<table>
<thead>
<tr>
<th>Title</th>
<th>Homogeneous SPC/E water nucleation in large molecular dynamics simulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Angélil, Raymond; Diemand, Jürg; Tanaka, Kyoko K; Tanaka, Hidekazu</td>
</tr>
<tr>
<td>Citation</td>
<td>The Journal of Chemical Physics, 143: 064507</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2015-08-11</td>
</tr>
<tr>
<td>Doc URL</td>
<td><a href="http://hdl.handle.net/2115/60180">http://hdl.handle.net/2115/60180</a></td>
</tr>
<tr>
<td>Rights</td>
<td>Copyright  American Institute of Physics. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the American Institute of Physics. The following article appeared in Journal of Chemical Physics, vol.143, pp.064507 and may be found at <a href="http://dx.doi.org/10.1063/1.4928055">http://dx.doi.org/10.1063/1.4928055</a>.</td>
</tr>
<tr>
<td>Type</td>
<td>article</td>
</tr>
<tr>
<td>File Information</td>
<td>Angelil2015.pdf</td>
</tr>
</tbody>
</table>

*Note: The table contains the metadata of the article.*
Homogeneous SPC/E water nucleation in large molecular dynamics simulations

Raymond Angélil,1 Jürg Diemand,1 Kyoko K. Tanaka,2 and Hidekazu Tanaka2
1Institute for Computational Science, University of Zurich, 8057 Zurich, Switzerland
2Institute of Low Temperature Science, Hokkaido University, Sapporo 060-0819, Japan

(Received 5 June 2015; accepted 24 July 2015; published online 11 August 2015)

We perform direct large molecular dynamics simulations of homogeneous SPC/E water nucleation, using up to \(4 \cdot 10^9\) molecules. Our large system sizes allow us to measure extremely low and accurate nucleation rates, down to \(10^{19} \text{ cm}^{-3} \text{ s}^{-1}\), helping close the gap between experimentally measured rates \(10^{17} \text{ cm}^{-3} \text{ s}^{-1}\). We are also able to precisely measure size distributions, sticking efficiencies, and cluster internal densities. We introduce a new functional form to implement the Yasuoka-Matsumoto nucleation rate measurement technique (threshold method). Comparison to nucleation models shows that classical nucleation theory over-estimates nucleation rates by a few orders of magnitude. The semi-phenomenological nucleation model does better, under-predicting rates by at worst a factor of 24. Unlike what has been observed in Lennard-Jones simulations, post-critical clusters have temperatures consistent with the run average temperature. Also, we observe that post-critical clusters have densities very slightly higher, \(5\%\), than bulk liquid. We re-calibrate a Hale-type \(J\) vs. \(S\) scaling relation using both experimental and simulation data, finding remarkable consistency in over 30 orders of magnitude in the nucleation rate range and 180 K in the temperature range. © 2015 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4928055]

I. INTRODUCTION

The vapor-to-liquid transition of water is a common phenomenon in nature, relevant to many areas of technology and science. Attempts to predict the rate of homogeneous water nucleation often fail because of the lack of understanding of the properties of the tiny seeds of the intermediate phase, which are not necessarily large enough to have reached the bulk liquid properties. The relevant properties of the tiny clusters which affect predicted nucleation rates include surface tension, temperature, and density. Molecular dynamics (MD) simulation has proven to be a powerful test of thermodynamic analytical nucleation models, now that codes are efficient enough and computers, fast enough. Realistic, atmospheric nucleation rates are too low to be possible in direct computer simulations, due to the large number of molecules required. The lowest water nucleation rates performed in simulations and reported in the literature are \(10^{23-24} \text{ cm}^{-3} \text{ s}^{-1}\), usually beyond the spinodal limit. Laboratory water nucleation rates on the other hand are far lower—usually \(10^{10} \text{ cm}^{-3} \text{ s}^{-1}\), although a few experiments have managed to measure far higher rates \(10^{17} \text{ cm}^{-3} \text{ s}^{-1}\). Our simulations of homogeneous SPC/E water nucleation (Fig. 1), which we report on in this paper, manage to close the gap considerably, resolving nucleation rates down to \(10^{19} \text{ cm}^{-3} \text{ s}^{-1}\).

Nucleation models, which seek to provide explanations and predictions for nucleation rates, have a long history of falling short when compared to experimental results.14-14 For the case of water, rate predictions from the classical nucleation theory (CNT) disagree with experimental measurements by factors of \(10^3-10^4\).13,5-7,10-17 These models also have difficulty when predicting rates measured in numerical molecular dynamics nucleation simulation experiments. However, with molecular simulation, one can make measurements more detailed and accurate than what is possible in laboratory experiments. Size distributions, nucleation rates, cluster densities, temperatures, and even cluster pressures, shapes, angular momenta, and surface tension measurements are possible. Understanding the properties of the tiny yet complex, many-body clusters which form are vital for the development of a complete and successful thermodynamic description of the phase transformation.18 Simulations allow us to identify the...
shortcomings in the assumptions made by existing nucleation models and suggest ways they may be improved. Cluster properties are noisy, necessitating large systems with many millions of molecules. This demands costly compute power, and only recently have some of these direct measurement techniques become possible.19–21

Direct vapor-to-liquid molecular dynamics simulation for a Lennard-Jones fluid has become a popular exercise due to the computational accessibility of the short-range, single-site potential.20,22–26 Water is significantly more demanding. For the same system size, more complicated molecular interaction potentials like SPC/E27–29 and TIP4P23,30,31 necessitate a few orders of magnitude more computational power than a pure Lennard-Jones simulation. An exception is mW water, a comparatively simple monoatomic single-site water model.32–38 MD nucleation simulations of mW water have been carried out, yet only on small systems with relatively high nucleation rates.38,39 The monoatomic water model proposed by Zipoli et al.40 offers similar advantages. However, we found that short-range potentials require extremely long equilibration times to form the correct equilibrium abundance of small clusters (dimers, trimers, etc.) in a supersaturated vapor, because interactions are rare, especially the three body encounters required for dimer formation. This drawback makes it computationally expensive to simulate realistic, steady state vapor-to-liquid with such short-range potentials—despite their low cost per time step—and we do not use them in this work. Matsubara et al.28 simulate homogeneous vapor-to-liquid nucleation using the SPC/E water model and include a Lennard-Jones carrier gas, measuring rates down to 2.3 · 1025 cm−3 s−1. SPC/E simulations by Tanaka et al.1 manage to reach nucleation rates 3 · 1024 cm−3 s−1. Both efforts additionally measure critical cluster sizes, formation energies, size distributions, and sticking probabilities for systems in the $T = 300–390 \text{ K, providing ample opportunity for model comparison and development.}$

In this study, we continue in similar spirit, yet simulating the SPC/E water vapor-to-liquid phase change in even larger computational volumes using longer time integrations. This allows for the measurement of lower nucleation rates than previously possible by a few orders of magnitude, and for the first time, measurements of naturally formed SPC/E cluster density and temperature profiles. Our results provide opportunities for the verification and calibration of the standard assumptions which go into nucleation models, in a previously unexplored temperature and saturation regime.

II. SIMULATIONS

A. Simulation code, setup, and parameters

We use the molecular dynamics SPC/E27 water model. SPC/E is a rigid 3-site model, which registers Coulombic interactions, as well as polarization corrections to each site, and further adds a Lennard-Jones component to the oxygen atom potential.

The Large-scale Atomic/Molecular Massively Parallel Simulator (or LAMMPS) computer program,41 developed at the Sandia National Laboratories and distributed under the GPL license, was used to perform the SPC/E simulations. We have verified that our runs produce the same results as found in similar, yet smaller SPC/E numerical nucleation experiments,1 which used an independent molecular dynamics code. We cut the short range Lennard-Jones component to the force field off at 9.8 Å. For these forces, as well as the others, the interactions are computed directly on per atom. However, the SPC/E Coulombic interactions are long range, and so after 63 Å, the spectral solver takes over and the interactions computed in reciprocal space. LAMMPS uses a particle-particle/particle-mesh solver. The solver maps the atom charges onto a mesh, solves the Poisson equation (Maxwell’s equation for the electric field) by performing a 3D fast Fourier transform, then interpolates the electric fields on the mesh points back onto the atom positions.31–43 SPC/E molecule rigidity is ensured through the use of the SHAKE algorithm.44 We choose an integration time step of $\Delta t = 2$ fs, common for SPC/E water simulations.1 A typical simulation runs for 72 h on 1024 cores. Our largest simulation ran for 1000 h on 8192 cores on the Piz Daint supercomputer at Centro Svizzero di Calcolo Scientifico (CSCS), performing $3 \cdot 10^8$ integration time steps.

The simulation box has periodic boundary conditions. Initially, the molecules are given random non-overlapping positions and random velocities. This is done at 1000 K, after which the ensemble is cooled and the box size expanded until the simulation reaches the target temperature and pressure. The run continues in this state under NVT conditions, regulated by a Nose-Hoover thermostat45–47 with temperature damping timescales of 1000 fs.

At this stage the gas is allowed to equilibrate for a fixed amount of time—dependent on the run temperature (Refer to Table I for the chosen equilibration timescales $t_e$ at each temperature). During this phase, the subcritical cluster equilibrium distribution forms. Around this stage, we begin to make nucleation rate, size distribution, and cluster growth rate measurements. For most runs, the nucleation rate is low enough that unnatural effects from the interventions due to the thermostat are minimal. Our nucleation rates are low enough that the latent heat of transformation in the simulations is extremely small, resulting in only a faint influence from the thermostat. Our largest run sees a total energy increase of $\sim 0.1 \%$ over the steady-state phase, i.e., our simulations are very close to NVE (micro-canonical) ensembles.

<table>
<thead>
<tr>
<th>$T$ (K)</th>
<th>$P_v$ (dyn/cm²)</th>
<th>$\gamma$ (dyn/cm)</th>
<th>$\rho_b$ (g/cm³)</th>
<th>$r_0$ (10⁻⁸ cm)</th>
<th>$\eta$</th>
<th>$\xi$ (ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>8.9 · 10⁴</td>
<td>53.4</td>
<td>0.997</td>
<td>1.93</td>
<td>6.05</td>
<td>8.95</td>
</tr>
<tr>
<td>325</td>
<td>4.1 · 10⁴</td>
<td>50.1</td>
<td>0.982</td>
<td>1.94</td>
<td>5.29</td>
<td>7.47</td>
</tr>
<tr>
<td>350</td>
<td>1.49 · 10⁴</td>
<td>46.6</td>
<td>0.966</td>
<td>1.95</td>
<td>4.63</td>
<td>6.31</td>
</tr>
<tr>
<td>375</td>
<td>4.47 · 10⁴</td>
<td>42.9</td>
<td>0.946</td>
<td>1.97</td>
<td>4.02</td>
<td>5.38</td>
</tr>
</tbody>
</table>
The first few columns of Table II lists the runs which were carried out, their target temperatures, box sizes, number of molecules, and their run times.

### B. Simulation analysis

We use the simple Stillinger criterion (also known as the friends-of-friends method) to identify clusters. As the simulation runs, the cluster size distribution is regularly calculated and outputted, typically resulting in >1000 size distribution histograms per simulation. The linking length was set at 6 Å for all runs and was tested to yield stable size distributions under convergence tests. Furthermore, this choice yields a monomer-dimer number ratio consistent with what is expected from the second virial coefficient applied to the SPC/E interaction potential. The regularly outputted size distributions can then be converted into cluster threshold sizes, whose slopes in the steady-state regime are the nucleation rates. Refer to Section III for further details on the nucleation rate analysis, as well as the results. From the nucleation rate vs. supersaturation ratio landscape, we calculate the critical cluster sizes using the first nucleation theory. The size distributions also allow us to follow the growth rate of the largest clusters in each simulation, providing a measurement of the monomer-cluster interaction sticking efficiency (refer to Section VI).

The measurements of specific cluster properties and how they vary with cluster size is crucial to testing assumptions used in theoretical nucleation models. However, because of the noisy nature of many of these properties, one needs many millions of molecules per simulation in order to resolve interesting cluster properties. Due to this limitation, we perform cluster temperature and cluster density profile measurements only for our largest simulation, which contained \( \sim 4 \times 10^9 \) molecules. We perform this at the end of the simulation, well-within the steady-state nucleation regime. This calls for per-atom outputs of velocity and position information. Sections VIII and IX detail how the density profile and temperature measurements, respectively, are made and discuss the results.

### III. NUCLEATION RATES

We use a modified Yasuoka-Matsumoto method (threshold method) to measure nucleation rates. In the steady-state regime, the time rate of increase of the number of clusters above a certain size \( N \) is the nucleation rate. However, the simulations must equilibrate—form the sub-critical size distribution—and properly populate it before they reach the steady state nucleation regime. How long the simulations take to transition into the steady state regime is not known a priori. Nucleation rates estimated from the first nucleation event alone (e.g., mean first passage time or survival probability methods) can be orders of magnitude smaller than the true steady state nucleation rates. Thus, we use the following method: To the size-threshold curves, we fit the following function, which is able to capture the transition from the equilibration to the steady-state phase:

\[
N(> i) = J \cdot N(> i, t) + N(> i, 0),
\]

where

\[
N(> i, t) = \frac{\pi}{2} + \arctan \left( \frac{t - t_0}{t_r} \right) \cdot \frac{t - t_0}{\pi} \cdot V,
\]

where \( J \) is the nucleation rate, \( t_0 \) is the lag time, \( t_r \) is the relaxation timescale, and \( V \) is the volume of the simulation box. This function captures the system’s transition from the initial equilibration phase to the intermediate relaxation phase as the clusters begin to form, through to the steady-state regime. We
count clusters above a certain post-critical size \( N \), frequently throughout the simulation, and fit the count to this curve, allowing \( J, J_0 \), and \( i^* \) to vary (Fig. 1). Visual inspection of this approach is provided in the upper left panels of Figure 2 for run T375c.

Our nucleation rate measurements are listed in Table II. Figure 3 plots our simulations’ nucleation rates against supersaturation and includes comparison to earlier results,\(^1\) which used smaller simulations and were therefore restricted to lower nucleation rates. Estimates for the critical cluster sizes using the first nucleation theorem, via

\[
i^*_N = \left( \frac{\partial \ln J}{\partial \ln S} \right)_T - 1,
\]

are included as annotations. Our nucleation rate results can be split into two categories.

- **High temperature** (325 K, 350 K, 375 K): Runs at these temperatures have nucleation rates in the range \( \sim 10^{19–24} \text{ cm}^{-3} \text{s}^{-1} \). These runs have generally low errors on the nucleation rates. For the higher nucleation rates, there is an error on the supersaturation, as the pressure drops significantly due to the large number of clusters forming quickly.

- **Low temperature** (300 K): Here, we measure nucleation rates in the range \( \sim 10^{23–24} \text{ cm}^{-3} \text{s}^{-1} \). These runs suffer from extremely long equilibration periods, which continue while the initial large, stable clusters are already forming. In other words, the sub-critical distribution formation timescale \( t_r \) is longer than the nucleation timescale \( 1/(J \cdot V) \). This leads to large errors in both the nucleation rate measurements and the supersaturation measurements.

### IV. RATE COMPARISON WITH ANALYTICAL MODELS

Nucleation models endeavor to describe the phase change process in purely thermodynamic terms. The standard approach tries to find the balance between the Gibbs free energy gain and cost due to the creation of volume and surface. The CNT is the most basic of them all and forms the basis upon which
many appendages have since been added. In the CNT, the surface energy term in the Gibbs free energy is simply calculated using the planar surface tension, with no additional corrections. The CNT nucleation rate is

\[ J_{\text{CNT}} = \sqrt{\frac{32\pi\gamma}{9m}} r_0^3 \left( \frac{p_b}{k_BT} \right)^2 \exp \left[ \frac{256\pi^2 r_0^3}{27(k_BT)^3(\log S)} \right], \]  

where \( m \) is the molecular mass, \( \gamma \) is the planar surface tension at the run temperature, \( p_b \) is the monomer partial pressure in the simulation box (assuming an ideal gas) gas pressure, and \( S \) is the supersaturation,

\[ S = \frac{p_b}{p_c}, \]

where \( p_c \) is the equilibrium vapor pressure at the run temperature. \( r_0 \) is a characteristic molecular radius,

\[ r_0 = \left( \frac{3}{4\pi\rho_b} \right)^{1/3}, \]

where \( \rho_b \) is the bulk liquid density at the run temperature. Table I includes the thermodynamic variables for SPC/E water, which we use in our analysis and comparison to nucleation models. The CNT predictions for the nucleation rates of SPC/E water at our runs’ supersaturations are shown as solid curves in Figure 3. We find that the CNT predicts too-high nucleation rates by factors of \( 10^{1-2} \).

Various authors\(^{50,60-63}\) employ a 2-parameter, temperature dependent correction factor,

\[ J_{\text{corr}} = J_{\text{CNT}} \exp \left( \frac{A + B}{T} \right). \]  

The Manka et al.\(^6\) (see their Figure 4) laminar flow diffusion chamber experiments find that the parameter pair \( (A,B) = (-27.56,6500) \) K corresponds to a global fit of their results and previous experiments.\(^{37-44,17}\) With these parameters, the CNT rate prediction remains unchanged at a temperature of \(-B/A = 235.8 \) K, and they still increase with temperature (at a fixed \( S \)), but less strongly than in CNT; These corrected CNT predictions, when extrapolated to our supersaturations (dashed curves in Figure 3) under-predict our measurements by 1-3 orders of magnitude. Using our data at temperatures \( T = 325,350,375 \) K to determine the best-fit parameter pair, we find \( (A,B) = (-20.5,6100) \) K. With our parameter pair, the CNT rate prediction remains unchanged at a temperature of \(-B/A = 297.6 \) K, and they increase with temperature at a rate between CNT and the Manka et al. model. However, we note that the resulting curves (dotted lines in Figure 3) are not quite steep enough—casting doubt on whether a purely temperature-dependent correction is sufficient in this high supersaturation regime.

The Modified Classical Nucleation Theory (MCNT)\(^{23}\) implements a minor modification to the CNT, namely, it stipulates that the free energy of formation of a cluster of size one is zero. This results in a free energy shift for all cluster sizes. Like the CNT, the MCNT over-predicts the nucleation rates and here the differences are even slightly larger (factor of \( 10^{2-4} \)). Figure 4 shows the ratio between the MCNT model predictions (red markers) and the direct MD measurements. Refer to Table II for the MCNT model nucleation rate predictions.

\[ \frac{\ln S}{(T_c/T - 1)^{1.5}}. \]

The Semi-Phenomenological (SP) model\(^{50,60-63}\) attaches a \( -1/R \) (or \( -1/(T_0) \)) correction to the surface tension, where \( R \) is the cluster size under the assumption of sphericity. This radial dependence is functionally equivalent to that introduced by the Tolman length.\(^{11,64,65}\) although the motivation is different: the coefficient to this term is set by the second virial coefficient \( B_2 \) so that the dimer number density is correctly predicted. The nucleation rate predictions for the SP model relative to the measured values are plotted with red markers in Figure 4. The predictions at \( T = 300 \) K are somewhat accurate—within a factor of 5 of the measured values, although, as noted in Section III, the measurements at these temperatures carry significant uncertainty in the nucleation rate. At the higher temperatures, the SP model under-predicts the measured MD rates by factors of 4–80. Table II lists the SP model nucleation rate predictions.

V. NUCLEATION RATE SCALING

In this section, we examine the scaling of the nucleation rates. For the case of water, Hale\(^66\) (and similarly for Lennard-Jones in Hale)\(^67\) uses a scaling relation\(^{68}\) for experimentally measured nucleation rates over the range \( J = 10^{4-10} \) \( \text{cm}^{-3} \text{s}^{-1} \) of

\[ \frac{\ln S}{(T_c/T - 1)^{1.5}}. \]

Tanaka et al.\(^{25}\) showed that this scaling relation works well for large scale Lennard-Jones simulations and Argon laboratory experiments, albeit with an exponent of 1.3 instead of 1.5. We confirm that same scaling relation (8) applies well to our SPC/E water nucleation rate measurements. However, we find that the combined nucleation rates from both SPC/E
Nucleation rates as function of $\ln$ 

Upper panel: sticking probability measurements for our simulations

This article is copyrighted as indicated in the article. Reuse of AIP content is subject to the terms at: http://scitation.aip.org/termsconditions. Downloaded to IP: 133.87.217.28 On: Wed, 12 Aug 2015 01:54:31

FIG. 5. Nucleation rates as function of $\ln S/(T_c/T - 1)^{1.7}$ for MD simulations and experiments. Our simulations are filled circles and the previous simulations are “+” symbols. The solid curves show the SP model for various temperatures (210, 315, 350, and 380 K). For thermodynamic quantities such as the surface tension and the saturated vapor pressure, we use those of the SPC/E at 315, 350, and 380 K for the comparison with the MD results, while real water at 210 K (pink curve) for the comparison with the experiment.

Simulations and laboratory experiments with water are even better scaled by 

$$\frac{\ln S}{(T_c/T - 1)^{1.7}}.$$  \hspace{1cm} (9)

Figure 5 shows the nucleation rates as a function of (9). This empirical scaling relation seems to work well over a surprisingly wide nucleation rate range— from $J = 10^{-2}$ to $J = 10^{28}$ cm$^{-3}$ s$^{-1}$ for both MD simulations and experiments. The results from the MD simulations join smoothly with the experiments with the scaling by $\ln S/(T_c/T - 1)^{1.7}$ Figure 5 also shows, using solid curves, the nucleation rates predicted by the SP model for various temperatures. This scaling relation also works very well for the SP model.

VI. STICKING PROBABILITIES

The sticking probabilities $\alpha$ can be calculated from the rate at which large, stable clusters grow. For each run, we observe the size of the largest cluster and measure its growth rate $di/dt$ over the second half of the simulation. Early on in the simulations, before stable clusters have formed, the largest designation jumps between clusters. However, the first stable cluster to form is likely to remain the largest until the end of the simulation. The upper right panel of Figure 2 shows our cluster growth rate measurements for run T375c. We find the cluster size $i(t)$ to be strongly cubic within the steady-state regime. The cluster growth rate is therefore proportional to the surface area. This is consistent with what has been found in Lennard-Jones nucleation simulations.\cite{23,24} We may determine $\alpha$ \cite{24} from

$$\alpha = \frac{3}{4\pi r_0^2 n_0 n} \left(1 - \frac{1}{8}\right)^{1/3} \frac{di^{1/3}}{dt}.$$  \hspace{1cm} (10)

The supersaturation $S$ dependence includes the effect of the evaporation of molecules from the clusters into the gas. We list the measured sticking probability results in Table II. Sticking probability results for our low temperature $T = 300$ K runs are somewhat unreliable and can exceed unity due to the large number of dimers, trimers, and tetramers which also contribute to cluster growth. Eq. (10) considers the accretion and evaporation of monomers only. Our sticking probability measurements are consistent with those measured at slightly higher supersaturations in Tanaka et al.\cite{1} The upper panel of Figure 6 plots $\alpha$ against $S$. The sticking probability is a necessary prerequisite for performing the $\Delta G$ landscape reconstruction procedure for post-critical clusters (see Section VII).

While we are, due to computational constraints, unable to probe the low nucleation rates observed in laboratory experiments, it is possible to measure cluster growth rates under laboratory conditions. We have performed an additional simulation from the end state of T325c, in which we measured a sticking probability $\alpha = 0.26$. We target the temperature and saturation conditions found in Brus et al.\cite{14} Using a Nose-Hoover thermostat, we maintain the temperature and gently increase the box size until the supersaturation $S = 2.5$, after which we continue running for 60 ns. Under these low pressure conditions, no new clusters nucleate (Brus et al.\cite{14} report nucleation rates $\sim 10^3$ cm$^{-3}$ s$^{-1}$) due to our comparatively small and short-lived system. However, clusters which
had previously nucleated and then grown under the original conditions persist. Using the largest of these still-post-critical clusters, we measure a decreased growth rate: an \(i^{1/3}\) slope shallower by a factor of \(\sim 7\). Including this, and the reduced (by a factor \(\sim 3\)) monomer number density into Eq. (10) gives a sticking probability for these laboratory-like growth rates of \(\alpha = 0.15\). In nucleation models, the sticking efficiency is usually taken to be unity, entering linearly in the transition growth rate (typically denoted \(R^*\)), as a prefactor to the \(\Delta G_i\) exponent. We find that in the \(T = 325\) K and \(S = 2.5\) regime, the water monomer-cluster sticking efficiency is approximately one seventh of what is usually used in model predictions, implying an expected lowering of predicted nucleation rates by the same factor.

**VII. FREE ENERGY RECONSTRUCTION**

In this section, we evaluate the formation free energy of a cluster \(\Delta G_i(S)\) directly from our molecular dynamics simulations, even for post-critical cluster sizes. We obtain \(\Delta G_i(S)\) from the equilibrium size distribution of the cluster. The equilibrium size distribution can be obtained using the steady state size distribution, the accretion rate of molecule on a cluster, and the nucleation rate, all of which can be measured directly from the MD simulations. Refer to Tanaka et al.\(^{23}\) for a thorough explanation of the technique. The cluster size distributions are measured in the MD simulations and time-averaged over the steady state nucleation phase. In the accretion rate, we use the value of the sticking probability obtained from MD simulations. With the use of them, we reconstruct the full equilibrium size distribution (at all sizes \(i\), where we have good abundance estimates, including \(i > i^*\)) and then the entire free energy function \(\Delta G_i(S = 1)\). We can further derive \(\Delta G_i(S = 1)\), which is a surface term corresponding to the work required to form the vapor-liquid interface, by subtracting the volume term from \(\Delta G_i(S)\),

\[
\Delta G_i(S = 1) = \Delta G_i(S) + (i - 1) \ln S. \tag{11}
\]

Figure 7 shows \(\Delta G(S = 1)\) obtained from the MD results at 375 K and various supersaturations. Since \(\Delta G(S = 1)\) is supersaturation independent, the values from all runs should overlap at all sizes. However, our simulation data are only good enough for accurate abundance estimates below a certain cluster size, which depends on the run properties. The highest nucleation rates run of these (T375a) produced a large number of clusters over the entire plotted size range and allows the most reliable reconstruction of \(\Delta G_i(S)\). The results from the other runs are only accurate at smaller sizes, where they overlap with (T375a). Figure 7 also shows the surface energy \(\Delta G_i(S = 1)\) divided by that of the CNT, i.e., \(\Delta G_i(S = 1)/(\eta i^{2/3}/kT)\). In the figure, we also show the results of the SP model, given by

\[
\Delta G_i(S = 1)/(\eta i^{2/3}/kT) = 1 + (\xi/\eta)i^{-1/3} - (\xi/\eta)i^{-2/3}. \tag{12}
\]

The simulation results deviate from the SP model at 375 K. In Figure 7, we can fit the reconstructed \(\Delta G_i(S = 1)\) with

\[
\Delta G_i(S = 1)/(\eta i^{2/3}/kT) = 1 + Ai^{-1/3} - Ai^{-2/3}, \tag{13}
\]

using a fitting parameter \(A = 0.9\) for small clusters.

Figure 8 shows the ratios between the model and (13) \((A = 0.9, 1.0, 1.0,\) and \(1.5\) at 375, 350, 325, and 300 K, respectively) and the MD simulations for two cases: one in which

![Figure 7. Reconstructed Gibbs free energy curves shifted to \(S = 1\) (top panel). Runs at \(T = 375\) K at different supersaturations were used. To get to \(S = 1\), the CNT volume term was subtracted, and the resulting \(\Delta G_i(S = 1)\) can be interpreted as the surface term. The bottom panel shows the \(\Delta G_i(S = 1)\) divided by the surface term from CNT. The surface term from the SP model and a simple fitting function (Eq. (13)) are shown with dotted and solid lines.](image)

![Figure 8. Comparisons of nucleation rate from the MD simulations and several model predictions: the SP model (grey filled circles), our new surface term fit (Eq. (13)) with \(A = 1\) (filled circles) and using the \(\alpha\) values measured in the MD simulations (open circles).](image)
\( \alpha = 1 \) and the other in which \( \alpha \) is set to be value obtained directly from simulation. In Figure 8, the predictions from the SP model are also shown for comparison. We find the new model agrees with the simulations within one order of magnitude for all cases. At 375 K, this is no surprise, since these data were used to determine the parameters of our fitting function for surface term (13). The good agreement at the other temperatures is encouraging and might motivate using (13) also to predict nucleation rates at different temperatures and supersaturations.

VIII. CLUSTER DENSITIES

It has been shown\(^{69}\) that for spherical clusters, liquid-vapor interface densities are well-approximated by

\[
\rho(r) = \frac{1}{2} \left[ \rho_c + \rho_g - \left( \rho_c - \rho_g \right) \tanh \left( \frac{2r - R}{d} \right) \right], \tag{14}
\]

where \( \rho_c \) is the number density within the cluster, \( \rho_g \) is the gas number density, \( R \) is the interface position, and \( d \) is its width. In each cluster’s center-of-mass frame, we bin the spherical number density, using a bin size of 1.5 Å. The number density profiles for clusters of the same size are used to make ensemble averages, to which Equation (14) can be fit. This method of measuring internal cluster densities is robust only for clusters which are large enough to possess a constant density core. Clusters with \( i < 20 \)–30 are unlikely to have reached a shape well-describable by (14), and larger clusters’ density profiles on the other hand are well-suited to this functional form. Density profile measurements are noisy, and so particularly, large runs with many clusters in each size bin are necessary for the ensemble average to provide acceptable accuracy. For this reason, we perform the density profile measurements in our largest simulation, T325f, and we do so at the end of the run. Figure 9 plots \( \rho_c \) against \( R \) for clusters in T325f. We observe an over-density for clusters between 4–5.5 Å; however, the clusters approach the bulk liquid values as they grow, although there seems to be a weak overdensity indication of \( \sim 5\% \). This is in contrast to recent Lennard-Jones nucleation simulations\(^{20}\) which showed cluster densities significantly lower than the bulk liquid values. For post-critically sized clusters, this was attributable to the increased cluster temperatures, due to the residual latent heat which had not been efficiently redistributed back into the gas. We surmise that the good agreement our internal cluster densities have with the bulk liquid values to be due to the fact that they are in thermal equilibrium with the surrounding gas.

In the Lennard-Jones case,\(^{20}\) the lowered densities for clusters with \( i = i^* \) implied larger surface areas, and therefore larger-than-expected surface energies, resulting in an increased free energy cost to form a critical cluster, which lowered nucleation rates from model predictions. We suspect that nucleation rate predictions are more successful for SPC/E water than they are for Lennard-Jones because the assumption of small clusters possessing the bulk density for \( i = i^* \) is more realistic for the case of SPC/E water.

IX. TEMPERATURES

We define the temperature of an ensemble of atoms from their mean kinetic energy,

\[
kT = \frac{2}{3} \langle E_{\text{kinetic}} \rangle = \frac{1}{3N} \sum_{i=1}^{N} m v_i^2. \tag{15}
\]

Using full per-particle velocity information outputted at the end of the simulation, we are able to investigate the cluster size dependence of temperature. We find that sub-critical clusters are at the run average temperature, as observed in similar Lennard-Jones simulations.\(^{20}\) However, contrary to what has been observed in Lennard-Jones nucleation simulations, post-critical SPC/E water clusters possess temperatures consistent with the run average temperature. Figure 10 plots the ensemble average (at each cluster size \( i \)) of their temperatures against the density profile interface midpoint \( R \) (i.e., the cluster radius).

The latent heat from condensation has been efficiently dissipated back into the gas, leaving the post-critical clusters in thermodynamic equilibrium with their surroundings. This finding is consistent with the post-critical clusters density profile measurements, which finds their densities at the expected bulk density. Had the clusters significant latent heat retention, their densities would be correspondingly lower. We conjecture that the efficient kinetic energy exchange is effected.
by the long-range Coulombic interactions—even molecules deep within a cluster may exchange energy and angular momentum with members of the gas—resulting in kinetic energy equipartition on shorter timescales than cluster growth rates. A molecule impinging on a cluster imparts heat into the cluster-system, yet the heat does not linger. This may have implications for non-isothermal nucleation models, which include latent heat retention in the thermodynamic description of growing droplets.

X. CONCLUSIONS

We have performed molecular dynamics simulations of SPC/E water and significantly closed the nucleation-rate gap between simulation and experiment, measuring nucleation rates low as $\sim 10^{-19}$ cm$^{-3}$ s$^{-1}$. This is an hitherto unexplored saturation and temperature region for water nucleation experiments and water nucleation simulation. Nucleation rate results in this new regime will provide models with further testing comparison opportunities, to complement the already-existing lower nucleation rates from experiment, and higher nucleation rates from other simulations. We summarize our most significant contributions below.

- We introduce a new functional form, Eq. (1) in order to implement the Yasuoka-Matsumoto nucleation rate measurement. This modified version smoothly captures the system's transition between the lag phase, relaxation phase, and onto the steady-state regime.
- As expected, the CNT over-estimates nucleation rates by a few orders of magnitude. Empirical CNT correction factor $s$ when using the Manka et al. best-fit parameter values (calibrated in the low nucleation rate, low saturation regime) under-estimates our rates by a few orders of magnitude. When fitting their proposed correction function to our results, we find that the slopes are not steep enough. We conclude that this empirical and purely temperature-dependent correction factor to the CNT is not rich enough to reproduce the qualitative behavior we observe in our regime.
- The MCNT nucleation model continues to over-predict nucleation rates, by factors of up to $10^4$. The SP model on the other hand, does somewhat better, under-predicting rates at worst by a factor of 24. Despite these failings, we note that these model predictions are significantly more accurate than the corresponding predictions for the Lennard-Jones fluid vapor-to-liquid nucleation.
- Performing a cluster growth rate measurement simulation under laboratory conditions (those found in Brus et al. of $T = 325$ K and $S = 2.5$), we measure a sticking probability of $\alpha = 0.15$. This suggests that in this regime, nucleation rate predictions from models should be lowered by a factor of seven.
- We find the cluster size $i(t)$ to be strongly cubic within the steady-state regime. The cluster growth rate is therefore proportional to the surface area, a result new to water nucleation. This is consistent with what has been found in Lennard-Jones nucleation simulations.
- Unlike Lennard-Jones nucleation simulations, we find that post-critical clusters have temperatures consistent with the simulation average temperature: Growing clusters are in thermal equilibrium with their surroundings. Latent heat is not retained as the clusters grow, it is efficiently dissipated back into the gas. We suspect this efficiency is due to the long-range Coulombic interactions, not present in the Lennard-Jones case. This could have an impact on nucleation models which include non-isothermal processes into the thermophysical modeling of cluster properties.
- Post-critical clusters have densities consistent with what is expected from the bulk liquid. There is a possible indication of an over-density for clusters around the critical size. This would imply a lower-than-expected surface area, which lowers the total surface energy, decreasing the free energy cost to form a critically sized cluster and would result in higher-than-expected nucleation rates.
- The scaling relation $\ln(S/(T/T_c - 1)^{1/7}$ is remarkably successful in reducing the 3-parameter $T$ vs. $S$ vs. $J$ surface into a 2-parameter curve. It accurately links nucleation rates from simulation and experiment from over 30 orders of magnitude in the nucleation rate range, and a temperature range of 180 K.

ACKNOWLEDGMENTS

Computations were performed on Piz Daint at CSCS and on the zBox4 at UZH. We thank the entire CSCS team. J.D. and R.A. are supported by the Swiss National Science Foundation. K.C.T. is supported by JSPS KAKENHI Grant Nos. 2540054, 15K05015, and 15H05731. We thank the referees for their suggestions and comments.

D. Kashchiev, Nucleation (Butterworth-Heinemann, 2000).