Homogeneous ice nucleation in an ab initio machine-learning model of water

Pablo M. Piaggi[a,1] , Jack Weis[b], Athanassios Z. Panagiotopoulos[b] , Pablo G. Debenedetti[b] , and Roberto Car[a,c,1]

Molecular simulations have provided valuable insight into the microscopic mechanisms underlying homogeneous ice nucleation. While empirical models have been used extensively to study this phenomenon, simulations based on first-principles calculations have so far proven prohibitively expensive. Here, we circumvent this difficulty by using an efficient machine-learning model trained on density-functional theory energies and forces. We compute nucleation rates at atmospheric pressure, over a broad range of supercoolings, using the seeding technique and systems of up to hundreds of thousands of atoms simulated with ab initio accuracy. The key quantity provided by the seeding technique is the size of the critical cluster (i.e., a size such that the cluster has equal probabilities of growing or melting at the given supersaturation), which is used together with the equations of classical nucleation theory to compute nucleation rates. We find that nucleation rates for our model at moderate supercoolings are in good agreement with experimental measurements within the error of our calculation. We also study the impact of properties such as the thermodynamic driving force, interfacial free energy, and stacking disorder on the calculated rates.

water | ice nucleation | molecular dynamics | machine learning | density-functional theory

Ice crystallization from supercooled liquid water is one of the most emblematic phase transformations to be found in nature. It is of key importance in the regulation of our planet's climate (1) and in many applications, such as artificial cloud seeding, cryopreservation, and food processing. Molecular simulations have proven an invaluable tool to obtain insight into molecular-level details of this process and to make predictions at conditions not readily accessible to experiments. For instance, Lupi et al. (2) considered the effect of stacking disorder (i.e., the presence of alternate layers of hexagonal and cubic ice) on the nucleation rates, and Sanz et al. (3) used systems of more than 100,000 molecules in order to compute nucleation rates at low supercoolings.

However, simulations of ice nucleation carried out so far have employed relatively simple empirical models, such as the coarse-grained monoatomic model of water mW (4) or the four-site rigid TIP4P water models (5). A different route to study this phenomenon is using ab initio molecular dynamics (MD) (6). In this technique, the forces acting on the atomic nuclei are derived from electronic structure calculations. At variance with empirical models, the ab initio potential energy surface does not rely on empirical information, captures complex bonding behavior between atoms, and describes the formation and breaking of chemical bonds. The solution of the many-body electronic Schrödinger equation is, in general, not tractable, and a widely used approximation in this context is Kohn–Sham density-functional theory (7) (DFT). The application of ab initio MD, however, has been limited for several decades to the simulation of relatively small systems (∼1,000 atoms) and short times (∼100 ps) due to its high computational cost. This limitation has precluded the study of ice nucleation from first principles.

A solution to this conundrum has been the use of machine-learning algorithms that are able to learn the energies and forces derived from DFT data (8). The machine-learning interatomic models constructed in this fashion reproduce the ab initio potential energy surface with high fidelity, are several orders of magnitude faster than DFT, and also show linear scaling with the number of nuclei. Such models have recently been applied to the study of crystal nucleation in silicon (9) and gallium (10). Previous simulations using first-principles models, however, explored only relatively large supercoolings, for which systems of a few thousand atoms are able to contain the required crystalline cluster.

Here, we compute ice-nucleation rates using an ab initio machine-learning model of water. We employ the seeding technique (3) and systems of up to 300,000 atoms in order to obtain nucleation rates in a broad range of supercoolings. Our results allow us to compare predictions from a model derived from first principles with direct experimental measurements of nucleation rates. Although we only simulate explicitly clusters of hexagonal ice, we take into account the effect of stacking disorder using a model for the chemical potential of ice with stacking disorder.

Significance

Until recently, simulating ice nucleation with quantum accuracy was deemed impossible due to the prohibitive computational cost of quantum-mechanical calculations. Recent progress enabled by machine learning has made these calculations tractable and thus greatly extended the field of application of molecular dynamics based on ab initio quantum-mechanical theory. We apply these advances to predict the rate of formation of ice nuclei in supercooled water and to study other quantities relevant to nucleation without relying on empirical force fields, albeit invoking the organizing framework of classical nucleation theory. This work is a step toward modeling nucleation processes in more realistic environments and at conditions in which chemical reactions play an important role.

Author affiliations: [a]Department of Chemistry, Princeton University, Princeton, NJ 08544; [b]Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544; and [c]Department of Physics, Princeton University, Princeton, NJ 08544

[1]To whom correspondence may be addressed. Email: ppiaggi@princeton.edu or rcar@princeton.edu.

During homogeneous ice nucleation, an ice cluster is formed within bulk liquid water. Typically, this phenomenon takes place below the melting temperature, and, thus, there is a driving force for the formation of ice. However, the formation of an ice cluster in the liquid creates a liquid–solid interface with an associated energetic penalty. The competition between a favorable bulk term and an unfavorable surface term leads to a free-energy barrier that the system must surmount in order to proceed with the transformation. The existence of a free-energy barrier makes nucleation a rare event and severely hinders the ability to study the phenomenon directly by using molecular simulations. Although there have been attempts to study ice nucleation by using straightforward molecular simulations (11), in general, the problem must be tackled by using more sophisticated techniques.

A possible route to study ice nucleation on the computer is rare-event techniques, such as path sampling (2, 12), forward-flux sampling (13, 14), or metadynamics (15, 16). These approaches can provide valuable insights into the nucleation mechanism, albeit at a high computational cost. A simpler alternative is the seeding technique (3), which is based on performing a series of relatively short simulations at different temperatures, starting from a configuration that contains an ice cluster embedded in liquid water. The aim of these simulations is to find the temperature T^*, for which the chosen cluster is critical—that is, to say, at T^*, the cluster has equal probabilities of growing and thawing. This information is then used in combination with the equations of classical nucleation theory (CNT) (17) to calculate the nucleation rate. This approach has several potential pitfalls that can affect the calculated rates, such as the appropriate choice of an order parameter to calculate the cluster size, and the applicability of CNT to the nucleation process under study. These limitations have been carefully considered in the literature (18), and the seeding technique has been shown to provide nucleation rates in good agreement with other methods (19, 20).

Another crucial ingredient in the simulation of ice nucleation is an accurate description of the interatomic interactions. Here, we derive the forces between nuclei from first-principles calculations. In particular, we use DFT adopting the Strongly Constrained and Appropriately Normed (SCAN) (21) exchange and correlation functional. SCAN is arguably one of the best semilocal functionals available, and many properties of ice and water have been studied using this functional—e.g., in refs. 22 and 23. Driving the dynamics directly using DFT forces would be unduly costly, and, instead, we use a machine-learning model trained on DFT data. The model is based on deep neural networks and was constructed by using the deep potential methodology developed by Zhang et al. (24). Below, we refer to this model as SCAN-ML (i.e., SCAN-trained, machine-learning-based model). The SCAN-ML model was carefully trained to reproduce data over a vast region of the phase diagram of water (25). SCAN-ML has been used to provide evidence of the existence of a liquid–liquid transition at deeply supercooled conditions (26) and to study the ice I_h–ice XI transition (27). The thermodynamic properties of this model relevant to ice nucleation were thoroughly characterized in ref. 23. The model has a melting temperature of 312 K, around 40 K larger than the experimental value. The density change upon melting is 6% in the model, somewhat smaller than the 9% found in experiments. Another important property is the relative stability between ice I_h and ice I_c, which are the two competing polymorphs during ice nucleation at ambient pressure. The SCAN-ML model correctly predicts that ice I_h is more stable than ice I_c, in agreement with experiments. Ref. 23 also analyzed the ability of the SCAN-ML model to reproduce SCAN energies in configurations that contain atomic environments compatible with both liquid water and ice

Table 1. Melting temperature (T_m), densities of ice I_h and liquid water at coexistence (ρ_{ice} and ρ_l), and enthalpy of fusion (ΔH_f) of SCAN-ML, experimental water, and the empirical models TIP4P/Ice and mW

	T_m (K)	ρ_{ice} (g/cm^3)	ρ_l (g/cm^3)	ΔH_f (kJ/mol)
SCAN-ML (23)	312(1)	0.949(1)	1.002(3)	7.6(1)
Experiment	273.15	0.917	0.999	6.01
TIP4P/Ice (5)	270	0.906	0.985	5.40
mW (28, 29)	273	0.978	1.001	5.3

and found that the model is a faithful representation of SCAN with deviations of less than 1.3 meV per H_2O molecule. We provide in Table 1 a summary of the properties of the SCAN-ML model, and we compare them with experimental data and results using the empirical water models TIP4P/Ice and mW.

Before describing the results of our simulations, we briefly discuss the advantages of SCAN-ML over empirical models. SCAN-ML is an all-atom, fully flexible model at variance with empirical potentials such as mW, which is a coarse-grained model, and TIP4P/Ice, which is an all-atom rigid model. Since SCAN-ML reproduces the DFT potential energy surface, the flexibility of the OH bonds depends on the environments, while in flexible empirical models, such as TIP4P/2005f (30), the flexibility of the bonds is modeled by using simple functional forms and a few parameters that do not depend on the environment. Another property that depends on the environment is the dipole moment of the water molecule. For instance, the dipole moment is different in liquid water and ice (22), but can also exhibit more subtle changes with the environment (31, 32). SCAN-ML is polarizable and able to capture the effects connected to changes in dipole moment (27, 32). SCAN-ML is also fully reactive and can describe the proton-transfer process in water. This model captures many-body interactions beyond two- and three-body, while mW is limited to three-body interactions, and TIP4P/Ice is based only on two-body interactions. SCAN-ML and TIP4P/Ice can both describe an important feature of ice I_h—namely, proton disorder, which is absent in the coarse-grained mW due to the lack of protons.

Our simulations based on SCAN-ML also have several limitations. While the electronic degrees of freedom are treated quantum-mechanically, the dynamics of the nuclei are based on the equations of motion of classical mechanics. Therefore, we ignore nuclear quantum effects (NQEs) that could be modeled by using path-integral MD. Another disadvantage is that SCAN-ML is around 1 to 2 orders of magnitude more computationally expensive than empirical models. Also, the properties of SCAN-ML differ somewhat from experimental properties, and this shows the limitations of the SCAN functional in the description of water and ice. Lastly, the model is short-ranged, with an interaction cutoff of 6 Å. It thus cannot capture the long-range electrostatic interactions (present, for instance, in TIP4P models) or van der Waals forces beyond this range. Long-range electrostatic interactions could be modeled by using the recently introduced deep potential long-range scheme (33).

We now turn to discuss the results of the seeding simulations. We studied ice I_h clusters of around 200; 700; and 4,500 molecules embedded in liquid water, and the corresponding total number of water molecules in the simulation boxes were around 4,000; 12,000; and 100,000, respectively. The choice of system size is discussed in detail in *SI Appendix*. The initial, equilibrated configurations of such ice clusters are shown in Fig. 1 *A–C*. We refer the reader to *Materials and Methods* for information about the equilibration procedure. The clusters are nearly spherical,
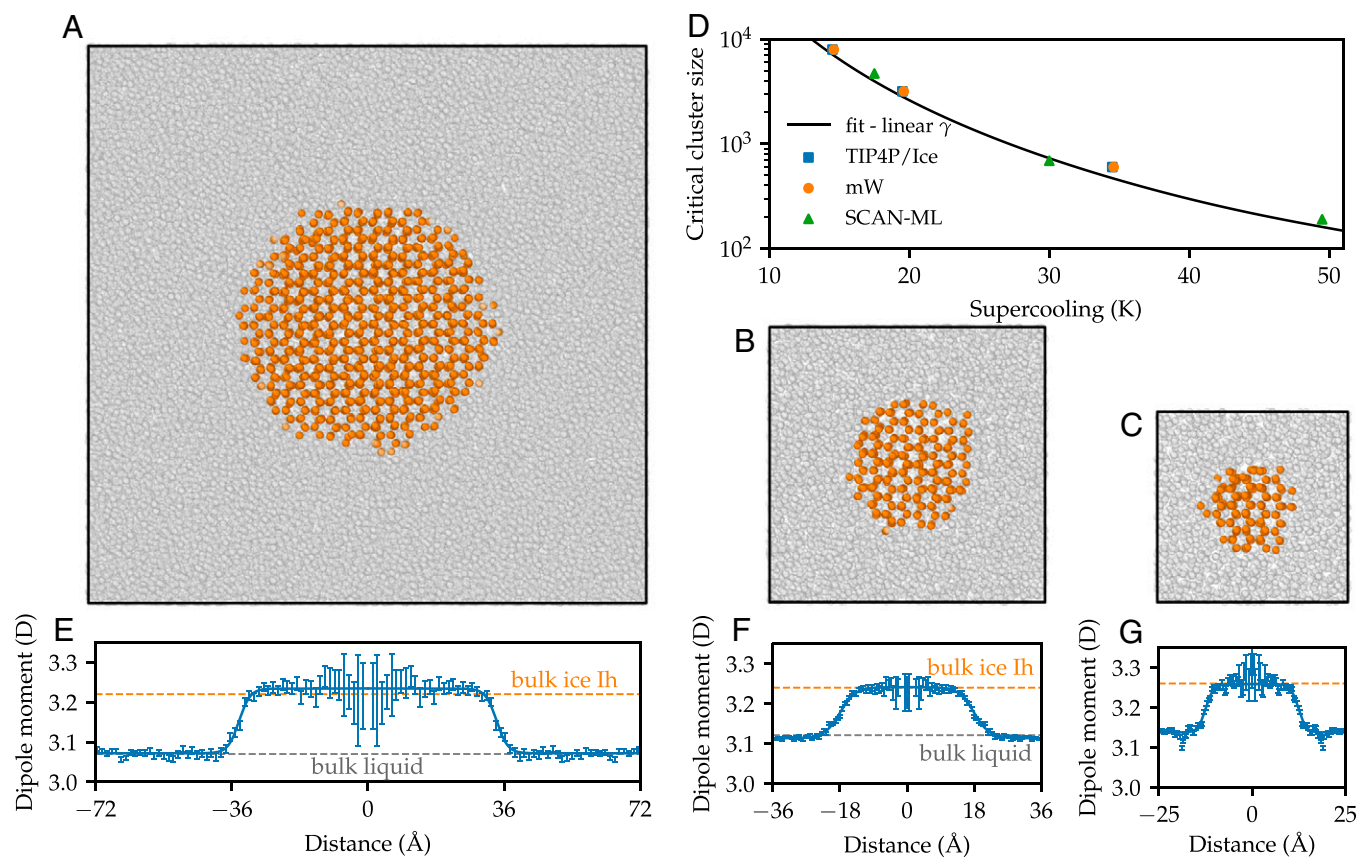
Fig. 1. Ice I_h clusters employed in the seeding simulations. *A–C* show snapshots of the equilibrated cluster configurations, in which only oxygen atoms are shown. Atoms with ice I_h-like environments (34, 35) are shown in orange, and atoms with liquid-like environments are shown in gray. *D* shows the supercooling for which each of the clusters is critical in the SCAN-ML model. The curve labeled fit-linear γ is based on the CNT formula $N^* = (32\pi\gamma^3)/(3\rho_{ice}^2|\Delta\mu|^3)$ and uses the linear fit to γ (see Fig. 3*A*). Results from ref. 36 for empirical models mW and TIP4P/Ice are also shown. *E*, *F*, and *G* show the quantum-mechanical average dipole moment of the water molecule as a function of the distance from the center of the ice clusters. Hyperbolic tangent functions were fit to the data and are shown in solid blue lines. Reference values for bulk ice I_h and liquid water were calculated at the equilibration temperatures (240 K, 275 K, and 290 K) and are shown with dashed orange and gray lines. The reference value for the liquid at 240 K is not provided due to the very long relaxation times at this temperature.

an observation that will be important when CNT is used to calculate several physical properties (see below). Some faceting of the clusters can be observed, and the hexagonal shape of the clusters is compatible with the sixfold symmetry of the basal plane of ice I_h.

MD simulations were performed at different temperatures, starting from the equilibrated configurations. The change in cluster size as a function of time is shown in *SI Appendix*, Fig. S1 for the three cluster sizes at different temperatures. From these simulations, we identified the temperatures T^* at which these clusters have equal probabilities of growing and thawing. In Fig. 1*D*, we show T^* for the three cluster sizes studied here with the SCAN-ML model. In order to determine the cluster size N^*, we must choose a local order parameter. The results depend somewhat on this choice (18), and, here, we employ a criterion similar to the one used by Espinosa et al. (19) in order to compare our results with the data reported therein (see *Materials and Methods* for details about our criterion to identify ice-like molecules). We also include in Fig. 1*D* the comparison with the results of Espinosa et al. (36) for two widely used empirical models—namely, mW and TIP4P/Ice. The results show that the critical cluster sizes are fairly independent of the model. At the highest supercooling studied here, around 50 K, the dynamics of liquid water are very slow, and thermal equilibration might not have been reached (see *SI Appendix* for a detailed discussion on the relaxation times of liquid water in the SCAN-ML model).

In order to illustrate the ability of SCAN-ML to capture subtle quantum-mechanical polarization effects, we calculated the

average dipole moment of the water molecule as a function of the distance from the center of the ice I_h clusters. The quantum-mechanical molecular dipole moment was computed according to the modern theory of polarization (37, 38), adopting the formulation in terms of Wannier centers (39). The dependence of the Wannier centers on the coordinates of the atoms in the system was described by a deep neural network, as described in refs. 31 and 32 (see *SI Appendix* for further details of the calculation). The results are shown in Fig. 1 *E–G*. The average dipole moment changes from around 3.25 D in the ice I_h cluster to around 3.1 D in the liquid water surrounding the cluster. We also show in Fig. 1 *E–G* the reference values for bulk ice I_h and liquid water (dashed lines), and the agreement with the dipoles in the cluster configurations is very good. There is also good agreement between the bulk dipole moments calculated here and reference values obtained with SCAN DFT (22). The experimental dipole moment of liquid water at 298 K is 2.9 ± 0.6 D and is reproduced relatively well by SCAN (22). We note that the average dipole moment of the water molecule is a function of the temperature. Since each cluster has been equilibrated at a different temperature, the reference bulk values differ in Fig. 1 *E–G*. Furthermore, in the configurations with the ice cluster embedded in liquid water, the average dipole moment transitions smoothly from the bulk ice I_h value to the bulk liquid value, and the water molecules at the interface have, on average, intermediate values of the dipole moment.

We now turn to assess the performance of the SCAN-ML model to describe ice nucleation. We calculate nucleation rates by

combining the information obtained from the seeding simulations with CNT. The predictions of CNT rest on various assumptions (17)—for instance, CNT assumes that clusters are spherical and show bulk ice I_h properties. Within CNT, the nucleation rate (nuclei per unit time per unit volume) is,

$$J = \rho_l Z f \exp(-\beta \, \Delta G^*), \qquad [1]$$

where ρ_l is the density of the liquid; Z is the Zeldovich factor, which represents the probability of a critical cluster to cross the energy barrier; f is the attachment rate; $\beta = 1/(k_B T)$; T is the temperature; and k_B is the Boltzmann constant. The nucleation free-energy barrier, ΔG^*, can be calculated by using the CNT formula,

$$\Delta G^* = \frac{|\Delta \mu| N^*}{2}, \qquad [2]$$

where $\Delta \mu$ is the difference in chemical potential between liquid water and ice I_h, and N^* is the number of water molecules in the critical cluster. Eq. **2** provides a convenient way to calculate rates from N^* and T^* obtained in the seeding simulations. ρ_l and $\Delta \mu$ of the SCAN-ML model were calculated in ref. 23, and further details about the determination of N^*, Z, f, ρ_l, and $\Delta \mu$ can be found in *Materials and Methods* and *SI Appendix*.

The nucleation rates thus calculated are shown in Fig. 2, together with results of experiments (40–46) and simulations using empirical models mW (36) and TIP4P/Ice (14, 16, 36). Most experiments are performed on micrometer-sized droplets and yield nucleation rates in the supercooling range 35 to 40 K (41, 42). However, experiments in the last 10 y have also used nano-sized droplets to reach much deeper supercoolings (40, 43). We have not included in our plot the experimental results of Laksmono et al. (49) since there are discrepancies between their rates and most measurements. Furthermore, it has been argued that more than one nucleus could have formed in those experiments (50). We have included in Fig. 2 a horizontal line representing the experimental homogeneous nucleation limit—i.e., the rate at which a micrometer-sized droplet freezes in 1 s (19). We note
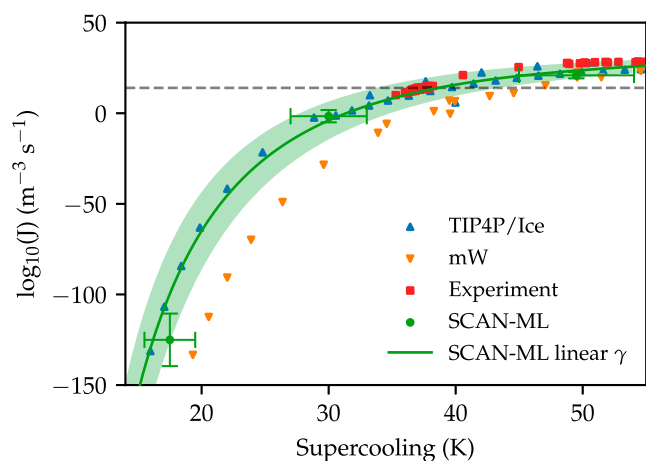

Fig. 2. Ice nucleation rates as a function of supercooling. Rates of the SCAN-ML model calculated in this work are compared with experimental data (40–46) and results from other works obtained using the models TIP4P/Ice (14, 16, 36) and mW (13, 36, 47, 48). We refer the reader to *SI Appendix*, Fig. S12 for further details about the computational techniques used to compute the rates of empirical models. The solid green line labeled SCAN-ML linear γ was obtained by using the CNT Eq. **1** and a linear fit to the interfacial free-energy data presented in Fig. 3. The green shaded area is an estimate of the error in this calculation. The experimental homogeneous nucleation limit (19) is shown as a horizontal gray dashed line and corresponds to $\log_{10}(J)(\mathrm{m}^{-3} \cdot \mathrm{s}^{-1}) = 14$. The calculation of error bars is described in *SI Appendix*.

that there are significant differences between the nucleation rates calculated from simulations using different methods. For instance, in the case of TIP4P/Ice, estimates from forward-flux sampling (14), metadynamics (16), and seeding (36) span about 10 orders of magnitude, which is, however, the typical error bar of the seeding technique. Nucleation rates of the SCAN-ML model are in good agreement with experimental measurements within the uncertainty of our calculation. Furthermore, the rates of SCAN-ML are intermediate between those of mW and TIP4P/Ice. Therefore, the performance of SCAN-ML is similar to that of the best available semiempirical models. As we shall see later, the inclusion of stacking disorder makes rates faster and reduces to some extent the discrepancy with experiment.

Another quantity that can be easily obtained from the seeding simulations is the interfacial free energy averaged over all orientations $\bar{\gamma}$. For this purpose, we employ the CNT expression,

$$\bar{\gamma} = \left(\frac{3 N^*}{32 \pi} \right)^{1/3} \rho_{ice}^{2/3} |\Delta \mu|, \qquad [3]$$

where the symbols have the same meaning as in Eq. **1** and Eq. **2**, and ρ_{ice} is the density of ice I_h. The results of this calculation are shown in Fig. 3A. Data for the mW and TIP4P/Ice models obtained from seeding simulations (36) are also shown. The dependence of $\bar{\gamma}$ on supercooling is a consequence of two different factors. The first is that the interfacial free energy of a flat interface depends on the temperature, and the second is that each cluster has a different size, and this will affect $\bar{\gamma}$, as shown by the Tolman equation (17).

In order to validate the results obtained using seeding simulations, we also calculated the interfacial free energy γ at coexistence for flat interfaces using advanced sampling simulations. For this purpose, we computed γ for the most relevant interfaces in ice I_h—namely, the prismatic ($1\bar{1}00$), secondary prismatic ($11\bar{2}0$), and basal (0001) planes. The method to compute γ for flat interfaces at coexistence is based on the reversible interconversion of the liquid and the respective liquid–ice I_h interface. This is achieved by a suitably designed bias potential that increases the probability of observing the high free-energy interfacial configuration. A schematic of the interface sampled during the simulation of the secondary prismatic plane is shown in Fig. 3B. Further details of this approach and its validation can be found in *Materials and Methods*. We also computed the interfacial free energy averaged over all orientations ($\bar{\gamma}$) as the mean of the three studied interfaces (19). The results of the free-energy calculations are summarized in Table 2, and $\bar{\gamma}$ is shown in Fig. 3A. As seen in this figure, the agreement between $\bar{\gamma}$ obtained from advanced sampling calculations and seeding is very good. For reference, we show in Table 2 results for the models mW and TIP4P/Ice, as reported in ref. 52.

We have also included in Fig. 3A and in Table 2 experimental results for $\bar{\gamma}$ at the melting temperature (53). There is no direct experimental measurement of $\bar{\gamma}$ at other temperatures, and estimates based on CNT differ significantly (53). For this reason, we have not included them in our analysis. The spread of the experimental results at the melting temperature is relatively large ($\sim$20 mJ/m^2) and has a mean value $\sim$31.5 mJ/m^2 after removing outliers. It has also been argued (19, 53) that the experiments of Hardy (51) based on the shape of the grain-boundary groove provide the most reliable estimate, with a value of 29.1 ± 0.8 mJ/m^2. $\bar{\gamma}$ for SCAN-ML is well within the region of uncertainty of the experimental measurements. However, the interfacial free energy of SCAN-ML is higher than the average experimental estimate. This behavior can be rationalized by taking into account that SCAN-ML has a melting temperature and enthalpy of fusion higher than both the
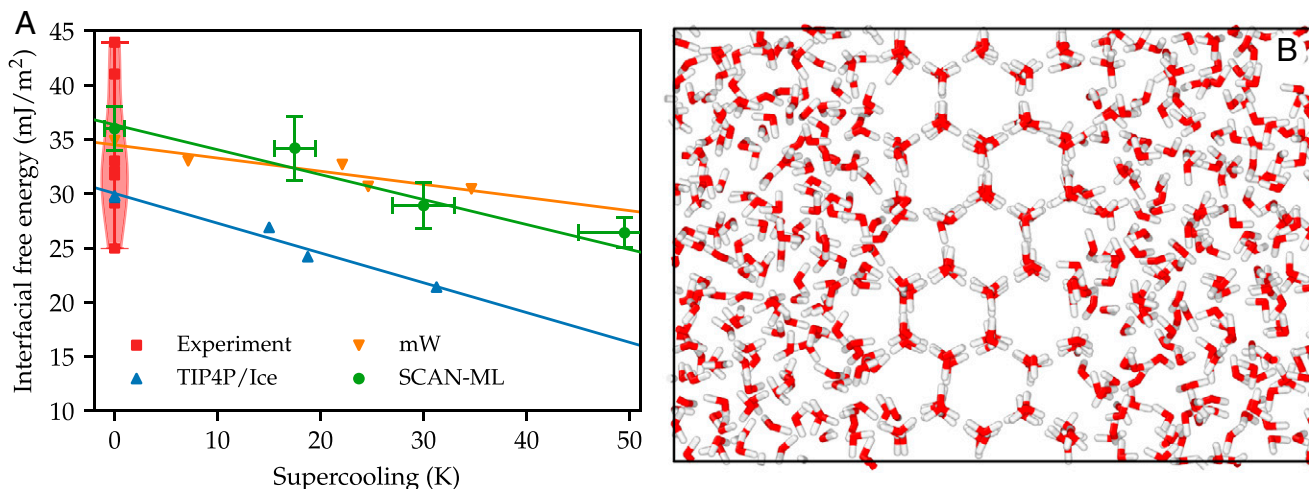
Fig. 3. Liquid water–ice I$_h$ interfacial free energy. (*A*) Interfacial free energy as a function of supercooling. Results for SCAN-ML at supercooling different from zero were obtained by using data from the seeding simulations and assuming the validity of CNT. We have included data from refs. 36 and 52 for the models mW and TIP4P/Ice and experimental measurements at the melting temperature (53). Linear fits to the results of the different models are shown as solid lines. The distribution of the experimental data is shown in red using a violin plot. The calculation of error bars is described in *SI Appendix*. (*B*) Planar interface between liquid water and the secondary prismatic plane of ice I$_h$. This configuration was extracted from an advanced sampling simulation at 312 K driven by SCAN-ML, during which the interface reversibly forms and melts.

corresponding numbers for real water (experiments) and TIP4P/Ice. Turnbull (54) observed that there is a strong correlation between the interfacial free energy and the enthalpy of fusion, and Laird (55) has made a similar observation for the correlation between the interfacial free energy and the melting temperature. It is thus expected that the interfacial free energy of SCAN-ML should be higher than in the experiment and in TIP4P/Ice [melting temperature ∼270 K (5)]. In *SI Appendix*, Fig. S4, we show that, indeed, the interfacial free energy correlates very well with the melting temperature in the TIP4P family and SCAN-ML. An option to account for the different melting temperatures of the models is to compare $\bar{\gamma}$ using units of $k_B T$ for the energy. A plot of $\bar{\gamma}$ in units of $k_B T/\text{m}^2$ vs. supercooling is shown in *SI Appendix*, Fig. S11. One could also estimate the value of $\bar{\gamma}$ in mJ/m^2 that SCAN-ML would have if its melting temperature were the experimental one. An appropriate rescaling of $\bar{\gamma}$ is $\bar{\gamma}' = \bar{\gamma} \, T_m^{\text{exp}}/T_m^{\text{SCAN-ML}}$, where T_m^{exp} and $T_m^{\text{SCAN-ML}}$ are the melting temperatures in the experiment and SCAN-ML, respectively. In this way, we obtain a interfacial free energy for SCAN-ML of $\bar{\gamma}' = 31.5$ mJ/m^2 at the melting temperature.

We now turn to analyze the thermodynamic properties of the models that affect nucleation rates by assuming the validity of CNT. The nucleation barrier ΔG^* controls nucleation rates at low and moderate supercoolings since it is exponentiated in Eq. **1**. The CNT expression for ΔG^* is,

Table 2. Interfacial free energy of ice I$_h$ with liquid water at coexistence

	Interfacial free energy (mJ/m^2)			
	$\gamma_{(1\bar{1}00)}$	$\gamma_{(11\bar{2}0)}$	$\gamma_{(0001)}$	$\bar{\gamma}$
SCAN-ML	36(2)	34(2)	37(2)	36(2)
Experiment (51)	—	—	—	29.1(8)
Experiment (avg.)	—	—	—	∼31.5
TIP4P/Ice (52)	31.6(8)	30.7(8)	27.2(8)	29.8(8)
mW (52)	35.1(8)	35.2(8)	34.5(8)	34.9(8)

We report results for the prismatic (1$\bar{1}$00), secondary prismatic (11$\bar{2}$0), and basal (0001) planes. The interfacial free energy averaged over all orientations $\bar{\gamma}$ is also reported. We have included experimental results (51, 53) (see text for details) and calculations using the mW and TIP4P/Ice models (52).
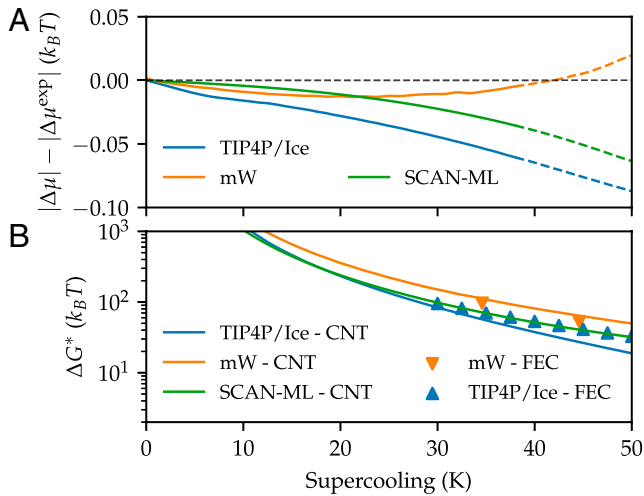
$$\Delta G^* = \frac{16\pi\bar{\gamma}^3}{3\rho_{\text{ice}}^2 |\Delta\mu|^2}. \qquad [4]$$

Therefore, the central physical quantities that govern nucleation rates at low and intermediate supercoolings are 1) the difference in chemical potential between liquid water and ice I$_h$ ($\Delta\mu$), 2) the interfacial free energy of ice I$_h$ with liquid water ($\bar{\gamma}$), and 3) the density of ice (ρ_{ice}). In the next paragraphs, we analyze these quantities for SCAN-ML, TIP4P/Ice, and mW.

In Fig. 4*A*, we show the difference between $|\Delta\mu|$ in different models and in the experiment $|\Delta\mu^{\text{exp}}|$ as a function of supercooling. $|\Delta\mu^{\text{exp}}|$ cannot be measured directly, and its calculation from experimentally measured heat capacities of liquid water and ice I$_h$ (56) is described in *SI Appendix*. $|\Delta\mu| - |\Delta\mu^{\text{exp}}|$ is reported in units of $k_B T$ in Fig. 4*A* in order to compare models with different melting temperatures. At 35 K of supercooling, $|\Delta\mu|$ is underestimated by 9% in SCAN-ML. The performance of SCAN-ML in describing this property is somewhat better than that of TIP4P/Ice, which underestimates $\Delta\mu$ by 17% at the same supercooling. The mW model is the most accurate among the models considered here, with $|\Delta\mu|$ at 35 K within 1% of the experimental value. However, mW changes from a underestimation of $|\Delta\mu|$ at low supercoolings to an overestimation at large supercoolings. This is a consequence of a much weaker deviation of $\Delta\mu$ from a linear dependence with temperature than the other models (*SI Appendix*, Fig. S8).

Results for the interfacial free energy are presented in Fig. 3*A*. There is limited experimental information to ascertain the deviation of the interfacial free energy with respect to experiments. However, the values of $\bar{\gamma}$ for TIP4P/Ice and SCAN-ML (adjusted for the different melting temperature) are in relatively good agreement with most experimental results and most likely within a 5% error. Instead, $\bar{\gamma}$ in the mW model is around 35 mJ/m^2, which is higher than the most reliable experimental estimates of $\bar{\gamma}$ and is most likely overestimated by around 10%. It is also possible to characterize the temperature dependence of $\bar{\gamma}$ using the interfacial entropy, $S_\gamma = -\partial\bar{\gamma}/\partial T$, that can be estimated from the slope of $\bar{\gamma}$, with respect to temperature in Fig. 3*A*. We observe that mW has a lower slope $\partial\bar{\gamma}/\partial T$ than SCAN-ML and TIP4P/Ice, and that the latter two models have a similar slope. This indicates that the interfacial entropy of the coarse-grained mW model is higher than

Fig. 4. Analysis of the influence of supersaturation on nucleation barriers. (*A*) Difference between the driving force for nucleation in computer models $|\Delta\mu|$ and in the experiment $|\Delta\mu^{\text{exp}}|$ as a function of supercooling. $|\Delta\mu|$ for the mW and TIP4P/Ice models was obtained from ref. 19. The calculation of $|\Delta\mu^{\text{exp}}|$ is based on experimental heat capacities (56). Above 38 K of supercooling results are shown as dashed lines to highlight the uncertainty in $|\Delta\mu^{\text{exp}}|$ due to the lack of experimental measurements of the heat capacity of liquid water at these conditions. See *SI Appendix* for further details on the calculation of $|\Delta\mu|$. (*B*) Nucleation barrier ΔG^* calculated using CNT (Eq. **4**— see text for details). Results from umbrella sampling and metadynamics free energy calculations (FEC) reported in refs. 16, 29, and 57 are also shown.

in the TIP4P/Ice and SCAN-ML all-atom models that include protons explicitly. Following ref. 28, the interfacial entropy can also be calculated by using,

$$S_\gamma \approx -\frac{\bar{\gamma}(T_m)\,\Delta C_p(T_m)}{\Delta H_f}, \tag{5}$$

where ΔH_f is the enthalpy of fusion, and $\Delta C_p(T_m)$ is the difference in heat capacity between liquid water and ice I_h at the melting temperature T_m. Using the values of ΔH_f and $\bar{\gamma}$, reported in Tables 1 and 2, and $\Delta C_p(T_m) = 49$ J/(mol · K) (23), we obtain $S_\gamma \approx -232\ \mu\text{J} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$ for SCAN-ML. This result is in good agreement with S_γ calculated from experimental data $(-215\ \mu\text{J} \cdot \text{m}^{-2} \cdot \text{K}^{-1})$ (28). A similar analysis for TIP4P/Ice gives $S_\gamma \approx -226\ \mu\text{J} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$, also in good agreement with the experiment. The mW model has a S_γ of $-44\ \mu\text{J} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$ (28) that is around a fifth of the experimental value. These results are in agreement with the discussion above based on the slopes of the lines in Fig. 3*A*. From this discussion, we deduce that an all-atom description seems essential to capture $\bar{\gamma}$ and its temperature dependence.

The density of ice I_h in the different models considered here is shown in Table 1 and in *SI Appendix*, Fig. S8 (data from refs. 29, 36, and 58). The density of SCAN-ML ice I_h is around 3% higher than in the experiment and, according to Eq. **4**, this would partially compensate the somewhat low $|\Delta\mu|$ in this model. For TIP4P/Ice, the density of ice I_h is around 1% lower than in the experiment, and we expect it to have a negligible effect compared to other errors. Finally, the mW model overestimates the density of ice I_h with respect to experiment by ~7%, and this might compensate in part for a large $\bar{\gamma}$.

We then computed the nucleation free-energy barriers using Eq. **4**, the values for $\Delta\mu$ and ρ_{ice} reported in *SI Appendix*, Fig. S8, and the linear fits to $\bar{\gamma}$ shown in Fig. 3*A*. The results are shown in Fig. 4*B*. We have included barriers from refs. 16, 29, and 57 obtained by using umbrella sampling and metadynamics free-energy calculations. For SCAN-ML, we expect that the barrier

should be overestimated since $|\Delta\mu|$ is underestimated. This is compatible with the nucleation rates being somewhat slower than the experiment (Fig. 2). In the TIP4P/Ice water model, $|\Delta\mu|$ is underestimated more than in SCAN-ML, and, therefore, we would expect an overestimation of the nucleation barrier and nucleation rates slower than in the experiment. At variance with this prediction, the seeding nucleation rates (36) of TIP4P/Ice seem to agree relatively well with the experiments. The rates calculated by Niu et al. (16) and Haji-Akbari and Debenedetti (14) for TIP4P/Ice, however, are slower than the experimental measurements. In the case of the mW water model, $\Delta\mu$ is in very good agreement with the experiment. For this reason, we surmise that the slow nucleation rates in this model can be traced back to an overestimation of $\bar{\gamma}$ not fully compensated by the overestimation of ρ_{ice}.

Another important aspect of ice nucleation is stacking disorder. There is significant experimental (59) and computational (2) evidence that nucleating ice clusters contain stacking faults—i.e., alternating layers of ice I_h and ice I_c—and the solid polymorph that exhibits this feature is called ice I_{sd} (60). The prevalence of stacking faults in ice at equilibrium depends on two thermodynamic properties—namely, the difference in chemical potential between ice I_h and ice I_c, $\Delta\mu_{I_h \to I_c}$, and the interfacial free energy between these two polymorphs, $\gamma_{I_h \to I_c}$. The experimental evidence on the value of $\Delta\mu_{I_h \to I_c}$ is limited due to the difficulty in obtaining pure ice I_c, although very recently, it has become possible to prepare samples with high structural purity (61). The available experimental data put $\Delta\mu_{I_h \to I_c}$ in the range from 0 to ~200 J/mol (see ref. 62 for a review). An alternative point of view is provided by Lupi et al. (2), who argue that the experimental results reported in ref. 63 put an upper limit to $\Delta\mu_{I_h \to I_c}$ at 16.5 ± 1.7 J/mol. On the computational side, the TIP4P/Ice and mW models have very small values of $\Delta\mu_{I_h \to I_c}$ of ~0 (64) and ~5 J/mol (29, 65), respectively. In ref. 23, we have found a $\Delta\mu_{I_h \to I_c}$ for the SCAN-ML model of 65 ± 37 J/mol. As we shall see, the precise value of $\Delta\mu_{I_h \to I_c}$ has an influence on rates, and further experimental and computational efforts are needed to shed light on its value.

We described the effect of stacking disorder using a model for the chemical potential of ice I_{sd} that rests on the following assumptions: 1) The entropy of mixing of ice I_c and ice I_h layers is ideal; 2) the interfacial free energy is negligible: and 3) stacking is only relevant in one direction—namely, the direction perpendicular to the basal plane of ice I_h. It can be shown that the first two assumptions give a lower bound for the chemical potential of ice I_{sd}. Since the effect of stacking disorder is more relevant when the chemical potential of ice I_{sd} is lower, then our model gives an upper bound for the possible effects of stacking disorder. A more sophisticated two-dimensional model has been used by Lupi et al. (2), and it was found that the simplified one-dimensional model underestimates the entropic stabilization due to stacking disorder. We also note that a similar model has been used by Pronk and Frenkel (66). Further details can be found in *Materials and Methods*. In Fig. 5*A*, we show the difference in chemical potential between ice I_{sd} and ice I_h, $\Delta\mu_{I_{sd} \to I_h}$, as a function of supercooling, as obtained from our model. The model takes as input the difference in chemical potential between ice I_c and ice I_h, $\Delta\mu_{I_h \to I_c}$. We used two different values for this quantity, one compatible with the free energy of the mW model, $\Delta\mu_{I_h \to I_c} = 5$ J/mol, and another one compatible with the SCAN-ML model, $\Delta\mu_{I_h \to I_c} = 65$ J/mol. In both cases, $\Delta\mu_{I_{sd} \to I_h}$ becomes negligible as the supercooling goes to zero and the critical cluster size goes to infinity. This reflects the fact that ice I_h is the most stable phase in the thermodynamic limit. At
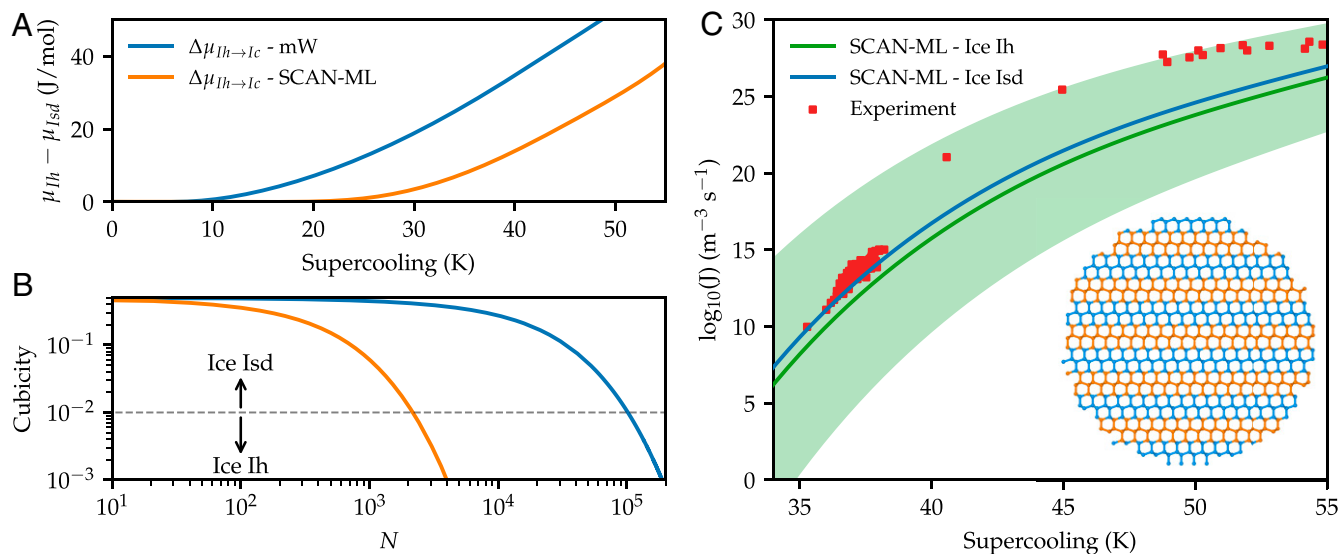
Fig. 5. Influence of stacking disorder on the rates. (*A*) Difference in chemical potential between ice I$_{sd}$ and ice I$_h$, $\Delta\mu_{I_{sd}\to I_h} = \mu_{I_h} - \mu_{I_{sd}}$, as a function of supercooling. (*B*) Cubicity as a function of the number of molecules N in a spherical ice cluster. The threshold to distinguish ice I$_{sd}$ from ice I$_h$ is shown with a dashed gray line and corresponds to a 1% cubicity. (*C*) Nucleation rates as a function of supercooling. Rates for ice I$_h$ are compared with experimental data and the results for ice I$_{sd}$ obtained with a model for stacking disorder. The green shaded region corresponds to the error in the rate of ice I$_h$ and was calculated as described in *SI Appendix*. Two values are considered for $\mu_{I_c} - \mu_{I_h}$ in Eq. **10**, one compatible with mW (5 J/mol) and another one compatible with SCAN-ML (65 J/mol).

larger supercoolings, the finite size effects start to be important, and ice I$_{sd}$ becomes progressively more stable against ice I$_h$. The magnitude of the stabilization, around 50 J/mol, is small compared with $|\Delta\mu|$, even at relatively large supercoolings.

It is also interesting to evaluate how $\Delta\mu_{I_h\to I_c}$ affects the cubicity of the nucleating clusters. We evaluated the cubicity as a function of the size of the cluster, and the results are reported in Fig. 5*B*. We found that for $\Delta\mu_{I_h\to I_c} = 5$ J/mol, the cubicity drops below 1% for clusters of around 100,000 molecules, in excellent agreement with findings of Lupi et al. (2) for the mW water model. Instead, for $\Delta\mu_{I_h\to I_c} = 65$ J/mol, the cubicity drops below 1% at around 2,000 molecules. Therefore, the extent to which stacking disorder is relevant in small clusters depends largely on $\Delta\mu_{I_h\to I_c}$.

$\Delta\mu_{I_{sd}\to I_h}$ can be used within CNT to estimate the nucleation rates of ice I$_{sd}$. In Fig. 5*C*, we show the nucleation rates of ice I$_{sd}$ calculated in this fashion. For clarity, we only show the results using a low value of $\Delta\mu_{I_h\to I_c}$—namely, 5 J/mol, since this gives the greatest effect for the rates and is easier to visualize. For this reason, it should be considered an upper bound for the effect, rather than the most reliable estimate. Despite the systematic choices we have made to obtain the maximum possible influence of stacking disorder on the rates, the effect of stacking disorder is around 2 or 3 orders of magnitude at deep supercoolings. This relatively small change, however, improves somewhat the agreement of SCAN-ML with the experiment. We also calculated nucleation rates using $\Delta\mu_{I_h\to I_c} = 65$ J/mol, and the results are shown in *SI Appendix*, Fig. S9.

This work shows that the latest advances in ab initio MD allow studies of complex phenomena such as ice nucleation from first principles. Our findings indicate that nucleation rates predicted based on SCAN DFT are in reasonably good agreement with experiment. The rates are similar to those estimated with the TIP4P/Ice model and somewhat faster than the rates of the mW model. The nucleation rate is a complex quantity that depends on many different properties of an atomistic model, such as the density of liquid water and ice, the water–ice interfacial free energy, and the difference in chemical potential between water and ice. We have performed a careful analysis of these properties, and

we have also compared them to the results of empirical models. SCAN-ML gives a balanced description of these properties that results in good agreement between the calculated rates and the experimental measurements.

We have also highlighted limitations of the SCAN functional in describing some of the properties of liquid water and ice. More accurate functional approximations and/or higher-level quantum chemical data are expected to improve the description of the properties of water. It has also been shown that the MB-pol model (67) based on Coupled Cluster CCSD(T) calculations reproduces experimental properties with high accuracy and is thus an interesting model to study in the future. Furthermore, in this work, we have neglected NQEs that may have an important impact on some properties. For instance, the difference in chemical potential between liquid water and ice is influenced by heat capacities, and the latter are affected significantly by NQEs (68). Despite their possible relevance, modeling NQEs through path-integral MD is still computationally impractical for nucleation simulations that employ large system sizes, such as the ones considered here. Understanding the impact of NQEs on ice nucleation is an interesting direction for future work. Finally, ab initio machine-learning models of water can be extended to describe a substrate in order to simulate heterogeneous ice nucleation, a process of direct relevance to atmospheric science and climate modeling. This would allow one to include hitherto-neglected phenomena, such as the effect of pH and the spontaneous hydroxylation of surfaces. For these reasons, we foresee continued progress in the simulation of ice nucleation from first principles and the prediction of rates that are in progressively improved agreement with experiments, as a result of an accurate description of the thermodynamic and kinetic properties of water and ice.

Materials and Methods

Molecular Dynamics. Simulations were performed by using LAMMPS (69) patched with the DeePMD-kit (70). The temperature was kept constant with the stochastic velocity-rescaling algorithm (71) using a relaxation time of 0.1 ps. A Parrinello–Rahman-type barostat was used to maintain the pressure at 1 bar,

and a relaxation time of 1 ps was employed. For the seeding simulations, an isotropic barostat was used, whereas for the advanced sampling simulations, only the pressure component in the direction perpendicular to the interface was controlled. The SCAN-ML model used with the DeePMD-kit was exactly the same as the one employed in ref. 23. The performance of the implementation that we used was around 1 ns/d using an optimal number of graphical processing units (GPUs) for a given system size. The performance with the latest version of DeePMD-kit would have been faster, at around 10 ns/d.

Seeding Simulations. Configurations for the seeding simulations were constructed in the following way. A simulation box with water molecules was prepared, and then a spherical cavity was carved from its center. This region was later filled with a seed of ice I_h with proton disorder created by using GenIce (72). This procedure was repeated for three different system sizes. These systems contained 3,934; 11,872; and 99,404 water molecules, respectively. Afterward, the energy was minimized to a relative accuracy 10^{-6}, and a 1-ns MD simulation for equilibration was performed at a temperature below the one for which the cluster is critical, in order to avoid partial melting of the cluster. The corresponding temperatures were 240; 275; and 290 K for the smallest, intermediate, and largest cluster, respectively. After equilibration, MD simulations were run at different temperatures to find T^*. The simulations for the largest system were run on the Summit supercomputer using 600 Nvidia V100 GPUs. Each of the simulations would have required ~5 y to be completed in a single GPU. The intermediate and small system sizes used 100 and 24 GPUs per simulation, respectively.

The size of the clusters was determined by using the local Steinhardt parameter $\bar{Q}_6$ proposed by Lechner and Dellago (73). $\bar{Q}_6$ was calculated by using the Freud (74) Python library (version 2.7.0). The threshold value of $\bar{Q}_6$ that separates liquid and ice I_h environments was determined for each temperature by using the criterion that an environment with the threshold value of $\bar{Q}_6$ has equal probabilities of being classified as liquid or ice I_h. Probability densities of $\bar{Q}_6$ for the liquid and ice I_h at different temperatures are shown in *SI Appendix*, Fig. S5A. The chosen thresholds as a function of temperature are shown in *SI Appendix*, Fig. S5B and show a linear correlation. We used a linear fit to these data to determine the threshold at any intermediate temperature. All seeding simulations were analyzed by using a threshold appropriate for the temperature of the simulation. We note that the overlap of the liquid and ice I_h distributions increase sharply upon crossing the Widom line. This can be expected, given that below the Widom line, liquid water resembles the low-density liquid (LDL) water phase and ice I_h is known to be more similar to LDL than to the high-density liquid. In Fig. 1, the classification in ice-like and liquid-like environments was performed with the Polyhedral Template Matching algorithm (34), as implemented in OVITO (35).

The prefactor in the CNT expression for the nucleation rates (Eq. **2**) requires the calculation of the Zeldovich factor Z and the attachment rate f. Z was calculated using,

$$Z = \sqrt{\frac{|\Delta\mu|}{6\pi k_B T N^*}},\qquad [6]$$

and f was calculated using (19),

$$f = \frac{\langle (N(t) - N(0))^2 \rangle}{2t},\qquad [7]$$

where $N(t)$ is the cluster size at time t. In practice, we computed f from the slope of $\langle (N(t) - N(0))^2 \rangle$ vs. $2t$. Results for f are shown in *SI Appendix*, Fig. S7.

Advanced Sampling Simulations. The calculation of the ice I_h–liquid-water interfacial free energy was performed with LAMMPS augmented with the PLUMED enhanced sampling plugin (75, 76). The initial configuration was made by using GenIce and consisted of 288 water molecules in the ice I_h structure with proton disorder. An equilibration of 1 ns at 312 K and 1 bar was then carried out, and the box dimensions were set to their average values during this run. In order to obtain simulation boxes adequate for the simulation of the prismatic, secondary prismatic, and basal interfaces, the box was replicated along one of the three main axes, and then the solid configuration was melted in a 1-ns run at 450 K, while only the direction along which the box was replicated was barostated.

Next, we performed an advanced sampling simulation for each interface, in which a bias potential was constructed using the On-the-fly Probability Enhanced Sampling (OPES) method (77). This method is an evolution of the well-known

metadynamics technique (15). The OPES bias potential was built as a function of a collective variable that counts the number of environments compatible with ice I_h in a region around an arbitrarily chosen atom (for instance, atom number 1). The number of environments compatible with ice I_h was calculated by using the environment similarity (78) metric, taking the four tetrahedral reference environments of ice I_h χ_i with $i = 1, .., 4$. As a result of the introduction of the bias potential, during the biased simulations, a slab of the ice I_h crystal is reversibly formed and melted. The free-energy difference between the liquid and the slab was calculated using,

$$\Delta G = -k_B T \log\left(\frac{Z^{\ddagger}}{Z^l}\right),\qquad [8]$$

where $Z^{\ddagger}$ and Z^l are the partition functions of the slab and the liquid. The interfacial free energy can then be calculated as,

$$\gamma = \frac{|\Delta G|}{2A},\qquad [9]$$

with A the cross-section of the interface. Further details are provided in *SI Appendix*.

This approach was validated by calculating the interfacial free energy of TIP4P/Ice that is known from literature (52). The interfacial free energy of TIP4P/Ice averaged all interfaces was found to be 31(1) mJ/m^2 in good agreement with the estimate from literature 29.8(8) mJ/m^2.

Model for Stacking Disorder. Stacking disorder was modeled by using the following expression for the difference between the chemical potential of ice I_{sd} and I_h:

$$\mu_{I_{sd}}(C, N) - \mu_{I_h} = C(\mu_{I_c} - \mu_{I_h}) - \frac{1}{N} T S_{mix} + \frac{1}{N} \sum_i \gamma_{sf} A_i,\qquad [10]$$

where C is the cubicity, N is the number of molecules, the index i runs through the ice I_h–ice I_c interfaces, γ_{sf} is the interfacial free energy of the stacking faults, and A_i is the area of the ith interface. The first term in Eq. **10** is the bulk contribution of ice I_h and ice I_c. The second term is the contribution from the entropy of mixing of the stacked layers. We assume that stacking is relevant only in one direction–i.e., the direction perpendicular to the basal plane of ice I_h. The last term in Eq. **10** takes into account the penalty to form an ice I_h–ice I_c interface. The second and third terms go to zero as $N \rightarrow \infty$, reflecting that stacking disorder is only relevant for finite systems.

Since the effect of stacking disorder on the rates and chemical potentials is small, we do the following approximations. First, we neglect the third term in Eq. **10** that is always positive. Second, we approximate the entropy of mixing with the ideal entropy of mixing,

$$S_{mix} \approx -N_l k_B (C \log(C) + (1 - C) \log(1 - C)),\qquad [11]$$

where N_l is the number of stacked layers. The ideal entropy of mixing is always larger than S_{mix}. These choices give a lower bound for $\mu_{I_{sd}}(C, N) - \mu_{I_h}$ and thus the greatest possible influence on the rates. We make the additional assumption that the cluster of N molecules is approximately spherical and calculate N_l using the expression,

$$N_l(N) = \frac{D}{d} = \left(\frac{6N}{\pi \rho_{ice}}\right)^{1/3} \frac{1}{d},\qquad [12]$$

where D is the diameter of the cluster and d is the distance between layers of the basal plane. The cubicity and chemical potential in equilibrium are found by minimizing $\mu_{I_{sd}}(C, N)$ with respect to C. In order to obtain $\mu_{I_{sd}}(C, N)$ as a function of temperature, we replace N with the number of molecules $N^*_{I_{sd}}(T)$ in a critical cluster with stacking disorder at a given temperature T. $N^*_{I_{sd}}(T)$ is not known, but can be approximated by the number of molecules $N^*_{I_h}(T)$ in a critical cluster of ice I_h that has been computed by using seeding simulations.

1. J. Vergara-Temprado et al., Strong control of Southern Ocean cloud reflectivity by ice-nucleating particles. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2687–2692 (2018).
2. L. Lupi et al., Role of stacking disorder in ice nucleation. *Nature* **551**, 218–222 (2017).
3. E. Sanz et al., Homogeneous ice nucleation at moderate supercooling from molecular simulation. *J. Am. Chem. Soc.* **135**, 15008–15017 (2013).
4. V. Molinero, E. B. Moore, Water modeled as an intermediate element between carbon and silicon. *J. Phys. Chem. B* **113**, 4008–4016 (2009).
5. J. L. Abascal, E. Sanz, R. García Fernández, C. Vega, A potential model for the study of ices and amorphous water: TIP4P/Ice. *J. Chem. Phys.* **122**, 234511 (2005).
6. R. Car, M. Parrinello, Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
7. W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133 (1965).
8. J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
9. L. Bonati, M. Parrinello, Silicon liquid structure and crystal nucleation from ab initio deep metadynamics. *Phys. Rev. Lett.* **121**, 265701 (2018).
10. H. Niu, L. Bonati, P. M. Piaggi, M. Parrinello, Ab initio phase diagram and nucleation of gallium. *Nat. Commun.* **11**, 2654 (2020).
11. M. Matsumoto, S. Saito, I. Ohmine, Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. *Nature* **416**, 409–413 (2002).
12. P. G. Bolhuis, D. Chandler, C. Dellago, P. L. Geissler, Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **53**, 291–318 (2002).
13. T. Li, D. Donadio, G. Russo, G. Galli, Homogeneous ice nucleation from supercooled water. *Phys. Chem. Chem. Phys.* **13**, 19807–19813 (2011).
14. A. Haji-Akbari, P. G. Debenedetti, Direct calculation of ice homogeneous nucleation rate for a molecular model of water. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10582–10588 (2015).
15. A. Laio, M. Parrinello, Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562–12566 (2002).
16. H. Niu, Y. I. Yang, M. Parrinello, Temperature dependence of homogeneous nucleation in ice. *Phys. Rev. Lett.* **122**, 245501 (2019).
17. V. Kalikmanov, *Nucleation Theory* (Lecture Notes in Physics, Springer, Dordrecht, Netherlands, 2012), vol. 860.
18. N. E. R. Zimmermann et al., NaCl nucleation from brine in seeded simulations: Sources of uncertainty in rate estimates. *J. Chem. Phys.* **148**, 222838 (2018).
19. J. Espinosa, E. Sanz, C. Valeriani, C. Vega, Homogeneous ice nucleation evaluated for several water models. *J. Chem. Phys.* **141**, 18C529 (2014).
20. C. P Lamas et al., Homogeneous nucleation of NaCl in supersaturated solutions. *Phys. Chem. Chem. Phys.* **23**, 26843–26852 (2021).
21. J. Sun, A. Ruzsinszky, J. P. Perdew, Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **115**, 036402 (2015).
22. M. Chen et al., Ab initio theory and modeling of water. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10846–10851 (2017).
23. P. M. Piaggi, A. Z. Panagiotopoulos, P. G. Debenedetti, R. Car, Phase equilibrium of water with hexagonal and cubic ice using the scan functional. *J. Chem. Theory Comput.* **17**, 3065–3077 (2021).
24. L. Zhang, J. Han, H. Wang, R. Car, E. Weinan, Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
25. L. Zhang, H. Wang, R. Car, E. Weinan, Phase diagram of a deep potential water model. *Phys. Rev. Lett.* **126**, 236001 (2021).
26. T. E. Gartner, 3rd et al., Signatures of a liquid-liquid transition in an ab initio deep neural network model for water. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26040–26046 (2020).
27. P. M. Piaggi, R. Car, Enhancing the formation of ionic defects to study the ice Ih/XI transition with molecular dynamics simulations. *Mol. Phys.* **119**, e1916634 (2021).
28. Y. Qiu, L. Lupi, V. Molinero, Is water at the graphite interface vapor-like or ice-like? *J. Phys. Chem. B* **122**, 3626–3634 (2018).
29. S. Prestipino, The barrier to ice nucleation in monatomic water. *J. Chem. Phys.* **148**, 124505 (2018).
30. M. A. González, J. L. Abascal, A flexible model for water based on TIP4P/2005. *J. Chem. Phys.* **135**, 224516 (2011).
31. L. Zhang et al., Deep neural network for the dielectric response of insulators. *Phys. Rev. B* **102**, 041121 (2020).
32. G. M. Sommers, M. F. Calegari Andrade, L. Zhang, H. Wang, R. Car, Raman spectrum and polarizability of liquid water from deep neural networks. *Phys. Chem. Chem. Phys.* **22**, 10592–10602 (2020).
33. L. Zhang et al., A deep potential model with long-range electrostatic interactions. *J. Chem. Phys.* **156**, 124107 (2022).
34. P. M. Larsen, S. Schmidt, J. Schiøtz, Robust structural identification via polyhedral template matching. *Model. Simul. Mater. Sci. Eng.* **24**, 055007 (2016).
35. A. Stukowski, Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool. *Model. Simul. Mater. Sci. Eng.* **18**, 015012 (2009).
36. J. R. Espinosa, C. Navarro, E. Sanz, C. Valeriani, C. Vega, On the time required to freeze water. *J. Chem. Phys.* **145**, 211922 (2016).
37. R. Resta, Theory of the electric polarization in crystals. *Ferroelectrics* **136**, 51–55 (1992).
38. R. D. King-Smith, D. Vanderbilt, Theory of polarization of crystalline solids. *Phys. Rev. B Condens. Matter* **47**, 1651–1654 (1993).
39. R. Resta, Macroscopic polarization in crystalline dielectrics: The geometric phase approach. *Rev. Mod. Phys.* **66**, 899 (1994).
40. A. J. Amaya, B. E. Wyslouzil, Ice nucleation rates near ∼225 K. *J. Chem. Phys.* **148**, 084501 (2018).
41. D. E. Hagen, R. J. Anderson, J. L. Kassner, Jr., Homogeneous condensation–freezing nucleation rate measurements for small water droplets in an expansion cloud chamber. *J. Atmos. Sci.* **38**, 1236–1243 (1981).
42. B. Krämer et al., Homogeneous nucleation rates of supercooled water measured in single levitated microdroplets. *J. Chem. Phys.* **111**, 6521–6527 (1999).
43. A. Manka et al., Freezing water in no-man's land. *Phys. Chem. Chem. Phys.* **14**, 4505–4516 (2012).
44. B. J. Murray et al., Kinetics of the homogeneous freezing of water. *Phys. Chem. Chem. Phys.* **12**, 10380–10387 (2010).
45. B. Riechers, F. Wittbracht, A. Hütten, T. Koop, The homogeneous ice nucleation rate of water droplets produced in a microfluidic device and the role of temperature uncertainty. *Phys. Chem. Chem. Phys.* **15**, 5873–5887 (2013).
46. C. A. Stan et al., A microfluidic apparatus for the study of ice nucleation in supercooled water drops. *Lab Chip* **9**, 2293–2305 (2009).
47. A. Haji-Akbari, R. S. DeFever, S. Sarupria, P. G. Debenedetti, Suppression of sub-surface freezing in free-standing thin films of a coarse-grained model of water. *Phys. Chem. Chem. Phys.* **16**, 25916–25927 (2014).
48. J. Russo, F. Romano, H. Tanaka, New metastable form of ice and its role in the homogeneous crystallization of water. *Nat. Mater.* **13**, 733–739 (2014).
49. H. Laksmono et al., Anomalous behavior of the homogeneous ice nucleation rate in "no-man's land". *J. Phys. Chem. Lett.* **6**, 2826–2832 (2015).
50. J. R. Espinosa, C. Vega, E. Sanz, Homogeneous ice nucleation rate in water droplets. *J. Phys. Chem. C* **122**, 22892–22896 (2018).
51. S. Hardy, A grain boundary groove measurement of the surface tension between ice and water. *Philos. Mag.* **35**, 471–484 (1977).
52. J. R. Espinosa, C. Vega, E. Sanz, Ice–water interfacial free energy for the TIP4P, TIP4P/2005, TIP4P/Ice, and mW models as obtained from the mold integration technique. *J. Phys. Chem. C* **120**, 8068–8075 (2016).
53. L. Ickes, A. Welti, C. Hoose, U. Lohmann, Classical nucleation theory of homogeneous freezing of water: Thermodynamic and kinetic parameters. *Phys. Chem. Chem. Phys.* **17**, 5514–5537 (2015).
54. D. Turnbull, Formation of crystal nuclei in liquid metals. *J. Appl. Phys.* **21**, 1022–1028 (1950).
55. B. B. Laird, The solid–liquid interfacial free energy of close-packed metals: Hard-spheres and the Turnbull coefficient. *J. Chem. Phys.* **115**, 2887–2888 (2001).
56. D. M. Murphy, T. Koop, Review of the vapour pressures of ice and supercooled water for atmospheric applications. *Q. J. Royal Meteorol. Soc.* **131**, 1539–1565 (2005).
57. B. Cheng, C. Dellago, M. Ceriotti, Theoretical prediction of the homogeneous ice nucleation rate: Disentangling thermodynamics and kinetics. *Phys. Chem. Chem. Phys.* **20**, 28732–28740 (2018).
58. K. Röttger, A. Endriss, J. Ihringer, S. Doyle, W. Kuhs, Lattice constants and thermal expansion of H2O and D2O ice Ih between 10 and 265 K. *Acta Crystallogr. B* **50**, 644–648 (1994).
59. A. J. Amaya et al., How cubic can ice be? *J. Phys. Chem. Lett.* **8**, 3216–3222 (2017).
60. T. L. Malkin, B. J. Murray, A. V. Brukhno, J. Anwar, C. G. Salzmann, Structure of ice crystallized from supercooled water. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1041–1045 (2012).
61. L. Del Rosso et al., Cubic ice Ic without stacking defects obtained from ice XVII. *Nat. Mater.* **19**, 663–668 (2020).
62. M. Nachbar, D. Duft, T. Leisner, The vapor pressure of liquid and solid water phases at conditions relevant to the atmosphere. *J. Chem. Phys.* **151**, 064504 (2019).
63. T. Hondoh, T. Itoh, S. Amakai, K. Goto, A. Higashi, Formation and annihilation of stacking faults in pure ice. *J. Chem. Phys.* **87**, 4040–4044 (1983).
64. A. Zaragoza et al., Competition between ices Ih and Ic in homogeneous water freezing. *J. Chem. Phys.* **143**, 134504 (2015).
65. D. Quigley, Communication: Thermodynamics of stacking disorder in ice nuclei. *J. Chem. Phys.* **141**, 121101 (2014).
66. S. Pronk, D. Frenkel, Can stacking faults in hard-sphere crystals anneal out spontaneously? *J. Chem. Phys.* **110**, 4589–4592 (1999).
67. V. Babin, C. Leforestier, F. Paesani, Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient. *J. Chem. Theory Comput.* **9**, 5395–5403 (2013).
68. C. Vega et al., Heat capacity of water: A signature of nuclear quantum effects. *J. Chem. Phys.* **132**, 046101 (2010).
69. S. Plimpton, Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
70. H. Wang, L. Zhang, J. Han, E. Weinan, DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput. Phys. Commun.* **228**, 178–184 (2018).
71. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).

72. M. Matsumoto, T. Yagasaki, H. Tanaka, GenIce: Hydrogen-disordered ice generator. *J. Comput. Chem.* **39**, 61–64 (2018).
73. W. Lechner, C. Dellago, Accurate determination of crystal structures based on averaged local bond order parameters. *J. Chem. Phys.* **129**, 114707 (2008).
74. V. Ramasubramani *et al.*, freud: A software suite for high throughput analysis of particle simulation data. *Comput. Phys. Commun.* **254**, 107275 (2020).
75. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
76. M. Bonomi *et al.*; PLUMED consortium, Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **16**, 670–673 (2019).
77. M. Invernizzi, M. Parrinello, Rethinking metadynamics: From bias potentials to probability distributions. *J. Phys. Chem. Lett.* **11**, 2731–2736 (2020).
78. P. M. Piaggi, M. Parrinello, Calculation of phase diagrams in the multithermal-multibaric ensemble. *J. Chem. Phys.* **150**, 244119 (2019).
79. P. Piaggi, J. Weis, A. Panagiotopoulos, P. Debenedetti, R. Car, Data from "Homogeneous ice nucleation in an ab initio machine learning model." DataSpace, Princeton University. https://dataspace.princeton.edu/handle/88435/dsp01x633f4197. Deposited 4 April 2022.
80. P. Piaggi, Data from "Homogeneous ice nucleation in an ab initio machine learning model of water." PLUMED-NEST, The public repository of the PLUMED consortium. https://www.plumed-nest.org/eggs/22/016/. Deposited 31 March 2022.