparameter optimization

TABLE V. Phases targeted by each AL iteration, and corresponding number of structures added for each.

Iteration	Explored phase	Structures added
1	<i>a</i> -Si:H	106
	<i>c</i> -Si:H	96
	<i>l</i> -Si:H	44
2	clusters	288
	<i>l</i> -Si:H	240
3	clusters	58
	surface	54
4	high-p	24
	molecules	300
5	<i>l</i> -Si:H	141
6	clusters	36
	<i>l</i> -Si:H	199
Total		1586

informed the choice of radial basis parameters and embedding. The full parametrization of SiH-ACE-25 is available at [113].

We use the SiH-GAP-22 training set as a validation set, relabeling the structures with our chosen DFT parameters. The interface structures of over 400 atoms were too large to label with our DFT parameters, and hence omitted; the reported errors do not include these interfaces.

We perform all MD simulations in LAMMPS [123,124]. The *a*-Si:H structures were generated by simulated melt-quenching, using a variable-rate protocol similar to that described in Ref. [15], which reduces computational cost compared to fixed-rate protocols. The structures were quenched in the *NpT* ensemble from 2000 to 1500 K at a rate of 10^{14} K/s, then from 1500 to 900 K at a rate of 10^{12} K/s, and finally to 500 K at a rate of 10^{13} K/s. The melt-quench protocol was run with an MD timestep of 0.1 fs and repeated ten times at each chosen value of hydrogen content.

For the structural analysis, we used the following cutoff distances, chosen from the tail of the first coordination peak of the RDF of *a*-Si:H: Si-Si=2.85 Å, Si-H=1.8 Å, and H-H=1 Å. For the liquid phase, we used: Si-Si=3.1 Å, Si-H=2 Å, and H-H=1 Å, similarly chosen from the RDF of the melt.

We use neutron scattering lengths of $b_{\text{Si}} = 0.41491 \times 10^{-14}$ m and $b_{\text{H}} = -0.3739 \times 10^{-14}$ m, taken from Ref. [125].

APPENDIX B: ACTIVE-LEARNING PROTOCOL

Table V details the configurational space explored in each round of active learning (AL). A total of 1586 new structures were added to the dataset through AL.

Iteration 1: Amorphous and crystalline structures with H content between 1 and 20 at.-% H were explored with *NpT* annealing runs at $T = 300, 600, 900, 1200$, and 1500 K for $t = 10$ ps, as well as *NVT* heating from 300 to 2000 K with a heating rate of 10^{13} K/s. Liquid configurations with H content between 1 and 20 at.-% H were sampled with *NVT* annealing runs at $T = 2000$ K for $t = 10$ ps, as well as *NVT* quenches from 2000 to 300 K with a quench rate of 10^{13} K/s.

Iteration 2: Liquid configurations with H content between 5 and 15 at.-% H and ρ between 2.2 and 2.8 g/cm³ were

TABLE VI. Force component RMSE values on the SiH-GAP-22-set for three model sizes, comparing a symmetric model parametrization to a custom asymmetric model shape.

Model	Symmetric ΔF_i (eV/Å)			Asymmetric ΔF_i (eV/Å)		
	Si	H	All	Si	H	All
cheap	0.17	0.34	0.19	0.14	0.29	0.16
medium	0.16	0.29	0.17	0.14	0.30	0.16
expensive	0.13	0.27	0.15	0.14	0.28	0.15

sampled with *NpT* annealing runs at $T = 2500$ and 5000 K for $t = 10$ ps, as well as *NVT* quenches from 2000 to 300 K and a quench rate of 10^{13} K/s. Clusters were obtained by random structure search (RSS) and relaxation [53].

Iteration 3: Surface structures (details on p. 2) were equilibrated in the *NpT* ensemble for $t = 10$ ps at $T = 500$ K. Clusters were obtained by RSS and relaxation [53].

Iteration 4: High pressure structures were obtained from compression simulations, where *a*-Si:H structures compression rate of 0.67 GPa/ps at $T = 500$ K. Molecules and radicals were placed at random in a box, then equilibrated in the *NpT* ensemble for 10 ps at 500 K.

Iteration 5: Liquid configurations with H content range between 10 and 50 at.-% H were explored with *NpT* quenches from 2000 to 300 K using a quench rate of 10^{14} K/s and a timestep of 0.1 fs.

Iteration 6: Liquid configurations with H content range between 25 and 50 at.-% H were explored with *NpT* annealing runs at $T = 1500$ K and *NpT* quenches from 2000 to 500 K with a quench rate of 10^{14} K/s and a timestep of 0.1 fs. Clusters were obtained by RSS and relaxation [53].

Simulations were stopped if they encountered a structure with $\gamma > 25$. The AL protocol was terminated when the potential no longer encountered configurations with $\gamma > 5$ when performing MD in relevant phase space, and showed stable dynamics in challenging simulations: *NpT* annealing of liquid silicon at 2000, 5000, and 7000 K for 1 ns, *NpT* annealing of liquid SiH with 50 at.-% H at 2000 K for 1 ns, and *NpT* melt-quench simulations from *l*-SiH to *a*-SiH with 50 at.-% H.

APPENDIX C: MODEL PARAMETRIZATION

To support our choice of a custom “asymmetric” model shape, we fit three models of the same total size as the models presented in Fig. 3(b), but with elemental basis blocks of the same size: that is 575, 900, and 1150 functions for each of the H, Si, SiH, and HSi blocks. Together, these blocks made respective total model sizes of 2300 functions (“cheap”), 3600 functions (“medium”), and 4600 functions (“expensive”). We refer to these models as “symmetric”.

We compare the numerical performance of these models to our main “asymmetric” models in Table VI, evaluating the models on the SiH-GAP-22-set. The models with an asymmetric potential shape perform marginally better than the symmetric ones for the “cheap” and “medium” model sizes. This can be rationalized from the greater relative number of

Si elemental basis functions in the asymmetric models than in the symmetric models: e.g., for the “cheap” models of 2300 total functions, the symmetric model has 575 Si functions whereas the asymmetric model has 1000 Si functions. As can be expected, a larger number of Si elemental functions will result in better Si force RMSE metrics for Si, but also for

the overall force error as there is a majority of Si environments. The symmetric models show a greater improvement in force errors with increasing number of functions than the asymmetric models, and the “expensive” symmetric model performs marginally better than the “expensive” asymmetric model.

-
- [1] R. Birney, J. Steinlechner, Z. Tornasi, S. MacFoy, D. Vine, A. S. Bell, D. Gibson, J. Hough, S. Rowan, P. Sortais, S. Sproules, S. Tait, I. W. Martin, and S. Reid, Amorphous silicon with extremely low absorption: Beating thermal noise in gravitational astronomy, *Phys. Rev. Lett.* **121**, 191101 (2018).
- [2] H. Lin, M. Yang, X. Ru, G. Wang, S. Yin, F. Peng, C. Hong, M. Qu, J. Lu, L. Fang, C. Han, P. Procel, O. Isabella, P. Gao, Z. Li, and X. Xu, Silicon heterojunction solar cells with up to 26.81% efficiency achieved by electrically optimized nanocrystalline-silicon hole contact layers, *Nat. Energy* **8**, 789 (2023).
- [3] B. Fischer, W. Beyer, A. Lambert, M. Nuys, W. Duan, K. Ding, and U. Rau, The microstructure of underdense hydrogenated amorphous silicon and its application to silicon heterojunction solar cells, *Solar RRL* **7**, 2300103 (2023).
- [4] R. A. Street, *Hydrogenated Amorphous Silicon*, Cambridge Solid State Science Series (Cambridge University Press, Cambridge, 1991).
- [5] S. R. Elliott, The structure of amorphous hydrogenated silicon and its alloys: A review, *Adv. Phys.* **38**, 1 (1989).
- [6] F. Wooten, K. Winer, and D. Weaire, Computer generation of structural models of amorphous Si and Ge, *Phys. Rev. Lett.* **54**, 1392 (1985).
- [7] M. Legesse, M. Nolan, and G. Fagas, Revisiting the dependence of the optical and mobility gaps of hydrogenated amorphous silicon on hydrogen concentration, *J. Phys. Chem. C* **117**, 23956 (2013).
- [8] K. Nepal, A. Gautam, C. Ugwumadu, and D. A. Drabold, New models of clean and hydrogenated amorphous silicon surfaces, *J. Non-Cryst. Solids* **660**, 123517 (2025).
- [9] A. Shah, Thin-film silicon solar cells, in *Practical Handbook of Photovoltaics (Second Edition)*, edited by A. McEvoy, T. Markvart, and L. Castañer (Academic Press, Boston, 2012), pp. 209–281.
- [10] D. L. Staebler and C. R. Wronski, Reversible conductivity changes in discharge-produced amorphous Si, *Appl. Phys. Lett.* **31**, 292 (1977).
- [11] Y. Ishikawa and M. B. Schubert, Flexible polycrystalline silicon solar cells with amorphous buffer layer, *Jpn. J. Appl. Phys.* **45**, 6812 (2006).
- [12] M. A. Caro, Machine learning based modeling of disordered elemental semiconductors: understanding the atomic structure of a-Si and a-C, *Semicond. Sci. Technol.* **38**, 043001 (2023).
- [13] 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).
- [14] V. L. Deringer, N. Bernstein, A. P. Bartók, M. J. Cliffe, R. N. Kerber, L. E. Marbella, C. P. Grey, S. R. Elliott, and G. Csányi, Realistic atomistic structure of amorphous silicon from machine-learning-driven molecular dynamics, *J. Phys. Chem. Lett.* **9**, 2879 (2018).
- [15] N. Bernstein, B. Bhattacharai, G. Csányi, D. A. Drabold, S. R. Elliott, and V. L. Deringer, Quantifying chemical structure and machine-learned atomic energies in amorphous and liquid silicon, *Angew. Chem. Int. Ed.* **58**, 7057 (2019).
- [16] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, and S. R. Elliott, Origins of structural and electronic transitions in disordered silicon, *Nature (London)* **589**, 59 (2021).
- [17] J. D. Morrow and V. L. Deringer, Indirect learning and physically guided validation of interatomic potential models, *J. Chem. Phys.* **157**, 104105 (2022).
- [18] J. D. Morrow, C. Ugwumadu, D. A. Drabold, S. R. Elliott, A. L. Goodwin, and V. L. Deringer, Understanding defects in amorphous silicon with million-atom simulations and machine learning, *Angew. Chem. Int. Ed.* **63**, e202403842 (2024).
- [19] L. A. M. Rosset, D. A. Drabold, and V. L. Deringer, Signatures of paracrystallinity in amorphous silicon from machine-learning-driven molecular dynamics, *Nat. Commun.* **16**, 2360 (2025).
- [20] D. Unruh, R. V. Meidanshahi, S. M. Goodnick, G. Csányi, and G. T. Zimányi, Gaussian approximation potential for amorphous Si : H, *Phys. Rev. Mater.* **6**, 065603 (2022).
- [21] A. Diggs, Z. Zhao, R. V. Meidanshahi, D. Unruh, S. Manzoor, M. Bertoni, S. M. Goodnick, and G. T. Zimányi, Hydrogen-induced degradation dynamics in silicon heterojunction solar cells via machine learning, *Commun. Mater.* **4**, 24 (2023).
- [22] L. C. Erhard, J. Rohrer, K. Albe, and V. L. Deringer, Modelling atomic and nanoscale structure in the silicon–oxygen system through active machine learning, *Nat. Commun.* **15**, 1927 (2024).
- [23] R. Drautz, Atomic cluster expansion for accurate and transferable interatomic potentials, *Phys. Rev. B* **99**, 014104 (2019).
- [24] Y. Lysogorskiy, C. v. d. Oord, A. Bochkarev, S. Menon, M. Rinaldi, T. Hammerschmidt, M. Mrovec, A. Thompson, G. Csányi, C. Ortner, *et al.*, Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon, *npj Comput. Mater.* **7**, 97 (2021).
- [25] A. Bochkarev, Y. Lysogorskiy, S. Menon, M. Qamar, M. Mrovec, and R. Drautz, Efficient parametrization of the atomic cluster expansion, *Phys. Rev. Mater.* **6**, 013804 (2022).
- [26] L. C. Erhard, C. Otzen, J. Rohrer, C. Prescher, and K. Albe, Understanding phase transitions of α -quartz under dynamic compression conditions by machine-learning driven atomistic simulations, *npj Comput. Mater.* **11**, 58 (2025).
- [27] Y. Zhou, D. F. Thomas du Toit, S. R. Elliott, W. Zhang, and V. L. Deringer, Full-cycle device-scale simulations of memory

- materials with a tailored atomic-cluster-expansion potential, *Nat. Commun.* **16**, 8688 (2025).
- [28] N. Leimeroth, L. C. Erhard, K. Albe, and J. Rohrer, Machine-learning interatomic potentials from a users perspective: a comparison of accuracy, speed and data efficiency, *Modelling Simul. Mater. Sci. Eng.* **33**, 065012 (2025).
- [29] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, Less is more: Sampling chemical space with active learning, *J. Chem. Phys.* **148**, 241733 (2018).
- [30] M. Kulichenko, K. Barros, N. Lubbers, Y. W. Li, R. Messerly, S. Tretiak, J. S. Smith, and B. Nebgen, Uncertainty-driven dynamics for active learning of interatomic potentials, *Nat. Comput. Sci.* **3**, 230 (2023).
- [31] Y. Lysogorskiy, A. Bochkarev, M. Mrovec, and R. Drautz, Active learning strategies for atomic cluster expansion models, *Phys. Rev. Mater.* **7**, 043801 (2023).
- [32] C. van der Oord, M. Sachs, D. P. Kovács, C. Ortner, and G. Csányi, Hyperactive learning for data-driven interatomic potentials, *npj Comput. Mater.* **9**, 168 (2023).
- [33] V. Zaverkin, D. Holzmüller, H. Christiansen, F. Errica, F. Alesiani, M. Takamoto, M. Niepert, and J. Kästner, Uncertainty-biased molecular dynamics for learning uniformly accurate interatomic potentials, *npj Comput. Mater.* **10**, 83 (2024).
- [34] S. Perego and L. Bonati, Data efficient machine learning potentials for modeling catalytic reactivity via active learning and enhanced sampling, *npj Comput. Mater.* **10**, 291 (2024).
- [35] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, The MLIP package: Moment tensor potentials with MPI and active learning, *Mach. Learn.: Sci. Technol.* **2**, 025002 (2021).
- [36] F. Grasselli, S. Chong, V. Kapil, S. Bonfanti, and K. Rossi, Uncertainty in the era of machine learning for atomistic modeling, *Digit. Discov.* **4**, 2654 (2025).
- [37] S. Acco, D. L. Williamson, P. A. Stolk, F. W. Saris, M. J. van den Boogaard, W. C. Sinke, W. F. van der Weg, S. Roorda, and P. C. Zalm, Hydrogen solubility and network stability in amorphous silicon, *Phys. Rev. B* **53**, 4415 (1996).
- [38] P. Danesh, B. Pantchev, B. Schmidt, and D. Grambole, Hydrogen solubility limit in hydrogenated amorphous silicon, *Semicond. Sci. Technol.* **19**, 1422 (2004).
- [39] A. Shkrebtii, Z. A. Ibrahim, I. Kupchak, T. Teatro, F. Gaspari, and D. Korbutiak, Dynamics and bonding of bond-centered hydrogen in amorphous hydrogenated Si: vibrational and optical signatures, in *2010 35th IEEE Photovoltaic Specialists Conference, Honolulu, HI* (IEEE, Piscataway, NJ, 2010), pp. 003663–003666.
- [40] K. Jarolimek, R. A. de Groot, G. A. de Wijs, and M. Zeman, First-principles study of hydrogenated amorphous silicon, *Phys. Rev. B* **79**, 155206 (2009).
- [41] M. Powell, The physics of amorphous-silicon thin-film transistors, *IEEE Trans. Electron Devices* **36**, 2753 (1989).
- [42] W. Y. Ching, D. J. Lam, and C. C. Lin, Electronic states and bonding configurations in hydrogenated amorphous silicon, *Phys. Rev. B* **21**, 2378 (1980).
- [43] L. McInnes, J. Healy, N. Saul, and L. Großberger, UMAP: Uniform Manifold Approximation and Projection, *J. Open Source Software* **3**, 861 (2018).
- [44] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, *Phys. Rev. B* **87**, 184115 (2013).
- [45] A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO – the Open Visualization Tool, *Model. Simul. Mater. Sci. Eng.* **18**, 015012 (2010).
- [46] E. V. Podryabinkin and A. V. Shapeev, Active learning of linearly parametrized interatomic potentials, *Comput. Mater. Sci.* **140**, 171 (2017).
- [47] Y. Eldar, M. Lindenbaum, M. Porat, and Y. Zeevi, The farthest point strategy for progressive image sampling, *IEEE Trans. on Image Process.* **6**, 1305 (1997).
- [48] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Comparing molecules and solids across structural and alchemical space, *Phys. Chem. Chem. Phys.* **18**, 13754 (2016).
- [49] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Machine learning unifies the modeling of materials and molecules, *Sci. Adv.* **3**, e1701816 (2017).
- [50] P. Rowe, V. L. Deringer, P. Gasparotto, G. Csányi, and A. Michaelides, An accurate and transferable machine learning potential for carbon, *J. Chem. Phys.* **153**, 034702 (2020).
- [51] T. Rocke and J. R. Kermode, Bayesian selection for efficient MLIP dataset selection, *Model. Simul. Mater. Sci. Eng.* **33**, 055020 (2025).
- [52] C. J. Pickard, and R. J. Needs, High-pressure phases of silane, *Phys. Rev. Lett.* **97**, 045504 (2006).
- [53] C. J. Pickard and R. J. Needs, *Ab initio* random structure searching, *J. Phys.: Condens. Matter* **23**, 053201 (2011).
- [54] T. C. Nicholas, D. F. T. d. Toit, L. A. M. Rosset, D. M. Proserpio, A. L. Goodwin, and V. L. Deringer, The structure and topology of an amorphous metal-organic framework, [arXiv:2503.24367](https://arxiv.org/abs/2503.24367) [cond-mat].
- [55] D. Abella, G. Franzese, and J. Hernández-Rojas, Many-body contributions in water nanoclusters, *ACS Nano* **17**, 1959 (2023).
- [56] S. Chong, T. Jiang, M. Domina, F. Bigi, F. Grasselli, J. Lee, and M. Ceriotti, Resolving the body-order paradox of machine learning interatomic potentials, *J. Chem. Phys.* **164**, 064121 (2026).
- [57] C. G. Van de Walle, Energies of various configurations of hydrogen in silicon, *Phys. Rev. B* **49**, 4579 (1994).
- [58] C. R. Hubbard, H. E. Swanson, and F. A. Mauer, A silicon powder diffraction standard reference material, *J. Appl. Crystallogr.* **8**, 45 (1975).
- [59] H. J. McSkimin, Measurement of elastic constants at low temperatures by means of ultrasonic waves—Data for silicon and germanium single crystals, and for fused silica, *J. Appl. Phys.* **24**, 988 (1953).
- [60] M. I. J. Probert and M. C. Payne, Improving the convergence of defect calculations in supercells: An *ab initio* study of the neutral silicon vacancy, *Phys. Rev. B* **67**, 075204 (2003).
- [61] A. J. Morris, C. J. Pickard, and R. J. Needs, Hydrogen/silicon complexes in silicon from computational searches, *Phys. Rev. B* **78**, 184102 (2008).
- [62] C. G. Van de Walle and J. P. Goss, Energetics and vibrational frequencies of interstitial H₂ molecules in semiconductors, *Mater. Sci. Eng., B* **58**, 17 (1999).
- [63] M. Barhoumi, D. Rocca, M. Said, and S. Lebègue, Elastic and mechanical properties of cubic diamond and silicon using

- density functional theory and the random phase approximation, *Solid State Commun.* **324**, 114136 (2021).
- [64] F. Corsetti and A. A. Mostofi, System-size convergence of point defect properties: The case of the silicon vacancy, *Phys. Rev. B* **84**, 035209 (2011).
- [65] G. G. DeLeo, M. J. Dorogi, and W. B. Fowler, Bond-centered interstitial hydrogen in silicon: Semiempirical electronic-structure calculations, *Phys. Rev. B* **38**, 7520 (1988).
- [66] C. G. Van de Walle, P. J. H. Denteneer, Y. Bar-Yam, and S. T. Pantelides, Theory of hydrogen diffusion and reactions in crystalline silicon, *Phys. Rev. B* **39**, 10791 (1989).
- [67] <https://github.com/lamrosset/SiH-ACE-25>.
- [68] R. Bellisent, A. Menelle, W. S. Howells, A. C. Wright, T. M. Brunier, R. N. Sinclair, and F. Jansen, The structure of amorphous Si:H using steady state and pulsed neutron sources, *Physica B* **156–157**, 217 (1989).
- [69] A. C. Wright, A. C. Hannon, R. N. Sinclair, T. M. Brunier, C. A. Guy, R. J. Stewart, M. B. Strobel, and F. Jansen, Neutron scattering studies of hydrogenated, deuterated and fluorinated amorphous silicon, *J. Phys.: Condens. Matter* **19**, 415109 (2007).
- [70] A. Stukowski, Computational analysis methods in atomistic modeling of crystals, *JOM* **66**, 399 (2014).
- [71] S. C. Moss and J. F. Graczyk, Evidence of voids within the as-deposited structure of glassy silicon, *Phys. Rev. Lett.* **23**, 1167 (1969).
- [72] L. J. Lewis, Fifty years of amorphous silicon models: The end of the story? *J. Non-Cryst. Solids* **580**, 121383 (2022).
- [73] Y. Liu, Y. Zhou, R. Ademuwagun, L. Walterbos, J. George, S. R. Elliott, and V. L. Deringer, Medium-range structural order in amorphous arsenic, *J. Am. Chem. Soc.* **148**, 9400 (2026).
- [74] A. H. Mahan, Y. Xu, D. L. Williamson, W. Beyer, J. D. Perkins, M. Vanecek, L. M. Gedvilas, and B. P. Nelson, Structural properties of hot wire *a*-Si:H films deposited at rates in excess of 100 Å/s, *J. Appl. Phys.* **90**, 5038 (2001).
- [75] J. Melskens, S. W. H. Eijt, M. Schouten, A. S. Vullers, A. Mannheim, H. Schut, B. Macco, M. Zeman, and A. H. M. Smets, Migration of open volume deficiencies in hydrogenated amorphous silicon during annealing, *IEEE J. Photovoltaics* **7**, 421 (2017).
- [76] B. von Roedern, L. Ley, and M. Cardona, Photoelectron spectra of hydrogenated amorphous silicon, *Phys. Rev. Lett.* **39**, 1576 (1977).
- [77] J. E. Graebner, B. Golding, L. C. Allen, D. K. Biegelsen, and M. Stutzmann, Solid hydrogen in hydrogenated amorphous silicon, *Phys. Rev. Lett.* **52**, 553 (1984).
- [78] W. E. Carlos and P. C. Taylor, Molecular hydrogen in *a*-Si: H, *Phys. Rev. B* **25**, 1435 (1982).
- [79] H. v. Löhneysen, H. J. Schink, and W. Beyer, Direct experimental evidence for molecular hydrogen in amorphous Si: H, *Phys. Rev. Lett.* **52**, 549 (1984).
- [80] J. B. Boyce and M. Stutzmann, Orientational ordering and melting of molecular H₂ in an *a*-Si matrix: NMR studies, *Phys. Rev. Lett.* **54**, 562 (1985).
- [81] M. W. C. Dharma-wardana, D. D. Klug, and R. C. Remsing, Liquid-liquid phase transitions in silicon, *Phys. Rev. Lett.* **125**, 075702 (2020).
- [82] D. J. Eaglesham, A. E. White, L. C. Feldman, N. Moriya, and D. C. Jacobson, Equilibrium shape of Si, *Phys. Rev. Lett.* **70**, 1643 (1993).
- [83] S. D. Solares, S. Dasgupta, P. A. Schultz, Y.-H. Kim, C. B. Musgrave, and W. A. Goddard, Density functional theory study of the geometry, energetics, and reconstruction process of Si(111) surfaces, *Langmuir* **21**, 12404 (2005).
- [84] A. A. Stekolnikov, J. Furthmüller, and F. Bechstedt, Absolute surface energies of group-IV semiconductors: Dependence on orientation and reconstruction, *Phys. Rev. B* **65**, 115318 (2002).
- [85] A. Ramstad, G. Brocks, and P. J. Kelly, Theoretical study of the Si(100) surface reconstruction, *Phys. Rev. B* **51**, 14504 (1995).
- [86] H. Neergaard Waltenburg and J. Yates, Surface chemistry of silicon, *Chem. Rev.* **95**, 1589 (1995).
- [87] D. Vanderbilt, Model for the energetics of Si and Ge (111) surfaces, *Phys. Rev. B* **36**, 6209 (1987).
- [88] R. Zhachuk, S. Teys, and J. Coutinho, Strain-induced structure transformations on Si(111) and Ge(111) surfaces: A combined density-functional and scanning tunneling microscopy study, *J. Chem. Phys.* **138**, 224702 (2013).
- [89] G.-H. Lu, M. Huang, M. Cuma, and F. Liu, Relative stability of Si surfaces: A first-principles study, *Surf. Sci.* **588**, 61 (2005).
- [90] D. Vanderbilt, Absence of large compressive stress on Si(111), *Phys. Rev. Lett.* **59**, 1456 (1987).
- [91] L. Hu, B. Huang, and F. Liu, Atomistic mechanism underlying the Si(111)–(7×7) surface reconstruction revealed by artificial neural network potential, *Phys. Rev. Lett.* **126**, 176101 (2021).
- [92] E. K. K. Abavare, J.-I. Iwata, A. Yaya, and A. Oshiyama, Surface energy of Si(110)- and 3C-SiC(111)-terminated surfaces, *Phys. Status Solidi B* **251**, 1408 (2014).
- [93] M. Dürr and U. Höfer, Hydrogen diffusion on silicon surfaces, *Prog. Surf. Sci.* **88**, 61 (2013).
- [94] M. Dürr and U. Höfer, Dissociative adsorption of molecular hydrogen on silicon surfaces, *Surf. Sci. Rep.* **61**, 465 (2006).
- [95] A. A. Stekolnikov, J. Furthmüller, and F. Bechstedt, Adatoms, dimers, and interstitials on group-IV(113) surfaces: First-principles studies of energetical, structural, and electronic properties, *Phys. Rev. B* **67**, 195332 (2003).
- [96] R. J. Bondi, S. Lee, and G. S. Hwang, First-principles study of the mechanical and optical properties of amorphous hydrogenated silicon and silicon-rich silicon oxide, *Phys. Rev. B* **81**, 195207 (2010).
- [97] T. Jiang, F. Khabaz, A. Marne, C. Wu, R. Gearba, R. Bodepudi, R. T. Bonnacaze, K. M. Liechti, and B. A. Korgel, Mechanical properties of hydrogenated amorphous silicon (*a*-Si:H) particles, *J. Appl. Phys.* **126**, 204303 (2019).
- [98] R. Kuschnerreit, H. Fath, A. A. Kolomenskii, M. Szabadi, and P. Hess, Mechanical and elastic properties of amorphous hydrogenated silicon films studied by broadband surface acoustic wave spectroscopy, *Appl. Phys. A* **61**, 269 (1995).
- [99] E. Johlin, N. Tabet, S. Castro-Galnares, A. Abdallah, M. I. Bertoni, T. Asafa, J. C. Grossman, S. Said, and T. Buonassisi, Structural origins of intrinsic stress in amorphous silicon thin films, *Phys. Rev. B* **85**, 075202 (2012).
- [100] L. Papargyri, P. Papanastasiou, and G. E. Georgiou, Effect of materials and design on PV cracking under mechanical loading, *Renew. Energy* **199**, 433 (2022).

- [101] X. Jiang, B. Goranchev, K. Schmidt, P. Grünberg, and K. Reichelt, Mechanical properties of *a*-Si:H films studied by Brillouin scattering and nanoindenter, *J. Appl. Phys.* **67**, 6772 (1990).
- [102] D. L. Young, P. Stradins, Y. Xu, L. M. Gedvilas, E. Iwaniczko, Y. Yan, H. M. Branz, Q. Wang., and D. L. Williamson, Nanostructure evolution in hydrogenated amorphous silicon during hydrogen effusion and crystallization, *Appl. Phys. Lett.* **90**, 081923 (2007).
- [103] Y. J. Chabal and C. K. N. Patel, Molecular hydrogen in *a*-Si:H, *Rev. Mod. Phys.* **59**, 835 (1987).
- [104] V. P. Bork, P. A. Fedders, R. E. Norberg, J. B. Boyce, and M. Stutzmann, Proton T1 for solid H₂ in *a*-Si:H, *J. Non-Cryst. Solids* **77–78**, 711 (1985).
- [105] E. Gericke, J. Melskens, R. Wendt, M. Wollgarten, A. Hoell, and K. Lips, Quantification of nanoscale density fluctuations in hydrogenated amorphous silicon, *Phys. Rev. Lett.* **125**, 185501 (2020).
- [106] T. A. Abtew, F. Inam, and D. A. Drabold, Thermally stimulated H emission and diffusion in hydrogenated amorphous silicon, *Europhysics Letters (EPL)* **79**, 36001 (2007).
- [107] T. A. Abtew and D. A. Drabold, Hydrogen dynamics and light-induced structural changes in hydrogenated amorphous silicon, *Phys. Rev. B* **74**, 085201 (2006).
- [108] W. Beyer and U. Zastrow, Solubility and diffusion of hydrogen in hydrogenated crystalline and amorphous silicon, *J. Non-Cryst. Solids* **227–230**, 880 (1998).
- [109] C. Chen and S. P. Ong, A universal graph deep learning interatomic potential for the periodic table, *Nat. Comput. Sci.* **2**, 718 (2022).
- [110] B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel, and G. Ceder, CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling, *Nat. Mach. Intell.* **5**, 1031 (2023).
- [111] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, F. Bigi, S. M. Blau, V. Cărare, M. Ceriotti, S. Chong, J. P. Darby, S. De, *et al.*, A foundation model for atomistic materials chemistry, *J. Chem. Phys.* **163**, 184110 (2025).
- [112] G. Beckett, J. Beech-Brandt, K. Leach, Z. Payne, A. Simpson, L. Smith, A. Turner, and A. Whiting, ARCHER2 Service Description, <https://doi.org/10.5281/zenodo.14507040> (2024).
- [113] L. A. M. Rosset and V. L. Deringer, Research data for “Atomic cluster expansion potential for the Si-H system”, <https://doi.org/10.5281/zenodo.18857006> (2026).
- [114] G. Kresse and J. Hafner, *Ab initio* molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium, *Phys. Rev. B* **49**, 14251 (1994).
- [115] G. Kresse and J. Furthmüller, Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set, *Comput. Mater. Sci.* **6**, 15 (1996).
- [116] P. E. Blöchl, Projector augmented-wave method, *Phys. Rev. B* **50**, 17953 (1994).
- [117] 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).
- [118] J. Sun, A. Ruzsinszky, and J. P. Perdew, Strongly constrained and appropriately normed semilocal density functional, *Phys. Rev. Lett.* **115**, 036402 (2015).
- [119] J. Sun, R. C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu, M. L. Klein, and J. P. Perdew, Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional, *Nat. Chem.* **8**, 831 (2016).
- [120] R. C. Remsing, M. L. Klein, and J. Sun, Dependence of the structure and dynamics of liquid silicon on the choice of density functional approximation, *Phys. Rev. B* **96**, 024203 (2017).
- [121] T. Head, M. Kumar, H. Nahrstaedt, G. Louppe, and I. Shcherbatyi, scikit-optimize/scikit-optimize (v0.9.0), Zenodo (2021), <https://doi.org/10.5281/zenodo.5565057>.
- [122] D. F. Thomas du Toit, Y. Zhou, and V. L. Deringer, Hyperparameter Optimization for Atomic Cluster Expansion Potentials, *J. Chem. Theory Comput.* **20**, 10103 (2024).
- [123] A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott, and S. J. Plimpton, LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, *Comput. Phys. Commun.* **271**, 108171 (2022).
- [124] J. R. Gissinger, I. Nikiforov, Y. Afshar, B. Waters, M.-k. Choi, D. S. Karls, A. Stukowski, W. Im, H. Heinz, A. Kohlmeyer, and E. B. Tadmor, Type label framework for bonded force fields in LAMMPS, *J. Phys. Chem. B* **128**, 3282 (2024).
- [125] V. F. Sears, Neutron scattering lengths and cross sections, *Neutron News* **3**, 26 (1992).