

Liquid water contains the building blocks of diverse ice phases

Bartomeu Monserrat ^{1,2}, Jan Gerit Brandenburg^{3,4}, Edgar A. Engel² & Bingqing Cheng ^{2,5} 

Water molecules can arrange into a liquid with complex hydrogen-bond networks and at least 17 experimentally confirmed ice phases with enormous structural diversity. It remains a puzzle how or whether this multitude of arrangements in different phases of water are related. Here we investigate the structural similarities between liquid water and a comprehensive set of 54 ice phases in simulations, by directly comparing their local environments using general atomic descriptors, and also by demonstrating that a machine-learning potential trained on liquid water alone can predict the densities, lattice energies, and vibrational properties of the ices. The finding that the local environments characterising the different ice phases are found in water sheds light on the phase behavior of water, and rationalizes the transferability of water models between different phases.

¹Department of Materials Science and Metallurgy, University of Cambridge, 27 Charles Babbage Road, Cambridge CB3 0FS, UK. ²Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge CB3 0HE, UK. ³Interdisciplinary Center for Scientific Computing, University of Heidelberg, Im Neuenheimer Feld 205A, 69120 Heidelberg, Germany. ⁴Digital Organization, Merck KGaA, Frankfurter Str. 250, 64293 Darmstadt, Germany. ⁵Accelerate Programme for Scientific Discovery, Department of Computer Science and Technology, 15 JJ Thomson Ave, Cambridge CB3 0FD, UK. ✉email: bc509@cam.ac.uk

The number of experimentally observed and theoretically predicted phases of water seems to be ever growing¹. The most ubiquitous phase on Earth, liquid water, has many intriguing properties, including a density maximum at 4 °C and ambient pressure, volume expansion upon freezing, unusually high surface tension, melting, and boiling point². Liquid water exhibits no long-range order and its local structure is difficult to quantify and yet intricately related to its unique properties^{2–5}. Beside the liquid, the various ice phases in the complex phase diagram of water are made from distinct local atomic environments¹, which lead to a large spread in their densities, lattice energies, and other thermodynamic as well as kinetic properties^{1,6}. Apart from the direct connection with physical properties, the local structures in water are also related to the transition paths between the phases^{7,8}.

One intriguing question thus is the structural relationship between the different ice phases, and between ice and liquid water. This is not an easy topic to investigate, however, due to the structural complexity of the liquid^{4,9,10} and the large number of ice phases¹. In this work, we exploit state-of-the-art advances in machine learning (ML) for chemistry and materials, in order to compare the local environments in various phases of water in a general and systematic manner. More specifically, we first curate a dataset consisting of 54 representative phases of ice including all the known phases (see “Methods” section), whose densities range from 0.7 to 1.4 g/mL. Then we demonstrate that the local atomic environments found in liquid water cover the ones observed in all these ice phases, using a universal and automated framework for comparing the local similarities. As a consequence of this inclusion, a machine-learning potential (MLP)¹¹ that is only trained on liquid water accurately reproduces the ice properties including lattice energies, mass densities, and phonon densities of states.

Results

Curated dataset of diverse water phases. We first select representative atomistic configurations of diverse crystalline and liquid phases. We start from 57 ice crystal structures, which include all the experimentally known ices. These were screened from an extensive set of 15,859 hypothetical ice structures using a generalized convex hull construction (an algorithm for identifying promising experimental candidates)^{12,13}. After rigorous geometry optimizations at zero pressure (see “Methods” section), we eliminate three defective phases and the very high pressure phase X, and added the originally missing ice IV. The “Methods” section describes the dataset of the remaining 54 ice phases in more detail. Note that some structures (with particular hydrogen arrangements) represent both a proton-ordered and a proton-disordered form: for example, one ice structure prototypes both ice Ih and XI. We consider the respective minimum potential energy configurations of the ice phases, because they provide reasonable and reproducible approximations to the physical properties of ice, and serve as starting points for computing thermodynamic properties.

Compiling a set of representative structures for liquid water is less straightforward, since the liquid persists over a wide range of temperatures and pressures. We consider 1000 diverse 64-molecule snapshots of liquid water, which have previously served in training a recent MLP¹¹. They were originally prepared using a three-step process. Bulk liquid systems of 64 water molecules were first equilibrated at high temperatures and densities between 0.7 and 1.2 g/mL. The resulting (de-correlated) configurations were then quenched using a steepest decent optimization. Finally, the 1000 most structurally diverse structures were extracted from all the collected liquid configurations using a farthest point sampling algorithm. There are two reasons why the 1000

configurations are representative of liquid water. First, they were constructed in order to cover a large part of the configurational space of possible atomic environments in liquid water. Second, the MLP trained using these structures reproduces many properties of water very well, including the density isobar and radial distribution functions at ambient pressure¹¹, which means that the training set contains the necessary information for describing liquid water at ambient pressure in a data-driven manner.

Direct comparison of the local environments. We employ the smooth overlap of atomic positions (SOAP)¹⁴ local descriptors to represent the atomic environments (i.e., the displacements of all the neighbors within a cutoff radius r_c around the central atom). More details regarding the representations are provided in the “Methods” section. For each structure, we then compute its global descriptors by taking the average of the local ones of all the atomic environments in that structure¹⁵. As the global descriptors are high-dimensional, we use principal component analysis (PCA) to build a two-dimensional embedding to visualize the relative difference (i.e., distances) between the structures. Essentially, a PCA map is the linear projection that best preserves the variances of the high-dimensional Cartesian distances of the dataset. Because only linear operations are involved throughout, the local and the global descriptors can be meaningfully projected onto the same PCA map.

We use these methodologies to analyse the 54 ice phases and the 1000 snapshots of the liquid. Figure 1a shows the PCA map of the global descriptors of all the structures: similar structures stay close on this map, while distinct ones are farther apart. The horizontal principal axis is strongly correlated with density, suggesting that density variance is a dominant feature of the

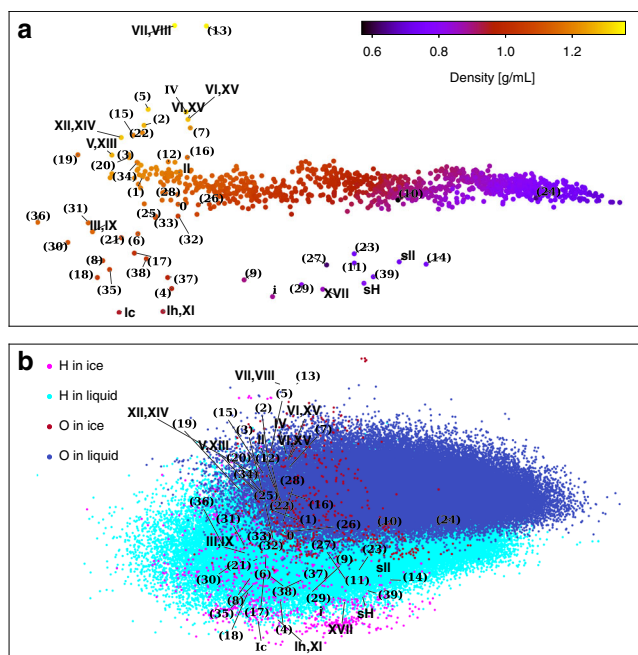


Fig. 1 PCA maps for the 54 ice phases and the 1000 liquid water configurations. The geometries of the ice phases have been optimized by HSE-3c. If known, each ice structure is labeled by the name of the proton ordered and disordered phase, and otherwise by a number. **a** Each dot indicates a structure, which is described by global descriptors. **b** Each small dot indicates the projection of a oxygen or hydrogen-centered local environment.

dataset. The ice phases and the liquid structures are separated on the map, while ice structures that are commonly considered to be similar (e.g., ice Ic and Ih) stay close together. The distinction between ices and liquids in the PCA map is to be expected considering the absence of long-range order in the latter. More structure–property relationships may be extracted from this PCA map, and we provide an interactive explorer of the ice and liquid water dataset in the Supplementary Data 1 (“Supplementary Data 1”), which runs in web browser and is made using Chemscope¹⁶.

However, the key message here is the comparison between the PCA map of the whole structures (Fig. 1a) and the one of individual atomic environments (Fig. 1b). The latter is the projection of local descriptors of all the atomic environments ($r_c = 6 \text{ \AA}$), onto the same PCA map as in Fig. 1a. Contrary to the clear distinction between ice and liquid shown in Fig. 1a, the local environments in liquids and ices are similar on Fig. 1b. Furthermore, the crystalline nature of the ices leads to comparatively few distinct atomic environments, and they are almost completely covered by the continuum realized in liquid water. In other words, liquid water in simulations prototypes all atomic environments pertinent to the 54 ice phases.

Predictions of the MLP on ices. The PCA maps provide a simple and general way of comparing and understanding the structural similarities, but choice of the representations and the linear dimensionality reduction inevitably lead to information loss and distortion. As an alternative similarity comparison, we explore how well a MLP¹¹ that is only trained on reference calculations for liquid water configurations describes diverse crystalline phases.

This MLP is based on revPBE0^{17,18} hybrid functional density functional theory (DFT) calculations with the semiclassical D3 dispersion correction¹⁹. The training set contains 1593 configurations: the first 1000 are classical configurations as described above, the remaining 593 originate from path-integral molecular dynamics (PIMD) simulations at ambient conditions. We omit those PIMD configurations in the PCA analysis above because nuclear quantum effects²⁰ complicate the direct comparisons with

classical water, and also because those configurations had a very minor effect on the training of the MLP in our previous work¹¹. The MLP uses an artificial neural network constructed according to the framework of Behler and Parrinello²¹. The total energy of the system is expressed as the sum of the individual contributions from the atom-centered environments of radius 6 \AA .

Crucially, the success of the MLP hinges on the notion of “nearsightedness”: energy and forces associated with a central atom are largely determined by its neighbors, and the long-range interactions can be approximated in a mean-field manner without explicitly considering the far-away atoms. This notion underlies many atomic and molecular force-field as well as most common MLPs²². From this point of view, to capture the energetics and dynamics of a phase of water, the key is to predict the local atom-centered contributions to the total energy and forces (Generalizations to systems with nontrivial long-range interactions like strong ionic liquids have to be done carefully as those might pose a challenge to the “nearsightedness”). In practice, this means the training set of the MLP needs to contain the essential local atomic environments of the particular phase. Following this logic, we postulate that, if the liquid water contains all the local environments of the ice phases, an MLP trained exclusively on snapshots of liquid water should also be able to describe the ice phases.

To verify our hypothesis, we benchmark the performance of the MLP against reference DFT calculations and experimental results (see more details in “Methods” section). The DFT references comprise (i) revPBE0-D3 using CP2K with similar numerical settings as the calculations performed to generate the training reference of the MLP, (ii) revPBE0-D3 using VASP and converged numerical settings, and (iii) the screened exchange hybrid HSE-3c as implemented in CRYSTAL17. Note that the multiple DFT references also provide an estimate on the intrinsic errors in these DFT calculations due to the choices on the specific hybrid functionals, numerical settings and the use of different software packages.

In Fig. 2a we show the comparison between the lattice energies and the densities (Fig. 2b) of the 54 ice phases. Note that for the

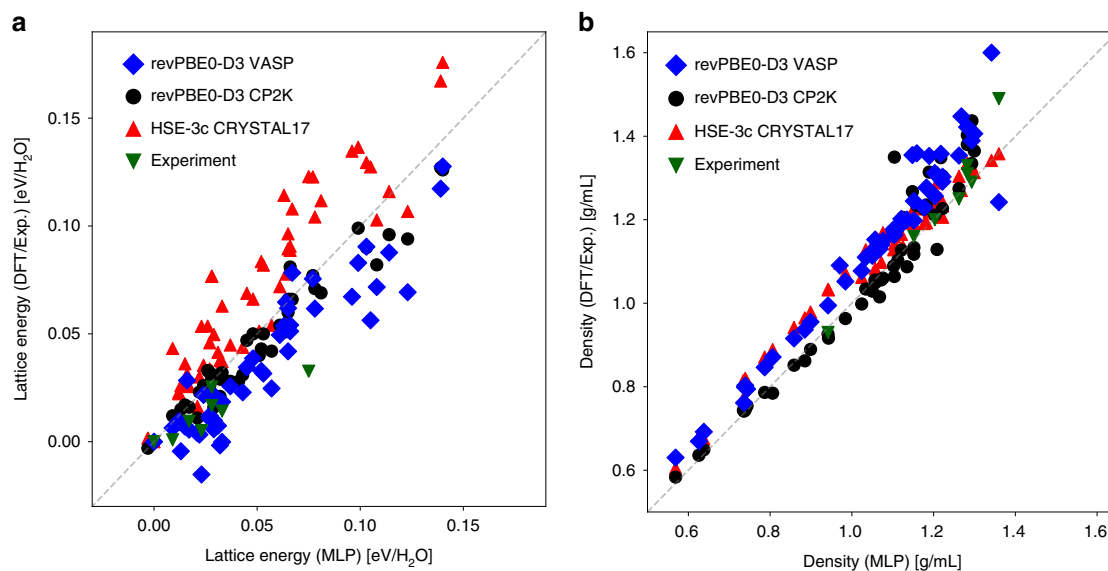


Fig. 2 A comparison between the lattice and energy of the 54 ice phases. The x-axes show the values computed using the MLP based on revPBE0-D3. **a** Results for lattice energy, and **b** shows the values for the densities of the ice phases. In both panels, the y-axes show the results from DFT calculations using the revPBE0-D3 functional employing VASP, the revPBE0-D3 functional employing CP2K, and the HSE-3c functional employing CRYSTAL17. The experimental references are taken from ref. 59: the densities were measured at between 0 and 90 K and ambient pressure, and the experimental lattice energies were extrapolated to 0 K, and zero-point energies were removed. Source data are provided as a Source Data file.

lattice energies different theories or experiments have different baselines, so for each set of calculation or measurement we use the energies the ice Ih/XI structure as a reference. In general, we find an excellent agreement between the MLP predictions and all four references, particularly for the densities. In particular, the differences between the MLP results and the ab initio references are on par with the DFT differences introduced by the details of the first principles calculations. For instance, the Pearson correlation coefficient R between the MLP densities and the CP2K values is 0.96 and the root mean square error (RMSE) is 0.06 g/mL, and the corresponding metrics between CP2K and VASP are 0.96 and 0.08 g/mL. For the lattice energies, the R and RMSE between the MLP and the CP2K predictions are 0.95 and 9 meV/H₂O, respectively, compared with 0.91 and 15 meV/H₂O between the CP2K and the VASP values.

The curvature of the potential energy surface around a local minimum relates to the harmonic frequencies at which the atoms in a crystal vibrate. To investigate the performance of the MLP for this quantity, we have calculated the phonon frequencies for the considered ice structures using the MLP as well as for a subset of the ice structures using revPBE0-D3 DFT calculations with both VASP and CP2K. Figure 3 provides a detailed comparison of the phonon density of states (DOS) for three representative structures with distinct densities. Figure 3a is for the ice VI phase and its proton-ordered counterpart, XV. Figure 3b corresponds to a structure that represents the ice Ih phase as well as the proton-ordered XI. Figure 3c shows the results for a low-density hypothetical phase. All the DOS show excellent agreement in both the low-energy region, corresponding to long-range dispersive crystal vibrations, and the high-energy region, corresponding to localized molecular vibrations. The small shift of low frequency phonons may be induced by the lack of long-range interactions of the MLP. A comparison across all structures can be found in the Methods, and it provides remarkable agreement between the MLP and the first principles methods across all structures in the entire energy range of vibrations.

Discussion

The similarities between the local environments in solid and liquid phases shed light on the structure of liquid water. There have been many efforts to develop a molecular understanding of water, in terms of orientational and translational order⁴, hydrogen bond networks⁹ and spontaneously forming dendritic voids¹⁰. Our approach of using local environments observed in ice as landmark points is a new way of interpreting liquid water as a mixture of ice structures. It is worth noting that, almost as the other side of the same coin, the idea of inferring the long-range ice order from good water potentials has been discussed by Rice and coworkers in the 1980s^{23,24}.

On the flip side, the similarity also suggests that the liquid and ice structures are distinct in Fig. 1a not because of the difference in local environments, but due to the presence of long-range order. The conclusion that liquid water contains all the ice environments explains why the MLP trained on liquid describes the ice phases well. This generalization is not specific to this MLP. Indeed, many water models, such as the coarse grained mW²⁵ model, the empirical water models SPC²⁶ and the TIPnP series^{27,28}, the polarizable AMOEBA²⁹, and the MB-pol water potential^{30,31} that are fitted to ab initio reference, qualitatively reproduce large parts of the phase diagram^{23,32,33}, despite having been developed primarily to simulate the liquid phase. In particular, MB-pol correctly reproduce the properties of water from the gas to the condensed phases³¹. For the MLP used here, the melting point of ice Ih and the relative stabilities between Ih and Ic have been computed¹¹.

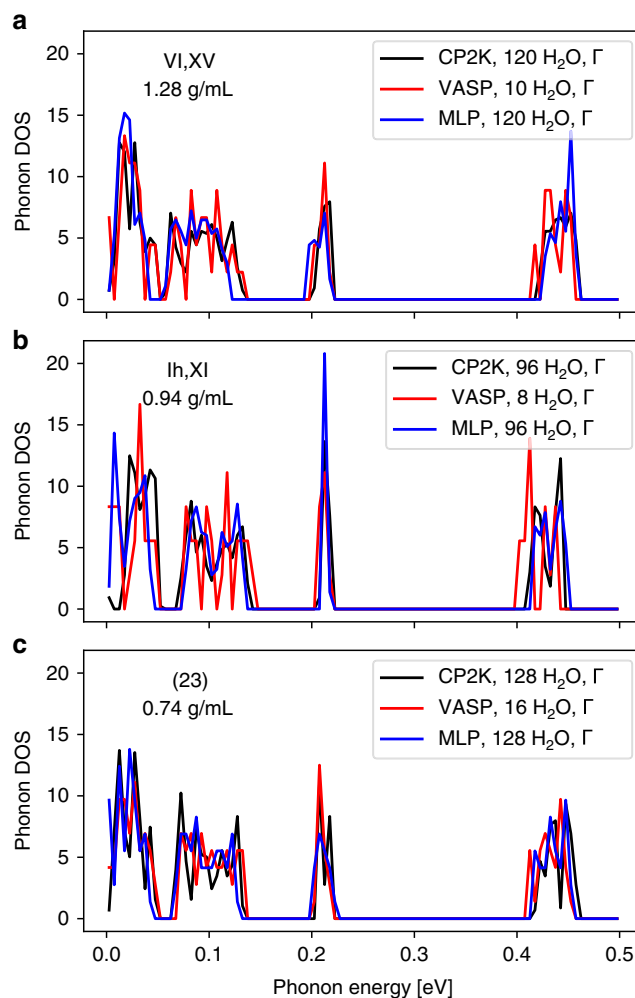


Fig. 3 A comparison between the phonon density of states (DOS) for three ice structures. **a–c** Results for three indicated ice phases with distinct densities. The three sets of DOS in each panel were computed using revPBE0-D3 DFT calculations employing VASP, revPBE0-D3 DFT calculations employing CP2K, and the MLP, all at the Gamma (Γ) point. The densities predicted at the MLP level are indicated. Source data are provided as a Source Data file.

In addition, the general and agnostic comparison of the local environments can be easily extended to study amorphous ice³⁴, interfaces, and water under confinement³⁵. It is also interesting to investigate how nuclear quantum fluctuations^{35–38} influence the distribution of the atomic environments in various phases.

Furthermore, our results illustrate the immense promise of employing MLPs in materials modeling. For instance, the MLP used here provides an accurate description of the static and vibrational properties of ice phases at a fraction of the cost of the corresponding DFT calculations. For example, the DFT calculation for the Γ -point phonons of a 4-molecule structure takes 264 CPU hours, and that for a 52-molecule structure takes 16,000 CPU hours, compared with just a few minutes for both on a laptop using the MLP. Besides, the fact that one can only train a MLP on one liquid phase and apply the potential to other phases evince the extend of its “extrapolability”, which significantly facilitates the constructions of the potentials. It is worth mentioning that, the MLP for water is stable enough to run MD and PIMD for all the phases. Despite the success of the MLP trained on the liquid water, we want to caution that for systems where long-range interactions are important, such as strongly ionic

systems, the short-range nature of the MLP may pose a limit on the overall accuracy. For a given system, it is also not certain if its melt contains sufficient local ordering to reproduce the atomic environments found in the solid phases.

Last but not least, using MLPs as tools for comparing atomic environments offers a new approach of analyzing complex atomic systems in an agnostic and general manner. Our analysis is, of course, not restricted to the chosen DFT level and can be extended to incorporate new developments in the field of ab initio methods^{39–42}.

To summarize, we compare the local environments in various crystalline ice phases and liquid water, using two ML-based approaches. We demonstrate that liquid water contains all the local atomic environments in diverse ice phases. Our conclusion provides a new and fundamental perspective on the understanding of liquid water and ices, and guides future efforts for modeling water.

Methods

SOAP representations for atomic environments. Numerous representations of atomic environments have been developed^{14,43–45}, and here we use the SOAP representation¹⁴. SOAP encodes the local environment \mathcal{X} around a central atom using a smooth atomic density function

$$\rho_{\mathcal{X}}^{\alpha}(\mathbf{r}) = \sum_{|r_i| < r_c} \exp\left(-\frac{|\mathbf{r} - \mathbf{r}_i|^2}{2\sigma^2}\right), \quad (1)$$

by summing over Gaussians centered on each atom i of species α (here hydrogen or oxygen) within a given cutoff distance r_c of the central atom. The density $\rho_{\mathcal{X}}^{\alpha}(\mathbf{r})$ is then expanded in a basis of orthonormal radial functions $g(|r|)$ and spherical harmonics $Y_{lm}(\hat{r})$ as

$$\rho_{\mathcal{X}}^{\alpha}(\mathbf{r}) = \sum_{nlm} c_{nlm}^{\alpha} g_n(|r|) Y_{lm}(\hat{r}). \quad (2)$$

Finally, the power spectrum is taken as

$$k_{nn'l}^{\alpha}(\mathcal{X}) = \sqrt{\frac{8}{2l+1}} \sum_m (c_{nlm}^{\alpha})^* c_{n'l'm}^{\alpha}, \quad (3)$$

which characterizes \mathcal{X} in a translation, permutation, and rotation invariant form^{14,15}. The vector $\{k_{nn'l}^{\alpha}\}$ constructed in this way up to certain cutoffs l_{\max} and n_{\max} can then be used as the local fingerprint $\Psi(\mathcal{X})$. We set the radius of the atomic environment to be $r_c = 6$ Å, so that it includes the second hydration shell of water molecules, and expand the SOAP descriptor up to $l_{\max} = 6$ and $n_{\max} = 6$. PCA maps that were constructed using SOAP with different hyper-parameters are in the Supplementary Figs. 1 and 2. In practice, we use the Dscribe Python package for constructing descriptors⁴⁶, and the ASAP Python package for the subsequent analysis (provided as Supplementary Software 1).

Choice of the ice configurations. The initial 57 structures are based on an extensive survey of ice¹³ that generated 15,859 configurations, by exploiting the isomorphism between ice, experimentally known zeolites and theoretically-enumerated four-connected SiO₂ networks. The resulting ice-like configurations were subsequently locally relaxed, before a generalized convex hull construction¹² was employed to screen for the ice structures that may be stable under certain thermodynamic conditions. These structures include the experimentally known phases of ice except for ice IV, which we then add back into the selection. Figure 4 shows the PCA map of the locations of the selected phases. Notably, many ice

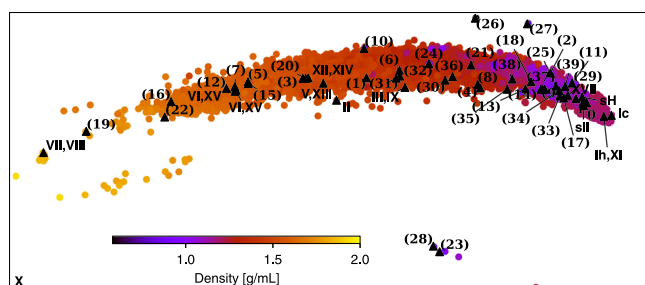


Fig. 4 PCA Map for 15,859 ice phases. The phases selected for the current study, as well as the phase X (outside the map), are marked on the map.

phases come in pairs of a low-temperature proton-ordered form and a higher-temperature proton-disordered form: Ih and XI, III and IX, V and XIII, VI and XV, VII and VIII, and XII and XIV. In this work, we focus on the particular proton-ordered realizations of these phases made available with ref. ¹³.

Initial geometry optimization of the ice structures using HSE-3c. In ref. ¹³ the ice structures were optimized at the PBE DFT level of theory using a coarse k -point grid and plane wave basis, trading accuracy for computational efficiency. For this study, we have therefore performed well-converged geometry optimization for the structures, by running a few cycles of local geometry optimizations followed by identifying and imposing crystal space group symmetries. These local optimizations are performed with the screened exchange hybrid functional HSE-3c⁴⁷ using tight optimization thresholds as implemented in CRYSTAL17^{48,49}. The Brillouin zone is sampled with a Γ -centered Monkhorst-Pack grid that has been converged individually for every system to yield a lattice energy accuracy well below 1 meV. HSE-3c has been shown to yield excellent molecular and intermolecular geometries as well as good noncovalent interaction energies^{50–52}, and in particular, suitable for water and ices. As all ice structures considered here were variable-cell geometry optimized at zero pressure, we were able to directly compare them in the subsequent analysis. Note also that the proton arrangements are typically “locked in” during optimizations, so the proton (dis) ordering of each optimized ice structure is determined by the initial configuration and does not necessarily represent the ground state. These relaxed geometries of the 54 ice phases, together with their distributions of the oxygen–oxygen distances and the oxygen–oxygen–oxygen angles are included in the Supplementary Figs. 3 and 4.

Geometry optimization using vasp. The geometries of all ice structures were further refined with revPBE0-D3 using the vasp package^{53,54}. The equilibrium volumes critically depend on the energy cutoff, and we used a relatively high value of 1200 eV to obtain converged results. The k -point sampling grids were the same as those determined for the HSE-3c calculations described above. Structures were constrained to their initial symmetries throughout the geometry optimization, and convergence was achieved with forces below 10^{-3} eV/Å and stress components below 10^{-2} GPa.

Geometry optimization using cp2k. We computed the equilibrium densities and the lattice energies of the ice structures using the cp2k code⁵⁵ with the revPBE0-D3. The computational details of the calculations are identical with refs. ^{11,56}, although the planewave cutoff energy was increased to 800 eV, to obtain smooth volume-energy curves. Despite a considerable amount of effort, the geometry optimization for three structures did not converge to reasonable values, so these calculations were discarded.

Phonon calculations. For the phonon calculations using the MLP, we first computed the Hessian matrix for the 54 geometry-optimized ice phases using finite displacements of 0.01 Å of each atom from its equilibrium position along x , y , and z axes. Then the Hessian matrix was diagonalized to obtain the phonon frequencies as the square root of the eigenvalues. We performed those phonon calculations for the ice systems in both their original cell taken from ref. ¹³, and in supercells of this original cell, obtained by repeating the original cell along all three crystallographic directions so that each dimension of the supercell is longer than 8 Å.

The phonon calculations using cp2k follow the same approach, and the DFT settings are identical to those for geometry optimization. Presumably due to numerical issues of the specific DFT setup that we used (e.g., cp2k only supports Γ -point sampling for hybrid functionals), a number of the ice phases contain imaginary phonons at the Γ -point, even after several rounds of geometry optimization. We discarded the CP2K phonon DOS for these phases, and only show the ones with real frequencies in Fig. 5.

The vasp phonon calculations were performed using the structures optimized with vasp with the same parameters described above. We used the finite displacement method⁵⁷ in conjunction with nondiagonal supercells⁵⁸, and commensurate k -point grids were used to sample the electronic Brillouin zones of the supercells. The Hessian of a given nondiagonal supercell was calculated by displacing each atom from its equilibrium position by 0.01 Å in symmetry-inequivalent directions and calculating the force constants by finite differences. The dynamical matrix for a given \mathbf{q} -point grid of the vibrational Brillouin was determined by combining the results from multiple nondiagonal supercell calculations as described in the ref. ⁵⁸. The resulting dynamical matrix was diagonalized to obtain the phonon frequencies and eigenvectors.

In all sets of phonon calculations, imaginary phonons appear in multiple structures at various \mathbf{q} -points in the Brillouin zone. This reflects the fact that the protons in many ice structures are disordered, and when we attempt to model them as periodic ordered structures using the unit cells from ref. ¹³, we are artificially constraining them to a saddle point of the potential energy surface rather than to a local minimum. In some of the structures, instabilities appear even at the Γ -point, and this is caused by the symmetrization step in preparing the structures, which again can place them at a saddle point. The imaginary phonons in this case break some of the imposed symmetries to lower the energy. These problems can be resolved by replicating the original simulation cell and re-relaxing the atomic positions of the supercell to allow for the appearance of disorder that lowers the overall energy, or by re-relaxing the primitive cell without imposing symmetry in

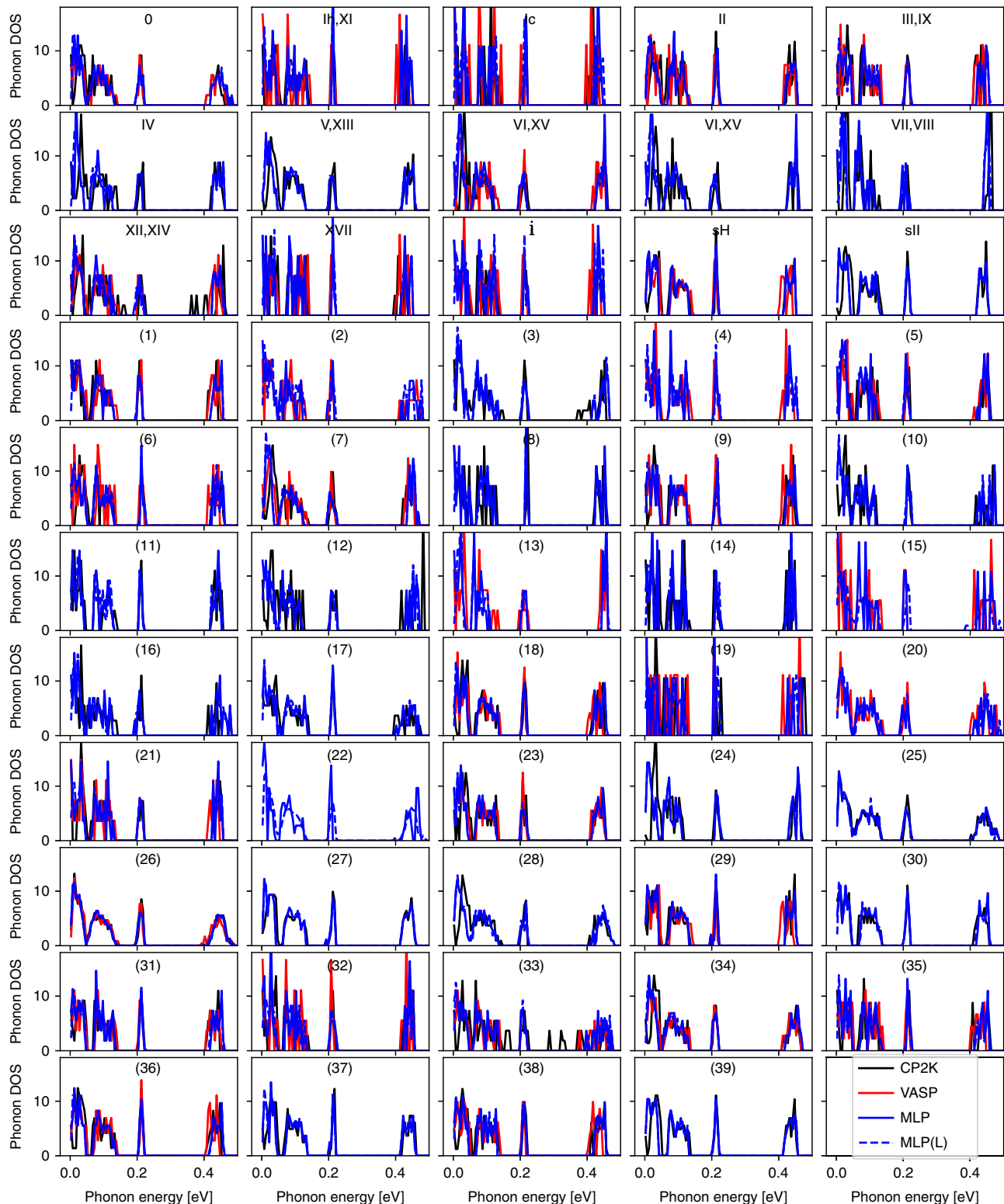


Fig. 5 A comparison between the phonon DOS at the Gamma point for 54 ice phases. The DOS were computed using revPBE0-D3 DFT calculations employing VASP, revPBE0-D3 DFT calculations employing CP2K, MLP using the original supercell, and MLP using replicated cell (MLP(L)). For the VASP and CP2K calculations, only those that exhibited no imaginary phonons are shown here. Source data are provided as a Source Data file.

the case of Γ -point phonons. This additional step is computationally trivial for the MLP calculations, but computationally extremely costly for the hybrid functional calculations using DFT. As a consequence, it is computationally prohibitive to accurately calculate the phonon densities of states for all ice structures at the DFT level, and we only consider a subset in Fig. 5.

Data availability

The 54 ice structures correspond to classical 0 K structures without external pressure, the CRYSTAL17, CP2K, and VASP input files, the Python notebook for analysis are provided in the Supplementary Data 1 (“Supplementary Data 1”), and available on <https://github.com/BingqingCheng/ice-in-water>. Source data are provided with this paper.

Code availability

We used the ASAP code for most of the analysis, which is available at: <https://github.com/BingqingCheng/ASAP>. The Python notebook for making Fig. 1 is included in the Supplementary Data 1 (“Supplementary Data 1”).

Received: 17 July 2020; Accepted: 23 October 2020;

Published online: 13 November 2020

References

- Salzmann, C. G. Advances in the experimental exploration of water’s phase diagram. *J. Chem. Phys.* **150**, 060901 (2019).
- Brini, E. et al. How water’s properties are encoded in its molecular structure and energies. *Chem. Rev.* **117**, 12385–12414 (2017).
- Marx, D., Tuckerman, M. E., Hutter, J. & Parrinello, M. The nature of the hydrated excess proton in water. *Nature* **397**, 601–604 (1999).
- Errington, J. R. & Debenedetti, P. G. Relationship between structural order and the anomalies of liquid water. *Nature* **409**, 318–321 (2001).
- Santra, B., DiStasio, R. A., Martelli, F. J. & Car, R. Local structure analysis in ab initio liquid water. *Mol. Phys.* **113**, 2829–2841 (2015).
- Santra, B. et al. Hydrogen bonds and van der Waals forces in ice at ambient and high pressures. *Phys. Rev. Lett.* **107**, 185701 (2011).
- Fitzner, M., Sosso, G. C., Cox, S. J. & Michaelides, A. Ice is born in low-mobility regions of supercooled liquid water. *Proc. Natl Acad. Sci.* **116**, 2009–2014 (2019).
- Del Rosso, L. et al. Cubic ice Ic without stacking defects obtained from ice xvii. *Nat. Mater.* **19**, 1–6 (2020).
- Ohmine, I. & Saito, S. Water dynamics: fluctuation, relaxation, and chemical reactions in hydrogen bond network rearrangement. *Acc. Chem. Res.* **32**, 741–749 (1999).
- Ansari, N., Laio, A. & Hassanali, A. Spontaneously forming dendritic voids in liquid water can host small polymers. *J. Phys. Chem. Lett.* **10**, 5585–5591 (2019).
- Cheng, B., Engel, E. A., Behler, J., Dellago, C. & Ceriotti, M. Ab initio thermodynamics of liquid and solid water. *Proc. Natl Acad. Sci.* **116**, 1110–1115 (2019).
- Anelli, A., Engel, E. A., Pickard, C. J. & Ceriotti, M. Generalized convex hull construction for materials discovery. *Phys. Rev. Mater.* **2**, 103804 (2018).
- Engel, E. A., Anelli, A., Ceriotti, M., Pickard, C. J. & Needs, R. J. Mapping uncharted territory in ice from zeolite networks to ice structures. *Nat. Commun.* **9**, 2173 (2018).
- Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
- De, S., Bartók, A. P., Csányi, G. & Ceriotti, M. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.* **18**, 13754–13769 (2016).
- Fraux, G., Cersonsky, R. & Ceriotti, M. Chemiscope: Interactive structure-property explorer for materials and molecules. *J. Open Source Softw.* **5**, 2117 (2020).
- Zhang, Y. & Yang, W. Comment on “generalized gradient approximation made simple”. *Phys. Rev. Lett.* **80**, 890–890 (1998).
- Goerigk, L. & Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **13**, 6670 (2011).
- Grimme, S., Hansen, A., Brandenburg, J. G. & Bannwarth, C. Dispersion-corrected mean-field electronic structure methods. *Chem. Rev.* **116**, 5105–5154 (2016).
- Markland, T. E. & Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem.* **2**, 1–14 (2018).
- Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
- Behler, J. Perspective: machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
- Morse, M. D. & Rice, S. A. Tests of effective pair potentials for water: predicted ice structures. *J. Chem. Phys.* **76**, 650–660 (1982).
- Townsend, M., Rice, S. A. & Morse, M. D. A test of an effective pair potential for liquid water. *J. Chem. Phys.* **79**, 2496–2498 (1983).
- Molinero, V. & Moore, E. B. Water modeled as an intermediate element between carbon and silicon. *J. Phys. Chem. B* **113**, 4008 (2009).
- Berendsen, H., Grigera, J. & Straatsma, T. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
- Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **112**, 8910 (2000).
- Laury, M. L., Wang, L.-P., Pande, V. S., Head-Gordon, T. & Ponder, J. W. Revised parameters for the amoeba polarizable atomic multipole water model. *J. Phys. Chem. B* **119**, 9423–9437 (2015).
- Babin, V., Medders, G. R. & Paesani, F. Development of a “first principles” water potential with flexible monomers. II: trimer potential energy surface, third virial coefficient, and small clusters. *J. Chem. Theory Comput.* **10**, 1599–1607 (2014).
- Reddy, S. K. et al. On the accuracy of the mb-pol many-body potential for water: interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys.* **145**, 194504 (2016).
- Yagasaki, T., Matsumoto, M. & Tanaka, H. Phase diagrams of TIP4P/2005, SPC/E, and TIP5P water at high pressure. *J. Phys. Chem. B* **122**, 7718 (2018).
- Dhabal, D., Chakravarty, C., Molinero, V. & Kashyap, H. K. Comparison of liquid-state anomalies in Stillinger-Weber models of water, silicon, and germanium. *J. Chem. Phys.* **145**, 214502 (2016).
- Limmer, D. T. & Chandler, D. Theory of amorphous ices. *Proc. Natl Acad. Sci.* **111**, 9413–9418 (2014).
- Rossi, M., Ceriotti, M. & Manolopoulos, D. E. Nuclear quantum effects in h+ and oh–diffusion along confined water wires. *J. Phys. Chem. Lett.* **7**, 3001–3007 (2016).
- Marx, D., Tuckerman, M. E. & Parrinello, M. Solvated excess protons in water: quantum effects on the hydration structure. *J. Phys.* **12**, A153 (2000).
- Li, X.-Z., Walker, B. & Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl Acad. Sci.* **108**, 6369–6373 (2011).
- Paesani, F., Iuchi, S. & Voth, G. A. Quantum effects in liquid water from an ab initio-based polarizable force field. *J. Chem. Phys.* **127**, 074506 (2007).
- Mardirossian, N. & Head-Gordon, M. b97m-v: a combinatorially optimized, range-separated hybrid, meta-gga density functional with vv10 nonlocal correlation. *J. Chem. Phys.* **144**, 214110 (2016).
- Wang, Y. et al. M06-sx screened-exchange density functional for chemistry and solid-state physics. *Proc. Natl Acad. Sci.* **117**, 2294–2301 (2020).
- Riera, M., Lambros, E., Nguyen, T. T., Götz, A. W. & Paesani, F. Low-order many-body interactions determine the local structure of liquid water. *Chem. Sci.* **10**, 8211–8218 (2019).
- Sharkas, K. et al. Self-interaction error overbinds water clusters but cancels in structural energy differences. *Proc. Natl Acad. Sci.* **117**, 11283–11288 (2020).
- Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
- Faber, F. A., Christensen, A. S., Huang, B. & Von Lilienfeld, O. A. Alchemical and structural distribution based representation for universal quantum machine learning. *J. Chem. Phys.* **148**, 241717 (2018).
- Sadeghi, A. et al. Metrics for measuring distances in configuration spaces. *J. Chem. Phys.* **139**, 184118 (2013).
- Himanan, L. et al. DScribe: library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **247**, 106949 (2020).
- Brandenburg, J. G., Caldeweyher, E. & Grimme, S. Screened exchange hybrid density functional for accurate and efficient structures and interaction energies. *Phys. Chem. Chem. Phys.* **18**, 15519–15523 (2016).
- Erba, A., Baima, J., Bush, I., Orlando, R. & Dovesi, R. Large-scale condensed matter dft simulations: performance and capabilities of the crystal code. *J. Chem. Theory Comput.* **13**, 5019–5027 (2017).
- Dovesi, R. et al. Quantum-mechanical condensed matter simulations with CRYSTAL. *WIREs Comput. Mol. Sci.* **8**, e1360 (2018).
- Caldeweyher, E. & Brandenburg, J. G. Simplified dft methods for consistent structures and energies of large systems. *J. Phys.* **30**, 213001 (2018).
- Rösel, S. et al. London dispersion enables the shortest intermolecular hydrocarbon h···h contact. *J. Am. Chem. Soc.* **139**, 7428–7431 (2017).
- Donà, L., Brandenburg, J. G. & Civalleri, B. Extending and assessing composite electronic structure methods to the solid state. *J. Chem. Phys.* **151**, 121101 (2019).
- Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
- Kresse, G. et al. Software vasp, vienna, 1999; g. kresse, j. furthmüller. *Phys. Rev. B* **54**, 169 (1996).
- Lippert, G., Hutter, J. & Parrinello, M. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theor. Chem. Acc.* **103**, 124 (1999).
- Marsalek, O. & Markland, T. E. Quantum dynamics and spectroscopy of ab initio liquid water: the interplay of nuclear and electronic quantum effects. *J. Phys. Chem. Lett.* **8**, 1545 (2017).
- Kunc, K. & Martin, R. M. Ab initio force constants of GaAs: a new approach to calculation of phonons and dielectric properties. *Phys. Rev. Lett.* **48**, 406–409 (1982).
- Lloyd-Williams, J. H. & Monserrat, B. Lattice dynamics and electron-phonon coupling calculations using nondiagonal supercells. *Phys. Rev. B* **92**, 184301 (2015).

59. Brandenburg, J. G., Maas, T. & Grimme, S. Benchmarking dft and semiempirical methods on structures and lattice energies for ten ice polymorphs. *J. Chem. Phys.* **142**, 124104 (2015).

Acknowledgements

We thank Alexandre Tkatchenko, and Francesco Paesani for critically reading the manuscript and giving helpful feedback. B.M. acknowledges support from the Gianna Angelopoulos Programme for Science, Technology, and Innovation, and from the Winton Programme for the Physics of Sustainability. B.C. acknowledges funding from the Swiss National Science Foundation (Project P2ELP2-184408). B.M. and B.C. acknowledge the resources provided by the Cambridge Tier-2 system operated by the University of Cambridge Research Computing Service (<http://www.hpc.cam.ac.uk>) funded by EPSRC Tier-2 capital grant EP/P020259/1. B.C. acknowledges allocation of CPU hours by CSCS under Project ID s957.

Author contributions

B.M., J.G.B., and B.C. designed research; B.M., J.G.B., and B.C. performed research; B.C. analyzed data; B.M., J.G.B., E.A.E., and B.C. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-19606-y>.

Correspondence and requests for materials should be addressed to B.C.

Peer review information *Nature Communications* thanks Davide Donadio and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020