Unbiased exploration of the transition region in ice nucleation using the NpH ensemble

Sebastian Falkner, Nadine Schwierz, Pablo Montero de Hijes

Angaben zur Veröffentlichung / Publication details:

Falkner, Sebastian, Nadine Schwierz, and Pablo Montero de Hijes. 2026. "Unbiased exploration of the transition region in ice nucleation using the NpH ensemble." *The Journal of Chemical Physics* 165 (5): 054504. https://doi.org/10.1063/5.0336532.

RESEARCH ARTICLE | AUGUST 05 2026

Unbiased exploration of the transition region in ice nucleation using the *NpH* ensemble

Special Collection: Festschrift in honor of Christoph Dellago: Exploring Paths and Barriers in Statistical Mechanics

Sebastian Falkner ; Nadine Schwierz ; Pablo Montero de Hijes

J. Chem. Phys. 165, 054504 (2026)
https://doi.org/10.1063/5.0336532

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Unbiased exploration of the transition region in ice nucleation using the *NpH* ensemble

Cite as: J. Chem. Phys. 165, 054504 (2026); doi: 10.1063/5.0336532
Submitted: 31 March 2026 • Accepted: 17 July 2026 •
Published Online: 5 August 2026

Sebastian Falkner,[1,2] Nadine Schwierz,[1] and Pablo Montero de Hijes[2,a)]

AFFILIATIONS

[1] Institute of Physics, University of Augsburg, 86159 Augsburg, Germany
[2] Faculty of Physics, University of Vienna, A-1090 Vienna, Austria

Note: This paper is part of the Special Topic Festschrift in Honor of Christoph Dellago: Exploring Paths and Barriers in Statistical Mechanics.
[a)]**Author to whom correspondence should be addressed:** pablo.montero.de.hijes@univie.ac.at

ABSTRACT

In this work, we employ the isenthalpic–isobaric (NpH) ensemble to sample the transition region of ice nucleation without any external bias, thereby avoiding potentially artificial memory effects introduced by projections onto collective variables. Within this framework, we identify relevant degrees of freedom that expose the intrinsically non-Markovian nature of the largest nucleus size, indicating that it is not sufficient on its own to describe nucleation dynamics. The NpH ensemble leads to long-lived nuclei through the coupling between latent heat release or absorption and temperature fluctuations. As a result, nuclei persist over extended timescales and undergo a slow internal evolution, which we refer to as aging. A signature of this behavior is the emergence of hysteresis in the largest cluster size–temperature plane. To quantify these effects, we perform a structural analysis based on a high-dimensional set of descriptors, which we project onto a low-dimensional latent space using a neural network-based autoencoder. This approach reveals the existence of structurally distinct classes of nuclei with similar sizes and temperatures. Finally, we compare the nucleus sizes obtained with this approach with those from previous studies employing different methodologies, finding good agreement.

I. INTRODUCTION

Molecular dynamics simulations follow Newton's equations of motion and are deterministic. When defined in full phase space, the dynamics is Markovian as the system's evolution is fully determined by its instantaneous state. Homogeneous ice nucleation is a rare event that typically requires using specialized computational techniques relying on projections onto a reduced set of collective variables. Examples include metadynamics,[1] forward-flux sampling,[2–4] lattice-mold,[5] umbrella sampling,[6] and variational umbrella sampling.[7] Ideally, the projected dynamics should remain Markovian; however, projection onto incomplete collective variables can lead to apparent memory effects. Indeed, several commonly used descriptors in ice nucleation have been shown to fail to produce Markovian dynamics.[8] Such effects are not unique to ice nucleation and have been observed in a variety of systems.[9–14]

For ice nucleation, the coarse-grained mW water model[15] exhibits rich structural behavior while maintaining a relatively low computational cost. Fivefold defective motifs, stacking disorder of cubic and hexagonal ice, and heterogeneous dynamics in the supercooled liquid have been shown to play a central role.[3,16–22] Interestingly, umbrella sampling studies reported predominantly cubic nuclei without fivefold motifs,[23] whereas annealing of previously nucleated clusters produced multitwinned ice nanocrystals.[16] This diversity suggests that, at a given thermodynamic state, ice nuclei may be better described as a family of structurally distinct configurations that depend on the nucleation pathway. Such variability can give rise to non-Markovian behavior in commonly used collective variables, such as the largest nucleus size n. This makes the mW model a suitable testbed for identifying descriptors that provide a more complete description of the nucleation process.

Beyond projection-based approaches, an alternative strategy to sample the transition region in nucleation consists of exploiting thermodynamic ensembles in which nuclei can be stabilized without any external bias. In the NVT ensemble, for certain combinations of N (number of components), V (volume), and T (temperature),

two-phase states with curved interfaces can be stabilized due to a feedback mechanism: fluctuations in nucleus size modify the properties of the surrounding liquid, thereby altering its pressure p and chemical potential μ.[24–44] Under these conditions, clusters stabilized in the NVT ensemble correspond to critical nuclei under an ensemble crossover to NpT, a correspondence established across a wide range of nucleation phenomena.[45–54]

While the modified-liquid-droplet (MLD) model provides a useful framework to identify stable nuclei in fluid–fluid equilibrium in the NVT ensemble,[35,55,56] its application to crystallization remains unaddressed. An alternative, much less explored possibility is the NpH ensemble, where stabilization is mediated by latent heat, which modifies T and, consequently, μ of the surrounding phase.[57,58]

In this work, we employ the NpH ensemble to investigate the stability of ice nuclei in mW water. We show that finite nuclei can become long-lived under NpH conditions, while still undergoing a slow aging process associated with internal structural transformations. We refer to these nuclei as aging or quasi-stable. Here, "quasi-stable" follows from the fact that these NpH nuclei behave analogously to *stable* nuclei in the NVT ensemble. That is, they remain finite for much longer than the growth or melting time, and when configurations sampled from these long-lived trajectories are transferred to NpT simulations at the corresponding T, they behave as critical nuclei. This provides a route to extend investigations of the interface between a critical ice nucleus and liquid water.[59]

Furthermore, hysteresis in the n–T plane confirms the presence of additional degrees of freedom beyond the largest nucleus size, indicating that n alone is an incomplete descriptor of the nucleation dynamics. We perform a structural analysis to characterize the variability of nucleus configurations. Finally, we compare the sizes of the long-lived nuclei obtained in the NpH ensemble, here denoted as quasi-stable nuclei, with those reported in previous nucleation studies. The central result is the ability of the NpH feedback mechanism to generate long-lived finite nuclei that explore the NpT critical region over extended time scales, allowing us to explore the evolution of slow variables relevant to ice nucleation.

II. METHODS

Molecular dynamics simulations of the coarse-grained mW water model[15] were performed using LAMMPS.[60] All simulations contained $N = 12\,800$ water molecules in periodic cubic simulation boxes at constant pressure $p = 1$ bar. Simulations were performed with the `fix nph` command in LAMMPS, using isotropic pressure control and default settings. This command integrates Nosé–Hoover-style non-Hamiltonian equations of motion.[61–64]

This implementation does not use a thermostat and does not reset the kinetic energy to a target temperature. The only global coupling is the deterministic coupling to the finite-mass barostat. This coupling is spatially homogeneous at the level of the cell degrees of freedom, but it is not equivalent to instantaneous thermostating of the system; local redistribution of latent heat occurs through the subsequent molecular dynamics. Moreover, previous simulations of ice growth showed that growth rates obtained without a thermostat were, within statistical uncertainty, consistent with those obtained in thermostated simulations.[65] Therefore, we do not expect the extended NpH dynamics used here to interfere significantly with the aging processes analyzed in this article.

In this extended-system implementation of the NpH ensemble, the conserved quantity is the extended enthalpy H^*, rather than the thermodynamic enthalpy of the physical system alone. The physical enthalpy is

$$H = U_p + U_k + pV, \tag{1}$$

where U_p and U_k are the potential and kinetic energies, respectively. The conserved extended enthalpy may be written schematically as

$$H^* = H + H_{\mathrm{baro}},$$

where H_{baro} denotes the finite-mass barostat and auxiliary extended-variable contributions. Consequently, H is not itself the exact conserved invariant of the extended-system dynamics and may display a systematic time dependence during long trajectories, whereas H^* should remain constant up to numerical integration error. We monitored H^* throughout the simulations to ensure that the observed evolution of the molecular system was not caused by numerical energy drift. A very small drift in H^* was observed over hundreds of nanoseconds: for the employed time step of 0.001 ps, the extended enthalpy drifted by $\sim 3 \times 10^{-5}$ eV/(ns N).

The value of H^*, and therefore the initial state of the NpH trajectory, is determined by the starting configuration and the assigned initial velocities. The starting configuration gives the initial values of U_p and V, while the imposed pressure determines the initial pV contribution. The initial kinetic energy U_k is set by the initial particle velocities, which were drawn from a uniform distribution corresponding to a chosen initial temperature T_i. After this initialization step, U_k evolves deterministically according to the extended NpH equations of motion. Since the long-time growth and aging dynamics studied here occur on time scales much longer than microscopic velocity decorrelation,[66,67] we do not expect the detailed shape of the initial velocity distribution to significantly affect the results, provided that the same initial kinetic energy is imposed.

Because the structure and size of a quasi-stable nucleus corresponding to a given set of N, p, and H are not known *a priori*, simulations were initialized from seeded configurations. Starting from a limited set of closely related initial configurations also allows us to investigate the extent to which their subsequent evolution diverges into distinct dynamical pathways. Additional details about the full set of trajectories are provided in the supplementary material. Figure 1 shows a screenshot of a nucleus before aging. To reduce structural bias while maintaining consistency, all seeds were extracted from a single spontaneous nucleation trajectory, exhibiting stacking disorder between hexagonal and cubic ice.

Several quantities were computed to characterize the nucleus and its interfacial region. The CHILL+ algorithm[68] implemented in OVITO[69] (also employed for atomic visualization) was used to label molecules as hexagonal ice (Ih), cubic ice (Ic), or liquid. The largest nucleus size n was defined by clustering Ih and Ic molecules within a cutoff distance of 3.5 Å and counting the number of molecules in the largest cluster, $n = n_{\mathrm{Ih}} + n_{\mathrm{Ic}}$. OVITO was further used to compute the molecular volume via Voronoi tessellation v, the radius of gyration R_g, and the center-of-mass position of the nucleus. Using R_g and the center of mass, two spatial regions were defined: the nucleus core and the interfacial region.

Additional per-particle quantities were computed in LAMMPS. These include bond orientational order Steinhardt parameters,[70] the

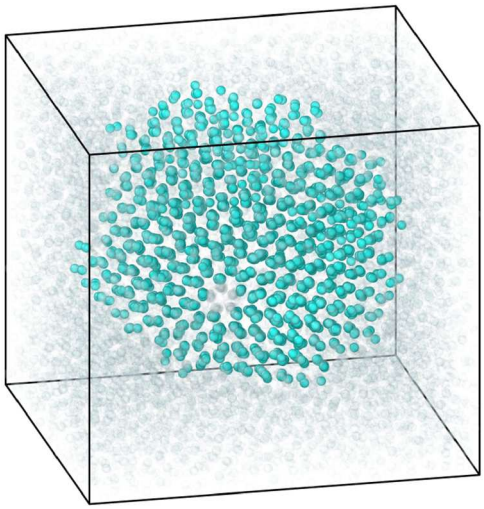

FIG. 1. Screenshot of an ice nucleus formed of hexagonal and cubic local ice environments exhibiting a fivefold defect. Liquid-like molecules are shown with transparency.

coordination number, and the per-particle partition of potential energy. The tetrahedrality parameter was computed as in Ref. 71. All per-particle quantities were first locally averaged and subsequently averaged within each region to construct region-based global collective variables. Further details on the specific Steinhardt parameters and the local averaging procedure are provided in the supplementary material.

To analyze the high-dimensional set of structural descriptors, we employ a time-lagged variational autoencoder[72,73] with the aim of constructing a low-dimensional latent representation of the system. Further details on the network architecture, training procedure, and feature preprocessing are provided in the supplementary material.

III. RESULTS

For certain combinations of N, V, and T, two-phase states with curved interfaces can be stabilized through a feedback mechanism: fluctuations in nucleus size modify the properties of the surrounding liquid, thereby altering its pressure p and chemical potential μ.[24–44] Under these conditions, clusters stabilized in the NVT ensemble correspond to critical nuclei under an ensemble crossover to NpT as established for cavitation,[45–47] condensation,[46–49] crystallization,[50–53] biomolecular condensates,[54] micellation,[74] and precipitation.[49]

In the NpH ensemble, stabilization arises instead, mostly, from the coupling between latent heat and temperature: growth or melting of the nucleus dynamically perturbs T and, therefore, the chemical potential of the surrounding liquid. This provides a feedback mechanism similar to that of the NVT ensemble, enabling the sampling of the transition region within a single unbiased simulation. The coupling mechanism in the NpH ensemble is, however, more intricate than in the NVT ensemble. In NpH, both T and V evolve dynamically. In contrast, in NVT simulations, T is regulated by the thermostat. We refer to the Appendix for a thermodynamic connection between ensembles.

More generally, stabilization is governed by the response of the chemical potential μ of the parent phase: perturbations of the nucleus—through growth, melting, or internal transformations—must induce a sufficiently strong change in μ such that the system relaxes back to a stationary state. This condition typically requires the nucleus size n to be sufficiently large relative to the total system size N.

Previous studies in the NVT ensemble have typically reported a single stationary nucleus for each set of global parameters, characterized by reaction coordinates such as nucleus size or radius, which fluctuate around a constant mean. Although structural rearrangements may occur, they are often sufficiently subtle that they do not perturb this apparent stability. However, in complex phase spaces with multiple minima and competing pathways, this picture may be oversimplified. Homogeneous ice nucleation in the mW model provides an ideal system to examine this scenario, as a wide variety of competing crystal structures have been reported.[3,16–23]

A. Aging of ice nuclei

First, we consider a configuration with $N = 12\,800$ mW water particles containing an ice seed of size $n_i \sim 200$, simulated in the NpH ensemble at 1 bar. The configuration is extracted from a spontaneous nucleation event and, therefore, contains both hexagonal and cubic ice. Since the value of H that stabilizes a critical nucleus is not known *a priori*, we explore different initial temperatures T_i, which correspond to different values of H. For this configuration, the most illustrative case corresponds to initial velocities drawn from a uniform distribution at $T_i = 268$ K.

As summarized in Table I, the system first undergoes rapid thermal relaxation, during which the temperature drops sharply from T_i within a few picoseconds before reaching thermal equilibrium. Figure 2(a) shows the time evolution of the nucleus size, n, on a logarithmic timescale throughout the subsequent stages summarized in Table I. As shown, the nucleus initially grows over the nanosecond timescale, followed by a quasi-stationary regime lasting ~10 ns. On longer timescales, however, a slow drift becomes apparent. As shown in Fig. 2(b), n decreases approximately linearly with time over nearly 250 ns. We refer to this long-lived regime as *aging*. This is followed by a temporary rebound, during which n reaches ~1700 before decreasing again. We emphasize that this rebound arises from the

TABLE I. Summary of characteristic timescales and approximated nucleus sizes n and temperatures T for the main trajectory discussed in Sec. III A. The subnanosecond thermal relaxation and equilibration regimes were probed from an independent trajectory saved with a higher output frequency than the production trajectory. The arrows indicate change, and the accompanying number is the ending value of n and T of the process. In this case, the initial temperature corresponds to 268 K.

Time (ns)	n	T (K)	Regime
10^{-3}	~200	→ 230	Thermal relaxation
10^{-1}	~200	230	Equilibration
10^{0}	→ 1700	→ 250	Growth
10^{2}	→ 600	→ 246	Aging
10^{2}	Rebound	Rebound	Rebound
10^{0}	→ 0	→ 237	Melting

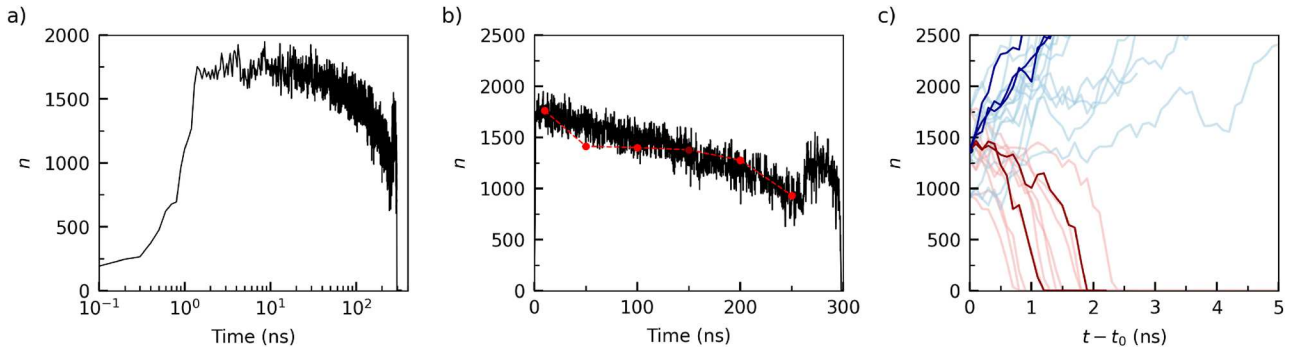

FIG. 2. Time evolution of the largest cluster size n. The time axis in panel (a) is shown on a logarithmic scale to resolve the different dynamical regimes of the *NpH* trajectory. Panel (b) focuses on the aging regime and shows the six configurations selected to initialize the *NpT* seeding trajectories. The solid circles connected by a dashed line indicate the instantaneous value of n of each selected configuration. Panel (c) shows the time evolution of n in the *NpT* trajectories after shifting the time origin of each set of trajectories to its corresponding initial time t_0. Five independent trajectories are initialized from each selected configuration. The trajectories initialized from the fourth selected configuration, highlighted in dark red in panel (b), are shown in blue when they grow and in red when they melt. In contrast, the remaining trajectories are shown with transparency, although following the same color code.

coupling between the nucleus and the finite liquid reservoir through latent heat. It is, therefore, system-dependent and not a general feature of the melting process. Eventually, the nucleus undergoes an abrupt melting event.

The time evolution of n is naturally coupled to that of T through the release and absorption of latent heat. Ice growth releases latent heat, increasing T, whereas ice melting absorbs heat, lowering it. Figure 3 shows the time evolution of T on logarithmic and linear timescales in panels (a) and (b), respectively. Although the overall evolution of T closely follows that of n, the variations are considerably smaller. For example, during aging, T changes by only a few K, whereas n varies by approximately a factor of 2.

The temperature after complete melting remains elevated relative to its value prior to growth. This increase may stem from residual structural ordering, differences between the conserved extended enthalpy in the NpH ensemble and the thermodynamic enthalpy of the atomic subsystem, and non-equilibrium initialization with velocities assigned at 268 K and relaxed to 230 K. Nevertheless, our focus is on the aging regime, where dynamics occur on longer timescales. Next, we test the equivalence of aging nuclei with critical nuclei in the *NpT* ensemble using the seeding approach.[75–78]

Six configurations extracted at different times over a 240 ns interval within the aging regime are used to initialize five independent trajectories each. The temperature imposed in the *NpT* simulations is taken as the time-averaged temperature over a symmetric ±5 ns window around each configuration. In Fig. 2(b), we superimpose, as solid circles connected by dashed lines, the instantaneous value of n for each configuration selected to initialize the seeding trajectories. We emphasize that n corresponds to the instantaneous value of the selected configuration, whereas the imposed T is obtained from the time average over the corresponding ±5 ns window. For example, the second set of trajectories starts from a large fluctuation, yielding a smaller instantaneous value of n than would be obtained by averaging over the same time window used for T.

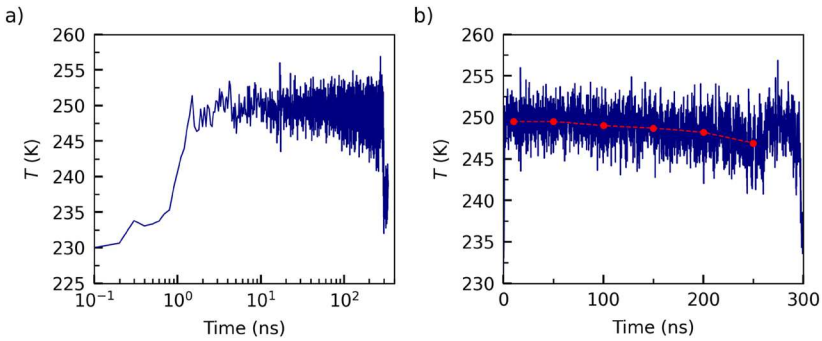

FIG. 3. Time evolution of the temperature T. The time axis in panel (a) is shown on a logarithmic scale to resolve the different dynamical regimes of the *NpH* trajectory. Panel (b) focuses on the aging regime and shows the temperatures used to initialize the *NpT* seeding trajectories. The solid circles connected by a dashed line indicate the imposed temperatures in the *NpT* simulations, obtained as time averages over symmetric ±5 ns windows around each selected configuration. These temperatures are superimposed on the temperature evolution of the corresponding *NpH* trajectory.

In Fig. 3(b), we analogously show the initial temperatures of the seeding trajectories. In Fig. 2(c), we show the time evolution of n after shifting the time origin of each set of seeding trajectories to its respective initial time t_0. The five trajectories initialized from the fourth selected configuration, highlighted in dark red in Fig. 2(b), are shown in highlighted blue when they grow and in highlighted red when they melt. In contrast, the remaining trajectories are also shown, albeit with transparency.

As shown in Fig. 2(c), the aging nuclei commit with approximately equal probability to growth or melting, and they do so on very short timescales. Due to the large difference between the aging timescale and the growth/melting timescale, the corresponding changes in n would appear almost as vertical lines in Fig. 2(b). This behavior is consistent with that expected for critical nuclei in the corresponding NpT ensemble. Although the seeding procedure is not a rigorous committor calculation, we note that, in ice nucleation, even modest displacements away from the critical region can drive the growth probability rapidly toward either zero or one. This reflects the steep and high nucleation barrier in the NpT ensemble, where a relatively small change in nucleus size may correspond to a displacement from the barrier top of several k_BT.[1,7,17,23,79,80] Consistently, for the second set, which starts from a large downward fluctuation in n, four trajectories melt, whereas only one grows, indicating that the imposed, averaged temperature is slightly too high for the instantaneous nucleus size of the selected configuration. The number of trajectories launched from each configuration is, therefore, intended to provide seeding-level statistics, as commonly used to assess whether a nucleus grows or melts under given thermodynamic conditions,[75,76,78] rather than a statistically converged estimate of the committor.

Therefore, the approximately balanced growth and melting outcomes provide a strong indication that the aging configurations are close to the critical region, while not constituting a fully converged committor calculation. These observations suggest that nuclei in the aging stage behave as critical nuclei under an appropriate ensemble crossover. However, the observed drift indicates that the initially formed cluster is not the most thermodynamically stable configuration. Instead, evolution in the NpH ensemble allows slow degrees of freedom—operating on timescales much longer than growth or melting—to relax. This interpretation may appear difficult to reconcile with the evolution of the nucleus along the trajectory. For instance, a nucleus of size $n \sim 600$ is initially strongly biased toward growth, whereas after aging, a nucleus of the same size instead evolves toward melting.

As shown in Fig. 4(a), there is a clear hysteresis in the n–T relation. During the growth stage starting from $n \sim n_i$, the temperature increases by ~20 K (from ~230 to ~250 K). In contrast, during melting, when the nucleus returns to $n \sim n_i$, the temperature has decreased by roughly only 10 K (to ~240 K). Consequently, a nucleus of size n_i at 230 K is committed toward growth, whereas a nucleus of the same size at 240 K evolves toward melting. This is consistent with the fact that the critical nucleus size is significantly larger at 240 K than at 230 K.[7] The remaining question is whether part of this hysteresis can be attributed to the nucleus becoming structurally distinct as it ages.

To this end, Fig. 4(b) shows the relation between n and the system volume V. In contrast to the n–T behavior, the hysteresis in V is much less pronounced. Since the simulations are performed at 1 bar,

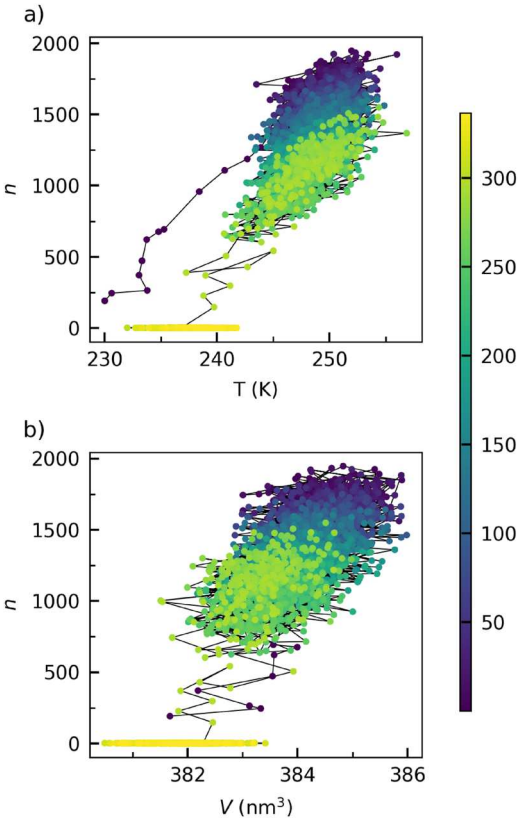

FIG. 4. Relation of n with (a) T and (b) total system V along the example trajectory. Color legend represents time t in ns. A clear correlation between n and T is observed over short time intervals, which can be attributed to latent heat release/absorption during partial growth and melting events typical of the fluctuations of a crystal nucleus in a liquid. However, the distribution does not collapse onto a single curve and instead shifts over time, displaying a pronounced hysteresis in the n–T plane. This indicates that n alone is insufficient to describe the state of the nucleus. Since the volume V shows comparatively smaller hysteresis and the pressure contributing to the pV term of the enthalpy H is just 1 bar, the observed behavior must originate from internal structural degrees of freedom of the nucleus, affecting the potential energy U_p, such as its polymorphic composition or defect structure, which evolves along the trajectory.

the pV term in Eq. (1) is expected to be small. Thus, the increase in kinetic energy must be compensated primarily by changes in potential energy and/or by energy exchange with the barostat degrees of freedom to conserve the extended enthalpy H^*. While this redistribution involves the entire system, it remains unclear to what extent, it is coupled to structural changes within the nucleus itself. For instance, in Fig. 4(a), at fixed temperature (e.g., $T = 250$ K), the nucleus size spans values differing by more than 100% over time. Therefore, while T and V help rationalize the change in commitment at fixed n, they are insufficient to fully characterize the state of the nucleus. This supports the interpretation that additional slow collective variables, not captured by n alone, influence its stability. In Sec. III B, we analyze aging nuclei using several descriptors to clarify this point.

Regarding this example trajectory, we conclude by examining whether the lifetime of the aged nucleus can be extended. To address this, we recall from hard-sphere (HS) studies in the *NVT* ensemble that only specific sets of global parameters (namely N and V, since T plays no role) allow the stabilization of particular values of n.[52] It is, therefore, possible that the aged nucleus—which is probably characterized not only by n but also by additional, yet unidentified variables—has left the range of stability associated with the chosen *NpH* conditions.

To test this hypothesis, we reset H by relaunching the simulation from the configuration at $t = 275$ ns, assigning velocities corresponding to initial temperatures in the range of 230–255 K. In Fig. 5(a), we show the original trajectory in black and the continuations after resetting H in different colors. We observe that the highest T_i (255 K) leads almost immediately to melting, whereas the lowest values (230 and 235 K) induce renewed growth followed by the formation of cylindrical structures. For intermediate values of T_i, the lifetime of the quasi-stable nucleus can be extended by ~50 ns; however, melting ultimately occurs in these cases. These results indicate that, for $N = 12\,800$ and $p = 1$ bar, it is difficult to find a choice of H that leads to a truly stable nucleus.

Therefore, in an attempt to further enhance stability, we modify the system size by reducing the number of particles, removing 42 molecules from the liquid, and repeating the procedure by initializing trajectories at different T_i. In Fig. 5(b), we show two representative trajectories ($T_i = 245$ and 247 K), which yield nuclei that remain quasi-stable for more than 500 ns. Over longer timescales, these trajectories exhibit a slight upward drift in n. In total, that means approximately sampling the transition region of ice nucleation for about 800 ns in an unbiased simulation.

Overall, the *NpH* ensemble enables the quasi-stabilization of ice nuclei that subsequently undergo aging, evolving not only in n but also in additional, yet unidentified, degrees of freedom. Additional continuations and trajectories initialized with $N = 12\,800$ and seed sizes of $n_i \sim 430$ and $n_i \sim 790$ exhibit qualitatively similar behavior. All quasi-stable nuclei observed in this study fall within a relatively narrow size range, indicating that, for fixed N and p, the set of nucleus sizes that can be stabilized is strongly constrained. Taken together, these trajectories (see the supplementary material for details) provide the basis for a more detailed analysis aimed at identifying the features that distinguish freshly formed nuclei from aged ones.

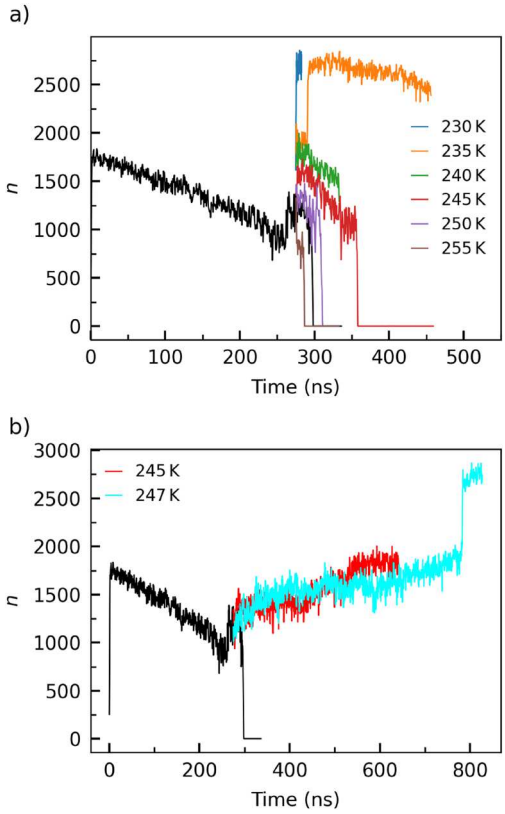

FIG. 5. Time evolution of n for the example trajectory, shown in black, and for continuations obtained after resetting (a) H and (b) both H and N simultaneously. Colored curves show the restarted continuations; the legends indicate the initial target temperature used to sample the velocities. The system rapidly loses memory of these initial velocities, and the temperature subsequently evolves according to the cluster dynamics. Continuations develop cylindrical structures after crossing $n \sim 2500$.

B. Collective variable search

To characterize the nucleus and its time evolution, we compute a broad set of features that could potentially serve as collective variables, all of which describe both the nucleus and the surrounding interfacial water (see the supplementary material for a complete list). These descriptors capture topological, energetic, and kinetic information. In addition, we include the temperature as a descriptor. In the *NpH* ensemble, T evolves over time and exhibits both correlation and hysteresis with respect to n, indicating that additional features beyond n and T are required to describe the state of the system.

We then train an autoencoder[73,81] to reduce the dimensionality of the nucleation data and find relevant degrees of freedom. An autoencoder is a neural-network tool that learns to summarize a large set of input variables by passing them through a small number of intermediate variables forming a bottleneck. These intermediate variables form the so-called latent space, which is learned in an unsupervised way. Machine learning unsupervised methods are well-established tools for discovering relevant structural features in complex systems.[82–84] The autoencoder employed here has two parts: first, an encoder that maps the input descriptors of each nucleus to the latent space; second, a decoder that attempts to reconstruct descriptors from the latent space representation. The network parameters are adjusted during training such that the reconstructed descriptors are as close as possible to the original ones measured by the mean squared error.

Since we want to focus on quantities relevant to the evolution of the system, we here use a time-lagged variational autoencoder.[72] In contrast to the standard autoencoder, the time-lagged version is given the task of reconstructing the descriptors at time $t + \tau$ using the descriptors at a previous time t. This forces the latent variables to include information on slowly evolving degrees of freedom of the system.[72] In the variational formulation, each configuration is mapped to a distribution in latent space. The loss includes a Kullback–Leibler regularization term relative to a simple prior,

which discourages isolated configurations from being encoded in arbitrarily separated regions of latent space. All descriptors that were used as input features after normalization are listed in Table S2 (see the supplementary material).

Given the trained autoencoder, we project all configurations onto the two-dimensional latent space and analyze the emerging patterns across all initial simulations and their continuations. In Fig. 6, we identify two dominant degrees of freedom, $\bar{Q}_4^n$ (top) and T (bottom), which span approximately orthogonal directions in the latent space. $\bar{Q}_4^n$ corresponds to the locally averaged Steinhardt fourth-order parameter[70] (following the approach of Lechner and Dellago[85]), computed over the nearest 16 neighbors and then averaged over all nucleus particles. All other input collective variables exhibit some degree of correlation with one of these two descriptors (see the supplementary material). Apart from $\bar{Q}_4^n$, the fraction of hexagonal ice within the nucleus, x_{Ih}, also displays a clear orthogonality with respect to T. However, we employ $\bar{Q}4^n$ as the structural descriptor in the following analysis, as it provides a continuous measure of local orientational order. In contrast to x_{Ih}, which relies on a discrete structural classification, $\bar{Q}_4^n$ varies smoothly with atomic positions and does not depend on a specific identification of crystalline motifs. This makes it more suitable as a collective variable for capturing structural fluctuations, despite the need to define the nucleus region over which it is averaged.

The latent space representation of Fig. 6 supports the observation that, at a given T, multiple nucleus compositions can be realized in the NpH ensemble, which are clearly distinguished by $\bar{Q}_4^n$. This

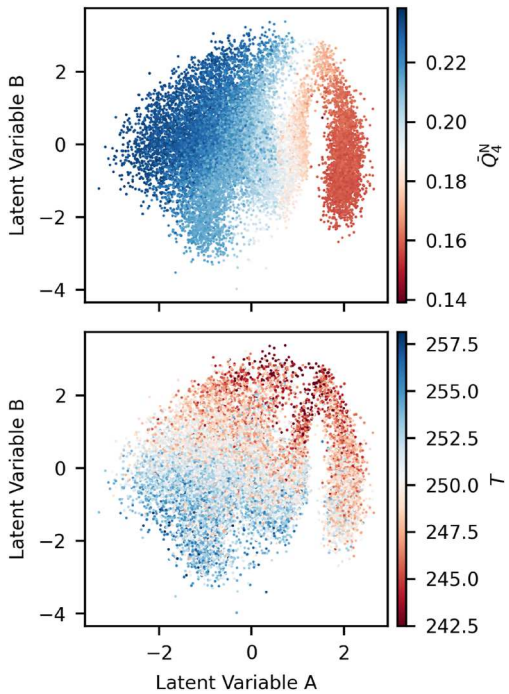

FIG. 6. Latent space representation of all nucleus configurations sampled in the NpH ensemble. In the top panel, points are colored according to the $\bar{Q}_4^n$ value of the nucleus before encoding; analogously, the bottom panel is colored by the instantaneous temperature of all atoms.

further suggests that n alone is not a sufficient reaction coordinate to capture the nucleation dynamics.

Moreover, the latent space exhibits a structured organization, with distinct branches and specific connectivity between them. When projecting the system dynamics onto this space, we observe that different simulations do not necessarily traverse the same regions, nor do they follow identical pathways. In the first row of Fig. 7, this is illustrated by three representative trajectories, each exploring a different region of the latent space. This suggests that structurally distinct nuclei can act as transition configurations, which subsequently evolve along different aging pathways.

As shown in the second row of Fig. 7, the freshly formed quasi-stable nuclei after growth exhibit clear structural differences. Although all configurations contain both Ih and Ic, in one case (left), planar stacking of Ih and Ic appears in relatively thick slabs, whereas in the other cases, fivefold defects disrupt this symmetry. In the middle case, multiple fivefold defects are present (only one is visible), while the rightmost case contains a single fivefold defect.

In the third row, the same nuclei are shown after aging. A consistent trend emerges: fivefold defects are progressively relaxed, promoting an increase in cubicity. This observation is consistent with the explanation of Reinhardt and Doye,[23] who argued that different sampling approaches can lead to distinct nucleus structures. In particular, methods that allow for local equilibration may yield different structures than strongly driven techniques, such as forward flux sampling. This behavior is also reminiscent of the annealing of fivefold nanoparticles observed experimentally.[86] In configurations with Ih/Ic stacking, the stacking becomes more finely distributed into thinner slabs.

In summary, a relatively small set of trajectories (see the supplementary material) already reveals a wide variety of ice nuclei, highlighting the complexity of nucleation pathways in the mW model. This behavior is consistent with the scenario proposed by Prestipino,[17] who suggested that multiple types of critical nuclei may coexist in a multidimensional collective-variable space, as previously reported for Lennard-Jones systems.[87] However, contrasting results have also been reported,[21] where it was argued that an appropriate definition of the largest cluster size can serve as an effective reaction coordinate.

C. Nucleation

Homogeneous ice nucleation in mW water has been extensively studied in the literature. Estimates of the critical nucleus size have been obtained using a variety of techniques, including forward flux sampling,[3,18] seeding,[75] umbrella sampling,[17,23] variational umbrella sampling,[7] path sampling,[21] and spontaneous nucleation simulations.[88]

It is important to note that the definition of the largest cluster size is not unique, and different studies employ different criteria. For example, the umbrella sampling study of Ref. 17 reports three distinct critical nucleus sizes at the same temperature, depending on the cluster definition used. Furthermore, as discussed above, the landscape of ice nucleation in mW water is highly diverse, allowing for multiple types of nuclei. Different biasing techniques and/or collective variables may, therefore, sample different pathways, further contributing to the spread in reported critical sizes. Therefore, the

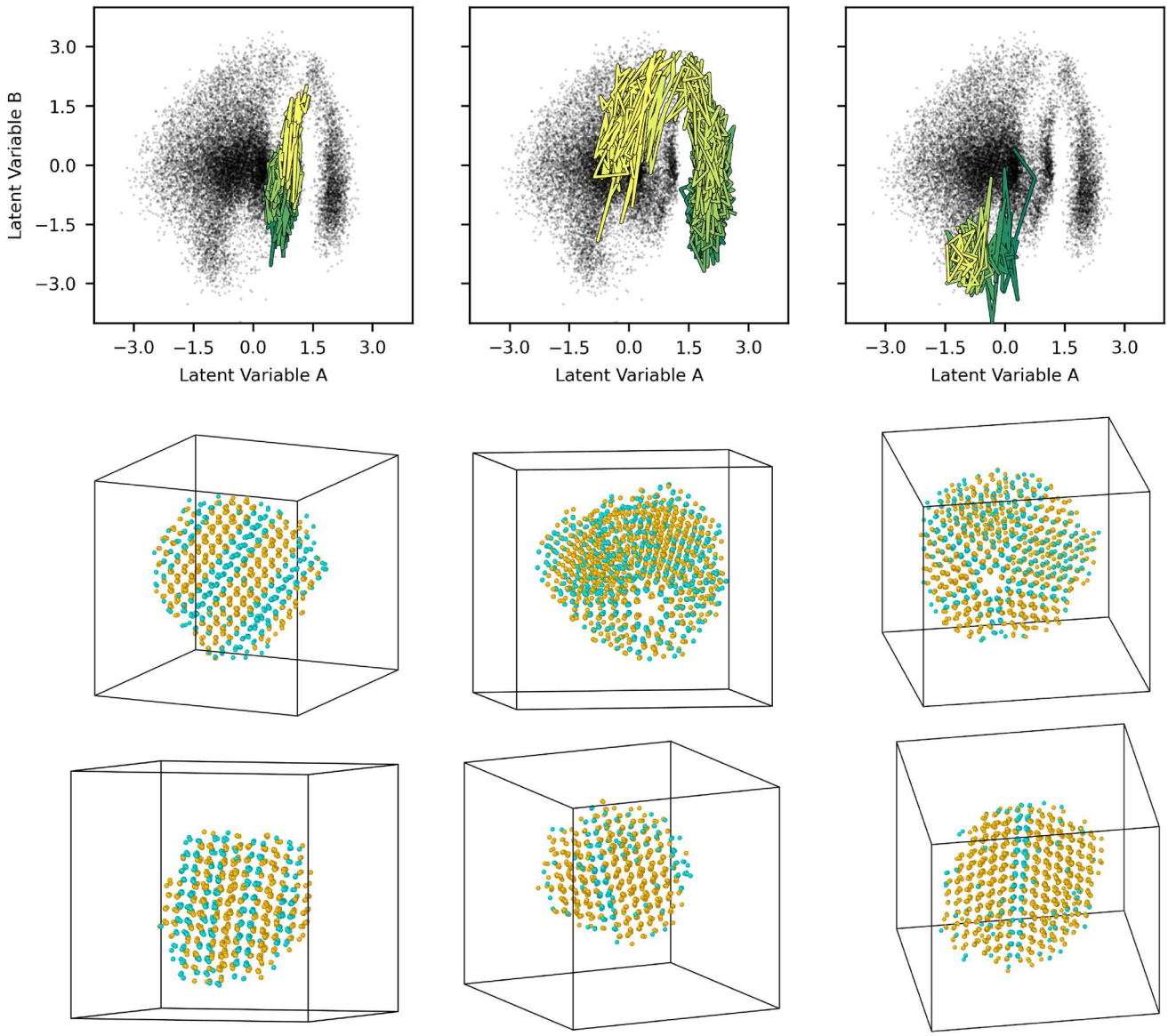

FIG. 7. Each column corresponds to a different *NpH* trajectory. In the first column, a quasi-stable nucleus ages over 80 ns, decreasing from $n \sim 1850$ to ~1100 before suddenly melting. In the second column, a quasi-stable nucleus ages over 250 ns, decreasing from $n \sim 1700$ to ~900 before melting. This one corresponds to the regime with linear decay in n of the illustrative trajectory used in Sec. III A. In the third column, a quasi-stable nucleus ages over only 25 ns at approximately constant size, $n \sim 2400$, before forming a cylindrical structure. The top-row panels show the trajectories projected onto a latent space. The lines are colored from green to yellow according to the relative time of the corresponding configuration along each trajectory, with early nuclei shown in green and aged nuclei in yellow. The middle-row panels show the corresponding quasi-stable nuclei before aging, while the bottom-row panels show them after aging. In the nucleus visualization, surrounding liquid water is not shown, and cyan particles correspond to Ih, whereas yellow particles correspond to Ic.

comparison with literature values should not be interpreted as a one-to-one quantitative comparison. Rather, it is used as a consistency check: despite the differences in nucleus definitions and simulation methodologies, the sizes obtained from the *NpH* trajectories fall within the range of previous estimates. This agreement supports the interpretation that the long-lived *NpH* nuclei approximately sample the transition region.

In what follows, we compare the critical nucleus sizes reported in the literature with the sizes of the quasi-stable nuclei obtained here as a function of temperature. In addition, the procedure used to identify quasi-stable nuclei naturally provides access to critical nucleus sizes in a manner analogous to seeding.[75] This is possible because trajectories are initialized from smaller seeds ($n_i \sim 200$, ~430, and ~790) and evolved at different initial temperatures T_i. We

observe that, for a specific T_i, the system briefly equilibrates, after which the nucleus grows in approximately half of the realizations and melts in the other half.

In this regime, the *NpH* ensemble—similarly to the *NVT* ensemble[89]—enables a seeding-like approach to estimate critical nuclei, in addition to stabilizing them. We refer to these configurations as nuclei obtained via seeding. In addition to the three initial seeds discussed throughout this work, we include a fourth seed ($n_i \sim 1000$), which did not produce sufficiently stable nuclei for aging but is suitable for this analysis. For the seeding data, both T and n correspond to their equilibrated values prior to the system committing to either melting or growth, rather than to the initial values T_i and n_i.

Figure 8 shows the critical nucleus sizes as a function of temperature reported in different studies, together with our results from aging and seeding nuclei. The agreement for the seeding nuclei is very good. For the aging nuclei, we represent the data as a density distribution, reflecting that these configurations occupy a relatively narrow range of sizes and temperatures but display a finite spread arising from the aging process. As shown, the quasi-stable nuclei are in close agreement with available data at comparable temperatures.[3,75]

Small discrepancies may arise from differences in cluster definitions and, to a lesser extent, from finite-size effects in our simulations, since stabilization requires relatively large values of n compared to N. However, these effects are expected to be minor, given the good agreement with previous work, and are certainly outweighed by the ability of the *NpH* ensemble to provide direct insights into the dynamics of nuclei in the transition region through simple, unbiased molecular dynamics simulations.

Here, we compare the critical nucleus size n with previous estimates primarily as a validation of the method. The good agreement, despite several possible sources of discrepancy, such as the cluster-size definition, finite-size effects, and the specific method used to study nucleation, suggests that future applications of this approach may provide further insights into ice-nucleation pathways. The main strength of the approach lies in generating and following long-lived nuclei in the transition region. Nevertheless, a more quantitative characterization of this region, including a more systematic assessment of finite-size effects, may require combining *NpH* trajectories with multidimensional free-energy calculations or committor analyses in the *NpT* ensemble, possibly using selected *NpH*-generated nuclei embedded in larger liquid reservoirs. Our structural analysis is intended to show that additional degrees of freedom may play a role in ice nucleation, while a full characterization of the relationships among these variables, and of their specific contributions to the crystallization barrier, is left for future work. This approach provides an efficient way of approximately sampling the ice nucleation transition region in an unbiased manner over very long times, allowing slow degrees of freedom to relax naturally.

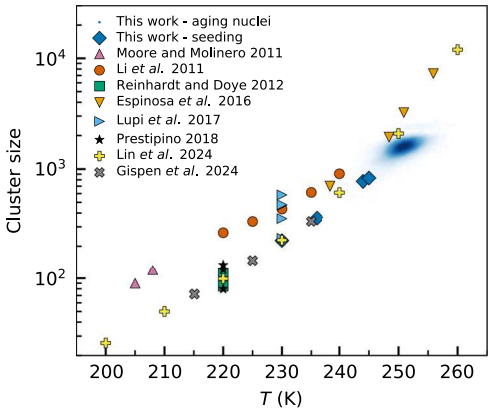

FIG. 8. Cluster size as a function of temperature T for ice nuclei obtained from different computational studies.[3,7,17,18,21,23,75,88] Colored symbols represent different results from the literature, except diamonds, which are obtained in this work. The blue background density map shows the distribution of cluster sizes obtained from aging trajectories, where color encodes the local point density estimated via kernel density estimation in the (T, $\log_{10}$n). The cluster size axis is shown on a logarithmic scale.

IV. CONCLUSIONS

The *NpH* ensemble provides a useful framework to explore the transition region of ice nucleation, enabling the generation and observation of long-lived nuclei without imposing an explicit bias on a prescribed collective variable. Within this framework, we find that the stabilized nuclei undergo a slow aging process, characterized by structural relaxation on time scales much longer than those associated with growth or melting. We, therefore, refer to these nuclei as quasi-stable, since their lifetimes span orders of magnitude beyond the characteristic growth and melting time scales.

This aging behavior is accompanied by hysteresis and history-dependent effects, indicating that the size of the largest nucleus alone is insufficient to describe the underlying dynamics. Our structural analysis further supports the existence of distinct classes of nuclei, highlighting the role of additional, slowly evolving degrees of freedom in nucleation. In particular, we observe that nuclei tend to increase their cubicity and relax fivefold defects during aging. Different internal structures are observed to emerge from the same initial configuration when the system is initialized with slightly different kinetic energies. This behavior illustrates the variability inherent to ice-nucleation pathways. The aim of the present work is primarily to validate the method and to expose the internal evolution of long-lived nuclei, thereby showing that the nucleus size n alone is not a sufficient collective variable. A more complete characterization of the transition region would benefit from a systematic assessment of the role played by the initial nucleus structure, which would require extending the present study from a small set of related initial seeds to a broader and more structurally diverse set of starting configurations.

The sizes of the quasi-stable nuclei remain consistent with the previous estimates of critical nucleus sizes, supporting the validity of the approach. Moreover, an ensemble crossover to the *NpT* ensemble shows that the aging nuclei behave as critical nuclei under the corresponding *NpT* conditions, confirming that we explore the transition region. However, a full characterization of nucleation pathways, and of the relationship between slow degrees of freedom and nucleus size, may require combining *NpH* simulations with free-energy methods or committor analyses.

SUPPLEMENTARY MATERIAL

The supplementary material encompasses a summarized description of the features, trajectories, and neural network parameters, and training is provided.

ACKNOWLEDGMENTS

This work is dedicated to Christoph Dellago, in recognition of his integrity, academic excellence, and outstanding mentorship. This research was funded in part by the Austrian Science Fund (FWF) through Grant No. SFB TACO 10.55776/F8100, available via https://www.fwf.ac.at/en/discover/research-radar. This work was also supported by the research support program (Forschungspotenziale besser nutzen!) of the University of Augsburg. For open access purposes, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission. Computer resources and technical assistance were provided by the Austrian Scientific Computing (ASC).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Sebastian Falkner: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Nadine Schwierz**: Conceptualization (supporting); Formal analysis (equal); Funding acquisition (equal); Investigation (supporting); Validation (equal); Writing – original draft (supporting); Writing – review & editing (equal). **Pablo Montero de Hijes**: Conceptualization (lead); Data curation (equal); Formal analysis (equal); Investigation (lead); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (lead); Writing – review & editing (equal).

DATA AVAILABILITY

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

APPENDIX: THERMODYNAMICS

We consider the reversible work W^E required to form a nucleus along a thermodynamic path in the ensemble E.[24] The system consists of a one-component, inhomogeneous state formed by a spherical ice nucleus embedded in supercooled liquid water. This configuration is taken as the reference state. The same reference state can be reached from different homogeneous parent states depending on the ensemble constraints, so the reversible work of formation depends on the ensemble along which the nucleus is created.

In the NpT ensemble, the intensive variables p and T are fixed externally, and the chemical potential μ of the parent phase remains constant during nucleus formation. This follows from the Gibbs–Duhem relation,

$$Nd\mu = -SdT + Vdp, \tag{A1}$$

together with the fact that the imposed pressure equals that of the external phase (see the supplementary material of Ref. 51). The reversible work is, therefore, the change in the Gibbs free energy,

$$W^{NpT} = \Delta G. \tag{A2}$$

In the μVT ensemble, the chemical potential is explicitly fixed, and the reversible work is given by the change in the grand potential,

$$W^{\mu VT} = \Delta\Omega. \tag{A3}$$

In both ensembles, the nucleus cannot modify the chemical potential of the surrounding phase. In contrast, in the NVT ensemble, the growth or melting of a nucleus can alter the density of the surrounding liquid, thereby modifying the pressure p and consequently the chemical potential μ. The reversible work is then given by the change in the Helmholtz free energy,

$$W^{NVT} = \Delta F. \tag{A4}$$

The reversible works in these ensembles can be related.[24,28,48,90] In particular,

$$W^{\mu VT} = W^{NpT} \tag{A5}$$

and

$$W^{NVT} = W^{NpT} + N\Delta\mu^{NVT} - V\Delta p^{NVT}, \tag{A6}$$

where $\Delta\mu^{NVT}$ and Δp^{NVT} denote the differences in μ and p between the homogeneous parent phase and the reference two-phase state along the NVT path. These additional terms in the NVT path lead to a minimum in the free energy.

In this work, we do not attempt to extend this connection rigorously to the NpH ensemble, as this is not required for our purposes and would be less straightforward due to its formulation in the entropy representation of thermodynamics. Its thermodynamic relation reads

$$dS = -\frac{\mu}{T}dN - \frac{V}{T}dp + \frac{1}{T}dH. \tag{A7}$$

However, we assume that the NpH ensemble is thermodynamically equivalent to the NpS ensemble, since fixing N, p, and H also fixes S. In the energy representation provided by the NpS ensemble, the enthalpy H plays the role of thermodynamic potential, with differential

$$dH = TdS + Vdp + \mu dN. \tag{A8}$$

The reversible work of nucleus formation in the NpH (or, as assumed, equivalently NpS) ensemble is, therefore,

$$W^{NpH} \equiv W^{NpS} = \Delta H. \tag{A9}$$

Using the identity $H = G + TS$, the reversible work in NpH can be related to that in NpT,

$$W^{NpH} = W^{NpT} + N\Delta\mu^{NpH} + S\Delta T^{NpH}, \quad \text{(A10)}$$

where $\Delta\mu^{NpH}$ and ΔT^{NpH} are the differences in μ and T between the homogeneous parent phase and the reference two-phase state along the NpH path. Again, two additional terms arise, like in the NVT–NpT crossover. A related expression connecting the NpH and NVT ensembles is given as follows:

$$W^{NpH} = W^{NVT} + V\Delta p^{NVT} + S\Delta T^{NpH} + N\left(\mu^{NpH} - \mu^{NVT}\right), \quad \text{(A11)}$$

where $\mu^{NpH} - \mu^{NVT}$ is the difference in chemical potential between the two parent states that lead to the same nucleus configuration along the NpH and NVT paths.

If the surrounding liquid is only weakly perturbed by the nucleus, the reversible work becomes equivalent in all ensembles. In this limit, the feedback mechanism required to stabilize a nucleus disappears, and minima in F or H are not expected. Stabilization, therefore, relies on finite-size effects that allow nucleus growth or melting to induce measurable changes in the intensive variables and shift the chemical potential, producing a stationary nucleus size.

Although a rigorous derivation of the reversible work in the NpH ensemble remains to be done, the present formulation already reveals the key mechanism of quasi-stabilization: nuclei are stabilized through the release or absorption of latent heat, which dynamically shifts the system temperature and, hence, necessarily the chemical potential.

REFERENCES

[1] H. Niu, Y. I. Yang, and M. Parrinello, "Temperature dependence of homogeneous nucleation in ice," Phys. Rev. Lett. **122**(24), 245501 (2019).
[2] A. Haji-Akbari and P. G. Debenedetti, "Direct calculation of ice homogeneous nucleation rate for a molecular model of water," Proc. Natl. Acad. Sci. U. S. A. **112**(34), 10582–10588 (2015).
[3] M. Lin, Z. Xiong, and H. Cao, "Bridging classical nucleation theory and molecular dynamics simulation for homogeneous ice nucleation," J. Chem. Phys. **161**(8), 084504 (2024).
[4] M. Chen, L. Tan, H. Wang, L. Zhang, and H. Niu, "Exploring homogeneous ice nucleation through molecular dynamics simulations with ab initio accuracy," Phys. Rev. B **111**(13), 134116 (2025).
[5] I. Sanchez-Burgos, A. R. Tejedor, C. Vega, M. M. Conde, E. Sanz, J. Ramirez, and J. R. Espinosa, "Homogeneous ice nucleation rates for mW and TIP4P/Ice models through lattice mold calculations," J. Chem. Phys. **157**(9), 094503 (2022).
[6] B. Cheng, C. Dellago, and M. Ceriotti, "Theoretical prediction of the homogeneous ice nucleation rate: Disentangling thermodynamics and kinetics," Phys. Chem. Chem. Phys. **20**(45), 28732–28740 (2018).
[7] W. Gispen, J. R. Espinosa, E. Sanz, C. Vega, and M. Dijkstra, "Variational umbrella seeding for calculating nucleation barriers," J. Chem. Phys. **160**(17), 174501 (2024).
[8] P. Montero de Hijes, S. Falkner, and C. Dellago, "Non-Markovian dynamics in ice nucleation," J. Chem. Phys. **164**(9), 094501 (2026).
[9] J. Kuipers and G. T. Barkema, "Non-Markovian dynamics of clusters during nucleation," Phys. Rev. E **79**(6), 062101 (2009).
[10] A. Kuhnhold, H. Meyer, G. Amati, P. Pelagejcev, and T. Schilling, "Derivation of an exact, nonequilibrium framework for nucleation: Nucleation is a priori neither diffusive nor Markovian," Phys. Rev. E **100**(5), 052140 (2019).
[11] S. Jungblut and C. Dellago, "Caveats of mean first-passage time methods applied to the crystallization transition: Effects of non-markovianity," J. Chem. Phys. **142**(6), 064103 (2015).
[12] J. Kuipers and G. T. Barkema, "Limitations of a Fokker–Planck description of nucleation," Phys. Rev. E **82**(1), 011128 (2010).
[13] S.-T. Tsai, Z. Smith, and P. Tiwary, "Reaction coordinates and rate constants for liquid droplet nucleation: Quantifying the interplay between driving force and memory," J. Chem. Phys. **151**(15), 154106 (2019).
[14] J. C. Barrett, "Energy dependent decay rates of Lennard-Jones clusters for use in nucleation theory," J. Chem. Phys. **126**(7), 074312 (2007).
[15] V. Molinero and E. B. Moore, "Water modeled as an intermediate element between carbon and silicon," J. Phys. Chem. B **113**(13), 4008–4016 (2009).
[16] X. Zhang, M. Matsumoto, Z. Zhang, and K. Mochizuki, "Multitwinned ice nanocrystals," ACS Nano **18**(45), 31021–31030 (2024).
[17] S. Prestipino, "The barrier to ice nucleation in monatomic water," J. Chem. Phys. **148**(12), 124505 (2018).
[18] T. Li, D. Donadio, G. Russo, and G. Galli, "Homogeneous ice nucleation from supercooled water," Phys. Chem. Chem. Phys. **13**(44), 19807–19813 (2011).
[19] F. Leoni, R. Shi, H. Tanaka, and J. Russo, "Crystalline clusters in mw water: Stability, growth, and grain boundaries," J. Chem. Phys. **151**(4), 044505 (2019).
[20] X. Huang, Z. Yuan, C.-H. Lo, H. Sun, L. Liao, H. Han, W. Li, W. Wang, Z. Xu, L. Liu *et al.*, "Kinetics of stacking-order evolution during heterogeneous ice formation," Phys. Rev. Lett. **136**(8), 088001 (2026).
[21] L. Lupi, A. Hudait, B. Peters, M. Grünwald, R. Gotchy Mullen, A. H. Nguyen, and V. Molinero, "Role of stacking disorder in ice nucleation," Nature **551**(7679), 218–222 (2017).
[22] M. Fitzner, G. C. Sosso, S. J. Cox, and A. Michaelides, "Ice is born in low-mobility regions of supercooled liquid water," Proc. Natl. Acad. Sci. U. S. A. **116**(6), 2009–2014 (2019).
[23] A. Reinhardt and J. P. K. Doye, "Free energy landscapes for homogeneous nucleation of ice for a monatomic water model," J. Chem. Phys. **136**(5), 054501 (2012).
[24] A. I. Rusanov, "The thermodynamics of processes of new-phase formation," Russ. Chem. Rev. **33**(7), 385 (1964).
[25] A. J.-M. Yang, "Free energy for the heterogeneous systems with spherical interfaces," J. Chem. Phys. **79**(12), 6289–6293 (1983).
[26] A. J.-M. Yang, "The thermodynamical stability of the heterogeneous system with a spherical interface," J. Chem. Phys. **82**(4), 2082–2085 (1985).
[27] J. K. Lee, J. A. Barker, and F. F. Abraham, "Theory and Monte Carlo simulation of physical clusters in the imperfect vapor," J. Chem. Phys. **58**(8), 3166–3180 (1973).
[28] D. J. Lee, M. M. Telo da Gama, and K. E. Gubbins, "A microscopic theory for spherical interfaces: Liquid drops in the canonical ensemble," J. Chem. Phys. **85**(1), 490–499 (1986).
[29] K. Binder, "Monte Carlo simulation of physical clusters of water molecules," J. Chem. Phys. **63**(5), 2265–2266 (1975).
[30] M. Rao, B. J. Berne, and M. H. Kalos, "Computer simulation of the nucleation and thermodynamics of microclusters," J. Chem. Phys. **68**(4), 1325–1336 (1978).
[31] V. Talanquer and D. W. Oxtoby, "Dynamical density functional theory of gas–liquid nucleation," J. Chem. Phys. **100**(7), 5190–5200 (1994).
[32] K. Binder, "Theory of the evaporation/condensation transition of equilibrium droplets in finite volumes," Physica A **319**, 99–114 (2003).
[33] K. Binder, B. J. Block, P. Virnau, and A. Tröster, "Beyond the Van Der Waals loop: What can be learned from simulating Lennard-Jones fluids inside the region of phase coexistence," Am. J. Phys. **80**(12), 1099–1109 (2012).
[34] L. G. MacDowell, P. Virnau, M. Müller, and K. Binder, "The evaporation/condensation transition of liquid droplets," J. Chem. Phys. **120**(11), 5293–5308 (2004).
[35] D. Reguera, R. K. Bowles, Y. Djikaev, and H. Reiss, "Phase transitions in systems small enough to be clusters," J. Chem. Phys. **118**(1), 340–353 (2003).
[36] S. K. Das and K. Binder, "Universal critical behavior of curvature-dependent interfacial tension," Phys. Rev. Lett. **107**(23), 235702 (2011).
[37] M. Schrader, P. Virnau, D. Winter, T. Zykova-Timan, and K. Binder, "Methods to extract interfacial free energies of flat and curved interfaces from computer simulations," Eur. Phys. J.: Spec. Top. **177**(1), 103–127 (2009).

[38]M. Schrader, P. Virnau, and K. Binder, "Simulation of vapor-liquid coexistence in finite volumes: A method to compute the surface free energy of droplets," Phys. Rev. E **79**(6), 061104 (2009).
[39]B. J. Block, S. K. Das, M. Oettel, P. Virnau, and K. Binder, "Curvature dependence of surface free energy of liquid drops and bubbles: A simulation study," J. Chem. Phys. **133**(15), 154702 (2010).
[40]A. Tröster, M. Oettel, B. Block, P. Virnau, and K. Binder, "Numerical approaches to determine the interface tension of curved interfaces from free energy calculations," J. Chem. Phys. **136**(6), 064709 (2012).
[41]A. Tröster and K. Binder, "Microcanonical determination of the interface tension of flat and curved interfaces from Monte Carlo simulations," J. Phys.: Condens. Matter **24**(28), 284107 (2012).
[42]A. Tröster, F. Schmitz, P. Virnau, and K. Binder, "Equilibrium between a droplet and surrounding vapor: A discussion of finite size effects," J. Phys. Chem. B **122**(13), 3407–3417 (2017).
[43]A. Statt, P. Virnau, and K. Binder, "Finite-size effects on liquid-solid phase coexistence and the estimation of crystal nucleation barriers," Phys. Rev. Lett. **114**(2), 026101 (2015).
[44]K. G. S. H. Gunawardana and X. Song, "Theoretical prediction of crystallization kinetics of a supercooled Lennard-Jones fluid," J. Chem. Phys. **148**(20), 204506 (2018).
[45]P. Rosales-Pelaez, I. Sanchez-Burgos, C. Valeriani, C. Vega, and E. Sanz, "Seeding approach to nucleation in the NVT ensemble: The case of bubble cavitation in overstretched Lennard Jones fluids," Phys. Rev. E **101**(2), 022611 (2020).
[46]I. Sanchez-Burgos, P. M. de Hijes, P. Rosales-Pelaez, C. Vega, and E. Sanz, "Equivalence between condensation and boiling in a Lennard-Jones fluid," Phys. Rev. E **102**(6), 062609 (2020).
[47]M. Camarillo, I. Sanchez-Burgos, C. P Lamas, P. Montero de Hijes, J. R Espinosa, and E. Sanz, "Condensation vs cavitation in water: A simulation study," J. Chem. Phys. **163**(4), 044505 (2025).
[48]A. Llamas-Jaramillo, I. Latella, and D. Reguera, "Free-energy reconstruction from stable clusters (FRESC): A new method to evaluate nucleation barriers from simulation," J. Chem. Phys. **164**, 034102 (2026).
[49]L. Li, F. Bachtiger, A. Finney, and M. Salvalaglio, "Computing nucleation rates from confined equilibria: The critical cluster equivalence principle," chemRxiv:2025-w9xfh/v4 (2025).
[50]P. Montero De Hijes, J. R. Espinosa, V. Bianco, E. Sanz, and C. Vega, "Interfacial free energy and Tolman length of curved liquid–solid interfaces from equilibrium studies," J. Phys. Chem. C **124**(16), 8795–8805 (2020).
[51]P. Montero de Hijes, K. Shi, E. G. Noya, E. Santiso, K. Gubbins, E. Sanz, and C. Vega, "The Young–Laplace equation for a solid–liquid interface," J. Chem. Phys. **153**(19), 191102 (2020).
[52]P. Montero de Hijes and C. Vega, "On the thermodynamics of curved interfaces and the nucleation of hard spheres in a finite system," J. Chem. Phys. **156**(1), 014505 (2022).
[53]M. de Jager, C. Vega, P. Montero de Hijes, F. Smallenburg, and L. Filion, "Statistical mechanics of crystal nuclei of hard spheres," J. Chem. Phys. **161**(18), 184501 (2024).
[54]L. Li, M. Paloni, A. R. Finney, A. Barducci, and M. Salvalaglio, "Nucleation of biomolecular condensates from finite-sized simulations," J. Phys. Chem. Lett. **14**(7), 1748–1755 (2023).
[55]D. Reguera and H. Reiss, "Extended modified liquid drop- dynamical nucleation theory (EMLD-DNT) approach to nucleation: A new theory," J. Phys. Chem. B **108**(51), 19831–19842 (2004).
[56]J. Wedekind, D. Reguera, and R. Strey, "Finite-size effects in simulations of nucleation," J. Chem. Phys. **125**(21), 214505 (2006).
[57]L. A. Zepeda-Ruiz, B. Sadigh, A. A. Chernov, T. Haxhimali, A. Samanta, T. Oppelstrup, S. Hamel, L. X. Benedict, and J. L. Belof, "Extraction of effective solid-liquid interfacial free energies for full 3d solid crystallites from equilibrium md simulations," J. Chem. Phys. **147**(19), 194704 (2017).
[58]Y. Wu, H. Cheng, H. Miao, K. Zhang, and J. You, "A concise, efficient and accurate method for calculating the solid–liquid interfacial free energy," Phys. Chem. Chem. Phys. **27**(11), 5459–5463 (2025).
[59]P. Montero de Hijes, K. Shi, C. Vega, and C. Dellago, "Comparing the mechanical and thermodynamic definitions of pressure in ice nucleation," J. Phys. Chem. Lett. **17**(8), 2367–2373 (2026).
[60]A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. In't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen *et al.*, "LAMMPS—A flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales," Comput. Phys. Commun. **271**, 108171 (2022).
[61]G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," J. Chem. Phys. **101**(5), 4177–4189 (1994).
[62]W. Shinoda, M. Shiga, and M. Mikami, "Rapid estimation of elastic constants by molecular dynamics simulation under constant stress," Phys. Rev. B **69**(13), 134103 (2004).
[63]M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, and G. J. Martyna, "A Liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal–isobaric ensemble," J. Phys. A: Math. Gen. **39**(19), 5629–5651 (2006).
[64]M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," J. Appl. Phys. **52**(12), 7182–7190 (1981).
[65]P. Montero de Hijes, J. R. Espinosa, C. Vega, and E. Sanz, "Ice growth rate: Temperature dependence and effect of heat dissipation," J. Chem. Phys. **151**(4), 044509 (2019).
[66]U. Balucani, J. P. Brodholt, and R. Vallauri, "Analysis of the velocity autocorrelation function of water," J. Phys.: Condens. Matter **8**(34), 6139–6144 (1996).
[67]P. Montero de Hijes, L. Neubeck, G. Kresse, and C. Dellago, "Dynamical properties of ab initio water from machine-learning potentials," arXiv:2606.18163 (2026).
[68]A. H. Nguyen and V. Molinero, "Identification of clathrate hydrates, hexagonal ice, cubic ice, and liquid water in simulations: The CHILL+ algorithm," J. Phys. Chem. B **119**(29), 9369–9376 (2015).
[69]A. Stukowski, "Visualization and analysis of atomistic simulation data with OVITO—The open visualization tool," Modell. Simul. Mater. Sci. Eng. **18**(1), 015012 (2009).
[70]P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," Phys. Rev. B **28**(2), 784 (1983).
[71]J. R. Errington and P. G. Debenedetti, "Relationship between structural order and the anomalies of liquid water," Nature **409**(6818), 318–321 (2001).
[72]C. Wehmeyer and F. Noé, "Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics," J. Chem. Phys. **148**, 241703 (2018).
[73]D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, 2014.
[74]A. K. Shchekin, K. Koga, and N. A. Volkov, "The effect of a finite number of monomers available for aggregation at nucleation and micellization in a fixed volume," J. Chem. Phys. **151**(24), 244903 (2019).
[75]J. R. Espinosa, C. Vega, C. Valeriani, and E. Sanz, "Seeding approach to crystal nucleation," J. Chem. Phys. **144**(3), 034501 (2016).
[76]V. Bianco, P. M. de Hijes, C. P. Lamas, E. Sanz, and C. Vega, "Anomalous behavior in the nucleation of ice at negative pressures," Phys. Rev. Lett. **126**(1), 015704 (2021).
[77]P. M. Piaggi, J. Weis, A. Z. Panagiotopoulos, P. G. Debenedetti, and R. Car, "Homogeneous ice nucleation in an ab initio machine-learning model of water," Proc. Natl. Acad. Sci. U. S. A. **119**(33), e2207294119 (2022).
[78]E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L. F. Abascal, and C. Valeriani, "Homogeneous ice nucleation at moderate supercooling from molecular simulation," J. Am. Chem. Soc. **135**(40), 15008–15017 (2013).
[79]C. R. C. Buhariwalla, R. K. Bowles, I. Saika-Voivod, F. Sciortino, and P. H. Poole, "Free energy of formation of small ice nuclei near the Widom line in simulations of supercooled water," Eur. Phys. J. E **38**(5), 39 (2015).
[80]D. Quigley and P. M. Rodger, "Metadynamics simulations of ice nucleation and growth," J. Chem. Phys. **128**(15), 154518 (2008).

[81] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "B-VAE: Learning basic visual concepts with a constrained variational framework," in 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, 2017.
[82] L. Kývala, P. Montero de Hijes, and C. Dellago, "Unsupervised identification of crystal defects from atomistic potential descriptors," npj Comput. Mater. **11**(1), 50 (2025).
[83] M. Ceriotti, "Unsupervised machine learning in atomistic simulations, between predictions and understanding," J. Chem. Phys. **150**(15), 150901 (2019).
[84] A. Offei-Danso, A. Hassanali, and A. Rodriguez, "High-dimensional fluctuations in liquid water: Combining chemical intuition with unsupervised learning," J. Chem. Theory Comput. **18**(5), 3136–3150 (2022).
[85] W. Lechner and C. Dellago, "Accurate determination of crystal structures based on averaged local bond order parameters," J. Chem. Phys. **129**(11), 114707 (2008).
[86] A. Mayoral, H. Barron, R. Estrada-Salas, A. Vazquez-Duran, and M. José-Yacamán, "Nanoparticle stability from the nano to the meso interval," Nanoscale **2**(3), 335–342 (2010).
[87] D. Moroni, P. R. Ten Wolde, and P. G. Bolhuis, "Interplay between structure and size in a critical crystal nucleus," Phys. Rev. Lett. **94**(23), 235703 (2005).
[88] E. B. Moore and V. Molinero, "Structural transformation in supercooled water controls the crystallization rate of ice," Nature **479**(7374), 506–508 (2011).
[89] D. Richard and T. Speck, "Crystallization of hard spheres revisited. II. Thermodynamic modeling, nucleation work, and the surface of tension," J. Chem. Phys. **148**(22), 224102 (2018).
[90] D. Reguera and H. Reiss, "Fusion of the extended modified liquid drop model for nucleation and dynamical nucleation theory," Phys. Rev. Lett. **93**(16), 165701 (2004).