Navigating at Will on the Water Phase Diagram

S. Pipolo, 1,∗ M. Salanne,2 G. Ferlat,1 S. Klotz,1 A.M. Saitta,1 and F. Pietrucci1,†

1Sorbonne Universités, UPMC Univ. Paris 06, CNRS UMR 7590, IRD UMR 206, MNHN, IMPMC, F-75005 Paris, France
2 Sorbonne Universités, UPMC Univ. Paris 06, CNRS, Laboratoire PHENIX, F-75005 Paris, France

(Dated: January 24, 2018)

Despite the simplicity of its molecular unit, water is a challenging system because of its uniquely rich polymorphism and predicted but yet unconfirmed features. Introducing a novel space of generalized coordinates that capture changes in the topology of the interatomic network, we are able to systematically track transitions among liquid, amorphous and crystalline forms throughout the whole phase diagram of water, including the nucleation of crystals above and below the melting point. Our approach, based on molecular dynamics and enhanced sampling / free energy calculation techniques, is not specific to water and could be applied to very different structural phase transitions, paving the way towards the prediction of kinetic routes connecting polymorphic structures in a range of materials.

PACS numbers:

Computational structure prediction methods [1, 2] have strongly contributed to the rapid increase of new predicted phases of materials with enhanced properties for applications (see, e.g., Ref. [3]). However, at present, no general approach has been developed for guiding experiments through the pathways connecting stable structures of condensed matter. Moreover, metastable phases are very often involved in phase transitions [4] and sometimes their kinetic stability is very high [5]. Thus, in order to recover the global minimum structure, one needs to find specific routes, by e.g. acting on pressure or temperature, in a way that is not at all trivial to guess [6].

A precise understanding of transition mechanisms and the corresponding kinetics is therefore the key to explain and control the behavior of matter. The case of water is emblematic because several experiments have disclosed connections between stable and metastable phases [7–10] and recently simulations have highlighted the importance of metastable states in understanding the mechanism of phase transitions and related transformations [11]. A classic example is the connection between the crystalline ice stable at ambient pressure (Ice I), and the low-density amorphous (LDA) and high-density amorphous (HDA) ices: by compressing Ice I up to 10 kbar at ≈ 80 K one obtains HDA instead of Ice VI,[12] which may be transformed into LDA by decompression of HDA at 130 K [10] or by heating recovered HDA at ambient pressure to beyond 130 K; [9] finally Ice I is recovered by heating up LDA. Similar connections between crystalline and amorphous ices are found in the high-pressure region of the water phase diagram where a very-high-density amorphous (VHDA) ice, plastic ices and crystalline structures with complex hydrogen-bond network (e.g. Ice VII) have been observed or predicted. [5, 13]

Molecular dynamics (MD), a simulation method that yields the atomic trajectories as a function of time at given thermodynamic conditions, is in principle able to track such transitions. The kinetic barriers are however generally too large to allow an efficient exploration of the configuration space within typical MD timescales. Hence, so far it has been necessary to introduce (i) simplistic interaction models [14] and/or (ii) seeding techniques [15]. Another approach consists in using enhanced sampling techniques that accelerate the occurrence of rare events by focusing on low-dimensional order parameters, also called collective variables (CV) [16]. Yet the CVs available to describe phase transitions are specifically designed for a given type of structural transformation [17–19], while no general CV scheme has been proven successful for a wide class of problems, in particular those involving amorphous systems. Recently, distance metrics developed for condensed matter [7, 20, 21, 23–25] have been proven to be successful in classifying structures in molecular or extended systems based on their atomic environment and/or interatomic network. Here we show that by combining enhanced sampling techniques with a novel CV based on a general metric we are able to define in an efficient way the topological space of transformations among liquid, amorphous, and crystalline forms of water. This allows exploring at will the phase diagram along pathways connecting minimum-energy structures with a single, general approach, capable to characterize mechanisms and energetics.

Our CV scheme relies on the concept of generalized distances between configurations of the system, as defined from relative atomic positions, and it only requires to postulate the initial and final states of the target transformation, without any assumption on the pathway. The transformation is represented in a two-dimensional space
\{s, z\} of path CVs \[26\], with \(s\) quantifying the progress of the transformation and \(z\) allowing to discriminate between different pathways and to represent transitions to states that are not the target ones. In such a CV scheme each configuration of the system is associated with a permutation invariant vector \([21]\) (PIV), built-up from interatomic Cartesian distances. The PIV is built starting from atom-type-specific ordered blocks, \(\mathbf{v}_{kk'}\), with elements

\[
\mathcal{v}_{kk'} = c_{kk'} \mathcal{S} \left( \frac{3}{\Omega_0} \sqrt{\frac{\Omega}{\Omega_0}} | \mathbf{r}_{kk} - \mathbf{r}_{kk'} | \right) .
\]

Here \(\mathbf{r}_{kk}\) is the position vector of the \(\beta\)-th atom of type \(k\) (oxygen or hydrogen), with \(\beta > \beta', k > k'\); \(c_{kk'}\) are coefficients that define the PIV variant (they are all equal to one in the original formulation); \(\Omega\) and \(\Omega_0\) are the volume of the simulation box and a reference volume (details below) respectively; \(\mathcal{S}\) is a switching function monotonically decreasing from one to zero as \(| \mathbf{r}_{kk} - \mathbf{r}_{kk'} |\) increases (see Ref. [27] for more details). In order to define the PIV, first the \(\mathcal{v}_{kk'}\) elements of each \(\mathbf{v}_{kk'}\) block are sorted in ascending ordered, then the different blocks are simply joined together resulting in a PIV of \(N_{\text{atoms}}(N_{\text{atoms}}-1)/2\) components, that we indicate with \(\mathbf{v}_s\). The sorting operation within each block introduces invariance upon permutation of identical atoms. The volume scaling factor (absent in Ref. [21]) was found to be important to avoid violent fluctuations of the cell parameters during meta-dynamics.

Distances between generic configurations \(X\) and \(Y\) are computed as squared Euclidean distances between the corresponding PIVs \(D_{XY} = \sum_{\alpha} (V_{Y\alpha} - V_{X\alpha})^2\), and used to map each configuration of the system \((X)\) into a point \((s_X, z_X)\) in the 2D space defined by the path CVs \([26]\), built starting from only two reference configurations \(A\) and \(B\) representing the initial and final state of the transformation:

\[
s_X = \frac{1 \cdot e^{-\lambda D_{AX}} + 2 \cdot e^{-\lambda D_{BX}}}{e^{-\lambda D_{AX}} + e^{-\lambda D_{BX}}} \quad (2)
\]

\[
z_X = -\lambda^{-1} \log \left( e^{-\lambda D_{AX}} + e^{-\lambda D_{BX}} \right) . \quad (3)
\]

The two coordinates track the progress from \(A\) to \(B\) and the distance from \(A\) and \(B\), respectively. We remark that this formulation does not contain any guess about the mechanism of the transformation, and that the freedom granted by the \(z\) coordinate allows to explore also the formation of unexpected metastable structures different from \(A\) and \(B\) (e.g., Ice VII-P in Figure 4-c). There is some freedom in the choice of the parameter \(\lambda\): in this work we adopted \(\lambda \approx 2.3/D_{AB}\), conveniently localizing the free energy basins of reference states \(A\) and \(B\) around \(s \approx 1.1\) and \(s \approx 1.9\), respectively, and leading to smooth transformation pathways and landscapes. A much larger \(\lambda\) would produce very irregular and discontinuous pathways, while a much smaller one would hamper the resolution of different phases. We used metadynamics \([28]\) for the discovery of continuum pathways between locally stable configurations, and umbrella sampling \([29]\) for the reconstruction of precise free energy landscapes. Computational details are given in Ref. [27].

In Figure 1-a we draw the pathways we followed on the phase diagram of a realistic model of water \([1]\), navigating within and across free-energy basins using standard MD and enhanced sampling techniques, respectively. Figure 1-b shows a two-dimensional map (see also Ref. [7]) of distances between the visited crystalline and amorphous structures: the metric employed to define the CVs is able to scatter the different phases in a way that recalls the topology of the phase diagram, representing kinetically connected phases as neighbors, and kinetically disconnected ones as far apart.

As a starting point we analyze the crystallization of Ice I both from the liquid and the LDA phases at \(P = 1\) bar and over a range of temperatures around the melting point \((T_m \approx 250\, \text{K}) for the adopted interatomic potential \([1]\)). The initial configurations have been respectively obtained by cooling down an equilibrium liquid phase from \(T = 300\, \text{K}\) and heating up a LDA structure from \(T = 100\, \text{K}\) \([27]\). Both crystallization transitions have been achieved multiple times in metadynamics simulations at \(T = 240\, \text{K}\) and \(T = 260\, \text{K}\) and they are all characterized by (i) a nucleation mechanism that we show in Figure 2 for the LDA–Ice I transformation at \(T = 240\, \text{K}\), (ii) the formation of a crystal nucleus of cubic symmetry (Ice Ic), and (iii) a final state with either a perfect cubic symmetry or made up of layers of cubic Ic and hexagonal Ili ice (see last snapshot in Figure 2-a). This last feature is in agreement with experimental findings \([31]\), and our results show that the formation of stacking disordered Ice I may also proceed via the merging of Ic-nuclei as well as via (i) random growth of Ic and Ili layers \([32]\) and (ii) direct formation of Ic-Ili nuclei \([19]\). In Figure 2-b we display the free energy profiles for the
FIG. 2: a, Sequence of snapshots ($\alpha, \beta, \gamma, \delta, \epsilon$) of the umbrella sampling simulation describing the progressive crystallization of LDA water into Ice I. Molecules in icelike environment, with an average local tetrahedral bond order parameter [17] of oxygen higher than 0.7, are shown in blue (in the box only). Snapshot $\gamma$ shows the Ice I nucleus at its critical size (saddle point in the free-energy profile). Snapshot $\epsilon$ shows stacking disordered ice I, with Ic cubic regions separated by Ih hexagonal layers (green lines).
b, Free-energy profile along the LDA-Ice I transformation pathway ($T = 240$ K, $P = 1$ bar) obtained via the weighted histogram analysis method applied to umbrella sampling trajectories. c, Comparison of the free-energy profiles, projected along the $s$ coordinate, for the crystallization transitions LDA – Ice I at $T = 240$ K and $P = 1$ bar and Liquid – Ice I at $T = 260$ K and $P = 1$ bar. The free energy minima of each simulation are arbitrarily set to zero ($N$ is the number of water molecules, here 800).

liquid–Ice I and LDA–Ice I transformations, respectively above and below the melting point. The calculated relative stabilities of the various phases agree with the phase diagram of the water model, and the order of magnitude of the free-energy profiles is consistent with free-energy differences calculated in previous works [33]. We remark that crystallizing the liquid above the melting temperature, in the bulk, without any seeds and with a very realistic water model, shows that our approach allows to perform very challenging transformations even in unfavorable conditions, reaching metastable states (here Ice I) starting from the global minimum.

In order to carry on our continuous journey towards high pressure, we follow the experimental routes by cooling down Ice I to $T = 100$ K, and then compressing it at $P = 10$ kbar using standard MD: in the neighborhood of this point of the phase diagram we explore (with enhanced sampling simulations) the transformation to HDA as a function of temperature and pressure. We find that the free-energy barrier decreases when temperature and pressure increase, as we show in Figure 3. Furthermore we note that although the free energy of Ice I is higher than the one of HDA already at $P = 10$ kbar and $T = 100$ K (which is expected since Ice I is not the stable phase in this region of the phase diagram), the free-energy barrier is nonzero. This result is in accordance with a common interpretation of the Ice I–HDA transformation [9, 34] that can be seen as an extrapolation to low temperatures and high pressures of the Ice I–Liquid

FIG. 3: a, Free-energy profiles projected along the path collective variable $s$ for the Ice I – HDA transformation at different thermodynamic conditions: $T = 100$ K and $P = 10$ kbar (black), $T = 120$ K and $P = 10$ kbar (red), $T = 100$ K and $P = 12$ kbar (blue). The free energy of HDA is arbitrarily set to zero. The inset shows the position of the three transformations in the phase diagram: at $T = 100$ K and $P = 10$ kbar Ice I is metastable (see Fig. 1-a) and separated from HDA by a barrier. b, Sequence of snapshots ($\alpha, \beta, \gamma$) along the amorphization process at $T = 100$ K and $P = 10$ kbar. They are taken from the umbrella sampling trajectories and their position along the $s$ coordinate is shown in panel a.
coexistence line. Even if this description is mostly qualitative, it certainly invigorates the idea that low-density crystalline ice is unstable with respect to denser disordered forms at high pressures, and that such instability occurs at higher pressures as temperature is decreased.

The HDA phase we obtain is then (i) decompressed at ambient pressure and transformed into LDA to close the loop of transitions at low pressure, (ii) compressed at $P = 12$ kbar and transformed into VHDA. Free-energy and density profiles for HDA-LDA and HDA-VHDA transformations are provided [27]. The HDA-VHDA transformation connects the low-pressure and high-pressure regions that we explore in this work. The VHDA phase is compressed and heated until it reaches $P = 50$ kbar and $T = 300$ K, and at this point we address its crystallization to Ice VII. While simulating the VHDA–Ice VII transformation via metadynamics we observe that the system visits two additional metastable configurations. This result demonstrates that our method does not constrain the system to sample configurations along a simple path connecting the two reference structures, but rather allows it to follow complex mechanisms and discover new free energy basins. The metastable structures differ markedly from the other phases, as shown in Figure 4 where we compare representative snapshots and the oxygen-oxygen radial distribution functions. The first metastable phase is identified as the plastic Ice VII-P, which had already been proposed by Himoto et al. [5]. Oxygen atoms are arranged in a rather ordered crystalline network, whereas the hydrogen bond network changes dynamically: the correlation decay time of molecular dipoles is more than one order of magnitude shorter than that of Ice VII at the same thermodynamic conditions. The second metastable phase (labeled here “Ice Y”) is characterized by a tetragonal oxygen lattice and stacked layers of hydrogen-bond networks. We leave a detailed investigation of this phase for future works (the atomic coordinates are provided [27]). The flexibility of our method in representing transformations among several states in a two-dimensional CV space is illustrated in the free energy landscape connecting the three crystalline phases (Ice VII, Ice VII-P and Ice Y) shown on Figure 4-c.

In conclusion, we propose a versatile method allowing the efficient simulation of phase transitions in condensed matter. We illustrated the approach by tackling the difficult problem of poly(a)morphism in water, including kinetically challenging amorphous-to-crystalline and liquid-to-crystalline transitions. In particular, we simulated transformations between the LDA, HDA, liquid and Ice I phases in the low-pressure region of the phase diagram and among VHDA and Ice VII in the high-pressure region. In both cases, important mechanistic informations could be extracted from the simulations, highlighting the role of metastable structures during phase transitions. Thanks to a very general formulation, the proposed approach is not restricted to specific transitions of a single material. Analysis of 50 experimental polymorphs belonging to 13 different materials (covalent, metallic, ionic, and molecular) indicates that the PIV-based metric is able to resolve all physically-distinct structures [27], suggesting a broad applicability of our simulation approach. The ability to discover transformation mechanisms, simulate nucleation events, and reconstruct free energy landscapes and kinetic barriers – all in a robust and systematic way – is highly complementary to structure prediction and materials discovery efforts.

Acknowledgments

FP thanks Grégoire A. Gallet for help with the implementation of a preliminary version of the PIV path collective variables. This work was funded within the Investissements d’Avenir program under reference ANR-11-IDEX-0004-02, within the framework of the cluster of excellence MATériaux Interfaces Surfaces Environnement (MATISSE) led by Sorbonne Universités. We acknowledge calculations performed on the Gnome cluster at UPMC, on the Occigen cluster at CINES, Montpellier, France, and on the Ada cluster at IDRIS, Orsay, France, under GENCI allocations 2015-091387 and 2016-091387.

[1] C. J. Pickard and R. Needs, J. Phys. Condens. Matter 23, 053201 (2011).
[2] C. W. Glass, A. R. Oganov, and N. Hansen, Comput. Phys. Commun. 175, 713 (2006).
[3] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, and R. Q. Snurr, Nat. Chem. 4, 83 (2012).
[4] R. E. Schreiber, L. Houben, S. G. Wolf, G. Leitus, Z.-L. Lang, J. J. Carbó, J. M. Poblet, and R. Neumann, Nat. Chem. 9, 369 (2016).
[5] K. Himoto, M. Matsumoto, and H. Tanaka, Phys. Chem. Chem. Phys. 16, 5081 (2014).
[6] A. V. Radha, L. Lander, G. Rousse, J. M. Tarascon, and A. Navrotsky, J. Mater. Chem. A 3, 2601 (2015).
[7] T. Bartels-Rausch, V. Bergeron, J. H. Cartwright, R. Escrivan, J. L. Finney, H. Grothe, P. J. Gutiérrez, J. Haspala, W. F. Kuhs, J. B. Pettersson, et al., Rev. Mod. Phys. 84, 885 (2012).
[8] J. C. Palmer, F. Martelli, Y. Liu, R. Car, A. Z. Panagiotopoulos, and P. G. Debenedetti, Nature 510, 385 (2014).
[9] O. Mishima and H. E. Stanley, Nature 396, 329 (1998).
[10] S. Klotz, T. Strässle, R. Nelmes, J. Loveday, G. Hamel, G. Rousse, B. Canny, J. Chervin, and A. Saitta, Phys. Rev. Lett. 94, 025506 (2005).
[11] J. Russo, F. Romano, and H. Tanaka, Nat. Mater. 13,
FIG. 4: a, Snapshots representing the oxygen atoms of the system in four different configurations: VHDA, Ice VII-P (plastic phase), Ice VII and Ice Y. All these structures are visited by the system while simulating the VHDA – Ice VII transformation at \( T = 300 \) K and \( P = 50 \) kbar via metadynamics. b, Oxygen–oxygen radial distribution functions as a function of the number of metadynamics simulation steps, for the stable configurations shown in panel a. The color code is consistent with panel a. c, Free energy landscape at \( T = 300 \) K and \( P = 50 \) kbar in the \( \{ s, z \} \) space obtained from umbrella sampling simulations. Equilibrated Ice VII and Ice Y are used to define the path collective variables. The simulation box includes \( N = 360 \) water molecules.

[29] G. M. Torrie and J. P. Valleau, J. Comput. Phys. 23, 187 (1977).
[1] J. L. Abascal and C. Vega, J. Chem. Phys. 123, 234505 (2005).
[31] T. L. Malkin, B. J. Murray, A. V. Brukhno, J. Anwar, and C. G. Salzmann, Proc. Natl. Acad. Sci. U.S.A 109, 1041 (2012).
[32] T. L. Malkin, B. J. Murray, C. G. Salzmann, V. Molinero, S. J. Pickering, and T. F. Whale, Phys. Chem. Chem. Phys. 17, 60 (2015).
[33] C. Vega, E. Sanz, J. Abascal, and E. Noya, J. Phys. Condens. Matter 20, 153101 (2008).
[34] J.-Y. Chen and C.-S. Yoo, Proc. Natl. Acad. Sci. U.S.A 108, 7685 (2011).
[35] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak, J. Chem. Phys. 81, 3684 (1984).
[36] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. 126, 014101 (2007).
[37] B. Roux, Comput. Phys. Commun. 91, 275 (1995).
[38] H. J. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. 91, 43 (1995).
[39] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Comput. Phys. Commun. 185, 604 (2014).
[40] M. Mendelev, S. Han, D. Srolovitz, G. Ackland, D. Sun, and M. Asta, Philos. Mag. 83, 3977 (2003).
[41] F. H. Stillinger and T. A. Weber, Phys. Rev. B 31, 5262 (1985).
[42] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc 118, 11225 (1996).
[43] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B 87, 184115 (2013).
[44] J. Chiu, F. W. Starr, and N. Giovambattista, J. Chem. Phys. 139, 184504 (2013).
Supplemental Material

Computational details

In all enhanced sampling simulations the PIV is defined employing the following switching function of interatomic distances $r_{ij}$:

$$S(r_{ij}) = \frac{1 - (r_{ij}/r_0)^n}{1 - (r_{ij}/r_0)^m}. \quad (4)$$

The decay range of the switching function is defined in order to maximize the distance between the two target phases using the initial volume of the box for each simulation as the reference volume $\Omega_0$. The set of parameters $\{r_0, n, m\}$ for eq. (4) used in simulations of Figures 2, 3 and 4 are respectively $\{0.7, 4, 12\}$, $\{0.4, 6, 20\}$ and $\{0.5, 8, 18\}$ ($r_0$ values are in nm). The first set of parameters is also used for the illustrative map in Figure 1-b, that is constructed, starting from random 2D positions of the representative points, by minimizing the difference between PIV distances and map distances with a Monte Carlo procedure, until obtaining errors smaller than 6%. $c_{OO} = 1.0$ and $c_{HH} = 0.2$ (eq.1) are used to construct the PIVs for the crystallization of Ice I, in all other transformations $c_{HH} = 0$ since oxygen-oxygen terms are sufficient to represent the transformation pathways.

The potential used to model interactions between water molecules is TIP4P/2005 [1], with periodically-repeated boxes of $N = 800$ molecules (except for the 2D umbrella sampling in Figure 4 where $N = 360$). MD and enhanced sampling simulations were performed in the $NPT$ ensemble. Time constants for pressure [2] and temperature [3] coupling of 0.02 ps and 0.2 ps were respectively used. Due to the sizable computational cost of PIV construction, enhanced sampling simulations are on average 10 (50) times slower than standard MD for a PIV built-up with oxygen (oxygen and hydrogen) atoms.

Metadynamics simulations, of typical duration between 10 and 100 ns, allowed to overcome kinetic barriers and quickly explore many different transformations, discovering mechanisms as well as unexpected ice phases. For instance, we could repeatedly simulate the crystallization of liquid water or of LDA water at different conditions above and below the melting point. Typical width, height, and deposition stride of the repulsive Gaussians are $\sigma_2 = 0.02$, $\sigma_x = 0.4$, $h = 1.0$ kJ/mol, $\tau = 2.0$ ps. To obtain free energy landscapes of high statistical precision we exploited umbrella sampling simulations, seeded from the structures explored with metadynamics. Each umbrella sampling window has a typical duration of 20 ns, with the last 5 ns (corresponding typically to more than 1000 times the autocorrelation decay time of the variables) employed to reconstruct the free energy landscape using the weighted histogram analysis method [4]. In the landscapes of Figures 2 and 4 we restricted the sampling to the relevant regions of the landscape, i.e., those containing the transition pathways. Statistical errors on free energies, estimated by bootstrapping or by cutting each trajectory in 5 segments and taking standard deviations, are smaller than $10^{-3}k_B T/N$. Periodic boundary conditions are applied and the 3D particle-mesh-Ewald approach is used for electrostatics with a real-space cutoff of 0.6 nm (the same cutoff is used for van der Waals interactions). GROMACS 5.1.2 [5] and a modified version of the PLUMED 2 plugin [6] (soon available from www.plumed.org or upon request) were employed to perform $NPT$ MD and enhanced sampling simulations.

PIV distances between polymorphs of different types of materials

In this section we analyze PIV distances in a set of 50 experimental polymorphs of ionic, molecular, metallic, and covalent solids, including elements, binary, ternary and organic compounds. The aim is to test whether our methodology – based on a general formulation that is not designed specifically for water – might be applied to several different classes of materials. All crystal structures are taken from the Crystallography Open Database (www.crystallography.net). In Tables I-XIII we report PIV distances between polymorphs of iron, silicon, sodium, carbon, phosphorus, sulphur, SiC, SiO₂, RbCl, Fe₂O₃, MgSiO₃, benzene, and paracetamol. In all cases, to avoid system-dependent fine tuning, we included all atoms in the PIV definition and we tested the same set of three different switching functions of interatomic distances, with $n = 6$, $m = 12$ and $r_0 = 2.5$, 3.5, or 4.5 Å, respectively. In the tables, PIV distances are scaled by the square root of the number of atoms, according to the definition in Ref. [7], to facilitate their comparison across different materials. Polymorphs are presented in order of decreasing density. To provide a reference allowing to appreciate the good separation between different polymorphs, we performed 100 ps-long MD simulations at 300 K of a single polymorph of iron (2000 atoms, embedded atoms potential [8]), silicon (2304 atoms, Stillinger-Weber potential [9]), and benzene (108 molecules, OPLS-AA potential [10]), finding a maximum value of PIV distance between snapshots of each system (i.e., a thermal spread) equal to 0.016, 0.004, and 0.005,
respectively (with $r_0 = 3.5\, \text{Å}$). With similar MD simulations we obtained reference structures for liquid iron and silicon, at 2500 K, also reported in Tables I and II for comparison.

It has been pointed out that, from a mathematical point of view, sets of points can be artificially built displaying a same set of sorted distances [11]. Nevertheless, as shown in this work as well as in Ref. [7], the PIV-based metric is able to resolve, in all the crystallographic cases considered so far, the structures corresponding to physically-distinct forms of a material (e.g., different polymorphs or different amorphous forms), whereas it does not separate the structures corresponding to physically-equivalent realizations of a same form (e.g., independent configurations belonging to a liquid, or to a same amorphous form). Not only the former, but also the latter is a very convenient feature of the present approach, leading to a compact and well-defined free energy basin even for liquid phases and amorphous forms. In the same spirit, invariance under permutation of identical atoms conveniently overlaps physically-equivalent structures differing only by the labelling of atoms.

In figure S1 we show 2D maps of PIV distances [7], generated using values in Tables I, II, X, and XI with $r_0 = 3.5\, \text{Å}$. Note that the points size is larger than the maximum value of intra-polymorph PIV distances reported above.

### TABLE I: Iron PIV distances between different polymorphs, and a liquid model from MD.

| $r_0 = 2.5\, \text{Å}$ | $r_0 = 3.5\, \text{Å}$ | $r_0 = 4.5\, \text{Å}$ |
|------------------------|------------------------|------------------------|
| Fm3m | liq | Imm3 | Fe6 | Pmma | Fm3m | liq | Imm3 | Fe6 | Pmma | Fm3m | liq | Imm3 | Fe6 | Pmma |
| Fm3m | - | 0.43 | 0.41 | 1.03 | 1.27 | - | 0.41 | 0.58 | 1.43 | 2.12 | - | 0.53 | 0.74 | 2.10 | 2.93 |
| liq | 0.43 | - | 0.29 | 0.84 | 1.11 | 0.41 | - | 0.34 | 1.18 | 1.90 | 0.53 | - | 0.41 | 1.74 | 2.62 |
| Imm3 | 0.41 | 0.29 | - | 0.80 | 1.06 | 0.58 | 0.34 | - | 1.18 | 1.92 | 0.74 | 0.41 | - | 1.68 | 2.58 |
| Fe6 | 1.03 | 0.84 | 0.80 | - | 0.28 | 1.43 | 1.18 | 1.18 | - | 0.83 | 2.10 | 1.74 | 1.68 | - | 1.15 |
| Pmma | 1.27 | 1.11 | 1.06 | 0.28 | - | 2.12 | 1.90 | 1.92 | 0.83 | - | 2.93 | 2.62 | 2.58 | 1.15 | - |

### TABLE II: Silicon PIV distances between different polymorphs, and a liquid model from MD.

| $r_0 = 2.5\, \text{Å}$ | $r_0 = 3.5\, \text{Å}$ | $r_0 = 4.5\, \text{Å}$ |
|------------------------|------------------------|------------------------|
| fcc | VII | VI | V | RT | liq | fcc | VII | VI | V | RT | liq | fcc | VII | VI | V | RT | liq |
| fcc | 0.28 | 0.93 | 0.97 | 1.47 | 1.33 | - | 0.40 | 0.68 | 1.09 | 1.86 | 1.63 | - | 0.71 | 0.68 | 1.12 | 2.00 | 1.95 |
| VII | 0.28 | - | 0.81 | 0.76 | 1.24 | 1.08 | 0.40 | - | 0.56 | 0.94 | 1.74 | 1.48 | 0.71 | - | 0.62 | 1.04 | 2.13 | 2.01 |
| VI | 0.93 | 0.81 | - | 0.63 | 0.98 | 0.90 | 0.68 | 0.56 | - | 0.55 | 1.46 | 1.22 | 0.68 | 0.62 | - | 0.64 | 1.76 | 1.70 |
| V | 0.97 | 0.76 | 0.63 | - | 0.87 | 0.66 | 1.09 | 0.94 | 0.55 | - | 1.34 | 1.06 | 1.12 | 1.04 | 0.64 | - | 1.44 | 1.34 |
| RT | 1.47 | 1.24 | 0.98 | 0.87 | - | 0.42 | 1.86 | 1.74 | 1.46 | 1.34 | - | 0.66 | 2.00 | 2.13 | 1.76 | 1.44 | - | 0.61 |
| liq | 1.33 | 1.08 | 0.90 | 0.66 | 0.42 | - | 1.63 | 1.48 | 1.22 | 1.06 | 0.66 | - | 1.95 | 2.01 | 1.70 | 1.34 | 0.61 | - |

### TABLE III: Sodium PIV distances between different polymorphs.

| $r_0 = 2.5\, \text{Å}$ | $r_0 = 3.5\, \text{Å}$ | $r_0 = 4.5\, \text{Å}$ |
|------------------------|------------------------|------------------------|
| hcp | bcc | hcp | bcc | hcp | bcc |
| hcp | - | 0.07 | - | 0.29 | - | 0.45 |
| bcc | 0.07 | - | 0.29 | - | 0.45 |

### TABLE IV: Carbon PIV distances between different polymorphs.

| $r_0 = 2.5\, \text{Å}$ | $r_0 = 3.5\, \text{Å}$ | $r_0 = 4.5\, \text{Å}$ |
|------------------------|------------------------|------------------------|
| cubic | lonsdaleite | diamond | graphite | cubic | lonsdaleite | diamond | graphite | cubic | lonsdaleite | diamond | graphite |
| cubic | - | 2.83 | 2.87 | 3.34 | - | 4.48 | 4.57 | 5.20 | - | 5.54 | 5.63 | 6.40 |
| lonsdaleite | 2.83 | - | 0.22 | 0.82 | 4.48 | - | 0.37 | 1.19 | 5.54 | - | 0.38 | 1.22 |
| diamond | 2.87 | 0.22 | - | 0.79 | 4.57 | 0.37 | - | 1.13 | 5.63 | 0.38 | - | 1.07 |
| graphite | 3.34 | 0.82 | 0.79 | - | 5.20 | 1.19 | 1.13 | - | 6.40 | 1.22 | 1.07 | - |
TABLE V: Phosphorus PIV distances between different polymorphs.

|         | $r_0 = 2.5 \text{ Å}$ | $r_0 = 3.5 \text{ Å}$ | $r_0 = 4.5 \text{ Å}$ |
|---------|------------------------|------------------------|------------------------|
| Pm3m    | - 0.32 0.50 0.52       | - 0.42 0.53 0.73       | - 0.38 0.59 0.94       |
| Cmca    | 0.32 - 0.50 0.53       | 0.42 - 0.53 0.73       | 0.38 - 0.51 0.90       |
| Imma    | 0.50 0.50 - 0.11       | 0.53 0.53 - 0.33       | 0.59 0.51 - 0.46       |
| P12c1   | 0.52 0.53 0.11 -       | 0.73 0.73 0.33 -       | 0.94 0.90 0.46 -       |

TABLE VI: Sulfur PIV distances between different polymorphs.

|         | $r_0 = 2.5 \text{ Å}$ | $r_0 = 3.5 \text{ Å}$ | $r_0 = 4.5 \text{ Å}$ |
|---------|------------------------|------------------------|------------------------|
| S ($\beta$) | - 1.20 1.21 1.23 1.25 | - 1.76 1.76 1.86 1.87 | - 2.50 2.51 2.60 2.61 |
| S7 (c) | 1.20 - 0.02 0.05 0.06 | 1.76 - 0.45 0.16 0.17 | 2.50 - 0.06 0.19 0.19 |
| S7 (n) | 1.21 0.02 - 0.06 0.06 | 1.76 0.04 - 0.17 0.16 | 2.51 0.06 - 0.19 0.19 |
| S12 | 1.23 0.05 0.06 - 0.03 | 1.86 0.16 0.17 - 0.08 | 2.60 0.19 0.19 - 0.10 |
| S8 | 1.25 0.06 0.06 0.03 - | 1.87 0.17 0.16 0.08 - | 2.61 0.19 0.19 0.10 - |

TABLE VII: SiC PIV distances between different polymorphs (M stands for Moissanite).

|         | $r_0 = 2.5 \text{ Å}$ | $r_0 = 3.5 \text{ Å}$ | $r_0 = 4.5 \text{ Å}$ |
|---------|------------------------|------------------------|------------------------|
| M-3C | - 0.10 0.22 - 0.35 0.68 | - 0.35 0.68 - 0.57 0.92 |
| M-P6 | 0.10 - 0.21 0.35 - 0.64 0.57 | 1.05 - 0.20 0.27 0.44 |
| M-6H | 0.22 0.21 - 0.68 0.64 | 0.92 1.05 - 0.31 0.71 |

TABLE VIII: SiO$_2$ PIV distances between different polymorphs.

|         | $r_0 = 2.5 \text{ Å}$ | $r_0 = 3.5 \text{ Å}$ | $r_0 = 4.5 \text{ Å}$ |
|---------|------------------------|------------------------|------------------------|
| Stishov. | - 0.77 0.82 1.03 0.91 | - 1.03 1.09 1.41 1.49 | - 1.61 1.42 1.87 2.04 |
| Coes. | 0.77 - 0.09 0.64 0.24 | 1.03 - 0.25 0.57 0.61 | 1.61 - 0.41 0.57 0.71 |
| Quartz | 0.82 0.09 - 0.63 0.19 | 1.09 0.25 - 0.50 0.51 | 1.42 0.41 - 0.63 0.77 |
| Tridym. | 1.03 0.64 0.63 | - 0.59 1.41 0.57 0.50 | - 0.35 1.87 0.57 0.63 |
| Crist. | 0.91 0.24 0.19 0.59 | - 1.49 0.61 0.51 0.35 | - 2.04 0.71 0.77 0.31 |

TABLE IX: RbCl PIV distances between different polymorphs.

|         | $r_0 = 2.5 \text{ Å}$ | $r_0 = 3.5 \text{ Å}$ | $r_0 = 4.5 \text{ Å}$ |
|---------|------------------------|------------------------|------------------------|
| Fm3m | - 0.23 - 0.84 - 1.19 | - 1.19 | - 1.19 |
| Pm3m | 0.23 - 0.84 - 1.19 | - 1.19 |

TABLE X: Fe$_2$O$_3$ PIV distances between different polymorphs.

|         | $r_0 = 2.5 \text{ Å}$ | $r_0 = 3.5 \text{ Å}$ | $r_0 = 4.5 \text{ Å}$ |
|---------|------------------------|------------------------|------------------------|
| Pbcn | - 0.39 0.41 0.53 0.58 | - 0.63 0.68 0.73 0.82 | - 1.05 1.20 1.10 1.37 |
| R3ch | 0.39 - 0.04 0.29 0.35 | 0.63 - 0.07 0.27 0.35 | 1.05 - 0.20 0.27 0.44 |
| C12c1 | 0.41 0.04 - 0.27 0.33 | 0.68 0.07 - 0.26 0.32 | 1.20 0.20 - 0.31 0.36 |
| Pna21 | 0.53 0.29 0.27 - 0.15 | 0.73 0.27 0.26 - 0.21 | 1.10 0.27 0.31 - 0.36 |
| P41212 | 0.58 0.35 0.33 0.15 | - 0.82 0.35 0.32 0.21 | - 1.37 0.44 0.36 0.36 |
In this section we present an analysis of the HDA-LDA and HDA–VHDA transformations. In Figure S2 we report density and free energy profiles along the pathways of the two transformations obtained from umbrella sampling simulations. Density changes smoothly with the progress of the transformation, the profiles show several local minima but free energy barriers are typically smaller than the ones of the transformations reported in the manuscript, especially for the HDA-VHDA transformation, where they are of the order of magnitude of the estimated errors (see Methods section). This is in agreement with previous computational studies on the compression of the HDA form [12], where no hysteresis behavior is observed during compression/decompression cycles. The set of parameters \( \{r_0, n, m\} \) used to define the switching function for the HDA–LDA and HDA–VHDA simulations of Figure S2 are respectively \( \{0.35, 8, 18\} \) and \( \{0.40, 6, 20\} \). For more computational details see the Methods section.

**Amorphous–amorphous transformations**

In this section we present an analysis of the HDA-LDA and HDA–VHDA transformations. In Figure S2 we report density and free energy profiles along the pathways of the two transformations obtained from umbrella sampling simulations. Density changes smoothly with the progress of the transformation, the profiles show several local minima but free energy barriers are typically smaller than the ones of the transformations reported in the manuscript, especially for the HDA-VHDA transformation, where they are of the order of magnitude of the estimated errors (see Methods section). This is in agreement with previous computational studies on the compression of the HDA form [12], where no hysteresis behavior is observed during compression/decompression cycles. The set of parameters \( \{r_0, n, m\} \) used to define the switching function for the HDA–LDA and HDA–VHDA simulations of Figure S2 are respectively \( \{0.35, 8, 18\} \) and \( \{0.40, 6, 20\} \). For more computational details see the Methods section.

**Comparison of LDA and liquid: densities and diffusion coefficients**

We report in Table XIV the densities and the diffusion coefficients obtained from four 10 ns-long MD simulations at \( P = 1 \) bar, of (i) LDA at \( T = 100 \) K (ii), LDA heated up at \( T = 240 \) K (iii), the liquid phase cooled down at \( T = 260 \) K, and (iv) the liquid phase at \( T = 300 \) K.
FIG. S1: PIV distance 2D maps for Fe, Si, MgSiO$_3$ and Fe$_2$O$_3$ ($r_0 = 3.5$ Å). Distances on the 2D map reproduce PIV distances, with an accuracy given by the following correlation coefficients: 0.9912 for Fe, 0.9905 for Si, 0.9964 for Fe$_2$O$_3$, and 0.9955 for MgSiO$_3$.

TABLE XIV: Properties of liquid and LDA water at different conditions.

| Temperature | Density (kg/m$^3$) | Diffusion Coeff. (10$^{-5}$ cm$^2$/s) |
|-------------|-------------------|----------------------------------------|
| LDA 100 K   | 987 ± 1           | 0.000041 ± 0.000008                    |
| LDA 240 K   | 980.9 ± 0.3       | 0.29 ± 0.03                            |
| liq 260 K   | 985.2 ± 0.1       | 0.8407 ± 0.0009                        |
| liq 300 K   | 976.3 ± 0.1       | 2.619 ± 0.008                         |

[1] J. L. Abascal and C. Vega, J. Chem. Phys. 123, 234505 (2005).
[2] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak, J. Chem. Phys. 81, 3684 (1984).
[3] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. 126, 014101 (2007).
[4] B. Roux, Comput. Phys. Commun. 91, 275 (1995).
[5] H. J. Berendsen, D. van der Spoel, and R. van Drunen, Comput. Phys. Commun. 91, 43 (1995).
[6] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Comput. Phys. Commun. 185, 604 (2014).
[7] F. Pietrucci and R. Martonák, J. Chem. Phys. 142, 104704 (2015).
[8] M. Mendelev, S. Han, D. Srolovitz, G. Ackland, D. Sun, and M. Asta, Philosoph. Mag. 83, 3977 (2003).
[9] P. A. Bartók and T. A. Weber, Phys. Rev. B 31, 5262 (1985).
[10] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc 118, 11225 (1996).
[11] A. P. Bartók, R. Kondor, and G. Csányi, Phys. Rev. B 87, 184115 (2013).
[12] J. Chiu, F. W. Starr, and N. Giovambattista, J. Chem. Phys. 139, 184504 (2013).
FIG. S2: a, Density (top) and free energy (bottom) profiles along the the HDA–LDA transformation pathway, represented by the path collective variable \( s \), at \( T = 100 \) K and \( P = 1 \) bar. \( \Delta G \) values are obtained via the weighted histogram analysis method applied to the umbrella sampling trajectories (last 5 ns of 20 ns-long simulation windows); density values are averaged over umbrella sampling trajectories (standard deviations are also reported). b, Same as panel a but for the HDA–VHDA transformation at \( T = 100 \) K and \( P = 12 \) kbar.