UC Berkeley
UC Berkeley Previously Published Works

Title

The microscopic mechanism of bulk melting of ice

Permalink

Journal

The Journal of Chemical Physics, 155(12)

ISSN

Authors

Moritz, Clemens
Geissler, Phillip L
Dellago, Christoph

Publication Date

2021-09-28

DOI

Peer reviewed

Clemens Moritz, Phillip L. Geissler and Christoph Dellago

View Online   **Export Citation**   **CrossMark**

The microscopic mechanism of bulk melting of ice

View Online   Export Citation   CrossMark

Clemens Moritz,[1] (iD) Phillip L. Geissler,[2,3] (iD) and Christoph Dellago[1,4,a)] (iD)

AFFILIATIONS

[1] Faculty of Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria
[2] Department of Chemistry, University of California, Berkeley, California 94720, USA
[3] Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
[4] Erwin Schrödinger Institute for Mathematics and Physics, Boltzmanngasse 9, 1090 Vienna, Austria

[a)]Author to whom correspondence should be addressed: christoph.dellago@univie.ac.at

ABSTRACT

We study the initial stages of homogeneous melting of a hexagonal ice crystal at coexistence and at moderate superheating. Our trajectory-based computer simulation approach provides a comprehensive picture of the events that lead to melting, from the initial accumulation of 5+7 defects, via the formation of L–D and interstitial–vacancy pairs, to the formation of a liquid nucleus. Of the different types of defects that we observe to be involved in melting, a particular kind of 5+7 type defect (type 5) plays a prominent role as it often forms prior to the formation of the initial liquid nucleus and close to the site where the nucleus forms. Hence, like other solids, ice homogeneously melts via the prior accumulation of defects.

I. INTRODUCTION

In this paper, we present a computer simulation study of the initial stages of melting of hexagonal water ice (ice Ih) at ambient pressure and superheating up to 11% above the melting point in a regime where the formation of a liquid nucleus of sufficient size is the rate limiting step. Our analysis of ensembles of trajectories yields a detailed, time-resolved picture of the different dynamical pathways that lead to the formation of such a liquid nucleus and the role that different types of defects play. We find that prior to melting, a number of the so-called 5+7 defects and larger defect structures in the hydrogen bond network accumulate in the volume that later become the liquid nucleus and that the size and number of these defects become larger as the degree of superheating is reduced.

The microscopic mechanisms that lead to melting of crystalline solids are a long-standing subject of solid state theory. In most situations, melting originates at the surfaces of a crystal;[1] however, under particular circumstances, a solid may melt from within the bulk of a crystal as opposed to its surfaces[2] (the so-called homogeneous melting). For example, micrometer-sized single crystal spheres made of silver melt homogeneously if their surfaces are covered by layers of gold atoms. This shell of atoms suppresses surface melting[3] by

having a higher melting point while forming a lattice that is compatible with the silver lattice. Using this method, substantial superheating of the Ag lattice can be achieved that is preempted by surface melting in a conventional setting.

A multitude of theories exist that put increasingly stringent limits on the amount of superheating a crystal can be subjected to before it becomes mechanically unstable.[4–10] However, in equilibrium, all these instabilities are preempted by thermal melting of the crystal via a nucleation and growth mechanism.[11] In many recent studies, preexisting defects have been found to play a major role[11–20] in melting mechanisms and to significantly influence the stability limits of crystals.

Melting of ice Ih—the ordinary form of ice that can be seen around the liquid nucleus in Fig. 1—is a particularly interesting example of a melting crystal, as its open structure is held together by a network of hydrogen bonds. At temperatures around the ambient pressure melting point, these bonds can rupture and form with relative ease compared to, e.g., covalent bonds, and hence, a multitude of hydrogen bonding defects occurs in ice under these conditions. While ice usually melts heterogeneously, homogeneous nucleation has been induced by internal heating. This is achieved by exciting the OH stretching mode of water with IR laser light.[21–25] Such
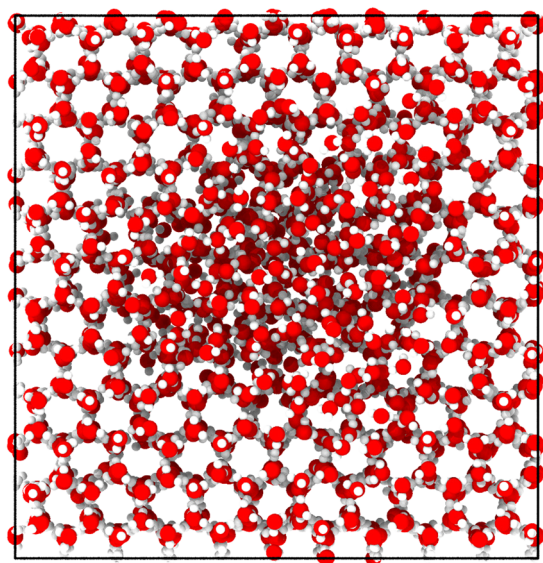
FIG. 1. Example configuration containing a spherical liquid cluster. For clarity, the potential energy of the configuration has been locally minimized. The configuration has been prepared by selecting the molecules in a sphere of radius 1.5 nm and melting them by heating. Afterward, the configuration is briefly equilibrated at a temperature of 303 K.

experiments strongly suggest that defects play a key role in determining the stability of ice crystals.[23]

Previous computer simulation studies of ice melting support the finding that defects play an important role. Donadio, Raiteri, and Parrinello[14] investigated the free energy landscape of melting at the melting point, finding that the so-called 5+7 defects form a minimum in the free energy landscape (see Sec. II for a description of the various defect types found in ice). Furthermore, they observed the formation of large defect structures in the hydrogen bond network that involve on the order of 50 molecules. Mochizuki, Matsumoto, and Ohmine[18] studied spontaneous melting under higher superheating (around 19% above melting temperature) where melting events occur spontaneously in simulations within a timeframe of a few nanoseconds. They identified the formation of separated defect pairs (either interstitial–vacancy or L–D pairs) as the controlling step in the melting process at these conditions and also observed 5+7 defects as part of the melting mechanism.

In this paper, we follow a similar approach to the one presented in Ref. 18 where an ensemble of trajectories is generated and the effect of defects is investigated. However, due to the lower degree of superheating, the critical step in the nucleation process is the formation of a liquid nucleus of sufficient size. At the temperatures we consider, the rate of formation of such critical nuclei is much lower than the rate of melting observed in Ref. 18. As a result, the required number of melting trajectories cannot be feasibly obtained from molecular dynamics (MD) simulations simply by waiting for spontaneous melting events to occur. In this paper, we generate unbiased melting trajectories based on the expectation of a nucleation-growth mechanism near phase coexistence, yielding hundreds of statistically independent samples. With the help of these trajectories, we then

assemble a detailed, time-resolved picture of the dynamic pathways that lead from a frozen crystal to a liquid nucleus at different degrees of superheating. This analysis yields a comprehensive picture of the roles that different defects play in the mechanism of homogeneous melting as a function of temperature.

The remainder of this paper is organized as follows: in Sec. II, we introduce the types of defects that we refer to throughout this paper. In Sec. III, we lay out our simulation methodology, and in Sec. IV, we present the results of our simulations. Section V summarizes and discusses our findings.

II. ICE DEFECTS

As part of our analysis, we classify different point defect types that can occur in ice. Here, we briefly introduce the types of defects we consider and summarize previous results on the role they play in melting. Higher dimensional defects such as dislocations and disclinations do not form spontaneously in the simulations presented in this paper, and hence, we limit this discussion to point defects only.

In order to discuss defects within the ice Ih structure, it is instructive to first reiterate some of the properties of hexagonal ice:[26] (1) the water molecules (H_2O) are laid out in a wurtzite structure,[27,28] (2a) each molecule takes part in four hydrogen bonds to its nearest neighbors, and (2b) two of the hydrogen bonds are donated to other molecules. Observations (2a) and (2b) together are known as the Bernal–Fowler ice rules.[27,29] (3) The hydrogen bond network of a defect-free ice Ih crystal can be decomposed into an array of six-membered rings. However, the direction of hydrogen bonds in the bond network is not uniquely determined, and consequently, there is no long-range proton order in ice Ih.[27,28]

Breaking rule (1), i.e., displacing a molecule far from its location in the perfect lattice, leads to an interstitial–vacancy (I–V) pair. It has been shown that translational diffusion within ice occurs by the movement of whole molecules[28,31] and that the concentration of ionic defects is low compared to that of molecular defects.[26,32] Hence, we expect that whole molecules also form the majority of I–V defects in the lattice (and not ionic defects). In our simulations, we only consider I–V defects where whole molecules are moved out of their lattice position. No ionic defects can occur in our simulations.

Breaking condition (2) while keeping condition (1) intact leads to the so-called Bjerrum or L–D pairs.[33] L and D defects are, respectively, characterized by a missing or an excess proton within a small region of space so that the ice rules are not satisfied and cannot be satisfied until a matching defect of the other type is encountered. Nevertheless, L–D pairs are often found bound to each other, forming an L+D complex that exhibits a lower potential energy than a separated L–D pair.[30] In this work, we do not distinguish between separated L–D pairs and L+D complexes and call all structures where the ice rules are broken L–D defects.

I–V and L–D defects will be referred to as *mobile defects* throughout this work because once a pair of these defects is formed and separated from each other, their movement through the system is relatively facile.[34] The authors of Ref. 18 discussed the critical role of mobile defect pairs in melting under high superheating conditions, where the separation of a mobile defect pair is found to drastically lower the free energy barrier that needs to be overcome in order for melting to occur.
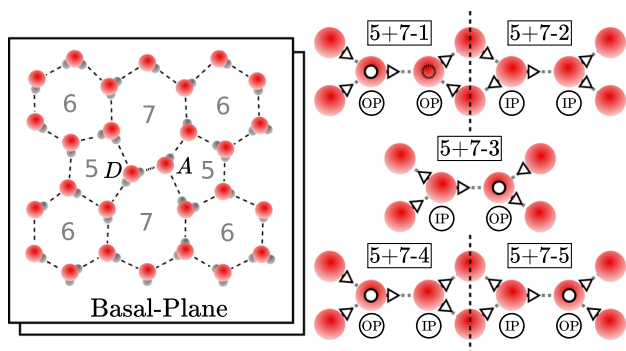
FIG. 2. Sketches of 5+7 defects. The red circles indicate oxygen atoms. Left: 5+7 defect in the basal plane of a hexagonal ice crystal. *D* and *A* mark the central molecules of the 5+7 defect that form a hydrogen bond that is part of two seven-rings. The numbers indicate how many molecules comprise the rings that enclose them. Right: classification of different 5+7 defect types according to the local arrangement of hydrogen bonds as defined in Ref. 30 and used in this paper. The white arrows point from donor to acceptor oxygen in hydrogen bonds. The circles on oxygens indicate hydrogen bonds that are formed along an orthogonal plane. IP and OP stand for in-plane and out-of-plane molecules, respectively, and indicate whether the plane that is spanned by the H–O–H triangle is parallel to the basal plane or orthogonal to it.

Finally, breaking of rule (3) while keeping rule (2) intact constitutes another class of defects, which (in line with Ref. 14) we will call *topological defects*. These are defects where the six-ring structure of the perfect lattice is broken, and instead, there are other ring combinations present. Note that the ice rules are still fulfilled in these defects and that the molecules that take part are only shifted by small distances from their positions in the perfect lattice. The most prominent of these defects is the so-called 5+7 defect, first found in simulations by Tanaka and Mohanty[35] and described in detail by Grishina and Buch.[30] In 5+7 defects, two five- and two seven-membered rings are found neighboring each other. These defects can then be further classified into different types according to the placement of protons around the central bond of the 5+7 defect (see
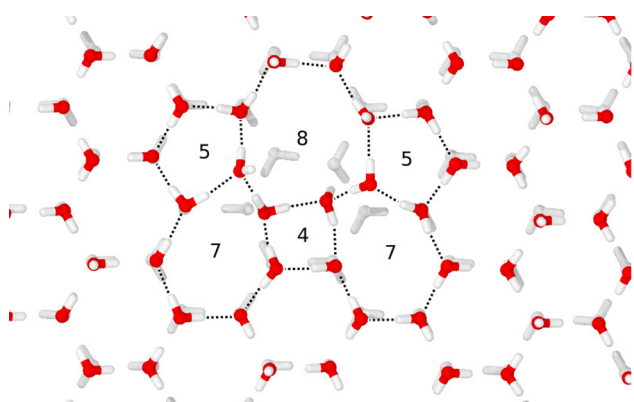

FIG. 3. Snapshot of a 455778 defect that is embedded into the basal plane of an ice Ih crystal. The numbers mark the four-, five-, seven-, and eight-membered rings in the H-bond network that make up the defect. The white molecules are located in an adjacent, defect-free plane.

Fig. 2) and according to the crystal plane they are formed in. 5+7 defects are readily observed in equilibrium simulations that use the TIP4P family of water models.[35,36]

Other combinations of ring sizes are also possible: for example, we encounter 455778 defects that center around a four-ring while still satisfying the ice rules (see Fig. 3). Larger defect structures (such as the ones observed in Ref. 14) are frequently observed in melting trajectories, and we will subsume all these structures under the name *extended topological defect* or E defect. All topological defects have in common that there is no efficient mechanism for these defects to move, which is why we will refer to them as *immobile defects* in the following.

In Sec. III, we present the simulation methodology used to generate trajectories that are then analyzed in terms of the defects that occur at various stages of melting.

III. METHODS

A. Generating parts of reactive trajectories

Our aim is to investigate the melting transition starting from configurations that contain an ice Ih crystal with possibly a few 5+7 defects (state $\mathcal{A}$ in Fig. 4) and ending in the liquid state (state $\mathcal{B}$). In particular, we are interested in the initial stages of these trajectories up to a state where a liquid nucleus has formed (state $\mathcal{S}$ in Fig. 4). Note that we include configurations that contain 5+7 defects into the definition of state $\mathcal{A}$. Such defect states occur readily in equilibrium trajectories under the conditions considered in this paper where the ice Ih crystal is metastable with respect to the liquid state.

Under these conditions, the two states $\mathcal{A}$ and $\mathcal{B}$ are separated by a free energy barrier so that the melting transition is a rare event. This means that the average waiting time τ_{AB} between preparing a system in an equilibrium frozen state in $\mathcal{A}$ and a melting event that leads the system from state $\mathcal{A}$ to the liquid state $\mathcal{B}$ exceeds the timescale of relaxation in state $\mathcal{A}$, τ_A, by orders of magnitude,

$$\tau_A \ll \tau_{AB}. \tag{1}$$
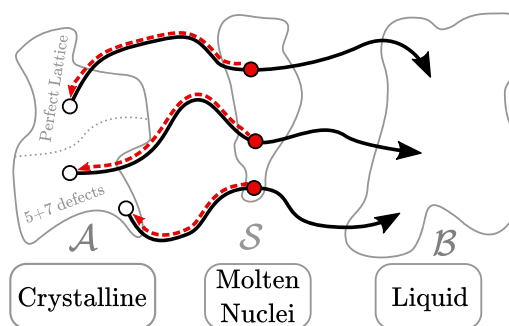

FIG. 4. Schematic representation of the method used to generate trajectories. Region $\mathcal{A}$ contains configurations that are completely solid with possibly a few 5+7 defects remaining. Region $\mathcal{B}$ contains configurations that are liquid, and region $\mathcal{S}$ contains molten nuclei prepared by the procedure laid out in Sec. III A. Our aim is to generate a sample of the early stages of melting trajectories (black curves with arrows). To do so, we integrate trajectories starting from configurations in $\mathcal{S}$ until they reach $\mathcal{A}$ (red dashed arrows) and subsequently invert the direction of time. Trajectories that reach $\mathcal{B}$ before $\mathcal{A}$ are discarded.

This so-called *separation of timescales* guarantees that the way melting events occur does not depend on the details of how frozen configurations are prepared and that, instead, we can think of the two states as being connected by an ensemble of melting trajectories that leave $\mathcal{A}$ and end in $\mathcal{B}$ without visiting $\mathcal{A}$ in the meantime. This ensemble of trajectories is called the *transition path ensemble*[37,38] (indicated by the black arrows in Fig. 4).

To generate the initial parts of melting trajectories, we take the following approach:

1. Pick a sample of configurations from equilibrium simulations of ice Ih performed at the chosen temperature and pressure.
2. Construct a liquid domain inside this configuration and locally equilibrate the resulting configuration. The resulting ensemble is denoted with $\mathcal{S}$.
3. Run molecular dynamics simulations using a time-reversible integration scheme starting from $\mathcal{S}$ until state $\mathcal{A}$ or $\mathcal{B}$ is reached. The resulting trajectories are referred to as *backward trajectories*.
4. Invert the time direction of the backward trajectories that end in $\mathcal{A}$ to get a sample of trajectories that lead from $\mathcal{A}$ to $\mathcal{S}$.

The use of a time-reversible integration scheme guarantees that the trajectories obtained by integrating backward in time and subsequently inverting the momenta are valid forward trajectories that lead from $\mathcal{A}$ to $\mathcal{S}$.[39]

Similar to the so-called seeding method,[40–43] we choose the liquid clusters to be spherical in shape. In reality, the shape of the liquid nuclei that form during homogeneous melting is likely not perfectly spherical due to slight differences in the interface tension associated with different crystal planes[42,44–46] and due to the different dynamics of crystal growth along the plane normals.[47–50] To assess the effect that different cluster shapes have on the early stages of melting trajectories, we perform additional simulations at the ice–liquid coexistence temperature that start from configurations with a slab-shaped liquid domain, such as the one shown in Fig. 5.

We refer to trajectories that are constructed from spherically shaped liquid domains as *spherical-geometry* trajectories and to the ones constructed from slab-shaped liquid domains as *slab-geometry* trajectories.
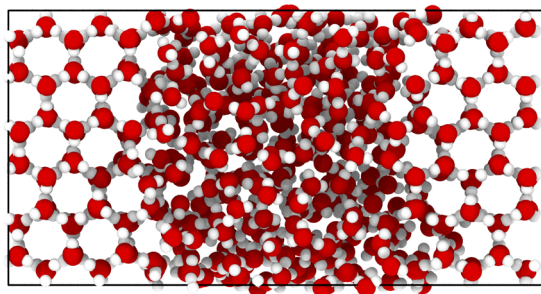

FIG. 5. Example of a configuration that contains a slab of liquid that consists of half the molecules in a 720 molecule water system. For clarity, the potential energy of the configuration has been locally minimized.

B. Simulation details

The simulations presented in this paper are obtained using the TIP4P/Ice water model[51] with a time-reversible rigid body integration scheme[52,53] as implemented in the LAMMPS simulation package.[54,55] Thermostats and barostats are implemented using Nosé–Hoover chains[56–58] where the x, y, and z directions are independently barostatted to a pressure of 1 bar. Long-range interactions are treated using a particle-mesh Ewald method (PPPM[59,60]) with an accuracy of 10^{-4}, and the time step is set to 1 fs. Snapshots are saved for analysis every 10 ps.

1. Spherical-geometry trajectories

MD simulations used to generate spherical-geometry trajectories are carried out with 2880 molecules in an almost cubic simulation box (see Fig. 1). The initial dimensions of the boxes are $44.9 \times 46.7 \times 44.0$ Å^3 in the directions orthogonal to the secondary prism plane, the prism plane, and the basal plane, respectively. To generate initial configurations for the backward trajectories, we follow the following procedure:

1. Construct a proton ordered ice XI lattice and randomly reorder the hydrogen bonds using a Monte Carlo procedure[61] in order to obtain ice Ih. Here, we require that the total dipole moment of the configuration is zero at the end of the procedure.
2. Equilibrate these configurations in a 30 ns parallel tempering[62–65] trajectory using replicas starting from a temperature of 258 K up to and including 328 K spaced 5 K apart.
3. Pick a sample of configurations from the parallel tempering simulation at the desired temperature.
4. Pick a random center for the liquid nucleus in each of the configurations, and find the molecules within a radius of 15 Å.
5. Heat the selected molecules using a thermostat while keeping the molecules outside the sphere fixed until the crystal structure in the selected region breaks down.
6. Equilibrate at the target temperature by first keeping the molten fraction fixed and propagating the molecules in the crystalline phase (for 20 ps) and then keeping the crystalline molecules fixed and propagating the molten molecules (for 25 ps). This procedure hinders the molecules from recrystallizing because the molecules inside and outside of the selected volume cannot collectively reorder into a frozen configuration.

It should be noted that given the initial volume of the liquid domain, the spherical shape is not the lowest free energy geometry. Instead, arranging the liquid domain in the shape of a cylinder that wraps around the periodic boundary conditions of the simulation box would lead to a lower overall surface area and, hence, a lower free energy.[66,67] These competing cluster shapes—which are an artifact of the periodic boundary conditions—are separated from each other by significant free energy barriers.[66,68,69] We see only small changes in the asphericity of the liquid domains in the generated trajectories, and no transitions to the cylinder geometry and therefore assume that the existence of the competing cylindrical phase has no effect on the generated trajectories.

2. Slab-geometry trajectories

Slab-geometry trajectories are generated using 720 molecules in an elongated box with initial dimensions $44.9 \times 23.3 \times 22.0$ Å³ (see Fig. 5 for an example configuration). The volume that contains the liquid is chosen so that the solid–liquid interfaces are parallel to the secondary prism face of the ice Ih crystal. This is the geometry that has been found to preferentially form in Ref. 70. No initial equilibration is performed because the system has sufficient time to equilibrate before the slab collapses into one of the two competing phases.

The resulting trajectories are quite different from the spherical-geometry trajectories at the same temperature of 268 K. The spherical liquid domains shrink rapidly and predictably due to surface tension. For the slab geometry, periodic boundary conditions eliminate contributions from the surfaces where the liquid domain wraps around the simulation box.[66,67] Surface tension therefore does not drive the growth of the slab-shaped liquid domains because a change in the volume of the liquid domain has no effect on the overall liquid–solid surface area.

The temperature of $T = 268$ K was chosen so that roughly half of the trajectories started with a slab-shaped domain melt and the others freeze, further reducing the thermodynamic force that drives the growth and shrinkage of the liquid domain. This temperature is slightly below the melting temperature at 1 bar of 272.2 K reported by Abascal et al.,[51] a discrepancy that is likely due to the non-negligible influence of the solid–liquid interfaces in the comparatively small simulation box with 720 molecules. The result are trajectories where the two solid–liquid interfaces that delimit the liquid domain diffuse in the direction of their surface normal until they are close enough that a fluctuation in the shape of the interfaces brings them into contact. The distance between the interfaces at which this contact occurs is small, roughly one layer of six-rings or 7 Å.

A larger simulation cell would yield a larger, and more realistic, contact distance. Increasing this distance considerably, however, comes at great computational cost for two reasons: first, the contact distance scales logarithmically with the size of the simulation box.[71] Second, increasing the size of the simulation box slows down the diffusion of the two interfaces, which, in turn, sharply increases the length of the required trajectories. The system size with interface areas of 23.3×22.0 Å² was chosen as a compromise between maximizing the contact distance between the two interfaces and minimizing the simulation time required to obtain trajectories.

3. Gathered data

In total, three sets of trajectories leading from the prepared configurations $\mathcal{S}$ to the frozen state $\mathcal{A}$ were generated: 108 slab-geometry trajectories at a temperature of $T = 268$ K, 504 spherical-geometry trajectories at $T = 268$ K, and 448 spherical-geometry trajectories at $T = 303$ K.

The temperature of 303 K is 11% superheated relative to the melting point[51] and has been chosen so that the spherical nuclei with a radius of 15 Å are slightly subcritical. Out of the 504 trajectories that were run, we observed 56 trajectories that melted before they could reach the frozen state $\mathcal{A}$. On average, $\mathcal{A}$ is reached in 13.6 ns (slab, 268 K), 5.1 ns (spherical, 268 K), and 2.3 ns (spherical, 303 K).

Data on equilibrium properties presented in this paper were obtained from the parallel tempering simulations that were also used to generate the initial configuration for the backward simulation runs. The data from different replicas are combined using the weighted histogram analysis method (WHAM[72,73]) as implemented in the PyEMMA package.[74]

We now invert the direction of time in these generated trajectories, which become examples of the early stages of melting. We also set $t = 0$ at the crystalline endpoint of each trajectory so that the system adopts a frozen configuration at time zero and melting commences as time increases. These time conventions will be used throughout the remainder of this paper.

C. Defect detection

To detect defects in the hexagonal ice structure, we adapt an array of different techniques previously used to investigate the properties of water and ice. These include, after locally minimizing the potential energy of a configuration, an analysis of the hydrogen bond network,[18,75] in particular the ring structures found therein,[14] and the analysis of configurations relative to a reference configuration to facilitate the detection of molecular interstitials and vacancies.[18] See Appendix A for a detailed description of the algorithms used.

These techniques yield an analysis, such as the one shown in Figs. 6 and 7. It includes counts of the number of four- and five-
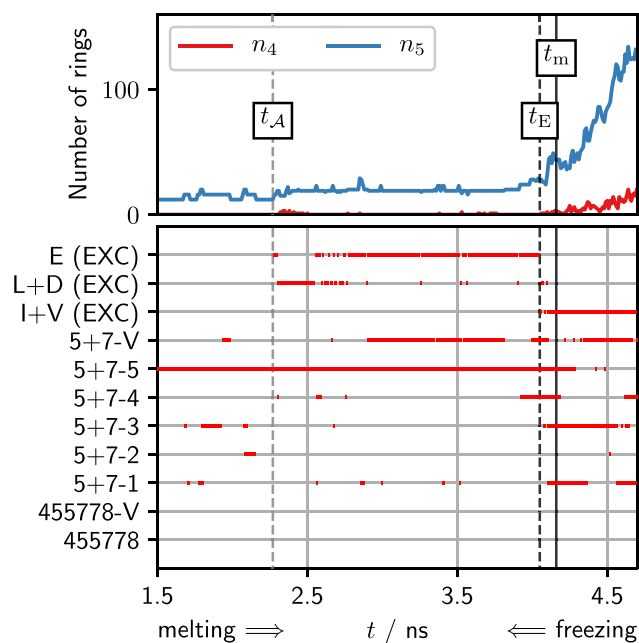

FIG. 6. Analysis of an example melting trajectory. At the beginning of the chosen timeframe, a number of 5+7-5 defects are present in an otherwise frozen configuration. The vertical lines indicate the times used to split the trajectory into the three stages described in Sec. IV. The red markings in the lower half of the plot indicate the presence of at least one defect of the given type in the system. Due to the algorithm used, only one of the defect types marked with (EXC) can be detected at a time (see Appendix A). The corresponding analysis for the other trajectories can be found in the supplementary material.
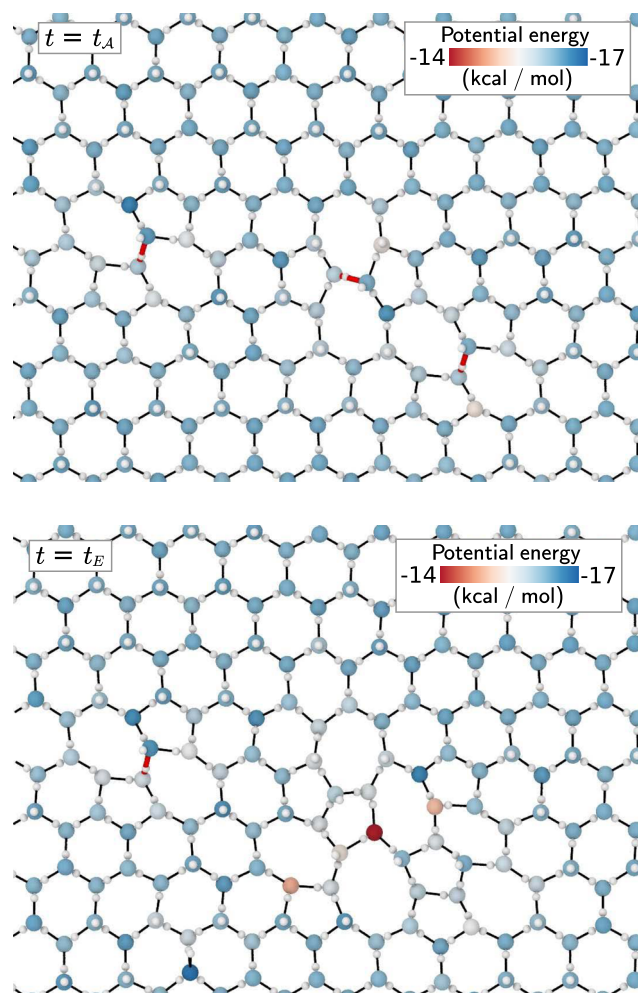
FIG. 7. Snapshots of the trajectory analyzed in Fig. 6 at time t_A (top) and at time t_E (bottom). A single basal plane that contains the 5+7 defects present at time t_A (red, thick bonds) is shown. The colors of molecules indicate their contribution to the total potential energy. The black lines indicate hydrogen bonds between molecules.

TABLE I. Thresholds in the growth of the number of five-rings over a 200 ps period, Δn_5, used to detect time t_m.

Geometry	Temperature (T) (K)	Δn_5
Slab	268	8
Sphere	268	8
Sphere	303	15

- t_A is the time when the system leaves state $\mathcal{A}$ for the last time, i.e., it is the last time where there are only 5+7 or 455778 defects present in the system.
- t_E is the last time when no mobile defects are present in the system. In the time between t_A and t_E, E defects are present in the system; L–D and I–V pairs may also form during this interval, but by definition, they must recombine before time t_E. If no E defect occurs along a trajectory, then t_A equals t_E.
- t_m is the time when the extended liquid domain forms. This event is associated with an increase in the rate at which five-membered rings appear, i.e., the slope of $n_5(t)$. Because this increase occurs on top of significant background fluctuations in $n_5(t)$, we look for an increase in $n_5(t)$ over a timespan of 200 ps that is larger than a given threshold Δn_5 (see Table I).

In the supplementary material, we include the defect analysis for each trajectory that is used in this study, including the times t_A, t_E, and t_m.

IV. THE THREE STAGES OF MELTING

Based on the times defined in Sec. III C, we now define three stages of the melting mechanism (cf. Fig. 8): stage I ($0 \leq t < t_E$), where only immobile defects are present in the system, stage II ($t_E \leq t < t_m$), where one or more mobile defects have formed, and stage III ($t \geq t_m$), where an extended liquid nucleus has formed. In Secs. IV A and IV B, we first discuss the spherical-geometry trajectories. Section IV C then discusses the differences found in slab-geometry trajectories.

A. Stage I: Topological defects ($t \leq t_E$)

In equilibrium and in stage I of melting trajectories, the bulk of defects that are present are of type 5+7. A prominent role in melting is played by 5+7-5 defects, as can be seen in Fig. 9, where we report the average numbers of defects found at time t_A compared to the average numbers found in equilibrium. The difference in defect numbers between these two scenarios is listed in Table II.

At both temperatures, the overall excess of defects found at time t_A is mostly accounted for by the excess of 5+7-5 defects. At $T = 268$ K, the number of 5+7-V defects is also significantly enhanced, while the number of 5+7-3 and 5+7-4 defects is slightly reduced. At $T = 303$ K, the average number of defects of type 5+7-V and types 5+7-1 through 5+7-4 is roughly equal to the numbers observed in equilibrium.

At time t_A, one of two things happens: either an E defect forms (*E-channel*) or a mobile defect forms directly (*57-channel*). Figure 8 shows the fractions of trajectories that pass through each of these
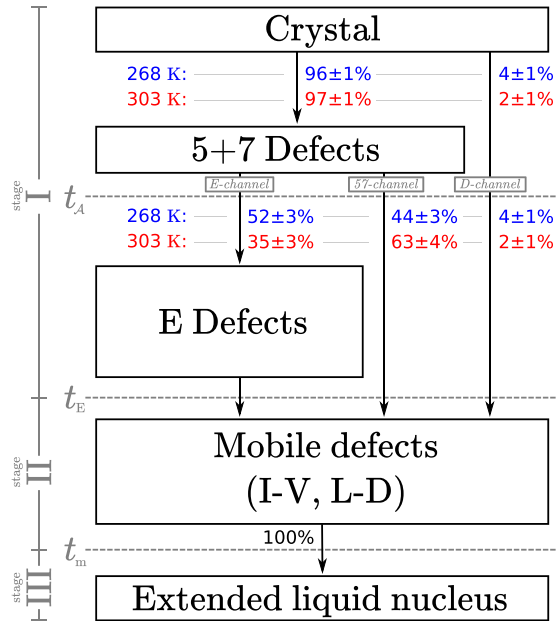
membered rings [$n_4(t)$ and $n_5(t)$, respectively] that are roughly proportional to the size of the liquid domain. It also indicates whether certain defects have been detected in the configuration. The defect types distinguished are I–V pairs, L–D pairs, and topological defects in the H-bond network: 5+7 defects, 455778 defects, and extended topological defect structures where the ice rules are fulfilled (E defects). The 5+7 defects are further split into the five horizontal types[30] (5+7-1 through 5+7-5) that are formed within the basal plane of the lattice and the vertical type (5+7-V) where the 5+7 defect is formed in a plane orthogonal to the basal plane.

Note that the algorithm used to detect defects of types I–V, L–D, and E does so in an exclusive fashion: if an I–V pair is present, L–D and E defects cannot be detected and if an L–D pair is present, E defects cannot be detected (see Appendix A, Fig. 19).

As part of the analysis, we determine three times along each melting trajectory: t_A, t_E, and t_m.

FIG. 8. Diagram of the stages observed during the melting of hexagonal ice crystals and the pathways that are taken by trajectories in which a spherical liquid nucleus forms. The percentages represent the fractions of trajectories that proceed along the indicated pathway relative to the total number of melting trajectories.
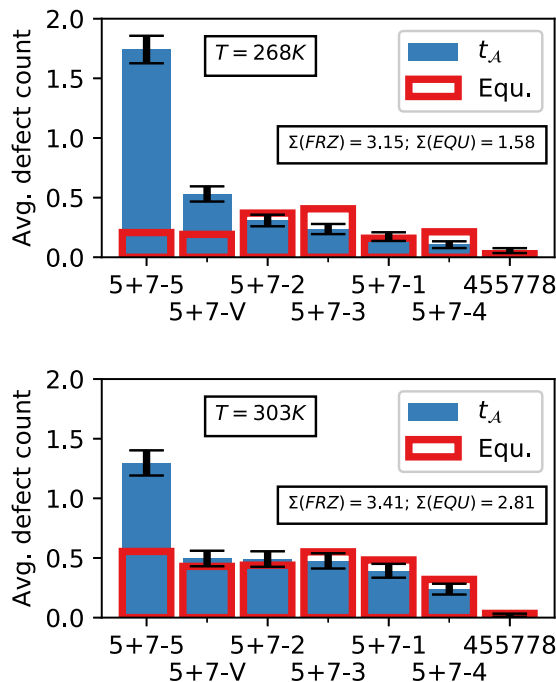
TABLE II. Average numbers of defects $\langle n \rangle$ at time t_A and in equilibrium, as well as the difference Δ between the two numbers, for different defect types and at different temperatures T. The reported errors are 95% confidence intervals calculated, assuming that the numbers of defects are Poisson distributed. This assumption is supported by an analysis of the defect statistics presented in Appendix C, Fig. 26.

$T = 268$ K	$\langle n \rangle (t = t_A)$	$\langle n \rangle$(Equ.)	Δ
5+7-5	1.74 ± 0.12	0.21	1.53
5+7-V	0.53 ± 0.06	0.19	0.34
5+7-2	0.31 ± 0.05	0.37	-0.06
5+7-3	0.24 ± 0.04	0.41	-0.17
5+7-1	0.17 ± 0.04	0.16	0.01
5+7-4	0.11 ± 0.03	0.21	-0.11
455778	0.06 ± 0.02	0.03	0.03
SUM	3.15 ± 0.16	1.58	1.57

$T = 303$ K	$\langle n \rangle (t = t_A)$	$\langle n \rangle$(Equ.)	Δ
5+7-5	1.30 ± 0.11	0.55	0.74
5+7-V	0.50 ± 0.07	0.43	0.07
5+7-2	0.49 ± 0.06	0.44	0.05
5+7-3	0.48 ± 0.06	0.55	-0.08
5+7-1	0.39 ± 0.06	0.48	-0.09
5+7-4	0.24 ± 0.05	0.32	-0.08
455778	0.02 ± 0.01	0.04	-0.02
SUM	3.41 ± 0.17	2.81	0.60

channels. As the temperature decreases from 303 to 268 K, the number of trajectories that involves an E defect increases from 35% to 52%.

In order to assess the role of 5+7 defects in the melting mechanism, we analyze the locations of defects relative to the center of the volume where the liquid nucleus eventually forms (the *nucleus volume*) as well as relative to the sites where mobile defects form at time t_E. To do so, we define the pair-correlation function between defects of reference type S and defects of target type T,

$$g_{ST}(r) = \frac{1}{c_S} \frac{H_{ST}(r, \Delta r)}{\rho_T V(r, \Delta r)}, \qquad (2)$$

where $H_{ST}(r, \Delta r)$ is a histogram of all pairwise distances between defects of type S and type T observed in a set of configurations, ρ_T is the equilibrium number density of defects of type T, and c_S is the total number of defects of type S observed. $V(r, \Delta r)$ is the volume of a spherical shell with inner radius $r - \Delta r/2$ and outer radius $r + \Delta r/2$, where Δr is the bin width of the histogram. Note that $g_{ST}(r)$ is not symmetric in S and T because $c_S \neq \rho_S$. Nevertheless, we expect $\lim_{r \to \infty} g_{ST}(r) = 1$.

We also define $n_{ST}(r)$ as the average number of defects T within a sphere of radius r centered on a defect of type S, i.e.,

$$n_{ST}(r) = \rho_T \sum_{r_i < r} V(r_i, \Delta r) g_{ST}(r_i). \qquad (3)$$

$g_T(r)$ and $n_T(r)$ are the analogous quantities where the reference points are the center of the spherical nucleus volume.

Figure 10 shows $g_T(r)$ and $n_T(r)$ obtained from configurations observed at time t_E (including trajectories that proceed through both
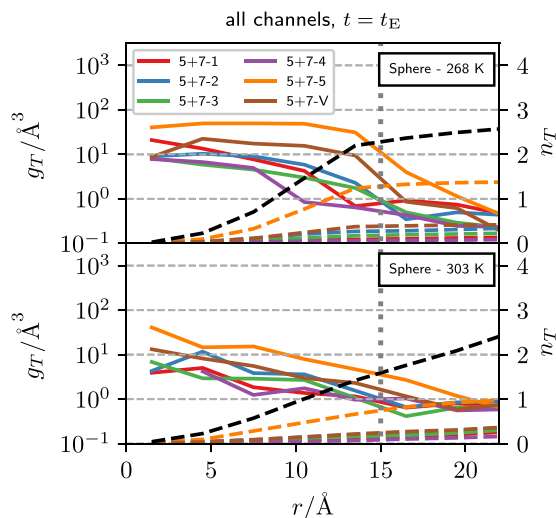

FIG. 9. Average number of 5+7 defects found in a cubic simulation box with 2880 molecules in equilibrium, alongside the probability of finding them in melting trajectories at time t_A. The results are shown in the top panel for $T = 268$ K and in the bottom panel for $T = 268$ K. The black error bars indicate a confidence level of 95%. $\Sigma(FRZ)$ and $\Sigma(EQU)$ indicate the average overall defect counts at time t_A and in equilibrium, respectively.

FIG. 10. Abundance of 5+7 type defects as a function of distance r from the center of the liquid nuclei that were used to seed the simulations. The gray vertical dashed line indicates the 15 Å radius of the liquid nucleus. The normalized defect densities $g_T(r)$ (solid lines) and the number $n_T(r)$ of defects within a spherical volume of radius r around the center of the nascent liquid nucleus (dashed lines) are shown. The black dashed line indicates the sum of n_T over all defect types, T.

the 57- and E-channels). At $T = 268$ K, there is up to a 50-fold excess over equilibrium in the density of 5+7 defects inside the radius of the liquid bubble that forms later on. The most abundant defect type inside this volume is the 5+7-5 defect with an average of 1.2 defects found within the radius of 15 Å, while the average total number of 5+7 defects within this volume is 2.2. The next most common defect within the nucleus volume is 5+7-V defects.

Under superheating conditions at $T = 303$ K, the excess is slightly less pronounced. Nevertheless, on average, we find 1.5 5+7 defects within 15 Å of the center of the forming bubble (5+7-5 defects: 0.7).

At $T = 268$ K, we observe a suppression of defects outside the nucleus volume relative to equilibrium. This suppression develops despite the fact that we used equilibrium configurations to seed the simulations, a procedure that enforces that the environment around the liquid domain is initially in equilibrium. The suppression of defect densities around the liquid nucleus indicates that the presence of the nucleus facilitates annealing of existing defects in the ice structure. We expect this effect to subside with increasing distance from the nucleus; however, the simulation box used in our simulations is not large enough to observe this return to average densities.

It is important to note here that precisely at coexistence, the radius of the critical nucleus is macroscopically large and that the nucleus radius of 15 Å used to seed our simulations was chosen for comparison with simulations performed under superheating. At the physically more realistic temperature of 303 K, the nucleus size used to seed simulations is chosen close to the critical nucleus size at this temperature. This allows us to extrapolate these simulation results to larger system sizes. Notably, under these conditions, there is no significant suppression of defect densities outside the nucleus volume.

B. Stage II: Mobile defects ($t_E < t \leq t_m$)

In this next stage, a mobile defect, i.e., either an L–D or an I–V pair, has formed. Which defect type has formed as a function of time relative to t_E is shown in Fig. 11 for different temperatures and cluster geometries. Just over half of the configurations (56%) observed at temperature $T = 303$ K contain an L–D pair, while the other ones contain an I–V pair. The fraction of configurations where an I–V pair is present then rapidly increases, reaching 98% 0.5 ns later. Similar behavior can be observed at $T = 268$ K regardless of the geometry of the liquid nucleus, giving us confidence that the observed timescale of roughly 0.5 ns for the time between forming an L–D pair and forming an I–V pair is largely independent of the geometry of the nucleus that forms and of temperature.

We have already shown in Sec. IV A that 5+7 defects tend to form within the eventual volume of the liquid nucleus. Figure 12 assesses the analogous behavior for mobile defects. While at $T = 268$ K, the first mobile defect that leads to melting forms inside the volume of the future liquid bubble 80% percent of the time, at $T = 303$ K, this share has declined to 64%. In both cases, the formation site is correlated with the center of the liquid nucleus [i.e., $g(r)$ is not flat]; however, the mobile defects are formed outside the nucleus volume in a significant number of trajectories.

To investigate the role of 5+7 defects in the creation of mobile defects, Fig. 13 shows an analysis of the defect densities around the site where a mobile defect forms. Only trajectories that proceed via the 57-channel are included in this analysis. Note that due to the sampling frequency of $(10\,\text{ps})^{-1}$ 5+7 defects with a lifetime smaller than 10 ps may not be detected in this analysis.

At both temperatures, we find that the density of 5+7 defects is enhanced close to the site where a mobile defect forms and we find the closest 5+7 defect within 6 Å in 71% of trajectories at 268 K (303 K: 57%). The average number of defects within 6 Å is 0.87 (303 K: 0.70). While at $T = 303$ K, all 5+7 defect types roughly
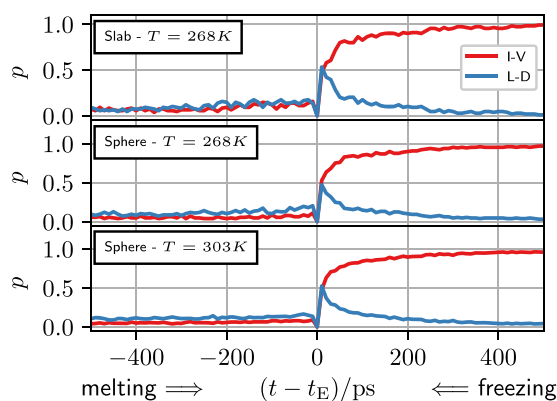

FIG. 11. Probabilities of finding the system with defects of type I–V or L–D, as functions of time. To calculate the time-dependent averages, trajectories have been aligned on time t_E. The states shown are detected in a mutually exclusive fashion, i.e., if there exists at least one I+V defect in the system, it is considered to be in the I+V state regardless of the number of L+D defects in the system. Note the similarity between the data obtained with the slab and the spherical geometry; the probabilities are largely independent of the shape of the cluster that is formed and of temperature. The probabilities of finding the system in states I+V and L+D by definition vanish at t_E and sum to one for $t > t_E$.
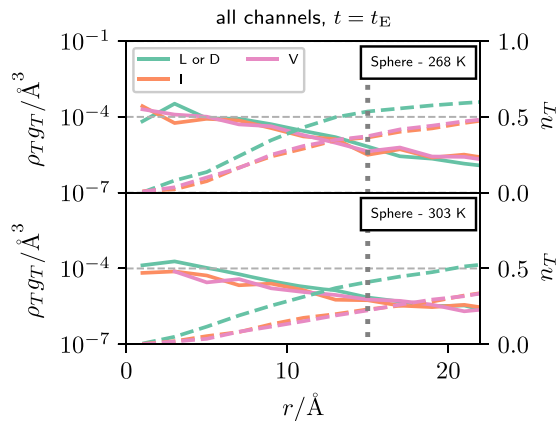
FIG. 12. Analysis of locations where L–D pairs and I and V defects form at time t_E relative to the center of the future liquid nucleus. The densities $\rho_T g_T(r)$ (solid lines) and the number of defects within a spherical volume of radius r around the center, $n_T(r)$ (dashed lines) are shown. The gray dotted line indicates the 15 Å radius of the nucleus volume.

contribute equally to the nearby defect population, at $T = 268$ K, there is a preference for 5+7-5 defects.

In trajectories that pass through the E-channel, mobile defects can also form close to the E defect. Hence, due to volume exclusion,
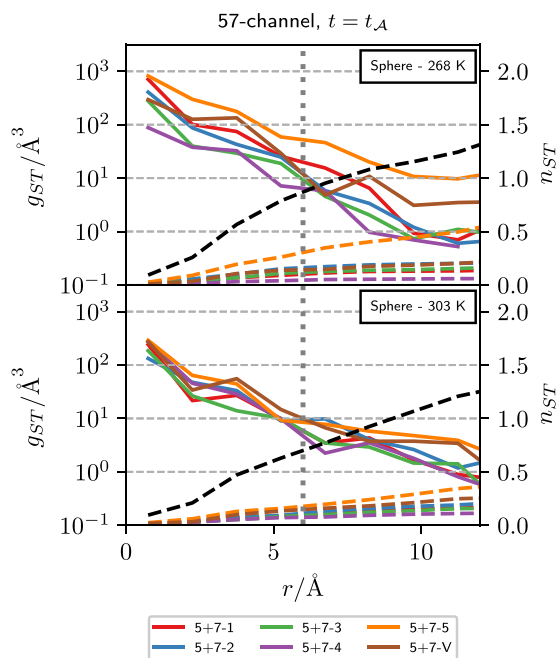

FIG. 13. Analysis of the positions of immobile defects around the site where a mobile defect pair of type I–V or L–D forms at time t_A. The pair-correlation functions g_{ST} (solid lines) and the average numbers of defects within a sphere of radius r, n_{ST} (dashed lines) are shown. Here, the reference defect type S are the defect types L, D, I, and V combined. The black lines are the sum over all defect types T. For clarity, the densities are shown on a semi-logarithmic scale, while n_{ST} is shown on a linear scale. Included in the analysis are configurations observed at time t_A in trajectories that proceed through the 57-channel. The corresponding analysis for trajectories that pass through the E-channel can be found in Appendix C (Fig. 27).
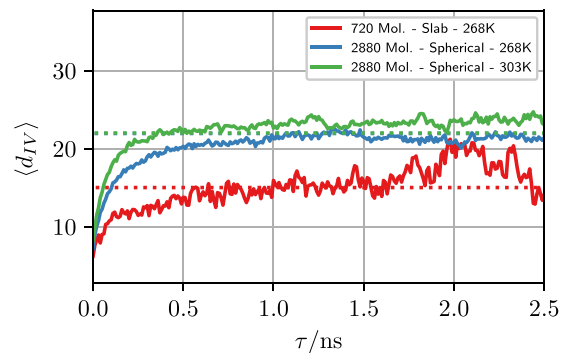

FIG. 14. Average distance between interstitial and vacancy as a function of time after their first creation in melting trajectories. Different simulation box and cluster geometries and temperatures are shown. The data shown have been calculated, considering configurations with a single I–V pair because the assignment of defects into pairs is unambiguous in this case. The dashed horizontal lines indicate the average distance between two randomly chosen points in the simulation box of the respective geometry.

the density of 5+7 defects around the site where a mobile defect forms is suppressed relative to the densities found in trajectories that pass through the 57-channel (see Appendix C, Fig. 27).

After an I–V defect pair has formed, we can track the motion of the two defects through the system. In Fig. 14, we show the average distances between the I and the V defects as a function of the elapsed time since the I–V pair has formed. The average is calculated over all configurations where a single I–V pair is present and each trajectory is included until time t_m. Within roughly 0.5 ns, the average distance between I and V defects approaches the value expected if one were to randomly place two particles in the simulation box (dotted lines in Fig. 14). In Fig. 15, we show the mean-square-displacements (MSDs), $\langle \mathbf{r}^2(t) \rangle$, of I and V defects during the same timespan. After a subdiffusive regime that lasts roughly 100 ps, $\langle \mathbf{r}^2(t) \rangle$ is close to linear, indicating that the defects freely diffuse through the system.
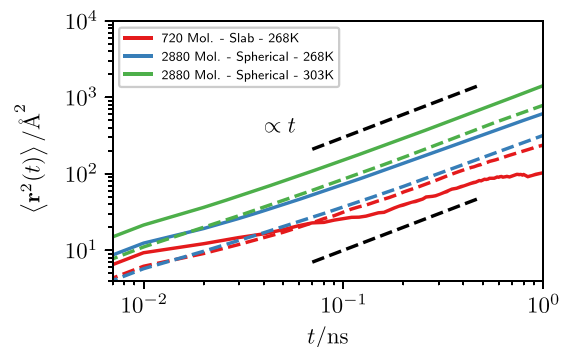

FIG. 15. Mean squared displacement of interstitial (solid lines) and vacancy defects (dashed lines) as observed in between the time of their first formation in melting trajectories up to the time where an extended liquid cluster forms, t_m. Different colors represent different system sizes and temperatures, and the black dashed lines are a guide to the eye, indicating linear behavior. Refer to Fig. 29 in Appendix C for a breakdown of $\langle \mathbf{r}^2(t) \rangle$ into different directional components $\langle r_i^2(t) \rangle$.

At 268 K, the ratio of the self-diffusion constants of interstitials and vacancies, D_I/D_V, equals ~2 (303 K: 1.75).

Together, these two datasets suggest that in the timespan between t_E and t_m both defect types independently diffuse through the simulation box. Indeed, we find that the non-equilibrium pair-correlation function between interstitials and vacancies obtained from configurations observed in this timespan is flat for distances larger than 10 Å (see Appendix C, Fig. 28).

To further investigate the role of I and V defects in the formation of a liquid nucleus, Fig. 16 shows the position of the I–V defect pair that is closest to the center of the nucleus volume at time t_m. We find that at 268 K in 98% of trajectories, there is at least a single I or V defect present inside the volume that later becomes the liquid nucleus. In 82% of trajectories, both the closest I and V defects are found in this volume. No significant imbalance between the two types of defects can be detected.

Note that Fig. 16 shows the probabilities of finding the I and V defect closest to the center of the liquid nucleus. Even for an ideal gas, the analogous distribution of the closest gas atom to a given location is not flat but has a maximum at a distance that is determined by the density of the gas. In Appendix B, we calculate the expected shape of this distribution and compare it to the data shown in Fig. 16. We find that there is a strong excess of I and V defects that are close to the liquid nucleus' center at time t_m relative to the expected distances based on uncorrelated density fluctuations.

At 303 K, the distribution of defect positions at time t_m is broader. Here, we find that in 90% of trajectories, at least one I or V defect is close to the center of the liquid nucleus at time t_m. V defects are found outside of the nucleus volume slightly more often than I defects. Note that the change in the slope of $n_5(t)$ at time t_m is less pronounced at $T = 303$ K than at 268 K, making the determination of t_m less 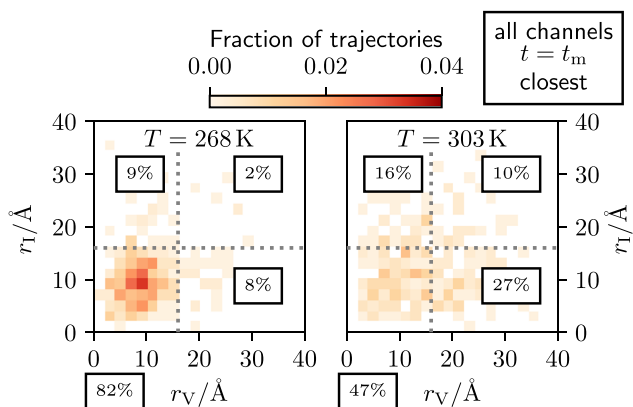precise. The broader distribution of defect positions is at least, in part, a consequence of this reduced precision. Nevertheless, here we also find a strong excess of I and V defects close to the center of the nucleus volume compared to a uniform distribution of defect positions.

C. Analysis of trajectories with slab-shaped clusters

The results obtained for slab-geometry trajectories at $T = 268$ K are largely similar to the ones obtained for spherical clusters at the same temperature. The transition proceeds via the E-channel in 57% ± 7% of trajectories, via the 57-channel in 42% ± 6%, and via the D-channel in 1% ± 1% of trajectories. This is in good agreement with the data obtained using spherical nuclei. Furthermore, also in this case, we find that mobile defects are preferentially formed around 5+7 defects and, in particular, around 5+7-5 defects, as can be seen in Fig. 17. Figure 11 demonstrates that the dynamics of forming I–V pairs after t_E are essentially identical to the ones observed with spherical cluster geometries.

A notable difference can be observed in the mobility of interstitial defects (Fig. 15): while all I and V defects exhibit some sub-diffusivity at timescales below 100 ps regardless of cluster geometry and temperature, this effect is strongly enhanced in interstitials that form when the liquid slab decays. The same is not true for the corresponding vacancies. An analysis that separates the contributions to the MSD in different lattice directions (see Appendix C, Fig. 29) also suggests that the movement of interstitials is hindered in both the direction orthogonal to the secondary prism plane and the direction orthogonal to the basal plane. An analysis of the positions of interstitial defects in these trajectories shows that they are confined to the volume that later becomes the slab-shaped liquid domain. This suggests that interstitials are more likely to not leave the volume of the liquid domain between the time they are formed and t_m if the cluster is slab-shaped.

We also analyzed the distribution of waiting times between t_E and t_m (Fig. 18). Here, we find a strong dependence of the waiting time on the simulation geometry where (at the same temperature of 268 K) the decay time τ decreases from 4.1 ns in the sphere-geometry simulations to 0.7 ns in the slab geometry. This reduction in waiting time may, in part, be a consequence of the pinning of interstitials in
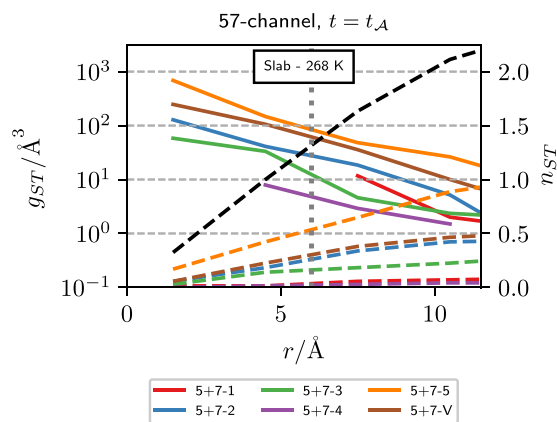

FIG. 16. Analysis of the locations of I–V defects at the time of formation of a liquid nucleus, t_m, using the center of the nucleus volume as reference. The fractions of trajectories in which the closest vacancy and the closest interstitial are found at distances r_V and r_I, respectively, are shown. The gray dotted lines indicate distances of 16 Å, just larger than the radius of the liquid nucleus volume of 15 Å. The percentages indicate the fraction of trajectories where the distances fall within the areas outlined by the gray dotted lines. The percentage of trajectories where both defects are within 16 Å of the center of the bubble is given in the lower left corner.

FIG. 17. Same analysis as shown in Fig. 13 for trajectories that end in a slab-shaped cluster in a 720 molecule system.
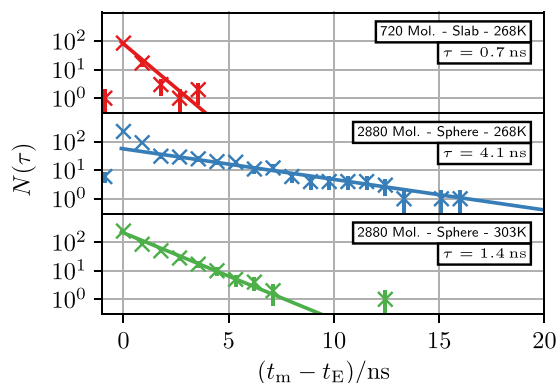
FIG. 18. Distributions of waiting time $t_m - t_E$ between the last time melting trajectories contain only defects of type 5+7, 455778 or extended defects of type E and the formation of an extended liquid cluster. Fits of the normalized exponential distribution $\tau^{-1} e^{-t/\tau}$ to all data points with $t_m - t_E > 0$ except for the simulations with a spherical cluster at $T = 268$ K where only the data points with $t_m - t_E > 1.5$ ns are used are shown.

the slab volume. However, based on calculations by Le Vot *et al.*[76] for one-dimensional systems, we would expect a significant dependence of waiting times on system size even in the case of equal diffusivities.

V. DISCUSSION AND OUTLOOK

We analyzed melting trajectories that were obtained using molecular dynamics simulations at coexistence and at 11% superheating. At superheating conditions, we observed the following sequence of events: (i) on average, 1.5 5+7 defects form within the volume that later become the liquid nucleus; 0.7 of these defects are 5+7-5 defects; (ii) in roughly 1/3 of trajectories, a larger topological defect structure that fulfills the ice rules (an E defect) forms; (iii) a mobile defect (an L–D or an I–V pair) forms close to either an E if one exists or close to a 5+7 defect; (iv) if an L–D pair has formed, an I–V pair forms within a timescale of 0.5 ns; (v) the I–V defects freely diffuse through the system; (vi) a liquid nucleus forms as a result of the interaction of the mobile defects with the 5+7 defects that are already present in the volume that is later occupied by the critical liquid nucleus.

Previous studies[14,18] have pointed out the role of 5+7 defects and larger defect structures in the melting mechanism of ice (modeled by the TIP4P water model). Our results show that the role of 5+7 defects is twofold: 5+7 defects are often found close to the site where the first mobile defect is formed, and second, 5+7 and E defects create a defective region that is susceptible to the formation of a liquid nucleus when a mobile interacts with it.

As the temperature is reduced, this defective region tends to become larger as demonstrated by the increase in melting trajectories that pass through the E-channel (52% at 268 K compared to 35% at 303 K) and the increase in the average number of 5+7 defects found close to the site where the liquid nucleus forms (2.2 compared to 1.5). This finding is consistent with the results of Mochizuki, Matsumoto, and Ohmine[18] who reported that at 19% superheating, 5+7 defects play a role in the formation of mobile defects, but no accumulation of these defects occurs prior to melting.

In Ref. 18, it was also shown that the formation of a separated interstitial–vacancy pair is the rate limiting step in the limit of high superheating. Under the conditions investigated in this paper, the rate limiting step is the formation of a liquid nucleus of critical size; however, also in this case, interstitials and vacancies play a crucial role in that they cooperate with 5+7 defects to form the initial liquid nucleus.

Prior to the formation of the liquid nucleus, the interstitials and vacancies diffuse freely through the simulation box. This has interesting implications when one wants to extrapolate simulation results to the thermodynamic limit. Since interstitials and vacancies are only weakly bound to each other, they are also not limited to form close to the accumulation of immobile defects they later interact with to form a liquid nucleus. This is underscored by the finding that even in the strongly confined environment of our simulation boxes, we find that a considerable number (36% at 303 K) of mobile defects form outside of the eventual volume of the liquid nucleus. This suggests that mobile defects may diffuse for considerable distances before they encounter an accumulation of other defects and form a liquid nucleus.

In summary, the results of our simulations further support the observation that interstitials and vacancies play an integral role in the microscopic mechanism of ice melting and establish that, as the degree of superheating becomes smaller, increasingly large immobile defect structures are present within the volume where the initial liquid nucleus forms prior to its formation. Interstitial and vacancy defects then interact with these immobile defects to form an initial liquid nucleus that later grows and melts the ice crystal. This adds ice to the list of solids with a melting mechanism that involves the prior accumulation of defects.

SUPPLEMENTARY MATERIAL

The defect analysis for the trajectories used in this study is available for download in the supplementary material.

APPENDIX A: THE DEFECT DETECTION SCHEME

Figure 19 presents the scheme used to detect defects in configurations throughout this paper. We start by minimizing the potential energy of a given configuration by setting the momenta of atoms to zero and annealing the system to a temperature of 1 K using a Langevin thermostat. The cooled configurations are then analyzed
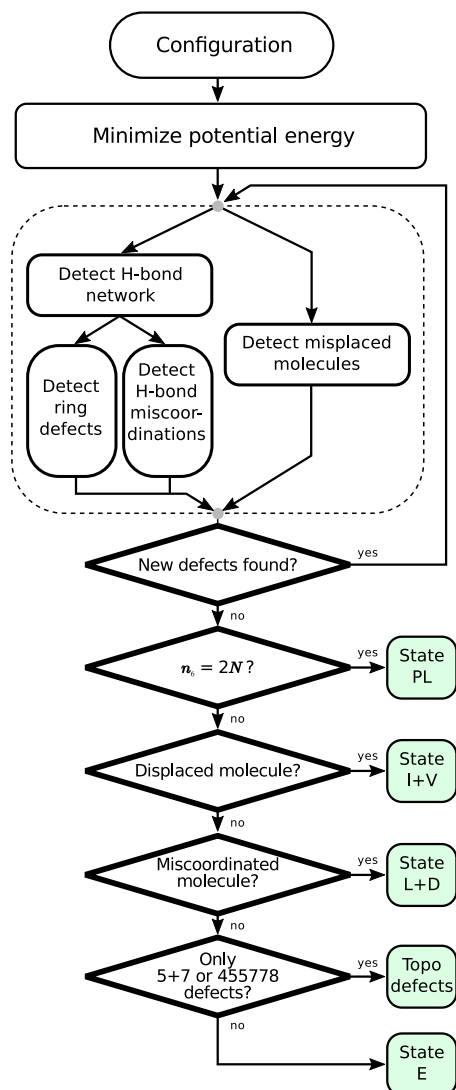
FIG. 19. Sketch of the algorithm used to detect defects in the ice Ih structure throughout this paper. It consists of two stages: in the first stage, the configuration is analyzed and deviations from the pristine ice Ih structure are marked using different approaches (see the descriptions in this appendix). In the second stage, a state is assigned based on the deviations found. n_6 and N are the number of six-rings detected and the number of molecules in the system, respectively.

using a number of algorithms, the details of which can be found below.

In the course of this analysis, deviations from a perfect hexagonal ice lattice are detected and, if possible, attributed to known defect structures. The attribution of anomalies in the H-bond network to a defect may reveal other known defect structures, and hence, the analysis is run iteratively until no new defects are found.

The last part of the detection scheme assigns each configuration to a global state based on the criteria shown in Fig. 19.

1. Detection of the hydrogen bond network and of L–D defects

Hydrogen bonds are detected in annealed configurations using the HB2 criterion proposed in Ref. 75 using a custom implementation for the LAMMPS simulation package. This criterion considers a set of two oxygen atoms and one hydrogen atom to form a hydrogen bond if the O–O distance is smaller than 3.5 Å and the O–H–O angle is larger than 140°. This analysis yields a network of HBs from which an undirected graph is constructed using the *networkx* python package.[77] This package provides a large number of graph analysis functions we use to carry out parts of the following analysis.

The simplest task is the count of the number of HBs each molecule donates (n_d) and accepts (n_a). In the perfect lattice, one expects $n_d = n_a = 2$ for all molecules. If either of these two criteria is not fulfilled for a given molecule, it is considered to be *miscoordinated*. This can be the result of both, an interstitial–vacancy and an L–D pair. If no I–V pair is found, an L–D pair is considered to be present. Note that L–D pairs where L and D stay bound to each other form frequently in hexagonal ice and that their separation is associated with an additional free energy barrier.[30] Here, we do not distinguish between bound and unbound L–D pairs.

2. Detection of topological defects in the hydrogen bond network

Defects in the ice Ih structure such as the 5+7 defect can present very similar molecular environments to what can be seen in the perfect lattice. The hydrogen bonds in the region of 5+7 defects fulfill the ice rules, and the changes in the bond angles are comparatively small. This makes it hard to detect defects based on the positions of molecules around a given molecule alone. In order to detect such defects, algorithms that are based on detecting the topology of the hydrogen-bond network have been proposed.[14,18] Here, we use a similar scheme that analyzes the rings that can be found in the network formed by H-bonds in a hexagonal ice crystal.

We define a ring in the H-bond network as a path that leads from a molecule to itself, where we move from one molecule to the next following H-bonds. We do not allow the path to cross itself, i.e., each intermediate molecule may only be visited once. In principle, with the use of periodic boundary conditions, a given molecule may be part of an infinite number of rings of arbitrary size. To make our ring definition unique, we restrict ourselves to a smaller set of rings: the set of shortest rings around each molecule that contain the molecule and two of its neighbors.

Consider a molecule O that is H-bonded to n_{hb} other molecules, as depicted in Fig. 20. In order to construct the shortest rings, we iterate through all $n_{hb}(n_{hb} - 1)/2$ pairs of neighbor molecules A and B. For each of these pairs, we look for the shortest path between A and B that does not pass through the molecule O. These paths are constructed using Dijkstra's algorithm[78] as implemented in networkx. There may be more than one path of the same length and—in order to make the result reproducible—we include all rings in the following analysis. Repeating this procedure for all molecules in the crystal yields a set of rings and a list of rings that pass through a given molecule, which we use to identify defect types.

5+7 defects[30,34] can be formed in the *c*-plane ("horizontally") or parallel to the prism or the secondary prism plane. The basic ideas
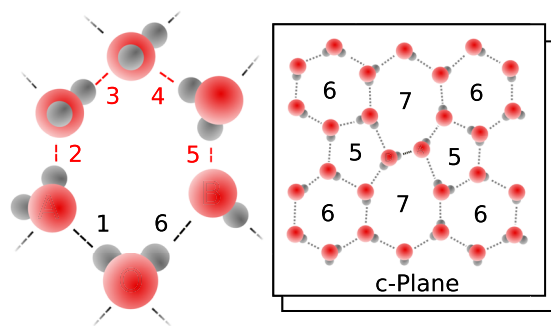
FIG. 20. Illustration of the ring-detection scheme used to detect topological defects in the ice Ih crystal lattice. Left: rings are detected by selecting a molecule O and all pairs of molecules it is connected to via hydrogen bonds, *A* and *B*. Rings are then found by finding all shortest paths from *A* to *B* that do not pass through edges O–*A* and O–*B* (H-bonds 1 and 6, respectively). Subsequently, the ring is closed by adding molecule O to the ring. Right: illustration of a 5+7 type 3 defect[30,34] as seen projected onto the *c*-plane of the crystal structure. The numbers within the rings count the number of molecules that are within the respective ring. *D* and *A* mark what we call the donor and the acceptor molecule of the 5+7 defect.

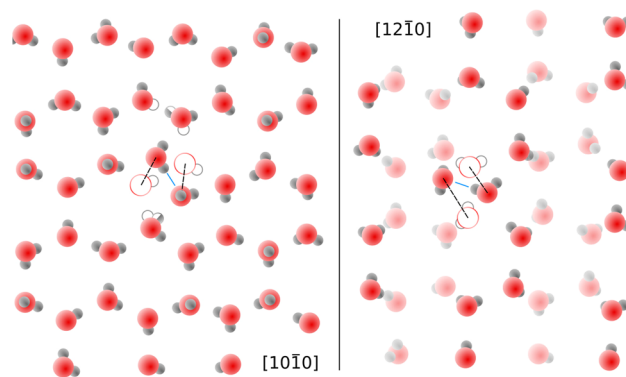**FIG. 21.** Sketches of vertical 5+7 defects seen projected onto the prism ([12$\bar{1}$0]) and the secondary prism plane. The molecules that are only outlined indicate the position of the respective molecule in the perfect lattice. The shaded molecules on the right are roughly 2 Å closer to the reader than the unshaded molecules.

for detecting these defects are demonstrated in the right panel of Fig. 20: the donor and the acceptor molecule of the 5+7 defect must be part of at least two seven-rings and one five-ring in the *c*-plane alone. If we also consider the rings that pass through molecules in the adjacent plane, donor and acceptor molecules of a horizontal 5+7 defect are part of eight seven-rings, four six-rings, and two five-rings. This ring fingerprint is used to pick out 5+7 defects from a given configuration.

In order to further distinguish the different types of 5+7 defects described in Ref. 30, we also take into account where the other H-bonds of donor and acceptor are pointing (cf. Table III). The HBs of each molecule can either both point toward a molecule in the same layer (in-plane, IP) or one of the HBs points toward a molecule in the same plane and the other one to a molecule in an adjacent plane (out-of-plane, OP). Hence, to detect these different types, one needs to associate each molecule with a layer of the configuration. Type 3 and type 5 defects are additionally distinguished by the direction the in-plane H-bonds of donor and acceptor are pointing to.

In addition to horizontal 5+7 defects with donor–acceptor pairs that are in the same *c*-plane, vertical 5+7 defects can also be observed. Figure 21 shows two examples of such defects where the donor–acceptor pair lies in the prism ([10$\bar{1}$0]) and the secondary prism-plane ([12$\bar{1}$0]). Together, these defects are referred to as 5+7-V defects.

The algorithm that assigns each molecule to a layer is sketched in Fig. 22. First, the configuration is sliced along the *x* axis into slices that contain two layers of molecules. Next, the configuration is squashed, leaving only the *z*-coordinates of the oxygen atom of each molecule. These *z*-positions are clustered using the *AgglomerativeClustering* hierarchical clustering algorithm implemented in the scikit-learn python package.[79,80] As metric we use the distances between the molecules taking periodic boundary conditions into account. *Average* linkage is used, and we set the number of clusters to be found to the number of *c*-planes expected (6 in the example
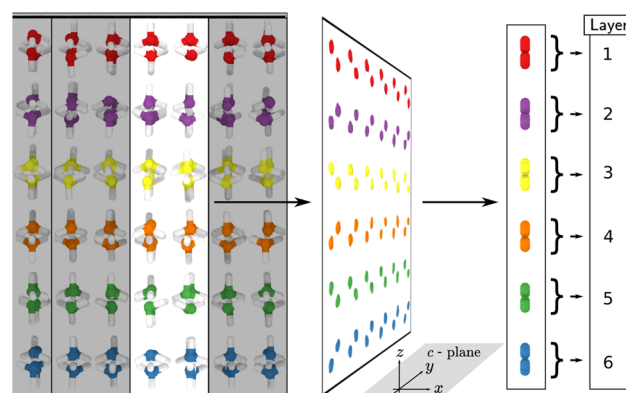
TABLE III. Table of horizontal 5+7 defect types. The 5+7 defects can be distinguished by the orientation of the donor and acceptor molecule. They can either be in-plane (IP, both outgoing hydrogen bonds are formed with molecules in the same *c*-plane) or out-of-plane (OP, one H-bond to a molecule in the same *c*-plane and one with a molecule in an adjacent plane).[30] To distinguish type 3 from type 5 defects, the orientation of the in-plane O–H vector of both molecules has to be taken into account.

Type	Donor	Acceptor	In-plane O–H vectors
1	OP	OP	⋯
2	IP	IP	⋯
3	IP	OP	Antiparallel
4	OP	IP	⋯
5	IP	OP	Orthogonal

FIG. 22. Schematic representation of the algorithm used to assign molecules to layers. The white area on the left indicates a single slice used for the analysis, which is carried out on each slice separately. The resulting layers are then matched across neighboring slices in order to arrive at a consistent numbering of layers throughout the system.

shown). This procedure is applied to each slice along the *x*-direction separately, and afterward, the resulting layers are matched across the slices using the center-of-mass of the layers found in each slice to identify which neighboring layers belong together. This slicing procedure is used to accommodate capillary waves that modulate the

layer positions as one moves along the *x*-direction. The result of this procedure is that each molecule is now associated with a number that identifies the layer the molecule is in.

The detection of defect structures is complicated if multiple defects are close to each other. In this case, the ring counts of
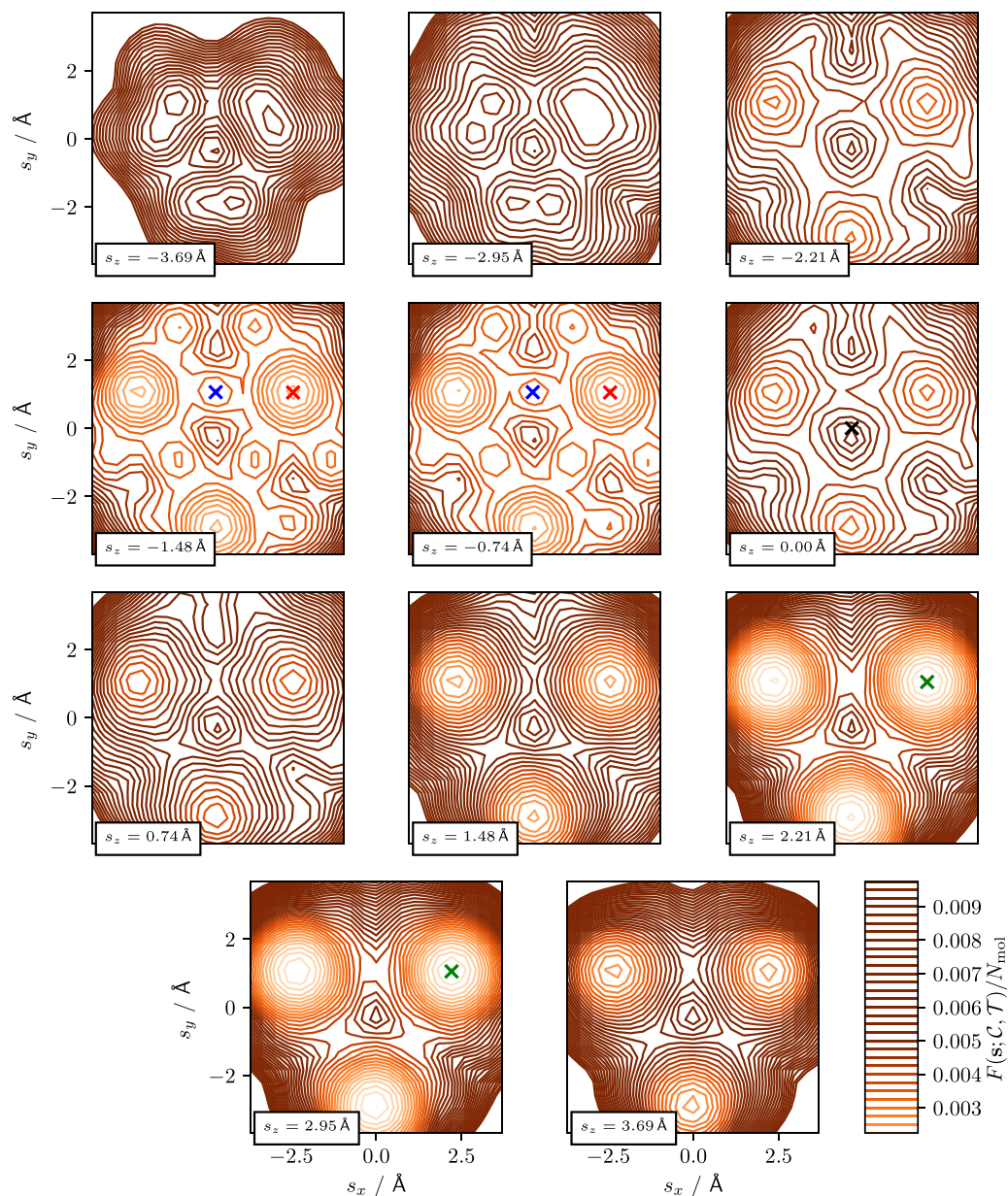

FIG. 23. Example contour plot of the summed mean-square deviation $F(\mathbf{s}; \mathcal{C}, \mathcal{T})$ as a function of the shift vector $\mathbf{s}$. The example shows the landscape used to align a configuration that contains 720 molecules that are arranged in a mostly pristine lattice except for a small number of defects. The black cross indicates the vector $\mathbf{s}$ at which the first optimization is started. The blue cross indicates the value of $\mathbf{s}$ after the first optimization ($s_z^{(1)} = -1.14$, in between the two levels shown), which corresponds to a shift of the template configuration relative to the optimal alignment along both the prism and the c axis. The red cross indicates another local minimum, which corresponds to an alignment where the crystal is shifted by exactly one layer in the *z*-direction (i.e., layers of type A are on top of type B layers). The green cross marks the value of $\mathbf{s}$ after the second optimization, which finds the best alignment of the two configurations ($s_z^{(2)} = 2.56$).

different defects can influence each other. To be able to detect 5+7 defects that are close to each other, we compiled a list of different ring fingerprints by inspecting a large number of configurations with unknown defect structures.

3. Detection of interstitial-vacancy defects

In order to identify interstitials and vacancies in configurations, we compare the minimized configuration with a template configuration that contains a perfect ice Ih crystal of the same size. This procedure is a simplified version of the method previously employed by Mochizuki, Matsumoto, and Ohmine[18] where we skip the calculation of the edit distance.

To do so, we first align the configuration that is analyzed (denoted by $\mathcal{C}$) with the template configuration ($\mathcal{T}$) using the following procedure:

1. Identify the set of molecules in the system, $\mathcal{I}$ whose contribution to the potential energy in the system is lower than -16 kcal mol^{-1}. These molecules are in ice-like configurations.
2. Align these molecules with their counterparts in the template configuration. To do so, use a global shift of the configuration, $\mathbf{s}$, to minimize the summed mean square displacement

$$F(\mathbf{s}; \mathcal{C}, \mathcal{T}) = \sum_{i \in \mathcal{I}} \left| \mathbf{x}_i^{\mathcal{C}} - \mathbf{X}(\mathbf{x}_i^{\mathcal{C}}; \mathcal{T} + \mathbf{s}) \right|^2, \quad (A1)$$

wherein $\mathbf{X}(\mathbf{x}_i^{\mathcal{C}}; \mathcal{T} + \mathbf{s})$ is the position of the closest neighbor of molecule i in $\mathcal{C}$ that can be found in $\mathcal{T}$ that has been shifted by $\mathbf{s}$. As the position of a molecule, we use the position of the oxygen. This yields an optimized vector $\mathbf{s}^{(1)}$ and a value of the MSD function $F_{\min}^{(1)}$.

3. If $F_{\min}^{(1)}/N_{\mathrm{mol}}$ is larger than a threshold value $f_{\max} = 0.03$ Å^2, shift the vector $\mathbf{s}^{(1)}$ by one layer distance in the z-direction and redo the optimization using this shifted vector as the initial condition.
4. If $F_{\min}^{(2)}/N_{\mathrm{mol}}$ is still larger than $f_{\max}$, again, shift the output vector of the previous configuration and redo the optimization. If this optimization does not succeed, the procedure fails.

The reason for this multi-step procedure becomes clear when we examine the landscape $F(\mathbf{s}; \mathcal{C}, \mathcal{T})$ shown in Fig. 23 that consists of multiple local and global minima. Figure 24 shows the alternating layers that a hexagonal ice crystal is comprised of. Depending on the initial shift of the template, an optimization of F can lead into the local minima marked by green and red dots in Fig. 23. These minima correspond to shifts where the template configuration is offset by exactly one layer width along the z-direction, i.e., such that the B-layers of the template is found on top of the A-layers of the configuration.

Fortunately, at the densities observed in the simulations presented here, there are no local minima if $\mathbf{s}$ is chosen such that the layers match. Hence, by performing multiple optimizations where $\mathbf{s}$ is shifted by one layer in the z-direction after the first optimization, we find the correct alignment of the two configurations. Only in some cases, the third optimization is required because the initial $\mathbf{s}$ of the first optimization finds a small local minimum. The optimizations are performed using the BFGS algorithm as implemented in the scipy python package.[81]
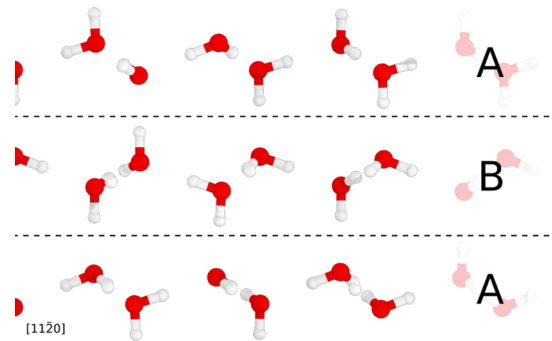

FIG. 24. Ice Ih structure as seen facing the prism face. The structure is comprised of alternating layers A and B with respect to the position of the oxygen atoms. Protons are arranged in a random pattern that fulfills the ice rules.

Given the optimal alignment of $\mathcal{T}$ to $\mathcal{C}$, we can now assign each molecule in $\mathcal{C}$ to the site in $\mathcal{T}$ that is its nearest neighbor and count the number of times that each site in $\mathcal{T}$ is found as a nearest site. In a defect-free crystal, this count, $n_i^{\mathcal{T}}$, is equal to one for each site. However, if there are interstitial–vacancy pairs in the system, some of $n_i^{\mathcal{T}}$s are different from one. These sites are marked as template mismatches (TMMs).

To avoid false positives that come about due to molecules translating only locally (e.g., to form a 5+7 defect), an additional step is performed where, if a site with an excess molecule and a site with a missing molecule are neighbors of each other (in the sense that they are H-bonded to each other in the template configuration), this pair of template mismatches is ignored.

The TMMs that remain are marked as interstitials if $n_i^{\mathcal{T}} > 1$ and as vacancies if $n_i^{\mathcal{T}} < 1$.

APPENDIX B: IDEAL GAS CLOSEST DISTANCE DISTRIBUTION

In Fig. 16 (Sec. IV), we presented the histograms of the distances of the closest interstitial and the closest vacancy to the center of the liquid domain. It is important to note that even for a homogeneous distribution of defects (i.e., an ideal gas), the distribution of the closest defect is not uniform but instead has a maximum that is determined by the density of the gas and the volume of the simulation box.[82]

Consider the probability of finding the atom closest to a given location at a distance that falls into the interval $[r, r + dr]$ for an ideal gas. Because the atom positions are not correlated with each other in the ideal gas, this probability can be decomposed into three factors: the probability that a specific atom is found at a distance inside $[r, r + dr]$ times the probability that none of the $N - 1$ other atoms are closer than $r + dr$ times the number of particles that can be picked. With the total system volume V and the sphere volume $v(r) = 4\pi r^3/3$, the probability density p_{id} can be written as

$$p_{\mathrm{id}}(r)\,dr = \left(\frac{V - v(r)}{V}\right)^{N-1} \frac{4\pi r^2}{V} N\,dr$$

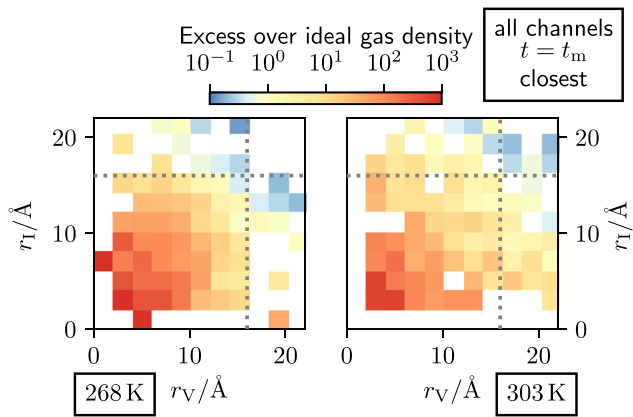$$= 4\pi \rho r^2 \exp\left[(N-1)\log(1 - v(r)/V)\right] dr. \quad (B1)$$

FIG. 25. Analysis of the locations of I–V defects at the time of formation of a liquid nucleus, t_m, using the center of the nucleus volume as reference. The ratios between the observed joint probability of the closest vacancy and the closest interstitial, $P(r_V, r_I)$, to the expected density, assuming that the defects follow ideal gas statistics $P_{id}(r_V, r_I)$, are shown. $P(r_V, r_I)$ is shown in Fig. 16 in the main part.

In the thermodynamic limit ($N \to \infty$, $V \to \infty$ such that $N/V = \rho$ = const.), this can be approximated by

$$p_{id}(r)\, dr \approx 4\pi\rho r^2 \exp\left[-4\pi\rho r^3/3\right]. \tag{B2}$$

To calculate the joint probability that the closest ideal gas-like interstitial is found at a distance in the interval $[r_I, r_I + dr_I]$ and the closest ideal gas vacancy in the interval $[r_V, r_V + dr_V]$, we multiply the two probabilities

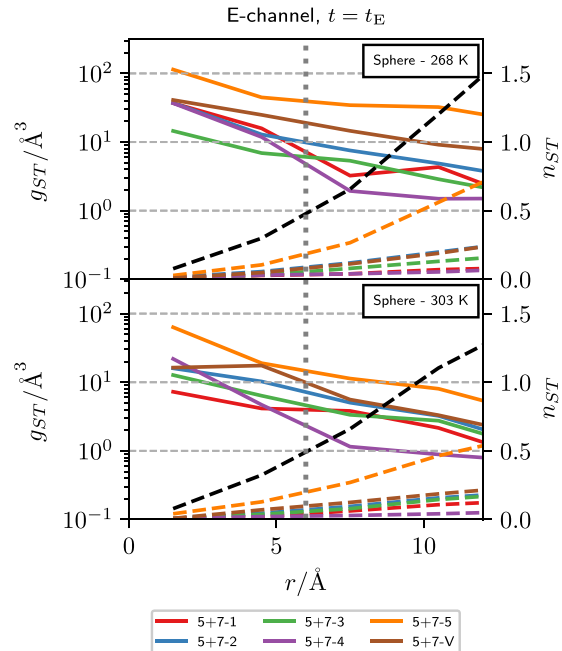$$p_{id}(r_V, r_I)\, dr_V\, dr_I = p_V(r_V)p_I(r_I)\, dr_V\, dr_I. \tag{B3}$$

FIG. 27. Analysis of the positions of the closest defect to the site where L–D or I–V defects form at time t_E in trajectories that pass through the E-channel. The pair-correlation functions g_{ST} (solid) and the average numbers of defects within a sphere of radius r, n_{ST} (dashed lines) are shown. The black lines are the sum over all defect types T. For clarity, the densities are shown on a semi-logarithmic scale, while n_{ST} is shown on a linear scale.

Integration over the size of a histogram bin, Δr, yields the probabilities

$$P(r_V, r_I) = \int_{r_V}^{r_V + \Delta r} \int_{r_I}^{r_I + \Delta r} dr_V\, dr_I\, p_V(r_V)p_I(r_I). \tag{B4}$$
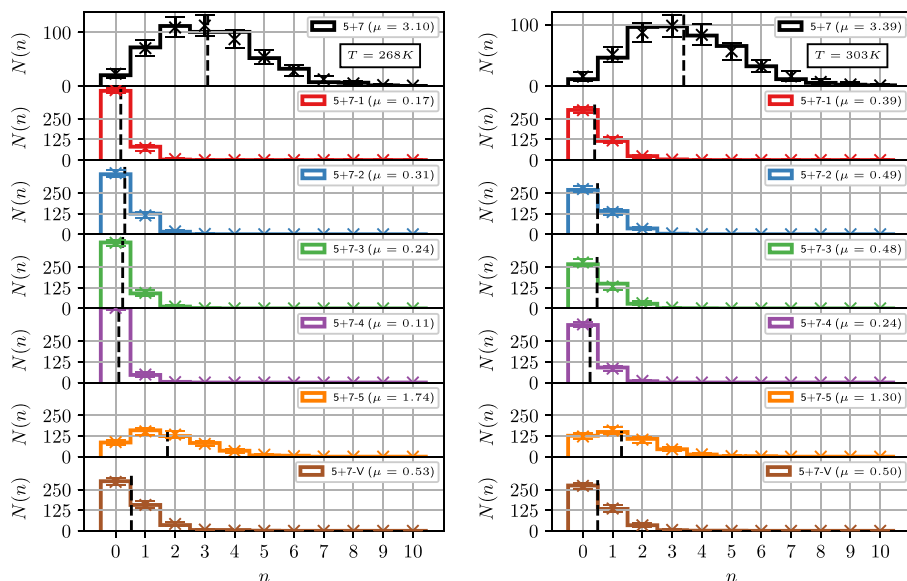

FIG. 26. Statistics of the number of immobile defects of given types found at time t_A in freezing/melting trajectories that end in a spherical nucleus at temperature $T = 268$ K (left) and $T = 303$ K (right). The histograms indicate the counts observed in trajectories, and the vertical dashed line indicates the average defect count, μ. The crosses represent the expected counts based on a Poisson distribution with average count μ, and the error bars indicate the interval into which 95% of counts would fall based on this distribution.
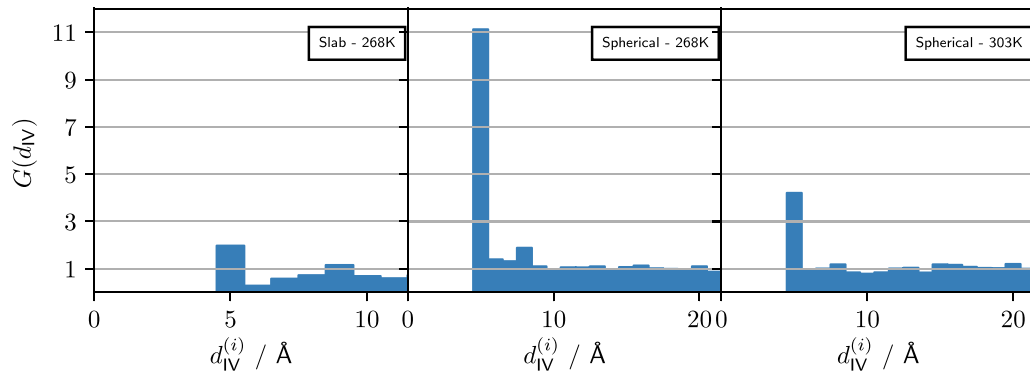
FIG. 28. Pair-correlation functions G between interstitial and vacancy defects found in melting trajectories between time t_E and t_m. Note that only configurations where a single defect pair is present are included in the analysis and that the trajectories in the underlying ensemble start at the creation of a defect pair. As a consequence, G is not an equilibrium property since finding defect pairs a short distance apart is guaranteed. The distribution has been normalized with the density $\rho = V^{-1}$, where V is the box volume resulting in an assumed number density of the defects of V^{-1}.

To compare this expected distribution to the ones obtained from melting trajectories, we need to determine the average densities ρ_I and ρ_V at time t_m. To do so, we count the total number of defects of type i observed in configurations at time t_m, n_i, and divide by the number of trajectories n and the average volume of the simulation box V.

Figure 25 shows the ratio between the observed probabilities, $P(r_V, r_I)$, and the ideal gas probabilities $P_{id}(r_V, r_I)$. There is a strong
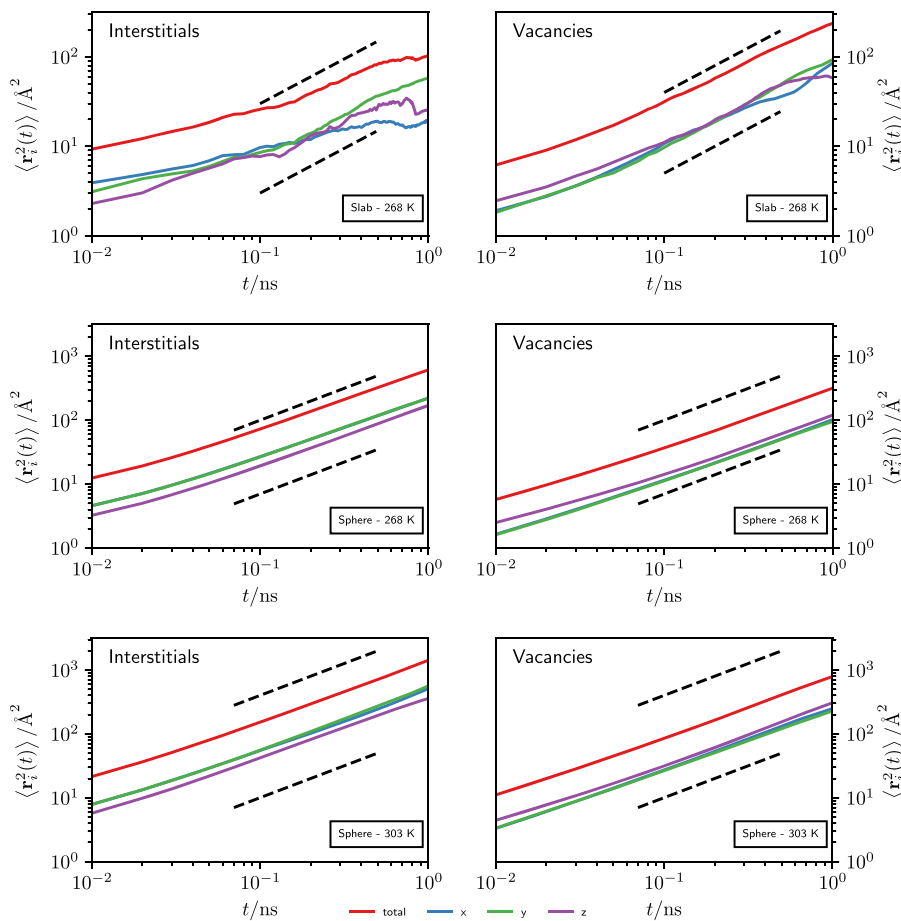

FIG. 29. Mean-square-displacements calculated for interstitials and vacancies at different temperatures and for different cluster geometries. The x, y, and z component are shown separately. The dashed lines indicate diffusive behavior where $r^2(t) \sim t$.

excess of configurations where interstitials and vacancies are close to the center of the liquid nucleus compared to what is expected in an ideal gas.

APPENDIX C: DATA SUPPLEMENT

Here, we present a number of additional graphs that were not included in the main part for clarity.

- Figure 26 shows the immobile defect counts found at time t_A in spherical-geometry trajectories obtained at 268 and 303 K.
- Figure 27 shows the defect densities of 5+7 defects around the site where a mobile defect forms for trajectories that pass through the E-channel. See Figs. 13 and 17 for the corresponding plots that include all trajectories.
- Figure 28 shows the non-equilibrium pair-correlation functions for I and V defects obtained in the timespan between t_E and t_m.
- Figure 29 shows the one-dimensional contributions to the mean-square-displacements of I and V defects in the same time period but only when there is a single I–V pair present.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] R. W. Cahn, Nature **323**, 668 (1986).
[2] Q. S. Mei and K. Lu, Prog. Mater. Sci. **52**, 1175 (2007).
[3] J. Daeges, H. Gleiter, and J. H. Perepezko, Phys. Lett. A **119**, 79 (1986).
[4] F. A. Lindemann, Phys. Z. **11**, 609 (1910).
[5] M. Born, J. Chem. Phys. **7**, 591 (1939).
[6] H. Mori, H. Okamoto, and S. Isa, Physica **73**, 237 (1974).
[7] H. J. Fecht and W. L. Johnson, Nature **334**, 50 (1988).
[8] J. L. Tallon, Nature **342**, 658 (1989).
[9] Z. H. Jin, P. Gumbsch, K. Lu, and E. Ma, Phys. Rev. Lett. **87**, 055703 (2001).
[10] A. B. Belonoshko, N. V. Skorodumova, A. Rosengren, and B. Johansson, Phys. Rev. B **73**, 012201 (2006).
[11] K. Lu and Y. Li, Phys. Rev. Lett. **80**, 4474 (1998).
[12] H. J. Fecht, Nature **356**, 133 (1992).
[13] M. Forsblom and G. Grimvall, Nat. Mater. **4**, 388 (2005).
[14] D. Donadio, P. Raiteri, and M. Parrinello, J. Phys. Chem. B **109**, 5421 (2005).
[15] S.-N. Luo, L. Zheng, A. Strachan, and D. C. Swift, J. Chem. Phys. **126**, 034505 (2007).
[16] X.-M. Bai and M. Li, Phys. Rev. B **77**, 134109 (2008).
[17] Z. Wang, F. Wang, Y. Peng, Z. Zheng, and Y. Han, Science **338**, 87 (2012).
[18] K. Mochizuki, M. Matsumoto, and I. Ohmine, Nature **498**, 350 (2013).
[19] A. Samanta, M. E. Tuckerman, T.-Q. Yu, and W. E, Science **346**, 729 (2014).
[20] S. Liang, L. Yi, and D. Liang, J. Phys. Chem. C **118**, 28542 (2014).
[21] H. Iglev, M. Schmeisser, K. Simeonidis, A. Thaller, and A. Laubereau, Nature **439**, 183 (2006).
[22] M. Schmeisser, A. Thaller, H. Iglev, and A. Laubereau, New J. Phys. **8**, 104 (2006).
[23] M. Schmeisser, H. Iglev, and A. Laubereau, Chem. Phys. Lett. **442**, 171 (2007).
[24] M. Schmeisser, H. Iglev, and A. Laubereau, J. Phys. Chem. B **111**, 11271 (2007).
[25] S. Fanetti, N. Falsini, P. Bartolini, M. Citroni, A. Lapini, A. Taschin, and R. Bini, J. Phys. Chem. Lett. **10**, 4517 (2019).
[26] M. de Koning and A. Antonelli, J. Chem. Phys. **128**, 164502 (2008).
[27] L. Pauling, J. Am. Chem. Soc. **57**, 2680 (1935).
[28] V. F. Petrenko and R. W. Whitworth, *Physics of Ice* (OUP Oxford, New York, 1999).
[29] J. D. Bernal and R. H. Fowler, J. Chem. Phys. **1**, 515 (1933).
[30] N. Grishina and V. Buch, J. Chem. Phys. **120**, 5217 (2004).
[31] B. Geil, T. M. Kirschgen, and F. Fujara, Phys. Rev. B **72**, 014304 (2005).
[32] M. Eigen and L. De Maeyer, Proc. R. Soc. London, Ser. A **247**, 505 (1958).
[33] N. Bjerrum, Science **115**, 385 (1952).
[34] R. Podeszwa and V. Buch, Phys. Rev. Lett. **83**, 4570 (1999).
[35] H. Tanaka and U. Mohanty, J. Am. Chem. Soc. **124**, 8085 (2002).
[36] J. Kolafa and M. Ončák, J. Phys. Chem. C **114**, 20518 (2010).
[37] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[38] P. G. Bolhuis, C. Dellago, and D. Chandler, Faraday Discuss. **110**, 421 (1998).
[39] C. Dellago, P. G. Bolhuis, and P. L. Geissler, in *Advances in Chemical Physics* (John Wiley & Sons, Inc., Hoboken, NJ, 2002), Vol. 123, pp. 1–78.
[40] X. M. Bai and M. Li, J. Chem. Phys. **122**, 224510 (2005).
[41] X.-M. Bai and M. Li, J. Chem. Phys. **124**, 124707 (2006).
[42] J. R. Espinosa, E. Sanz, C. Valeriani, and C. Vega, J. Chem. Phys. **141**, 18C529 (2014).
[43] J. R. Espinosa, C. Vega, C. Valeriani, and E. Sanz, J. Chem. Phys. **144**, 034501 (2016).
[44] R. Handel, R. L. Davidchack, J. Anwar, and A. Brukhno, Phys. Rev. Lett. **100**, 036104 (2008).
[45] R. L. Davidchack, R. Handel, J. Anwar, and A. V. Brukhno, J. Chem. Theory Comput. **8**, 2383 (2012).
[46] J. R. Espinosa, C. Vega, and E. Sanz, J. Phys. Chem. C **120**, 8068 (2016).
[47] H. Nada and Y. Furukawa, J. Phys. Chem. B **101**, 6163 (1997).
[48] H. Nada and Y. Furukawa, J. Cryst. Growth **283**, 242 (2005).
[49] D. Rozmanov and P. G. Kusalik, J. Chem. Phys. **137**, 094702 (2012).
[50] J. R. Espinosa, C. Navarro, E. Sanz, C. Valeriani, and C. Vega, J. Chem. Phys. **145**, 211922 (2016).
[51] J. L. F. Abascal, E. Sanz, R. García Fernández, and C. Vega, J. Chem. Phys. **122**, 234511 (2005).
[52] T. F. Miller, M. Eleftheriou, P. Pattnaik, A. Ndirango, D. Newns, and G. J. Martyna, J. Chem. Phys. **116**, 8649 (2002).
[53] H. Kamberaj, R. J. Low, and M. P. Neal, J. Chem. Phys. **122**, 224114 (2005).
[54] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).
[55] LAMMPS version has been used to generate trajectories, August 22, 2018.
[56] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[57] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, Mol. Phys. **87**, 1117 (1996).
[58] G. J. Martyna, M. L. Klein, and M. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).
[59] R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (CRC Press, 1988).
[60] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[61] R. B. Ayala and V. Tchijov, Can. J. Phys. **81**, 11 (2003).
[62] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992).
[63] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).
[64] T. Okabe, M. Kawata, Y. Okamoto, and M. Mikami, Chem. Phys. Lett. **335**, 435 (2001).
[65] Y. Mori and Y. Okamoto, J. Phys. Soc. Jpn. **79**, 074003 (2010).
[66] K. Leung and R. K. P. Zia, J. Phys. A: Math. Gen. **23**, 4593 (1990).
[67] A. Tröster, C. Dellago, and W. Schranz, Phys. Rev. B **72**, 094103 (2005).
[68] T. Neuhaus and J. S. Hager, J. Stat. Phys. **113**, 47 (2003).
[69] C. Moritz, A. Tröster, and C. Dellago, J. Chem. Phys. **147**, 152714 (2017).
[70] E. Małolepsza and T. Keyes, J. Chem. Theory Comput. **11**, 5613 (2015).

[71] C. Moritz, M. Sega, M. Innerbichler, P. L. Geissler, and C. Dellago, Phys. Rev. E **102**, 062801 (2020).

[72] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[73] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).

[74] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, and F. Noé, J. Chem. Theory Comput. **11**, 5525 (2015).

[75] A. Vila Verde, P. G. Bolhuis, and R. K. Campen, J. Phys. Chem. B **116**, 9467 (2012).

[76] F. Le Vot, S. B. Yuste, E. Abad, and D. S. Grebenkov, Phys. Rev. E **102**, 032118 (2020).

[77] A. A. Hagberg, D. A. Schult, and P. J. Swart, in Proceedings of the 7th Python Science Conference (SciPy 2008), edited by G. Varoquaux, T. Vaught, and J. Millman (2008), pp. 11–15.

[78] E. W. Dijkstra, Numer. Math. **1**, 269 (1959).

[79] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011).

[80] Scikit-learn version 0.18.2 is used.

[81] E. Jones, T. Oliphant, P. Peterson *et al.*, SciPy: Open source scientific tools for Python.

[82] P. Hertz, Math. Ann. **67**, 387 (1909).