Quantifying Chemical Structure and Machine-Learned Atomic Energies in Amorphous and Liquid Silicon

Noam Bernstein, Bishal Bhattarai, Gábor Csányi, David A. Drabold, Stephen R. Elliott, and Volker L. Deringer*

Abstract: *Amorphous materials are being described by increasingly powerful computer simulations, but new approaches are still needed to fully understand their intricate atomic structures. Here, we show how machine-learning-based techniques can give new, quantitative chemical insight into the atomic-scale structure of amorphous silicon (a-Si). We combine a quantitative description of the nearest- and next-nearest-neighbor structure with a quantitative description of local stability. The analysis is applied to an ensemble of a-Si networks in which we tailor the degree of ordering by varying the quench rates down to $10^{10}\,K\,s^{-1}$. Our approach associates coordination defects in a-Si with distinct stability regions and it has also been applied to liquid Si, where it traces a clear-cut transition in local energies during vitrification. The method is straightforward and inexpensive to apply, and therefore expected to have more general significance for developing a quantitative understanding of liquid and amorphous states of matter.*

The structure of amorphous silicon (*a*-Si) is widely approximated as a continuous random network with tetrahedral coordination,[1] but its details are much more intricate: defective environments, such as threefold-bonded "dangling bonds", as well as the degree of medium-range order, have been discussed.[2] Together with experimental probes,[3] atomistic computer simulations have been giving useful insight into *a*-Si for decades,[4] and large-scale simulation models now contain up to hundreds of thousands of atoms.[5] With the recent emergence of linear-scaling machine-learning(ML)-based interatomic potentials reaching accuracy levels close to quantum mechanics,[6] materials modeling is promising to become even more realistic—especially in describing amorphous solids,[7] as recently shown for *a*-Si.[8]

Still, there remains the more fundamental challenge of not only to describe amorphous structures but to truly understand them. Simple criteria are widely used, including atomic coordination numbers (here denoted as N) and bond angles, which both give information about short-range order (SRO),[9] or ring statistics as a representative for medium-range order (MRO).[10] However, we do not know of a previous simple and general numerical approach that can quantify SRO and MRO at the same time. And even more critically, these purely structurally-based indicators cannot give information about the energetic stability of individual environments.

Here, we describe a general, ML-based approach that quantifies local structures and local energies of all individual atoms in models of *a*-Si. We first introduce a structural coordinate that unifies the description of SRO and MRO environments and then combine this structural information with a second, stability coordinate in a two-dimensional plot. Both analyses rely on the "learning" of local structure, manifested in a mathematically well-defined framework without parametric terms. The ability to "machine-learn" local chemical knowledge is an emerging research theme throughout the discipline: ML-predicted atomic energies have been used to understand the stability and chemical nature of molecules[11] and crystal structures,[12] and to accelerate structural optimization.[13] Here, we transfer such analyses to the amorphous and liquid states, where there is an even more dire need for information about atomically resolved stabilities and properties.

Our object of study is an ensemble of *a*-Si networks that we created in parallel ML-driven molecular-dynamics (MD) simulations: 512-atom models of liquid Si were cooled to solidify into *a*-Si (Figure 1a).[8] Slower cooling yields more ordered networks;[8] hence, changing the cooling rate allows us to tailor the degree of order in the structures and to probe its influence on the properties. Remarkably, the most ordered structures we obtained (for quench rates of 10^{11} and $10^{10}\,K\,s^{-1}$), albeit still containing $\approx 1\%$ defects, are energetically more favorable by 0.02 eV/at. (at the DFT-PBE level) than a fully tetrahedral-like relaxed Wooten–Winer–Weaire (WWW) model,[1] which is currently considered a gold-standard model for *a*-Si (see Supporting Information).

[*] Dr. N. Bernstein
Center for Materials Physics and Technology
U.S. Naval Research Laboratory
Washington, DC 20375 (USA)

Dr. B. Bhattarai, Prof. D. A. Drabold
Department of Physics and Astronomy, Ohio University
Athens, OH 45701 (USA)

Prof. G. Csányi, Dr. V. L. Deringer
Department of Engineering, University of Cambridge
Cambridge CB2 1PZ (UK)
E-mail: vld24@cam.ac.uk

Prof. S. R. Elliott, Dr. V. L. Deringer
Department of Chemistry, University of Cambridge
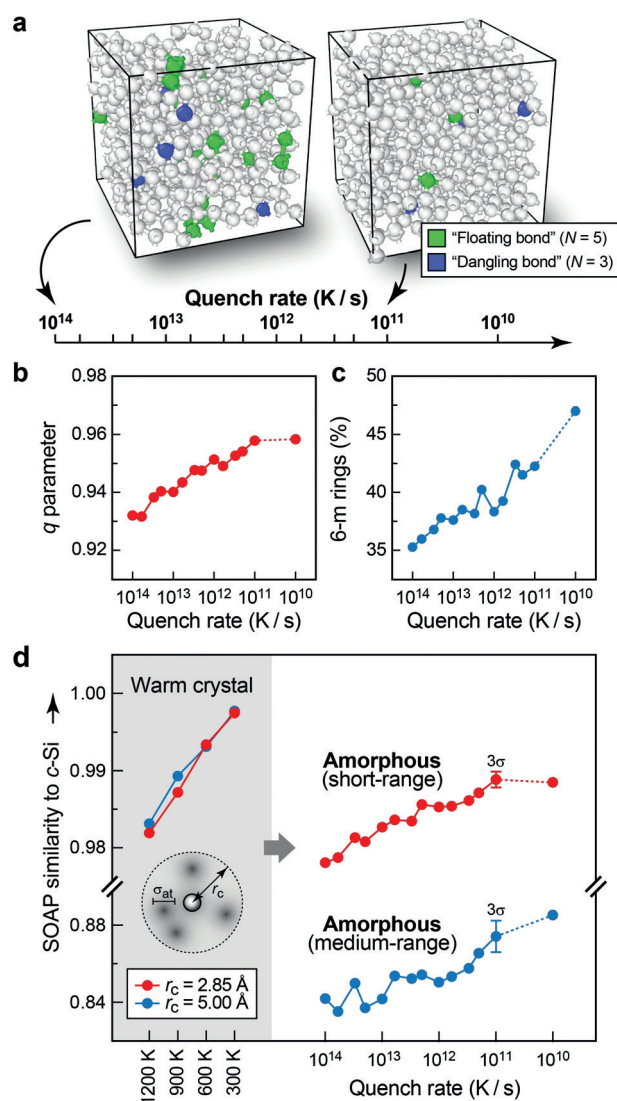Cambridge CB2 1EW (UK)

Figure 1. Progressively ordered *a*-Si networks from melt–quench simulations with an ML-based interatomic potential of quantum-mechanical quality. a) Scale of cooling rates and associated required simulation times (1 ps requires 1000 MD time steps). Each tick corresponds to one independent MD simulation. Between 10^{14} and 10^{11} Ks^{-1}, we cooled at the respective constant rate; for the much more demanding 10^{10} Ks^{-1} simulation, we varied the rate during the run (see Supporting Information). Two simulation cells are shown as examples and coordination defects are highlighted by coloring (green: over-coordinated "floating-bond" environments; blue: under-coordinated "dangling-bond" environments). b) Increasing short-range order (SRO) in these systems, quantified using an established order parameter that returns unity for ideal tetrahedral environments.[9] c) Increasing medium-range order (MRO), assessed by counting 6-membered rings.[10] d) Unified description of both length scales using SOAP analysis. We first calibrated the SOAP kernel parameters (Table 1) for NNs (red) and NNNs (blue) using samples of thermalized *c*-Si and then applied the method to our *a*-Si networks. Median values over all atoms in the cells are given for each system. Error bars are shown for the SOAP values at 10^{11} Ks^{-1} to estimate the scattering of the results; they indicate the threefold standard deviation for five additional, independent runs (see Supporting Information).

We start by illustrating the current state of the art. An established indicator for SRO in *a*-Si measures how similar

the atomic environments are to ideal tetrahedra, as probed via the bond angles.[9] This order parameter increases as expected with slower quenching (increasing ordering) and then converges in a way that the median results for our 10^{11} and 10^{10} Ks^{-1} structures are very similar (Figure 1b). We also look at MRO via the count of six-membered rings (Figure 1c)—which is 100% in diamond-type *c*-Si, where all atoms are in cyclohexane-like rings, but smaller in *a*-Si due to the presence of disorder. We stress again that these are two disjoint measures: bond angles cannot quantify MRO and the ring count does not give information about SRO.

To progress further, we now turn to the smooth overlap of atomic positions (SOAP) kernel,[14] a mathematical approach that has been used with success to fit ML potentials[15] and to analyze structures.[16] In the SOAP formalism, the neighborhood density of the *i*-th atom, smoothed by Gaussian functions with width σ_{at} and truncated by a cutoff function f_c, is expanded into an atom-centered basis of radial parts R_n and spherical harmonics Y_{lm},[14]

$$\rho_i(\mathbf{r}) = \sum_j \exp\left[-\frac{(r - r_{ij})^2}{2\sigma_{at}^2} \right] f_c(r_{ij}) = \sum_{nlm} c_{nlm}^{(i)} R_n(r) Y_{lm}(\widehat{\mathbf{r}}) , \qquad (1)$$

similar to how electronic wave functions are expanded in quantum chemistry. Based on the resulting combination coefficients, a similarity function or kernel can then be constructed, which provides a quantitative similarity measure on a scale from 0 to 1. However, the absolute value depends on the chosen cutoff radius and on the Gaussians that are placed on the atomic neighbors. To analyze both nearest-neighbor (NN) and next-nearest neighbor (NNN) environments, we here propose to combine two different SOAP kernels: that for the NN shell, making a sharp distinction between environments, and that for the NNN shell, being more tolerant to small structural changes. Therefore, we calibrate the fuzziness (via σ_{at}) using *c*-Si at $T > 0$ K as a reference by requiring that the NN and NNN range of the SOAP values are similar in an ordered network with only thermal fluctuations (shaded area in Figure 1d). We then apply the same pair of kernels to *a*-Si: there, the increasing order with decreasing quench rate, the convergence of SRO, and the further increase of MRO are all correctly described within the same conceptual framework.

The simplicity and power of this NN–NNN kernel pair can also be shown by applying it to crystalline allotropes of Si (Table 1). In the lonsdaleite(hexagonal diamond)-type form, the NN environments are as in the cubic diamond type (namely, ideal tetrahedra, with a similarity of 1.000), but the NNN environments differ, due to the rings being in boat rather than chair conformations, and hence the similarity to *c*-Si drops to 0.974. We next look at an open-framework Si allotrope, *o*S24, which was synthesized from Na$_4$Si$_{24}$ by sodium de-intercalation.[18] In *o*S24, the atoms are tetrahedrally coordinated, too, but with strong local distortions, and so the resulting NN values are comparable to those in *a*-Si; the NNN values drop further, because the open framework structure is remarkably different from *c*-Si. Finally, for β-tin-type Si with 4+2-type coordination, the NN environments are
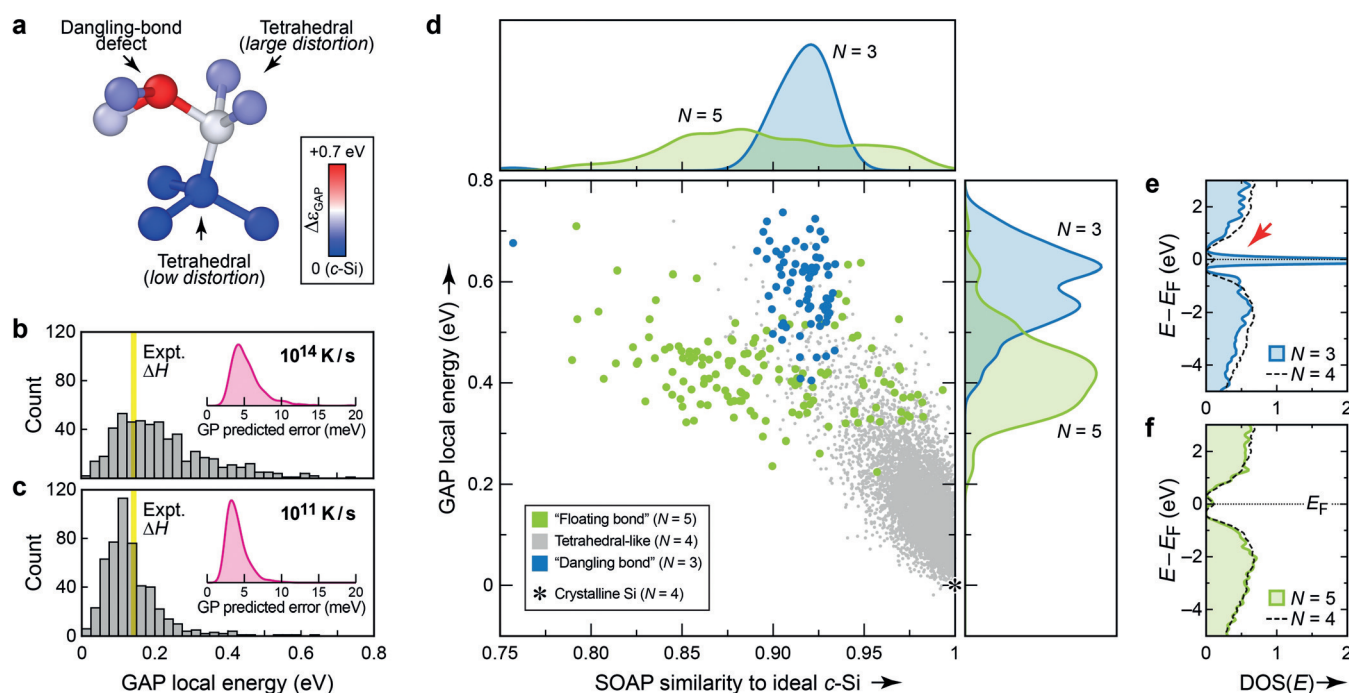
Figure 2. Machine-learned atomic energies in *a*-Si. a) Sample structural fragment, chosen to represent a dangling-bond defect with high energy (red), a distorted tetrahedral environment with intermediate energy (white), and a more favorable tetrahedral environment with only low distortion (blue). Atoms are color-coded according to their GAP atomic energy, ε_i, given relative to that in diamond-type *c*-Si. b) Histogram of atomic energies in the structure quenched at 10^{14} K s^{-1}. The experimental enthalpy of relaxed *a*-Si (from Ref. [19], also relative to *c*-Si) is indicated by yellow shading. The Gaussian-process(GP)-predicted error is indicated in the inset, showing a kernel-density estimate for all atoms in the structure (see Supporting Information for details). c) Same analysis as in (b) for the more ordered 10^{11} K s^{-1} structure. d) 2D plot revealing the connection between structural order (NN similarity to diamond-like *c*-Si; horizontal axis) and GAP local energy for the individual atoms (vertical axis). Results are collected for all 14 systems, that is, for all quench rates from 10^{14} to 10^{10} K s^{-1} (see Figure 1). Kernel-density estimates (smoothed histograms) are given for projections on both axes. e), f) Local electronic DOS for a structure quenched at 10^{11} K s^{-1} from Ref. [8], illustrating the very different electronic fingerprints of three- and fivefold-bonded coordination defects. DOS plots are normalized per atom; for comparison, the average local DOS for all fourfold bonded atoms in the same structure is given by dashed lines. The red arrow in (e) highlights the mid-gap states associated with dangling-bond defects.

clearly dissimilar to those in *c*-Si, and the NNNs even more so (Table 1).

We now consider the energies of the individual atoms, a crucial piece of information that cannot easily be obtained from DFT computations, which yield the total energy for the entire cell. In contrast, atomic energies are directly included in many ML-based interatomic potentials by construction.[6,18] In the Gaussian approximation potential (GAP) framework

Table 1: SOAP parameters[14] for the pair of kernels defined in this work and results for atomic sites in crystalline Si allotropes as obtained from both kernels.

		NN kernel	NNN kernel
Basis set size (n_{max}, l_{max})		(16, 16)	(16, 16)
Cutoff radius r_{cut} [Å]		2.85	5.00
Transition width r_Δ [Å]		0.30	0.60
Smoothness σ_{at} [Å]		0.30	0.60
Similarity of diamond-type *c*-Si and other allotropes			
[lonsdaleite]	Si1≡Si2 (2*b*)	1.000	0.974
*o*S24[17]	Si1 (8*f*)	0.980	0.792
	Si2 (8*f*)	0.994	0.908
	Si3 (8*f*)	0.987	0.747
[β-tin]	Si (4*a*)	0.723	0.342

we use here, the total energy is a sum of machine-learned atomic energies that generally read[18]

$$\varepsilon_i = \sum_j \alpha_j K_{ij}, \tag{2}$$

where the sum runs over a set of reference environments in the training database (index *j*) and two environments are compared using the similarity (kernel) function K_{ij}, for which we use the SOAP formalism here. Hence, our structural analysis, GAP-MD, and local energies all build on the same mathematical framework.

Figure 2a shows that, indeed, machine-learned atomic energies reveal the stability trends intuitively expected for *a*-Si, the interpretation being qualitative for now. In the structural fragment shown in Figure 2a, the dangling-bond defect (*red*) has a high local energy, whereas the two tetrahedral-like central atoms (white, blue) are more energetically favorable, depending on how strongly they are distorted. Histograms of these data, collected for a disordered, rapidly quenched structure (Figure 2b) and a more ordered, slowly-quenched structure (Figure 2c) reveal that the energetic center of gravity for the more ordered *a*-Si network does coincide with the experimentally determined stability.[19] We

note that such analyses are, in principle, possible with empirical interatomic potentials,[20] but can be limited by the parametric shape of the potential. In contrast, our approach depends only on the input data, is readily generalized, and combines accurate DFT input data with a high-fidelity ML fit whose uncertainty can be quantified[15d] to be in the meV range (insets of Figure 2b,c).

A 2D plot combining both quantities, the SOAP-based diamond-similarity and the local energy, is perhaps the most revealing form (Figure 2d). Distinct energy regions at around +0.4 and +0.6 eV above that of ideal c-Si are found for floating-bond ($N=5$) and dangling-bond ($N=3$) environments, respectively. The floating bonds show a wide structural variation within the NN shell, indicated by a large spread over the SOAP (x-axis) coordinate, whereas dangling bonds clearly peak around a similarity value of 0.92. In this respect, the dangling-bond atoms are clearly structurally closer to c-Si than the 4+2-type-coordinated atoms in the β-tin allotrope (Table 1), but they are considerably less diamond-like than the median result for any of our a-Si structures (Figure 1d), which are dominated by fourfold-bonded atoms. Looking at Figure 2d again, it appears that the data points for dangling bonds ($N=3$) lie in the tail of a continuation of the plot for tetrahedral-like environments ($N=4$); this is not the case for floating bonds ($N=5$). Finally, we note that the energies for $N=4$ atoms reach up to high values: their median is +0.14 eV, but the 98th percentile is at +0.42 eV and thus the remaining 2% fourfold-bonded atoms have energies that are higher than the median result for $N=5$ defects (+0.42 eV). This explains how our 10^{10} K s^{-1} structure, albeit having defects, can be lower in energy than the defect-free WWW model.[1]

The higher GAP atomic energy (that is, larger instability) of dangling bonds compared to floating bonds is not only in line with previous theories,[2a, 20] but it can also be confirmed by the electronic densities of states (DOS). For the energetically unfavorable dangling bonds ($N=3$), a large peak at the Fermi level, within the band gap, is seen from an atom-resolved projection of the DOS (Figure 2e). In contrast, the energetically more favorable floating bonds ($N=5$) make no pronounced mid-gap contributions to the DOS (Figure 2f).

This approach is expected to be general and to have wider significance. We show, as an example, an extension from the amorphous to the even more disordered liquid phase of silicon (Figure 3) which has been widely studied by simulations[22] and experiments.[23] Remarkably, the liquid appears to consist of atoms with a well-defined normal distribution of GAP local energies, around +0.6 eV/at. above that of c-Si. This distribution stays almost unchanged all the way until a narrow temperature window of approximately 1175 to 1195 K (Figure 3). There is a distinct transition with a bimodal distribution of atoms (center right panel in Figure 3) that our analysis suggests to be clearly either liquid-like or amorphous-like. Established empirical potentials[21] fail to capture the nature of this transition, presumably because they have not been fitted to include the diverse structures in the liquid state. Further work will deal with detailed mechanistic studies of liquid Si using our new approach and with the extension to other systems: the ideas described here could be transferred
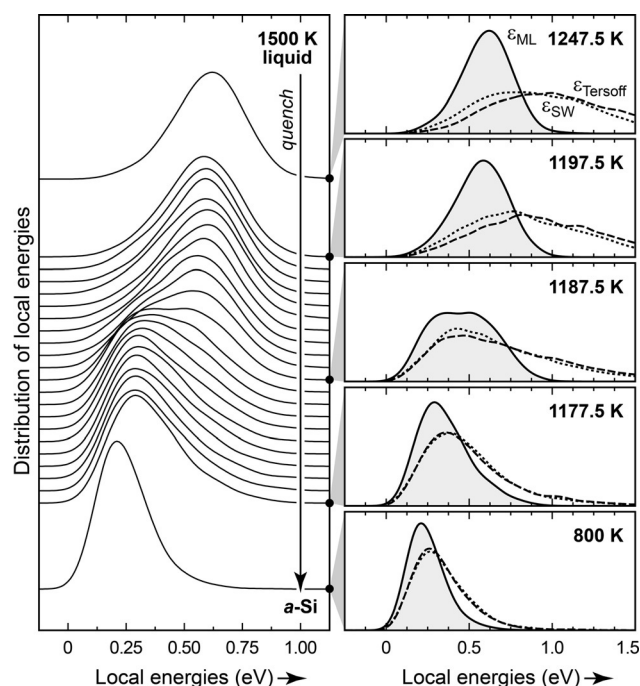

Figure 3. Machine-learned atomic energies in liquid Si and their evolution during quenching into an amorphous configuration. Left: Kernel density estimates of local energies in a 4096-atom system at selected points of an MD simulation (trajectory data from Ref. [8]). Right: Close-ups at given temperature values with results from two empirical potentials (Tersoff and Stillinger–Weber, SW; Ref. [21]), overlaid for the same structures, evidencing the qualitatively different interpretation given by our approach.

to other tetrahedral networks such as the homologous material germanium or the highly complex amorphous forms of carbon[7b] (on which work is ongoing), or even to crystalline, amorphous, and liquid states of water.[9b, 24]

Conflict of interest

The authors declare no conflict of interest.

[1] a) F. Wooten, K. Winer, D. Weaire, *Phys. Rev. Lett.* **1985**, *54*, 1392; b) B. R. Djordjević, M. F. Thorpe, F. Wooten, *Phys. Rev. B* **1995**, *52*, 5685.

[2] a) S. T. Pantelides, *Phys. Rev. Lett.* **1986**, *57*, 2979; b) R. Xie, et al., *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13250.

[3] a) W.-L. Shao, J. Shinar, B. C. Gerstein, F. Li, J. S. Lannin, *Phys. Rev. B* **1990**, *41*, 9491; b) P. Roura, J. Farjas, P. Roca i Cabarrocas, *J. Appl. Phys.* **2008**, *104*, 073521; c) A. M. R. De Graff, M. F. Thorpe, *Acta Crystallogr. Sect. A* **2010**, *66*, 22.

[4] Examples: a) D. A. Drabold, P. A. Fedders, O. F. Sankey, J. D. Dow, *Phys. Rev. B* **1990**, *42*, 5135; b) I. Štich, R. Car, M. Parrinello, *Phys. Rev. B* **1991**, *44*, 11092; c) N. Bernstein, J. L. Feldman, M. Fornari, *Phys. Rev. B* **2006**, *74*, 205202.

[5] R. Atta-Fynn, P. Biswas, *J. Chem. Phys.* **2018**, *148*, 204503.

[6] J. Behler, *Angew. Chem. Int. Ed.* **2017**, *56*, 12828; *Angew. Chem.* **2017**, *129*, 13006, and references therein.

[7] a) G. C. Sosso, G. Miceli, S. Caravati, F. Giberti, J. Behler, M. Bernasconi, *J. Phys. Chem. Lett.* **2013**, *4*, 4241; b) M. A. Caro, V. L. Deringer, J. Koskinen, T. Laurila, G. Csányi, *Phys. Rev. Lett.* **2018**, *120*, 166101; c) V. L. Deringer, C. Merlet, Y. Hu, T. H. Lee, J. A. Kattirtzi, O. Pecher, G. Csányi, S. R. Elliott, C. P. Grey, *Chem. Commun.* **2018**, *54*, 5988; d) N. Artrith, A. Urban, G. Ceder, *J. Chem. Phys.* **2018**, *148*, 241711; e) B. Onat, E. D. Cubuk, B. D. Malone, E. Kaxiras, *Phys. Rev. B* **2018**, *97*, 094106; f) V. Lacivita, N. Artrith, G. Ceder, *Chem. Mater.* **2018**, *30*, 7077.

[8] The computational protocol has been introduced and validated against experimental data in our previous work, which described structures at quench rates between 10^{14} and $10^{11}\,\mathrm{K\,s^{-1}}$: V. L. Deringer, N. Bernstein, A. P. Bartók, M. J. Cliffe, R. N. Kerber, L. E. Marbella, C. P. Grey, S. R. Elliott, G. Csányi, *J. Phys. Chem. Lett.* **2018**, *9*, 2879. Where appropriate, structures from these simulations are used in the present work; however, we added new data points that sample more finely spaced quench rates (Figure 1a) and we expanded the scope of the simulations to a slow quench rate of $10^{10}\,\mathrm{K\,s^{-1}}$ (see Supporting Information for details). The particular version of the potential used is described in Ref. [15d].

[9] a) P.-L. Chau, A. J. Hardwick, *Mol. Phys.* **1998**, *93*, 511; b) J. R. Errington, P. G. Debenedetti, *Nature* **2001**, *409*, 318.

[10] D. S. Franzblau, *Phys. Rev. B* **1991**, *44*, 4925.

[11] a) K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, A. Tkatchenko, *J. Phys. Chem. Lett.* **2015**, *6*, 2326; b) K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, A. Tkatchenko, *Nat. Commun.* **2017**, *8*, 13890;

c) X. Chen, M. S. Jørgensen, J. Li, B. Hammer, *J. Chem. Theory Comput.* **2018**, *14*, 3933.

[12] In β-rhombohedral boron, sites with partial vacancy formation were shown to be associated with high (unfavorable) GAP atomic energies: V. L. Deringer, C. J. Pickard, G. Csányi, *Phys. Rev. Lett.* **2018**, *120*, 156001.

[13] T. L. Jacobsen, M. S. Jørgensen, B. Hammer, *Phys. Rev. Lett.* **2018**, *120*, 026102.

[14] A. P. Bartók, R. Kondor, G. Csányi, *Phys. Rev. B* **2013**, *87*, 184115.

[15] Examples: a) A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, M. Ceriotti, *Sci. Adv.* **2017**, *3*, e1701816; b) V. L. Deringer, G. Csányi, *Phys. Rev. B* **2017**, *95*, 094203; c) D. Dragoni, T. D. Daff, G. Csányi, N. Marzari, *Phys. Rev. Mater.* **2018**, *2*, 013808; d) A. P. Bartók, J. R. Kermode, N. Bernstein, G. Csányi, *Phys. Rev. X* **2018**, *8*, 041048.

[16] a) S. De, A. P. Bartók, G. Csányi, M. Ceriotti, *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754; b) J. Mavračić, F. C. Mocanu, V. L. Deringer, G. Csányi, S. R. Elliott, *J. Phys. Chem. Lett.* **2018**, *9*, 2985; c) M. O. J. Jäger, E. V. Morooka, F. F. Canova, L. Himanen, A. S. Foster, *npj Comput. Mater.* **2018**, *4*, 37; d) M. A. Caro, A. Aarva, V. L. Deringer, G. Csányi, T. Laurila, *Chem. Mater.* **2018**, *30*, 7446.

[17] D. Y. Kim, S. Stefanoski, O. O. Kurakevych, T. A. Strobel, *Nat. Mater.* **2015**, *14*, 169.

[18] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, *Phys. Rev. Lett.* **2010**, *104*, 136403.

[19] S. Roorda, W. C. Sinke, J. M. Poate, D. C. Jacobson, S. Dierker, B. S. Dennis, D. J. Eaglesham, F. Spaepen, P. Fuoss, *Phys. Rev. B* **1991**, *44*, 3702.

[20] P. C. Kelires, J. Tersoff, *Phys. Rev. Lett.* **1988**, *61*, 562.

[21] a) "SW": F. H. Stillinger, T. A. Weber, *Phys. Rev. B* **1985**, *31*, 5262; b) "Tersoff": J. Tersoff, *Phys. Rev. Lett.* **1986**, *56*, 632.

[22] Examples: a) S. Sastry, C. A. Angell, *Nat. Mater.* **2003**, *2*, 739; b) V. Molinero, S. Sastry, C. A. Angell, *Phys. Rev. Lett.* **2006**, *97*, 075701; c) V. V. Vasisht, S. Saw, S. Sastry, *Nat. Phys.* **2011**, *7*, 549; d) L. Bonati, M. Parrinello, *Phys. Rev. Lett.* **2018**, *121*, 265701.

[23] Examples: a) A. Hedler, S. L. Klaumünzer, *Nat. Mater.* **2004**, *3*, 804; b) M. Beye, F. Sorgenfrei, W. F. Schlotter, W. Wurth, A. Föhlisch, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16772.

[24] Large-scale extended water networks are also becoming accessible to ML-based modeling. A recent study successfully described the melting point of water (to within 1 K): T. Morawietz, A. Singraber, C. Dellago, J. Behler, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8368.