I. INTRODUCTION

The aspiration of computational materials science is to autonomously predict structures with desirable properties and to design technologically relevant materials. This poses three main challenges: (i) comprehensively surveying the high-dimensional configuration space describing all possible structures, (ii) identifying experimentally and technologically relevant structures from a virtually infinite zoo of possible (meta)stable configurations, and (iii) designing experimental protocols to synthesize the structures of interest.

Numerous applications such as Refs.1–7 demonstrate how configuration spaces can be explored effectively by combining atomistic calculations with various structure searching techniques8–14, despite the exponential increase in the number of computationally (meta)stable structures with system size15. Meanwhile, the computational effort involved in (i) mapping phase diagrams using extensive Gibbs free energy calculations and (ii) determining possible synthetic pathways using methods such as forward flux sampling16 and enhanced sampling metadynamics17,18 prevents bulk calculations for large numbers of locally stable structures. One of the key steps on the path to computational materials design is thus the reliable identification of the manageably small number of compounds stabilized by diverse thermodynamic conditions, given that geometries and relative stabilities are only available for one particular set of conditions.

In the absence of kinetic effects19,20 a convex hull (CH) construction can be used to identify structures and compounds that are stable with respect to decomposition into two or more parent structures at fixed thermodynamic conditions. For instance, consider the volume-based CH. If two structures A and B with molar volumes \( V(A) \) and \( V(B) \) and free energies \( G(A) \) and \( G(B) \) are part of the hull, then any structure \( C \) with molar volume \( V(A) < V(C) < V(B) \) and a free energy that lies above the line joining A and B on a \( V-G \) plot will spontaneously decompose at constant volume into a mixture of A and B (see Fig. 1 (d), taking \( \phi = V \)).

CH constructions have proven useful in numerous structure searching applications such as Refs.3–7. However, the conventional form has some crucial limitations. The choice of one particular feature, such as molar volume, on which the CH is constructed, relies on experimental evidence or preconceived notions of which thermodynamic constraints may stabilize structures of interest. It limits which stabilizable structures are identified, and is generally insufficient to explore the structural diversity that can be accessed experimentally through complex thermodynamic constraints such as pressure, composition, doping with guest molecules, substitution of portions of organic compounds, electric or magnetic fields, etc. (for instance, see Ref.23). Furthermore, the conventional CH construction neglects inevitable inaccuracies in the computed (free) energies and geometries, which render the CH probabilistic in nature.

While the identification of experimentally-synthesizable compounds is the focus of this work, the generalized CH framework proposed in the following also translates (at negligible computational cost) input energies into a far better measure of structural stability, namely the energy relative to the GCH. The latter can be used in place of bare energies in diverse applications, such as experimental crystal structure determination protocols or as the fitness function driving structure searches in situ.
II. THE GENERALIZED CONVEX HULL

To overcome the above limitations we introduce a probabilistic generalized CH (GCH) framework for evaluating the probabilities of structures being stabilized by general thermodynamic constraints. A schematic representation of this framework is shown in Fig. [1]. It (i) quantifies the uncertainty arising from the inevitable errors in the underlying energies and structures, and (ii) rests on geometric fingerprints \( \Phi = \{ \Phi_i \} \), which reflect the full structural diversity of the dataset. While there is considerable freedom in choosing such fingerprints, they must exhibit an additive behavior, that guarantees that a macroscopic sample \( X \), which is a phase-separated mixture of different components \( X_k \) with molar fractions \( w_k \), has a fingerprint \( \Phi(X) = \sum_k w_k \Phi(X_k) \). A simple way to guarantee that \( \Phi_i \) fulfills this requirement is to choose descriptors that are consistent with an atom-based decomposition, \( \Phi(X) = \sum_{X \in X} \Phi(X')/N_X \). Here \( \Phi(X') \) are the fingerprints of the \( N_X \) atom-centered, local environments \( X \) within the structure \( X \). Additivity ensures that any structure with features inside a convex region of \( D \)-dimensional feature-space can be decomposed into a phase-separated mixture of the \( D+1 \) vertices of the convex region, without changing the corresponding \( D \) features of the fingerprint describing the system (although the resultant fingerprint may differ in the remaining features). By considering the molar free energy as a function of a set of \( D \) features \( \Phi_i \), one can thus generalize the CH construction to identify the structures that are stable with respect to decomposition subject to the abstract “thermodynamic constraint” defined by a given set of \( D \) features.

Data-driven structure fingerprints. For a given dataset \( \{ X_k \} \) we extract a small set of key data-driven features that captures most of its structural diversity by performing a kernel principal component analysis (KPCA) on a kernel measure of similarity \( K(X_k, X_l) \) between pairs of structures \( X_k \) and \( X_l \). That is, we compute the eigenvalues \( \lambda_i \) and eigenvectors \( u_i \) of the kernel matrix, \( K_{kl} = K(X_k, X_l) \), and evaluate the features of a structure \( X_k \) as

\[
\Phi_i(X_k) = \sum_l u_i^l \sqrt{\lambda_i} K_{kl}.
\]

These features are additive, provided that \( K(X_k, X_l) = \sum_{X_k \in X_k, X_l \in X_l} k(X_k, X_l)/N_{X_k}N_{X_l} \), where \( k(X_k, X_l) \) is a kernel measure of similarity between pairs of local environments \( X_k \) and \( X_l \).

In practice, we use the smooth overlap of atomic positions (SOAP) kernel, which is constructed around atom-
centered, local environments and thus additive, but is otherwise general and agnostic, and can be applied seamlessly to different kinds of materials.\textsuperscript{21–24} Crucially, SOAP fingerprints have proven reliable for both energy regression\textsuperscript{23} and structure classification\textsuperscript{24} for the systems discussed in the following. SOAP describes an atomic environment $X$ as a sum of atom-centered Gaussians
\begin{equation}
\langle \alpha r | X \rangle = \sum_{j \in X, \alpha} \exp \left( -\frac{(x_j - r)^2}{2\sigma^2} \right),
\end{equation}
where $x_j$ are the Cartesian coordinates of the atoms of chemical identity $\alpha$ (H, O, C, ... ) within a radial cutoff $r_c$. $\sigma$ specifies the width associated with each atomic probability distribution. The (rotationally averaged) power spectrum of the expansion of an environment $X$ on a basis of radial functions $R_n(r)$ and spherical harmonics $Y_{lm}(\hat{r})$,
\begin{equation}
\langle \alpha n\ell l | X \rangle \propto \sum_m \langle \alpha n\ell m | X \rangle \langle X | \alpha n\ell m \rangle,
\end{equation}

\begin{equation}
\langle \alpha n\ell l | X \rangle = \int dr R_n(r) Y_{lm}(\hat{r}) \langle X | \alpha r \rangle,
\end{equation}
provides the representation to define the environmental kernels
\begin{equation}
k(X, X') = \left[ \sum_{\alpha n\ell l} \langle \alpha n\ell l | X \rangle \langle \alpha n\ell l | X' \rangle \right]^2.
\end{equation}
Loosely speaking, the resultant KPCA features $\Phi_i(X_k)$ are orthonormal measures of the similarity of the structure $X_k$ to a particular combination of all structures in the dataset, dominated by the structurally most distinct configurations.

Feature selection and interpretation. The abstract nature of these KPCA features begs the question of (i) how to identify which among them have the potential to stabilize structures and should thus be included in the GCH construction, and (ii) how to relate them to experimentally realizable conditions. When no prior knowledge of the system is available the KPCA eigenvalue spectrum provides indication of the maximum intrinsic dimensionality (Fig. 2) of the structure data at hand.\textsuperscript{25} It can thus be used to choose the dimensionality of the GCH such that the full structural diversity of the dataset is explored. Even in this worst case scenario, the resultant pool of candidates is typically orders of magnitude smaller than the underlying structure database, rendering it possible to further investigate the relations between the features of the candidates and physical observables (or thermodynamic constraints) such as density, composition, etc. This can not only help to translate abstract structural features into practically realizable synthetic protocols, but also to refine the selection of features on which the GCH is constructed \textit{a posteriori} to those which couple strongly to experimentally realizable conditions and thus have the greatest potential for stabilizing structures.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig2.png}
\caption{KPCA eigenvalues for the applications we discuss in this work, namely: hydrogen (black), H$_2$O$_{1-x}$ (red), and pentacene (blue), obtained from SOAP similarity kernels with $r_c = 2$ Å, 5 Å, and 5 Å, respectively.}
\end{figure}

Probabilistic GCH and uncertainty quantification. So far, the GCH framework neglects the inevitable uncertainties in (computed) free energies, lattice parameters and atomic positions, and therefore in the determination of the hull vertices, making it necessary to use rather arbitrary confidence regions around the hull, and to manually remove duplicate structure.\textsuperscript{27} We therefore propose a probabilistic extension in which the GCH probability distribution is sampled by constructing many possible convex hulls based on free energies and geometries, which have been randomized according to their respective uncertainties. In practice we take the typical model errors on the energies $\epsilon$ and Cartesian coordinates (for example, due the choice of density functional in density functional theory (DFT) calculations or the absence of quantum nuclear effects) to be known from experience or benchmarks. We estimate the resultant errors in the energies \textit{relative to the instantaneous hull}, $\sigma_{G_k}$, exploiting structural correlations to account for correlations between the errors in $\{G_k\}$. In particular we ensure that $\sigma_{G_k}$ vanishes for the vertex structures and any structure that is a phase-separated mixture of the vertices of its associated simplex (its “parent phases”), while otherwise reflecting how different a given non-hull structure is from the parent phases. The rationale is that the typical errors are not random, but correlate with the structural features. Consider for instance a phase-separated mixture $X_k$ composed of molar fractions $w_{kl}$ of the parent phases $X_l$ with calculated energies $G_l + \epsilon_l$. Its calculated energy is identical to the corresponding combination of the energies of the parent phases, including their errors, $\sum_l w_{kl}(G_l + \epsilon_l)$. This is exactly the definition of the convex hull energy constructed on $G_l + \epsilon_l$, so that the energy of $X_k$ relative to the hull will be zero regardless of the errors. Hence, $\sigma_{G_k}$ should vanish.

Let us introduce a practical definition that satisfies this requirement. We estimate $\sigma_{G_k}$ as the fraction of the total error $\epsilon$ associated with the deviation of the features

FIG. 3. Maps of 7,594 hydrogen structures spanned by the two dominant KPCA features, $\Phi_1$ and $\Phi_2$. Due to their abstract nature (Eqs. (1) to (4)) the numerical value of $\Phi_1$ and $\Phi_2$ is not shown. Each point corresponds to a structure in the dataset. The maps on the left are colored according to (a) molar volume and (b) molar energy. One can see the clear correlation between the KPCA coordinates, and structural and energetic properties. (c) The larger map highlights structures with non-negligible probability $p_{\text{vertex}}$ of being part of the GCH built on the first four KPCA features, which is represented as a color scale. Candidates surviving an additional “coarse-graining” step down to the point where all remaining structures have $p_{\text{vertex}} = 1$ are labeled according to space group and number of atoms per unit cell. By comparison with the map colored according to molar energy, one sees that the convex hull identifies clusters of configurations that are low in energy and/or extremal in structure.

We learn by ridge regression from a machine-learning model of $G_k$, and $\sigma_{G_k}^2$ is the variance of $G$ over the entire dataset. Due to additivity, for a physical mixture, $\Phi_i(X_k) = \Phi_i^{\text{GCH}}(X_k)$ for all the features, including those that are not used for the GCH construction, which ensures that $\sigma_{G_k} = 0$. On the contrary, for each point that is not a physical mixture of hull points, only the features used to build the GCH will coincide with $\Phi_i^{\text{GCH}}(X_k)$. In this case, $\sigma_{G_k}$ scales with the residual structural diversity that is not captured by the GCH coordinates. Note that the dependence of the uncertainties $\sigma_{G_k}$ on the instantaneous hull implies that the hull distribution must be sampled “self-consistently”.

The randomization of the features $\Phi_i$ requires knowledge of how the uncertainty in the underlying atomic coordinates and lattice parameters (or “structure parameters”) propagates to uncertainties in the features, $\sigma_{\Phi_i}$. We estimate $\sigma_{\Phi_i}$ by randomizing the structure parameters of $n_s$ reference configurations $X_r$, that we choose by farthest point sampling. In practice, we randomize each reference structure $n_s$ times, compute the features for the randomized structures $\Phi_i(X_r^s)$, and evaluate

$$\sigma_{\Phi_i} = \sqrt{\frac{1}{n_s n_r} \sum_r \sum_s (\Phi_i(X_r^s) - \Phi_i(X_r))^2}.$$  

After extensive sampling of the GCH distribution the rate with which each structure occurs as a vertex $p_{\text{vertex}}(X_k)$ roughly quantifies how trustworthy the identification of the structure $X_k$ as stabilizable is and its average distance from the hull provides a measure of its (meta)stability.

**Coarse graining of the GCH vertices.** In cases where large numbers of very similar structures (for example owing to stacking faults or partial disorder) compete for stability each candidate exhibits a small individual probability of becoming stable. However, collectively such a cluster of structures represents a stable phase. For convenience we reduce the full list of potential vertices to representatives of each cluster, that is, of each stable phase. These are identified by sequentially eliminating the $N$ lowest probability candidates with a cumulative probability $\sum_{k=1}^{N} p_{\text{vertex}}(X_k) < 1$ (which guarantees that no complete cluster of structures that constitutes one stabilizable structure gets eliminated entirely in one step) from the dataset and resampling the GCH for the thus reduced dataset. This procedure is repeated until the lowest $p_{\text{vertex}}(X_k)$ is above a set threshold of 0.5.
This “coarse-graining” ensures that the surviving candidates correctly accumulate the probability of becoming stable associated with their respective clusters of similar structures. Even though we only consider these marginal probabilities, the GCH directly samples the full hull distribution, which can further be used to investigate for instance which structures compete for stability.

III. APPLICATIONS

To demonstrate the power of the GCH framework, we apply it to four problems of increasing complexity, namely a database of hydrogen solid phases at terapascal pressure, a set of oxygen-hydrogen binary crystal structures, a subset of this database for which we demonstrate how a GCH can predict oxygen phases that are stabilized by magnetism, and a set of crystalline polymorphs of pentacene for which we investigate chemical substitutions and demonstrate the stability of the GCH to errors in the input energy data. The respective structure databases are available as supplemental material.

Hydrogen at gigapascal pressure. As a first test, we analyze 7,964 locally stable hydrogen structures from an ab initio random structure search (AIRSS)\(^{125}\) at 500 GPa based on DFT geometry optimizations using the Perdew-Burke-Ernzerhof (PBE) functional\(^{29}\) where extensive experimental and theoretical literature\(^{50} - 63\) provides a detailed reference of stabilizable structures. Fig. 3 shows a representation of the GCH procedure when performed on the dominant two KPCA components resulting from a SOAP kernel \((r_c = 2\) Å). It is clear that the principal components correlate strongly with the cohesive energy and the molar volume of the structures, and that the GCH procedure identifies configurations that are extremal in geometry and/or particularly favorable energetically. While this two-dimensional map provides for a more intuitive visualization, the KPCA eigenvalue spectrum (see Fig. 2) suggests that the intrinsic dimensionality of the dataset is higher. We therefore consider for further analysis the GCH constructed on the top four components. We identify 81 candidate structures, and successfully recover the high-pressure molecular \(I4_1 amd\) and atomic \(R3m\) phases of hydrogen, as well as analogs of the lower-pressure phases II to IV (a comparison between the structures and their lower-pressure analogs is given in the SI). The latter are not stable at the simulated conditions, so being able to find very similar structures among the candidates is a testament to the long-sightedness of AIRSS and the predictive power of the GCH. To achieve the same feat using a conventional energy-volume CH, structures up to around 8 meV/atom above the CH have to be retained, leaving a disproportionately larger pool of more than 2,000 potentially stabilizable structures.

Oxygen-Hydrogen binary compounds. The next level of complexity in computational materials discovery involves the modeling of multi-component systems: in the case of 51,376 locally stable \(H_xO_{1-x}\) configurations from an ab initio random structure search (AIRSS)\(^{125}\) at 20 GPa (again based on DFT geometry optimizations using the PBE functional\(^ {20}\) the GCH framework must resolve the most stable stoichiometries, while at the same time recovering various hydrogen, ice and oxygen phases. The KPCA eigenvalues based on a SOAP kernel \((r_c = 5\) Å) decay by more than an order of magnitude after the first feature (see Fig. 2). This reflects the dominant role of composition in determining structural diversity and forecloses the identification of the first KPCA feature with composition (see Fig. 3 (a)). Along this principal axis, one identifies the expected stable oxygen, hydrogen, and ice structures, but also crystalline hydrogen peroxide, ice phases with different fractions of intercalated hydrogen molecules and crystalline molecular hydrogen and oxygen phases with guest water molecules. The latter are unstable in the absence of other stabilizing fields as highlighted by an energy-composition CH construction. Their stability on the GCH arises because the first KPCA feature (while predominantly describing composition) also measures molar volume as an additional stabilizing factor. When constructed on the first two
KPCA features the GCH framework identifies 171 stabilizable structures, differing in both stoichiometry and geometry (see Fig. 5 (b)). Among nine hydrogen structures are phase I, the P21-
24 candidate for phase II, and the Cmca-4 candidate for phase IV.55 Reassuringly, the three ice phases include the experimentally stable ice VII/VIII and the Pmc21 high-pressure candidate phase of Hermann et al.35

Magnetically-stabilized phases of oxygen. The six oxygen structures deserve a more detailed discussion, as they demonstrate that the GCH is capable of revealing subtle mechanisms of stabilization, which have barely been touched upon in literature, such as the stabilization of unconventional molecular oxygen phases by external magnetic fields. Using the nomenclature introduced in Refs. 36–38, the six oxygen structures include the conventional α/β and δ phases, in which the O₂ molecules align in the so-called “H”-state (Fig. 5 (b)). The GCH further detects α/β and δ phases with uniformly-tilted O₂ molecules (“S” state, Fig. 5(c)) and an α phase, in which the molecules display an alternating tilt pattern (“X” state, Fig. 5 (d)). Experimental evidence suggests that these may be stabilized by strong magnetic fields,36,40 which we further substantiate using spin-polarized DFT calculations using QUANTUM ESPRESSO (see Fig. 5 and SI Fig. S5). This demonstrates (i) that structural features do indeed correlate with subtle responses to manipulations of the electronic structure of a configuration and (ii) how one can verify the coupling between abstract structural coordinates and experimentally realizable thermodynamic constraints.

Nitrogen substitution in pentacene. As a final example, we analyze a database of 564 locally stable arrangements of pentacene molecules. This application beyond high-pressure physics demonstrates how the GCH can suggest suitable starting points for studies of chemical substitution. The configurations were obtained by a systematic structure search42 based on rigid, DFT-optimized molecular units interacting via the W99 force-field.43 In Ref.42, this structure search is accompanied by independent searches for 5A (see Fig. 6 and 5B nitrogen-substituted molecules, which are required because the stability of a given molecule is rarely a good predictor of the behavior of its substituted counterpart.14

We first perform a KPCA of the pentacene dataset using the same SOAP kernel (r_e = 5Å and σ = 0.3Å) which has previously proven suitable for energy regressions.53 Alongside conventional, energetically favorable herringbone configurations, such as the Campbell bulk phase (Fig. 6 (a)), the GCH constructed on the two dominant KPCA features identifies five energetically unfavorable configurations with planar, colinear arrangements of molecules as stabilizable. Whereas nitrogen substitution of the global minimum pentacene configuration leads to a high-energy, unstable structure, several of the GCH vertices that are much higher in energy, which would therefore be discarded in a conventional analysis, retain their geometry upon nitrogen substitution and relaxation (see Fig. 6). Moreover, they exhibit competitive energies compared to

FIG. 5. PCA projection of the subset of 84 pure oxygen structures onto Φ₁ and Φ₂ as obtained for the full dataset of 51,376 H₂O₁₋₅ structures. Diamagnetic molecular structures (solid circles) are colored according to ΔG/Δm. Atomic and ferromagnetic molecular structures are shown as open circles and crosses, respectively. The shaded regions highlight molecular structures in the H, S, and X configurations ((b) to (d)), and are colored according to the respective mean values of ΔG/Δm. This highlights the correlation between Φ₁, molecular tilts, and energetic response to magnetization ΔG/Δm as a proxy of the potential for stabilization by magnetic fields.

FIG. 6. Sublimation energies, E_{subl}, of different pentacene configurations in kJ/mol before (left) and after 5A nitrogen substitution (center), and after subsequent geometry optimization (right). (a) is among the most unstable pentacene configurations in the dataset. (e) is the most stable 5A substituted azapentacene configuration among 594 configurations from an independent structure search.42 The E_{subl} computed for the Campbell bulk phase (b) of 151.019 kJ/mol agrees with the experimental values of 154.34 and 156.9 ± 13.6 kJ/mol45 to within the errors.
the most stable 5A substituted configuration. The GCH framework has thus effectively identified pentacene configurations with potential for stabilization by nitrogen substitution.

**Sensitivity to errors in energetics.** The probabilistic sampling of the GCH does not only provide a robust strategy to eliminate redundant structures and for uncertainty quantification. It also significantly reduces the sensitivity to errors in input energies compared to conventional deterministic CH constructions. To assess how sensitive different CH constructions are with respect to the details of the input energies, we calculate DFT sublimation energies for the 564 pentacene configurations from Ref\(^\text{[22]}\), (i) The RMSE in kJ/mol in the W99 convex hull energies \(\{E_k^{\text{W99}}\}\) compared to reference DFT convex hull energies \(\{E_k^{\text{DFT}}\}\) (for the full dataset), and (ii) the distance \(d\) between the W99 and DFT based hulls as defined in Eq. (6).

\[
\text{RMSE}(\{E_k^{\text{W99}}\}) = 0.22, 0.11, 0.10, 0.07, 0.07, 0.0139, 0.0227, 0.0168, 0.0087, 0.0066, 0.0009
\]

|             | RMSE \(\{E_k^{\text{W99}}\}\) | \(d\) |
|-------------|-------------------------------|------|
| \(\text{E-\text{\(\rho\)}}\) CH | 0.22                          | 0.0139 |
| d-GCH (1D)  | 0.11                          | 0.0227 |
| GCH (1D)    | 0.10                          | 0.0168 |
| GCH (1D) cg | 0.07                          | 0.0046 |
| d-GCH (3D)  | 0.07                          | 0.0087 |
| GCH (3D)    | 0.07                          | 0.0066 |
| GCH (3D) cg | 0.07                          | 0.0009 |

TABLE I. Sensitivity analysis of the (conventional) energy-density \(\text{E-\text{\(\rho\)}}\) CH hull, and deterministic (d-GCH), and probabilistic hulls (GCH) constructed on the first (1D) and first three (3D) KPCA features (before and after coarse-graining (cg)). Different metrics of the similarity of different CH constructions are evaluated on the basis of W99 and DFT sublimation energies for the 564 pentacene configurations from Ref\(^\text{[22]}\).

The results of this analysis, shown in Table I, confirm that the GCH construction reduces the sensitivity of both the vertex selection and the measure of stability compared to a conventional construction. Increasing the dimensionality of the fingerprint space on which the hull is constructed, sampling probabilistically different realizations of the hull, and eliminating redundant structures in the database, all lead to a more robust determination of stabilizable structures that should be considered for further theoretical or experimental investigation.

### IV. CONCLUSIONS

These examples clearly evidence the wide spectrum of thermodynamic constraints which can be rationalized using the GCH framework and serve to showcase the remarkable versatility and transferability of the GCH framework. The construction is only weakly dependent on the details of the kernel, and its probabilistic nature renders it robust to errors in the determination of the (free)-energies of different phases, which is very important given the harsh compromises one has to make between the accuracy and thoroughness of high-throughput structure searches. Moreover, it is capable of eliminating redundant configurations in a physically meaningful way and of providing estimates of stability regimes in terms of experimentally realizable thermodynamic constraints. The GCH framework provides a robust, data-driven, method- and error-insensitive evolution of the convex hull construction, one of the most essential tools to predict and rationalize the stability of materials, and to identify experimentally stabilizable structures among large numbers of locally stable configurations.

**Acknowledgements.** M.C., A.A. and E.A.E. were supported by the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 677013-HBMAP). C.J.P. is supported by the Royal Society through a Royal Society Wolfson Research Merit award. Calculations were performed on the Archer facility of the United Kingdom’s national high-performance computing service (for which access was obtained via the UKCP consortium [EP/P022596/1]). We would like to thank G.M. Day and
J. Yang for sharing the W99 optimized configurations of pentacene and 5A crystals, and for insightful discussion.

1 C. J. Pickard, M. Martínez-Canales, and R. J. Needs, Physical Review B 85, 214114 (2012).
2 S. Azadi, B. Monserrat, W. M. C. Foulkes, and R. J. Needs, Physical Review Letters 112, 165501 (2014).
3 N. D. Drummond, B. Monserrat, J. H. Lloyd-Williams, P. López Ríos, C. J. Pickard, and R. J. Needs, Nature Communications 6, 7794 (2015).
4 I. Errea, M. Calandra, C. J. Pickard, J. R. Nelson, R. J. Needs, Y. Li, H. Liu, Y. Zhang, Y. Ma, and F. Mauri, Physical Review Letters 114, 157004 (2015).
5 A. P. Drozdov, M. I. Eremets, I. A. Troyan, V. Ksenofontov, and S. I. Shylin, Nature 525, 73 (2015).
6 M. Mayo, K. J. Griffith, C. J. Pickard, and A. J. Morris, Chemistry of Materials 28, 2011 (2016).
7 B. Monserrat, R. J. Needs, E. Gregoryanz, and C. J. Pickard, Physical Review B 94, 134101 (2016).
8 C. J. Pickard and R. J. Needs, Physical Review Letters 97, 045504 (2006).
9 C. W. Glass, A. R. Oganov, and N. Hansen, Computer Physics Communications 175, 713 (2006).
10 M. Ansli and S. Goedecker, Journal of Chemical Physics 133, 224104 (2010).
11 T.-Q. Yu and M. E. Tuckerman, Physical Review Letters 107, 015701 (2011).
12 Q. Zhu, A. R. Oganov, C. W. Glass, and H. T. Stokes, Acta Crystallographica B 68, 215 (2012).
13 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, Computational Materials Science 68, 314 (2013).
14 A. M. Reilly et al., Acta Crystallographica B 137, 439 (2016).
15 Stillinger, Phys. Rev. E 59, 48 (1999).
16 Allen et al., Journal of Chemical Physics 124, 024102 (2006).
17 Quigley and Rodger, Molecular Simulations 35, 613 (2016).
18 Giberti et al., IUCrJ 2, 256 (2015).
19 R. Malik, F. Zhou, and G. Ceder, Nature Materials 10, 587 (2011).
20 A. Pulido, L. Chen, T. Kaczorowski, D. Holden, M. A. Little, S. Y. Chong, B. J. Slater, D. P. McMahon, B. Bonillo, C. J. Stackhouse, A. Stephenson, C. M. Kane, R. Clowes, T. Hasell, A. I. Cooper, and G. M. Day, Nature 567, 543 (2017).
21 W. J. Szlachta, A. P. Bartók, and G. Csányi, Physical Review B 90, 104108 (2014).
22 V. L. Deringer and G. Csányi, Physical Review B 95, 094203 (2017).
23 A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Science Advances 3, e1701816 (2017).
24 F. Musil, S. De, J. Yang, J. E. Campbell, G. M. Day, and M. Ceriotti, Chemical Science 9, 1289 (2018).
25 S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Physical Chemistry Chemical Physics 18, 13754 (2016).
26 K. Fukunaga and D. R. Olsen, IEEE Transactions on Computers 20, 176 (1971).
27 E. A. Engel, A. Anelli, M. Ceriotti, C. J. Pickard, and R. J. Needs, Nature Communications 9, 2173 (2018).
28 C. J. Pickard, M. Martínez-Canales, and R. J. Needs, Physical Review B 86, 059902 (2012).
29 J. P. Perdew, K. Burke, and M. Ernzerhof, Physical Review Letters 77, 3865 (1996).
30 M. I. Eremets and I. A. Troyan, Nature Materials 10, 927 (2011).
31 J. M. McMahon and D. M. Ceperley, Physical Review Letters 106, 165302 (2011).
32 J. M. McMahon, M. A. Morales, C. Pierleoni, and D. M. Ceperley, Reviews of Modern Physics 84, 1607 (2012).
33 M. I. Eremets, I. A. Troyan, and A. P. Drozdov, (2016), arXiv:1601.04479 [cond-mat.mtrl-sci].
34 P. Dalladay-Simpson, R. T. Howie, and E. Gregoryanz, Nature 529, 63 (2016).
35 A. Hermann, N. W. Ashcroft, and R. Hoffmann, Proceedings of the National Academy of Sciences 109, 745 (2012).
36 G. C. DeFotis, Physical Review B 23, 4714 (1981).
37 M. C. van Hemert, P. E. S. Wormer, and A. van der Avourd, Physical Review Letters 51, 1167 (1983).
38 R. Kitaura, S. Kitagawa, Y. Kubota, T. C. Kobayashi, K. Kindo, Y. Mita, A. Matsuo, M. Kobayashi, H. Chang, T. C. Ozawa, M. Suzuki, M. Sakata, and M. Takata, Science 298, 2352 (2002).
39 Y. A. Freiman and H. J. Jodl, Physics Reports 401, 1 (2004).
40 T. Nomura, Y. H. Matsuda, S. Takeyama, A. Matsuo, K. Kindo, J. L. Her, and T. C. Kobayashi, Physical Review Letters 112, 247201 (2014).
41 P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Cresolli, G. L. Chiarotti, M. Coccoccioni, I. Dabo, A. Dal Corso, S. Fabris, G. Fratesi, S. de Gironcoli, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, Journal of Physics: Condensed Matter 21, 395502 (2009).
42 J. E. Campbell, J. Yang, and G. M. Day, Journal of Materials Chemistry C 5, 7574 (2017).
43 D. E. Williams, Journal of Molecular Structure 485, 321 (1999).
44 I. Giangreco, J. C. Cole, and E. Thomas, Crystal Growth and Design 17, 3192 (2017).
45 C. De Kruif, Journal of Chemical Thermodynamics 12, 243 (1980).
46 V. Oja and E. M. Suuberg, Journal of Chemical Engineering Data 43, 486 (1998).
47 R. B. Campbell, J. M. Robertson, and J. Trotter, Acta Crystallographica A 44, 705 (1961).
48 We use a plane-wave energy cutoff of 100 Rydberg, a Monkhorst-Pack k-point grid [19] spacing of less than 2π × 0.07 Å−1, and the ultrasoft C.pbe-n-kjpaw_psl.0.1.UPF, H.pbe-kjpaw_psl.0.1.UPF, and N.pbe-n-kjpaw_psl.0.1.UPF pseudopotentials from http://www.quantum-espresso.org.
49 H. J. Monkhorst and J. D. Pack, Physical Review B 13, 5188 (1976).