Monte Carlo simulations of critical cluster sizes and nucleation rates of water

Joonas Merikanto,[a] Hanna Vehkamäki, and Evgeni Zapadinsky
University of Helsinki, Department of Physical Sciences, P.O. Box 64, FIN-00014 University of Helsinki, Finland

(Received 27 October 2003; accepted 22 March 2004)

We have calculated the critical cluster sizes and homogeneous nucleation rates of water at temperatures and vapor densities corresponding to experiments by Wölk and Strey [J. Phys. Chem B **105**, 11683 (2001)]. The calculations have been done with an expanded version of a Monte Carlo method originally developed by Vehkamäki and Ford [J. Chem. Phys. **112**, 4193 (2000)]. Their method calculates the statistical growth and decay probabilities of molecular clusters. We have derived a connection between these probabilities and kinetic condensation and evaporation rates, and introduce a new way for the calculation of the work of formation of clusters. Three different interaction potential models of water have been used in the simulations. These include the unpolarizable SPC/E [J. Phys. Chem. **91**, 6269 (1987)] and TIP4P [J. Chem. Phys. **79**, 926 (1983)] models and a polarizable model by Guillot and Guissani [J. Chem. Phys. **114**, 6720 (2001)]. We show that TIP4P produces critical cluster sizes and a temperature and vapor density dependence for the nucleation rate that agree well with the experimental data, although the magnitude of nucleation rate is constantly overestimated by a factor of 2×10^4. Guissani and Guillot's model is somewhat less successful, but both the TIP4P and Guillot and Guissani models are able to reproduce a much better experimental temperature dependency of the nucleation rate than the classical nucleation theory. Using SPC/E results in dramatically too small critical clusters and high nucleation rates. The water models give different average binding energies for clusters. We show that stronger binding between cluster molecules suppresses the decay probability of a cluster, while the growth probability is not affected. This explains the differences in results from different water models.
 [DOI: 10.1063/1.1740754]

I. INTRODUCTION

Vapor–liquid nucleation processes play an important role in the formation of atmospheric aerosols. Homogeneous nucleation in the atmosphere, where low vapor pressure species form new aerosol particles, is a multicomponent process which is considered to occur mainly via the ternary sulphuric acid–water–ammonia route and at low temperatures, possibly via binary sulphuric acid–water route.[1] Also heterogeneous nucleation of water and organic vapors on the surface of nonsoluble aerosol particles has lately been under intense research, although currently it seems that organic vapors mainly participate in the growth of newly formed particles.[2]

A good description of complex atmospheric nucleation processes requires a thorough knowledge of the simplest form of nucleation such as unary homogeneous nucleation. However, there are still severe discrepancies between theoretical predictions and laboratory measurements of unary homogeneous nucleation rates. These problems are considered in the present study.

Nucleation occurs through the birth of small molecular clusters that form in sequential molecular collisions. Up to a certain size known as the critical size, the clusters are unstable and tend to evaporate. The instability makes the equilibrium cluster distribution fall steeply with the increasing size. As the density of the clusters in any real physical situation is low, we can normally consider that acquisitions of free monomers dominate in the condensation process. Evaporation is also believed to occur mainly through the escape of single monomers. The cluster formation mechanism is then described as a chain of reactions

$$(1)+(N-1) \underset{\alpha_N}{\overset{\beta_{N-1}}{\rightleftharpoons}} (N), \tag{1}$$

where the condensation rate constant β_{N-1} is smaller than the evaporation rate constant α_N when N, the number of molecules in the cluster, is small. The critical cluster N^* is a cluster for which $\beta_{N^*-1}=\alpha_{N^*}$. Only few clusters manage to reach the critical size and are able to grow indefinitely. The rate at which the critical clusters appear in a unit volume of vapor is called the nucleation rate.

To date the most widely used theory describing the nucleation phenomenon is called the ''classical nucleation theory'' (CNT). It was developed by Volmer and Weber,[3] Becker and Döring,[4] and Zeldovich.[5] It is based on approximate analytical solutions of kinetic equations describing the growth and decay of clusters in a metastable state. Importantly, it relies on a number of approximations, most significantly on the ''capillarity approximation'' which means the use of bulk properties of liquid to describe clusters having only a small number of molecules. Although in many cases

[a]Electronic mail: joonas.merikanto@helsinki.fi

CNT works reasonably well, it often fails to predict the right temperature dependence for the experimental nucleation rate. For many substances such as n-nonane,[6] ethanol,[7] methanol,[8] i-propanol,[9] and water[10] the theoretical homogeneous nucleation rates are too low at low temperatures and too high at high temperatures. To improve the situation, extensions to classical nucleation theory have been introduced by Lothe and Pound,[11] Oxtoby and Evans,[12] and Dillmann and Meier,[13] among others, but nucleation phenomenon still remains poorly understood.

There is a growing interest to study nucleation on a molecular level. As there are no experiments that could directly observe the molecular processes and the full mathematical solutions are too complicated, this has been done with the aid of molecular computer simulations. Since the pioneering Monte Carlo (MC) simulation by Lee, Barker, and Abraham[14] on argon clusters, nucleation has been extensively studied by molecular dynamics and Monte Carlo simulations.[15–22] As the validity of interaction potentials for small clusters with complex molecules is often questionable, argon is still frequently used as a model substance in nucleation simulations. However, nucleation experiments are difficult to carry out with argon, and a comparison between experiments and simulations has only been done by Garcia and Torroja.[15]

Aside from argon, water presents another popular choice for nucleation simulations. Experimental nucleation rates are available for water, but the only comparison between simulations and experiments, so far, has been done by Hale and DiMattio,[20] who achieved the experimental temperature dependency of the nucleation rate with a scaled expression for the Helmholtz free energy differences deduced from a simulation with the TIP4P[23] water model. In other recent Monte Carlo simulations, Kusaka *et al.*[16] have used the SPC/E[24] water model for the evaluation of equilibrium distribution of water clusters and Gao *et al.*[25] applied SPC/E to study the effect of an electric field on homogeneous nucleation. Yasuoka and Matsumoto[26] have carried out a direct molecular-dynamics simulation of the nucleation of water with TIP4P.

In our present work we apply the MC technique developed by Vehkamäki and Ford[22,27] to calculate the critical cluster sizes and nucleation rates of water at the temperatures and vapor densities corresponding to the experiments by Wölk and Strey.[10] We have studied three different water models, namely two rigid and nonpolarizable models SPC/E and TIP4P and a rigid and polarizable model developed by Guillot and Guissani.[28] Our MC technique has been previously used for cases of new phase nucleation in the Ising model of interacting spins[27] and Lennard-Jones atoms.[22] With this method, we allow a fixed number of water molecules, forming a single cluster, to evolve in a canonical MC simulation. We deduce the stability of the cluster by calculating the average grand canonical growth and decay probabilities from attempted creations and annihilations of single molecules. The critical cluster is a cluster whose growth and decay probabilities are equal. Here, we also introduce a new method to calculate of the work of formation of the critical cluster directly from the growth and decay probabilities of clusters from monomers upward. We use the work of formation with the classical nucleation theory to predict the nucleation rate.

In the next section we review the theoretical background to the simulation technique and introduce a new way of calculating the work of cluster formation. The computational details are described in Sec. III. Section IV gives a sensitivity analysis for the calculation of the work of formation with respect to the applied nearest neighbor distance in the definition of the cluster. The Sec. V describes the applicability of three different water models (TIP4P, SPC/E, and an interaction potential recently developed by Guillot and Guissani) to our type of nucleation simulation and presents the calculated critical cluster sizes and nucleation rates. Finally, the conclusions are given in Sec. VI.

II. THEORETICAL BACKGROUND TO THE SIMULATION TECHNIQUE

The probability density for a system to be in a state with N indistinguishable particles having coordinates $\{\mathbf{r}_i\}$ in a volume V is given by[29]

$$P(N,\{\mathbf{r}_i\}) = \frac{1}{\Xi} \frac{\Lambda^{-3N} \exp(\mu N/(kT)) \xi_{in}^N}{N!} \exp\left[\frac{-U_N(\{\mathbf{r}_i\})}{kT}\right], \tag{2}$$

where $\mathbf{r}_i$ is the position of particle i and $U_N(\{\mathbf{r}_i\})$ is the interaction energy of the N particle system, which depends on the configuration $\{\mathbf{r}_i\}$. The curly brackets represent the set of N particle positions, ξ_{in} is the internal partition function of one molecule, T is the temperature, k is the Boltzmann constant, μ is the chemical potential and $\Lambda = \sqrt{h^2/(2\pi mkT)}$ is the thermal de Broglie wavelength of the particles. Here m is the mass of the particle and h is Planck's constant. The classical grand canonical partition function Ξ of a system of N indistinguishable molecules is given by

$$\Xi = \sum_{N=0}^{\infty} \left\{ \frac{\Lambda^{-3N} \exp(\mu N/(kT)) \xi_{in}^N}{N!} \times \int_V \prod_{i=1}^{N} d\mathbf{r}_i \exp\left[\frac{-U_N(\{\mathbf{r}_i\})}{kT}\right] \right\}. \tag{3}$$

It is convenient to combine de Broglie wavelength, chemical potential and internal partition function in one parameter

$$\gamma = \Lambda^{-3} \exp(\mu/(kT)) \xi_{in}. \tag{4}$$

For the simulation purposes the upper limit for the number particles is usually set to $N_{\max}$ and $N_{\max} - N$ noninteracting distinguishable ghost particles are inserted at random positions.[30] Then, the grand canonical partition function is

$$\Xi = \sum_{N=0}^{N_{\max}} \int_V \prod_{i=1}^{N_{\max}} d\mathbf{r}_i \frac{\gamma^N}{V^{N_{\max}-N} N!} \exp\left[\frac{-U_N(\{\mathbf{r}_i\})}{kT}\right]. \tag{5}$$

The probability density for the system to be in a state with N particles having coordinates $\{\mathbf{r}_i\}$, and $N_{\max} - N$ noninteracting particles at arbitrary positions, can then be identified as

$$P(N,\{\mathbf{r}_i\}) = \frac{1}{\Xi}\frac{\gamma^N}{V^{N_{\max}-N}N!}\exp\left[\frac{-U_N(\{\mathbf{r}_i\})}{kT}\right]. \quad (6)$$

In the algorithm for grand canonical simulation presented by Yao *et al.*,[30] along with nontrivial creation and annihilation of attempts, the conventional Metropolis moves[31] are also present. In what follows the notation $\{\mathbf{r}_i\}\oplus\mathbf{r}_k$ stands for a cluster configuration $\{\mathbf{r}_i\}$ where a particle is added to position $\mathbf{r}_k$, and the notation $\{\mathbf{r}_i\}\ominus\mathbf{r}_j$ stands for a cluster configuration $\{\mathbf{r}_i\}$ where a particle from position $\mathbf{r}_j$ is taken away. $\{\mathbf{r}_i\}$, $\mathbf{r}_j\rightarrow\mathbf{r}_j'$ denotes configuration $\{\mathbf{r}_i\}$ where a particle from position $\mathbf{r}_j$ moves to position $\mathbf{r}_j'$.

The probability that one particle is created at position $\mathbf{r}_k$ is given by $\min[1,C(N,\{\mathbf{r}_i\}\oplus\mathbf{r}_k)]$,[30,32] where

$$C(N,\{\mathbf{r}_i\}\oplus\mathbf{r}_k) = \frac{P(N+1,\{\mathbf{r}_i\}\oplus\mathbf{r}_k)}{P(N,\{\mathbf{r}_i\})} = \frac{\gamma V}{N+1}\exp\left\{\frac{-[U_{N+1}(\{\mathbf{r}_i\}\oplus\mathbf{r}_k)-U_N(\{\mathbf{r}_i\})]}{kT}\right\}, \quad (7)$$

and the probability that a particle at position $\mathbf{r}_j$ is annihilated is $\min[1,A(N,\{\mathbf{r}_i\}\ominus\mathbf{r}_j)]$, where $A(N,\{\mathbf{r}_i\}\ominus\mathbf{r}_j)$ is given by

$$A(N,\{\mathbf{r}_i\}\ominus\mathbf{r}_j) = \frac{P(N-1,\{\mathbf{r}_i\}\ominus\mathbf{r}_j)}{P(N,\{\mathbf{r}_i\})} = \frac{N}{\gamma V}\exp\left\{\frac{-[U_{N-1}(\{\mathbf{r}_i\}\ominus\mathbf{r}_j)-U_N(\{\mathbf{r}_i\})]}{kT}\right\}. \quad (8)$$

The probability that a particle at position $\mathbf{r}_j$ moves to position $\mathbf{r}_j'$ is $\min[1,M(N,\{\mathbf{r}_i\},\mathbf{r}_j\rightarrow\mathbf{r}_j')]$, where

$$M(N,\{\mathbf{r}_i\},\mathbf{r}_j\rightarrow\mathbf{r}_j') = \exp\left[-\frac{\Delta U_N}{kT}\right], \quad (9)$$

and ΔU_N is the energy difference between new and old configurations. In order to produce a series of configurations representative of the grand canonical ensemble, the relative probabilities of attempted creation, α_C, and destruction, α_D, have to be equal. The probability of a translational move α_T is independent of α_C and α_D.

Starting from these formulas Vehkamäki and Ford[22,27] introduced growth and decay rates for the given configuration of the cluster. For configuration $\{\mathbf{r}_i\}$ the decay rate, that is, the total probability for any molecule to be annihilated in a Monte Carlo step, is given by

$$D_N(\{\mathbf{r}_i\}) = \frac{\alpha_D}{N}\sum_{j=1}^{N}\delta_{\text{clu}}\min[1,A(N,\{\mathbf{r}_i\}\ominus\mathbf{r}_j)], \quad (10)$$

where δ_{clu} sets the probability to zero if the annihilation would result in splitting the cluster, and $\delta_{\text{clu}}=1$ when the annihilation move satisfies the cluster criterion. The growth rate for a configuration $\{\mathbf{r}_i\}$ reads

$$G_N(\{\mathbf{r}_i\}) = \frac{\alpha_C}{N_{\max}-N}\sum_{k=1}^{N_{\max}-N}\delta_{\text{clu}}\min[1,C(N,\{\mathbf{r}_i\}\oplus\mathbf{r}_k)], \quad (11)$$

where δ_{clu} is zero if the molecule created would not be part of the cluster according to the cluster definition, and $\delta_{\text{clu}}=1$ when the cluster criterion is satisfied by the creation move. Note that the creation probability depends on both the cluster configuration $\{\mathbf{r}_i\}$ and the configuration of the noninteracting particles $\{\mathbf{r}_k\}$.

For the canonical ensemble containing an N-molecule cluster and $N_{\max}-N$ noninteracting distinguishable molecules the average value of arbitrary variable Y is given by

$$\langle Y\rangle = \frac{1}{Q_N V^{N_{\max}-N}}\int_V\prod_{i=1}^{N_{\max}}d\mathbf{r}_i Y\delta_{\text{clu}}\exp\left[\frac{-U_N(\{\mathbf{r}_i\})}{kT}\right], \quad (12)$$

where Q_N is the configuration integral of the N-molecule cluster defined as

$$Q_N = \int_V\prod_{i=1}^{N}d\mathbf{r}_i\delta_{\text{clu}}\exp\left[\frac{-U_N(\{\mathbf{r}_i\})}{kT}\right]. \quad (13)$$

Then, the canonical partition function of the N-molecule cluster is

$$Z_N = \frac{\Lambda^{-3N}\xi_{in}^N Q_N}{N!} = \frac{\gamma^N\exp[-N\mu/(kT)]Q_N}{N!}. \quad (14)$$

Vehkamäki and Ford[22,27] assumed that when canonical ensemble averages for decay and growth rates

$$\overline{D}_N \equiv \langle D_N\rangle \quad (15)$$

and

$$\overline{G}_N \equiv \langle G_N\rangle, \quad (16)$$

are equal the cluster is said to be critical. The critical cluster condition can formally be written as

$$\overline{G}_{N^*} = \overline{D}_{N^*}, \quad (17)$$

where N^* is the number of molecules in the critical cluster. We show that this assumption corresponds to defining the critical cluster as the location of the free energy maximum.

For the grand canonical Metropolis scheme, assuming $\alpha_C=\alpha_D$, the following detailed balance equation is valid:[33]

$$P(N,\{\mathbf{r}_i\})\delta_{\text{clu}}\min[1,C(N,\{\mathbf{r}_i\}\oplus\mathbf{r}_k)] = P(N+1,\{\mathbf{r}_i,\mathbf{r}_k\})\delta_{\text{clu}}\times\min[1,A(N+1,\{\mathbf{r}_i,\mathbf{r}_k\}\ominus\mathbf{r}_k)], \quad (18)$$

where creation and annihilation probabilities $C(N,\{\mathbf{r}_i,\mathbf{r}_k\})$ and $A(N+1,\{\mathbf{r}_i,\mathbf{r}_k\}\ominus\mathbf{r}_k)$ are given by Eqs. (7) and (8), respectively, and the probability density for a configuration $\{\mathbf{r}_i\}$ with N particles $P(N,\{\mathbf{r}_i\})$ is given by Eq. (6). Let us multiply both sides of the detailed balance equation by

$$\prod_{i=1}^{N_{\max}}dr_i, \quad (19)$$

and integrate over all coordinates. All configurations, where cluster particles are at the same positions and the ghost molecules only exchange their positions between each other, equally contribute to the integral. It allows us to group configurations such way that we can insert Eqs. (10) and (11) into the integration in both sides obtaining

$$\frac{1}{\alpha_c}\int_V \prod_{i=1}^{N_{\max}} dr_i P(N,\{\mathbf{r}_i\})G_N(\{\mathbf{r}_i\}) = \frac{1}{\alpha_d}\int_V \prod_{i=1}^{N_{\max}} dr_i P(N+1,\{\mathbf{r}_i\})D_{N+1}(\{\mathbf{r}_i\}). \quad (20)$$

It is more visual to illustrate this procedure representing integration as a summation. Let us assume that the system space is divided to $N_{\max}$ pointlike cells $\mathbf{c}_j$, $j=1,...,N_{\max}$ each of which can contain only one (real or ghost) particle. (In principle, $N_{\max}$ can be less than number of cells; we keep them to be equal in order to make illustration simpler). This is done in the spirit of counting density of states, and the space is considered fully packed when each of the cells contain one particle. We distribute N real particles and $N_{\max}-N$ particles to these cells, and sum equation (18) over all possible configurations. In the summation each $\mathbf{r}_i$ (real), $\mathbf{r}_k$ (ghost on the left-hand side and real on the right-hand side) and $\mathbf{r}_k'$ (ghost) goes through all the cells, but two particles are not allowed in the same cell.

$$\sum_{\mathbf{r}_i=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\sum_{\mathbf{r}_k=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\sum_{\mathbf{r}_{k'}=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\{P(N,\{\mathbf{r}_i\})\delta_{\text{clu}}\min[1,C(N,\{\mathbf{r}_i\}\oplus\mathbf{r}_k)]\} = \sum_{\mathbf{r}_i=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\sum_{\mathbf{r}_k=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\sum_{\mathbf{r}_{k'}=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\{P(N+1,\{\mathbf{r}_i\}) \times \delta_{\text{clu}}\min[1,A(N+1,\{\mathbf{r}_i,\mathbf{r}_k\}\ominus\mathbf{r}_k)]\}. \quad (21)$$

We first sum left-hand side over ghost particle positions. Performing the sum over the positions of all ghost particles $k'\neq k$ just leads to a factor of $(N_{\max}-N-1)!$, since the summand is independent of where the $(N_{\max}-N-1)$ ghost particles $k'\neq k$ are. The sum over the position of ghost particle k gives $(N_{\max}-N)/\alpha_C\cdot G_N(\{\mathbf{r}_i\})$ according to Eq. (11). On the right-hand side summing over ghost particles k' gives also a factor of $(N_{\max}-N-1)!$, summing over now a *real* particle k gives $(N+1)/\alpha_D\cdot D_{N+1}(\{\mathbf{r}_i\}\ominus\mathbf{r}_k)$ using Eq. (10). Now we are left with sums only over positions of real particles i and the detailed balance equations turns into

$$\sum_{\mathbf{r}_i=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\left\{\frac{(N_{\max}-N)!}{\alpha_C}P(N,\{\mathbf{r}_i\})G_N(\{\mathbf{r}_i\})\right\} = \sum_{\mathbf{r}_i=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\left\{\frac{(N_{\max}-N-1)!(N+1)}{\alpha_D}P(N+1,\{\mathbf{r}_i\}) \times D_{N+1}(\{\mathbf{r}_i\})\right\}. \quad (22)$$

If we want to return to sums over positions of all $N_{\max}$ particles we get

$$\sum_{\mathbf{r}_i=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\sum_{\mathbf{r}_k=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\sum_{\mathbf{r}_{k'}=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\left\{\frac{1}{\alpha_C}P(N,\{\mathbf{r}_i\})G_N(\{\mathbf{r}_i\})\right\} = \sum_{\mathbf{r}_i=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\sum_{\mathbf{r}_k=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\sum_{\mathbf{r}_{k'}=\mathbf{c}(1)}^{\mathbf{c}(N_{\max})}\left\{\frac{1}{\alpha_D}P(N+1,\{\mathbf{r}_i\}) \times D_{N+1}(\{\mathbf{r}_i\})\right\}, \quad (23)$$

where the factorials and $(N+1)$ have disappeared since the summations over $\mathbf{r}_{k'}$ and $\mathbf{r}_k$ correspond to $(N_{\max}-N)!$ identical terms on the left-hand side and $(N_{\max}-N-1)!(N+1)$ on the right-hand side. Note that $G_N(\{\mathbf{r}_i\})$ is in this context uniquely defined for a certain cluster configuration $\{\mathbf{r}_i\}$, because we have chosen $N_{\max}$ and the cells so that the ghosts fill the space entirely, and the configuration of the ghosts is unique for a given cluster configuration. The last equation is a discrete analog of continuous Eq. (20).

Further transformation of Eq. (20) is straightforward when $\alpha_C=\alpha_D$. We multiply left-hand side by the ratio of configuration integrals Q_N/Q_N and the right-hand side similarly by $(Q_{N+1})/(Q_{N+1})$, then taking into account the explicit form of $P(N,\{\mathbf{r}_i\})$ [Eq. (6)] and definitions (4), (12)–(16), we observe that the integration in Eq. (20) actually represents canonical averaging, producing the following relation:

$$\frac{Z_{N+1}}{Z_N}\exp(\mu/kT)=\frac{\overline{G}_N}{D_{N+1}}. \quad (24)$$

At this point we see that the method originally proposed by Vehkamäki and Ford[22,27] for evaluating the critical cluster in nucleation theory can also be used for computing free energy difference between two clusters of adjacent sizes. Using the conventional definition for Helmholtz free energy F_N

$$F_N=-kT\ln Z_N, \quad (25)$$

we can rewrite Eq. (24) as

$$F_{N+1}-F_N=\mu-kT\ln\frac{\overline{G}_N}{D_{N+1}}. \quad (26)$$

The majority of homogeneous nucleation theories consider imperfect gas as a mixture of ideal gases, each gas containing clusters of a certain size. Thus, the equilibrium cluster distribution is given by the mass action law[34,35]

$$\frac{\mathcal{N}_N}{Z_N}=\left(\frac{\mathcal{N}_1}{Z_1}\right)^N, \quad (27)$$

where $\mathcal{N}_N$ is the number of N-molecule clusters. According to Eq. (27) clusters are in equilibrium with monomers. The same is the case for the grand canonical Monte Carlo method. The simulated cluster is in equilibrium with ideal gas molecules. Hence, the input chemical potential is the chemical potential of the ideal gas and corresponds to the chemical potential of the monomer gas, which in turn is a component of the imperfect gas. For an ideal gas the chemical potential is given by[29]

$$\mu = kT \ln\left(\frac{\Lambda^3 \mathcal{N}_1}{\xi_{in} V}\right) = kT \ln\left(\frac{\Lambda^3}{\xi_{in}} n_v\right), \tag{28}$$

where $n_v = \mathcal{N}_1 / V$ is the monomer number density. Using Eqs. (14), (25), (28), and the Stirling approximation $\ln \mathcal{N}_1! \simeq \mathcal{N}_1 \ln \mathcal{N}_1 - \mathcal{N}_1$ Eq. (27) can be rewritten as

$$\mathcal{N}_N = \exp\left(-\frac{\Delta F_N}{kT}\right), \tag{29}$$

where $\Delta F_N = F_N - N_\mu$. Rewriting Eq. (26) as

$$\Delta F_{N+1} = \Delta F_N - kT \ln \frac{\overline{G}_N}{D_{N+1}}, \tag{30}$$

we can see that assumption (17) of Vehkamäki and Ford[22,27] correspond to critical size, being defined as the location of the energy maximum with accuracy ± 1 molecule. Equations (26) and (30) also describe a new method for calculating the cluster Helmholtz free energy. Indeed, starting from F_1 we can subsequently calculate the Helmholtz free energies of any size clusters.

We rewrite Eq. (29) as

$$\mathcal{N}_N = \mathcal{N}_1 \exp\left(-\frac{\Delta W_N}{kT}\right), \tag{31}$$

where ΔW_N is defined as

$$\Delta W_N = \Delta F_N + kT \ln(\mathcal{N}_1). \tag{32}$$

The value ΔW_N is usually called the reversible work of formation of an N-molecule cluster.[34] Using Eqs. (30) and (32) we can express ΔW_N as

$$\Delta W_N = \Delta F_1 + kT \ln(\mathcal{N}_1) - kT \sum_{j=2}^{N} \ln \frac{\overline{G}_{j-1}}{D_j}. \tag{33}$$

Thus by using Eqs. (14) and (25) we obtain $\Delta F_1 = -kT \ln(\mathcal{N}_1)$ reducing Eq. (33) to

$$\Delta W_N = -kT \sum_{j=2}^{N} \ln \frac{\overline{G}_{j-1}}{D_j}, \tag{34}$$

and the equilibrium cluster distribution (31) can be expressed as

$$\mathcal{N}_N = \mathcal{N}_1 \prod_{j=2}^{N} \frac{\overline{G}_{j-1}}{D_j}. \tag{35}$$

The last equation looks very natural if considering grand canonical growth and decay as a type of kinetic process. In the kinetic approach the detailed balance condition leads to equilibrium cluster distribution that is given by[4,36]

$$\mathcal{N}_N = \mathcal{N}_1 \prod_{j=2}^{N} \frac{\beta_{j-1}}{\alpha_j}, \tag{36}$$

where β_j and α_j are the condensation and evaporation rate constants appearing in Eq. (1). The fraction $\overline{G}_{j-1}/D_j$ can then be identified as β_{j-1}/α_j.

In this section we have justified the method proposed by Vehkamäki and Ford[37] for a quick estimation of the size of critical cluster. Based on their idea we have developed a tool for the calculation of Helmholtz free energy of clusters, allowing one to estimate the work of formation of clusters or the equilibrium cluster distribution. Although our method uses canonical averaging, it can be considered as an extreme case of the grand canonical Monte Carlo method developed by Kusaka *et al.*[16] for the estimation of equilibrium cluster distribution. They proposed to simulate the cluster distribution inside certain intervals of the cluster sizes, say between $N_{\min}$ and $N_{\max}$. The narrower this interval is, the better the accuracy that can be reached. Thus, Oh and Zeng[38] have used $N_{\max} - N_{\min} = 3$. In our case $N_{\max} - N_{\min}$ is equal to 1, which indicates the best possible accuracy for this type of simulation.

There are different Monte Carlo approaches to simulating vapor nucleation. They can be divided in two types. The first one is direct simulation of vapor to observe clustering.[17,21,39] The second type, similar to the method presented here, is the simulation of an isolated cluster to calculate the cluster free energy.[14,16,18] At the first sight the direct simulation seems to be more rigorous than the simulation of an isolated cluster. However, all cited methods of the first type assume validity of Eq. (31). From statistical mechanics we know that if equilibrium cluster distribution exists, then the cluster concentration can be calculated using the cluster partition function [Eq. (27)]. Then, the only more general feature of direct vapor simulation is taking the cluster–monomer and cluster–cluster interactions into account. As was shown by Oh and Zeng[38] their contributions are negligible for water. This gives us the equivalence of two types of vapor simulations. Simulations of the isolated cluster are much less time consuming since there are much less molecules in the simulated system. Our estimations have shown that the method presented here seems to be as computationally efficient as other methods of the second type. The advantage of our method lies in the fact that the canonical average of grand canonical growth and decay rates, as was shown above, can be related to the rate constants in Becker and Döring kinetic approach. It allows us to get a deeper insight into the nucleation process.

III. COMPUTATIONAL DETAILS

Our simulation is a semigrand canonical Metropolis[31] Monte Carlo simulation that evaluates the grand canonical growth and decay probabilities in a canonical ensemble consisting of a single cluster. One cluster size is studied at a time, and no track is kept of vapor or noninteracting molecules. The cluster–vapor interaction is neglected since it has been shown to be vanishingly small.[38] Two parameters that enter the simulation are temperature T and the monomer number density n_v related to the chemical potential according to Eq. (28). During MC moves we keep the cluster's center-of-mass at the center of a spherical cavity with volume V, which is large enough to assure that the cluster edges are always far from the cavity boundaries. We define the cluster according to the Stillinger[40] cluster definition as a network of connected neighbors, where the distance to the nearest neighbor is less than $r_{\text{neigh}} = 3.8$ Å. The cluster is created by grand canonical insertions of molecules around the

origin. We then let it relax in a sequence of $10^8 \cdot N$ canonical Monte Carlo moves, where N denotes the number of molecules in the cluster.

The growth probability for a specific molecular configuration is calculated by inserting molecules at random positions around the cluster. The creation probability is calculated for every insertion and its average gives the growth probability for the given configuration. The density $(N_{\max}-N)/V$ of the inserted molecules was $2\rho_l$ (ρ_l is the liquid density) for small clusters ($2 \leqslant n \leqslant 12$) and $1-1.5\rho_l$ for larger clusters. The dimer creation probabilities were calculated by inserting molecules with a density of $10^5\rho_l$ around the monomer located at the origin. In Eq. (11), $(N_{\max}-N)$ accounts for the total amount of attempted insertions in volume V, but according to the Stillinger cluster definition only molecules that are inserted in the volume $V'=4\pi r'^3/3$ closer than $r'=r_{\text{furthest}}+r_{\text{neigh}}$ to the origin can even in principle be connected to the cluster (r_{furthest} is the distance of the furthest molecule from the origin). Thus we save computer time by actually attempting creation only in volume V', but scaling the number of attempts by V/V' when calculating $(N_{\max}-N)$ in Eq. (11).

Likewise, the decay probability for a given configuration is the average of annihilation probabilities of each molecule. Only annihilations which do not split the remaining cluster into two parts are accounted for. The average growth and decay probabilities are canonical ensemble averages of single configuration probabilities. The canonical simulation consists of 3×10^8 Monte Carlo moves where every 150th configuration is sampled. The maximum displacement was adjusted to achieve the 50% acceptance ratio, the average shift being around 0.3 Å, depending on the cluster size. A random rotation angle between 0° and 17° around a randomly oriented axis was combined with the translational displacement. After each move, the center-of-mass of the cluster was moved to the origin. Several vapor densities were studied simultaneously, as the vapor density only affects γ in Eqs. (7) and (8), energies U being independent of n_v.

For small clusters the decay probability is greater than the growth probability, but an increase in size causes the molecules to become more tightly bound to the cluster and hence the cluster is less likely to decay. In the simulation the critical cluster size N^* at a given n_v and T is found when the averages of growth probability $\overline{G_{N^*-1}}$ and decay probability $\overline{D_{N^*}}$ are equal. If we calculate G_N and D_N for each cluster size from monomer upward, we can solve the nucleation rate. This can be done in two ways. If the critical cluster size is small ($N^* \leqslant 10$) we can use the exact summation for the nucleation rate given by[4,36]

$$J^{-1}=\frac{1}{\beta_1 n_v}\left[1+\sum_{i=2}^{\infty}\left(\prod_{N=2}^{i}\frac{\alpha_N}{\beta_N}\right)\right], \tag{37}$$

where for a cluster containing N molecules, α_N is the evaporation rate constant, and the condensation rate constant β_N is given by kinetic gas theory as

$$\beta_N=\frac{1}{1+\delta_{1,N}} n_v \sqrt{6kT}\left(\frac{3}{4\pi}\right)^{1/6} \times\left(\frac{1}{Nm_1}+\frac{1}{m_1}\right)^{1/2}[(Nv_1)^{1/3}+v_1^{1/3}]^2, \tag{38}$$

where m_1 is the molecular mass and v_1 is the molecular volume of the pure bulk liquid. We can now replace α_N/β_N by $\overline{D_N}/\overline{G_N}$. The expression for nucleation rate then becomes

$$J^{-1}=\frac{1}{\beta_1 n_v}\left[1+\frac{\overline{D_2}}{\overline{G_2}}+\frac{\overline{D_2}\,\overline{D_3}}{\overline{G_2}\,\overline{G_3}}+\cdots\frac{\overline{D_2}\,\overline{D_3}\cdots\overline{D_i}}{\overline{G_2}\,\overline{G_3}\cdots\overline{G_i}}+\cdots\right]. \tag{39}$$

The summation in Eq. (39) converges rapidly when $i \gg N^*$, but around half of the contribution comes from sizes larger than the critical size, and we have to extend our simulations to sizes of around $2N^*$ to use this equation. If the critical cluster size is large, it is more convenient to calculate the nucleation rate from the familiar classical expression

$$J=K_{N^*}\exp(-\Delta W^*/(kT)), \tag{40}$$

where the work of formation of the critical cluster ΔW^* is calculated from Eq. (34). The prefactor K_{N^*} contains the collision rate of monomers with a single critical cluster in a saturated vapor and the Zeldovich nonequilibrium factor Z_{N^*}. It can be expressed as

$$K_{N^*}=\beta_{N^*} n_v^{S=1} Z_{N^*}=\beta_{N^*}\frac{p_{\text{sat}}}{kT}\sqrt{\frac{\sigma}{kT}}\left(\frac{2v}{9\pi N^{*2}}\right)^{1/3}, \tag{41}$$

where v is the molecular volume of liquid water, S is the saturation ratio and σ is the bulk surface tension. In this paper we have used Eq. (40) instead of Eq. (39) in all calculations of the nucleation rate. The classical formation energy, which we compare our simulated formation energy with, is given by

$$\Delta W=A\sigma-NkT\ln S, \tag{42}$$

where A is the surface area of the cluster calculated assuming bulk liquid density and a spherical droplet. The classical critical cluster size is

$$N^*=\frac{32\pi\sigma^3}{3\rho_l^2(kT\ln S)^3}. \tag{43}$$

The saturation vapor pressure,[10] surface tension[10] and density[41] are given by

$$p_{\text{sat}}(T)=\exp(77.344\,912\,96-7235.424\,651/T-8.2\cdot\ln T+0.005\,711\,3\cdot T)\ \ [\text{Pa}], \tag{44}$$

$$\sigma(T)=93.6635+0.009\,133\cdot T-0.000\,275\cdot T^2\ \ [\text{mN/m}], \tag{45}$$

$$\rho_l(T)=\frac{1}{v}=\frac{1049.572-0.1763\cdot T}{m_w}\ \ [1/\text{m}^3], \tag{46}$$

where m_w is the mass of a water molecule.

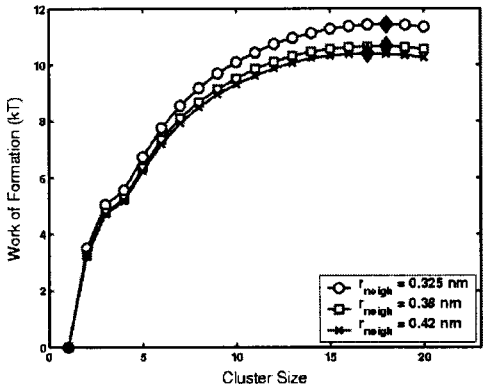

FIG. 1. The work of formation as a function of cluster size for three different limiting neighbor distances at 300 K and $n_v = 6.82 \times 10^{-3}\,\mathrm{nm}^{-3}$ for TIP4P potential model. The critical size is marked with a filled diamond in each case.

IV. THE EFFECT OF CLUSTER DEFINITION

The correct identification of a physical cluster is essential to molecular theories of nucleation. We shall not go into details here, but one should note that there are several competing ways to identify a physical cluster, and the applied cluster definition is often chosen according to simulation method. A good description of a variety of cluster definitions can be found in a paper by Reiss *et al.*[42]

One common cluster definition is presented by Lee, Barker, and Abraham,[14] where the molecules belong to the same cluster if they can be included in a sphere of fixed volume v, whose origin is at the center-of-mass of these molecules. The volume v is defined so that the thermodynamic properties of the cluster are nearly independent of it over a wide range of v. This cluster is usually called an LBA cluster. In our simulation we have chosen to use another common cluster definition presented by Stillinger,[40] where the cluster is defined as a group of molecules for which each member is connected to at least one other member within a minimum distance r_{neigh}. The cluster is then defined as a connected network of neighboring molecules. When applying the Stillinger cluster definition, the choice of r_{neigh} is somewhat arbitrary. For bulk liquid water the natural choice of r_{neigh} would be the first minimum of the oxygen–oxygen radial distribution function. For small clusters the density at the edge of the cluster is less than the bulk liquid density. As in the LBA cluster, it is then natural to allow molecules to be more loosely bonded than in bulk liquid. For our simulations we chose $r_{\mathrm{neigh}} = 3.8\,\text{Å}$, corresponding to the distance between the first minimum and the second maximum in the oxygen–oxygen radial distribution function of bulk liquid water.

Figure 1 shows how the work of formation curve is affected if r_{neigh} is varied. The curves correspond to r_{neigh} distances of 3.25 Å (first minimum of O–O radial distribution function), 3.8 Å (simulation), and 4.2 Å (second maximum of O–O radial distribution function). It can be seen, that not only the height of energy barrier is affected but the critical cluster size can also vary. However, the differences are not very significant.

V. RESULTS AND DISCUSSION

We now use the theoretical tools presented above for simulations of the nucleation of water. The simulation results for critical cluster sizes and nucleation rates are compared with experimental data by Wölk and Strey[10] and with classical nucleation theory (CNT). Our simulations are run with three different water models at constant temperature T and monomer density n_1. The tested models include two widely used stiff and unpolarizable potential models SPC/E[24] and TIP4P,[23] and a fairly recent stiff but polarizable potential model developed by Guillot and Guissani,[28] called GG model in this paper.

The method was first tested with SPC/E. This model has been used in the calculations of work of formation of small water droplet also by Kusaka *et al.*[16] We ran our simulations at the conditions used by Kusaka *et al.*,[16] $T = 298.15\,\mathrm{K}$ and $n_1 = 10^{-3}\,\mathrm{nm}^{-3}$ with SPC/E potential, and were able to exactly reproduce the free energy curve in their Fig. 8.

Wölk and Strey[10] have measured the homogeneous nucleation rates at five different temperatures T by varying the vapor pressures p_v. To compare our simulations with their data we need to relate our input parameter n_1 to p_v. There are two ways to do this. The first way is to input into simulation some value n_1 and calculate the resulting cluster distribution n_N by using Eq. (31). Assuming that imperfect gas is a mixture of ideal gases corresponding to each cluster size, p_v can be then be solved from

$$p_{\mathrm{V}} = kT \sum_{N=1}^{\infty} n_N . \tag{47}$$

In practice the summation only extends to some limiting size N, and larger clusters are artificially taken away from the system.[34] Therefore, N can be safely set to a size slightly greater than the critical size, as long as the critical size is fairly large. Another way is to calculate the monomer partial pressure contribution to the experimental value of p_v. This can be done using a virial expansion with its first term corresponding to the monomer gas.

The choice of method does not matter as long as the contribution of nonmonomer gases to vapor pressure is rather small. By carrying out the virial expansion one can indeed verify that during the experiments of Wölk and Strey the clusters give a negligible contribution to p_V. But as we shall see later, the SPC/E model produces very small critical cluster sizes, and the choice of N becomes difficult. Thus, for SPC/E, the calculation of p_v from Eq. (47) becomes questionable. For this reason we have chosen to represent all the results in terms of n_1.

Figure 2 shows simulation results for the reversible work of cluster formation as a function of cluster size at 259.9 K and $n_1 = 4.61 \times 10^{-4}\,\mathrm{nm}^{-3}$ for all three potential models and the classical value calculated with Eq. (42). At this temperature and vapor density the classical nucleation theory agrees quite well with the experimental data by Wölk and Strey and the classical work of formation curve can be used as a reference for the experimental free energy barrier. The figure emphasizes how dramatically the resulting work of formation depends on the choice of model potential. Compared to

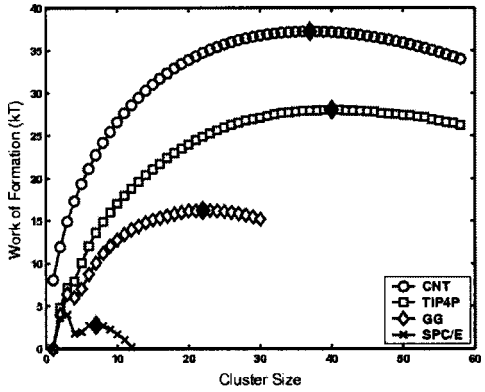

FIG. 2. The reversible work of cluster formation against cluster size at 259.9 K and $n_1 = 4.61 \times 10^{-4}\ \mathrm{nm}^{-3}$ for TIP4P, SPC/E, and GG models of water. Also the classical nucleation theory prediction is shown since the experimental nucleation barrier height is close to the classical one at these conditions.

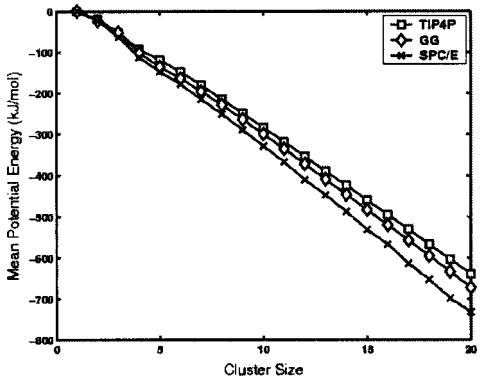

FIG. 3. The mean potential energy of the clusters against cluster size at 259.9 K and $n_v = 4.61 \times 10^{-4}\ \mathrm{nm}^{-3}$. Note that the cluster definition affects the mean potential energy in the clusters.

the classical curve, SPC/E especially seems to produce an energy barrier and a critical cluster size that are clearly too low. TIP4P produces a work of formation curve that agrees best with classical nucleation theory. We can also note that the critical cluster size with TIP4P is close to the one predicted by CNT. The GG model, that is the only polarizable water model, does not perform as well as TIP4P. Interestingly, all the water potentials produce a jump in the work of formation when the cluster has four molecules. This implies the existence of a magic number, representing a fairly stable tetramer structure.

The fact that all models produce too low work of formation curves must lie in the properties of the potential models. The models that we study here are relatively simple representations of the water molecule. They have been constructed to represent a set of some of the bulk properties of water while failing to accurately represent others. To this day there exists no universal water model that could accurately describe all the properties simultaneously with a reasonable accuracy, and even the growing complexity of the models has not improved the situation significantly. In a nucleation simulation one would like to use a water model capable of describing the interaction between molecules not only in the bulk liquid phase, but also inside the nucleating clusters containing only a few molecules. Indeed, if we look at binding energies of clusters produced by different models (Fig. 3), we notice that the stronger binding corresponds to lower work of formation. Kathmann *et al.*[36] have shown that TIP4P potential results in too much binding between cluster molecules compared to real water. This in turn decreases the work of formation of clusters. It is then understandable that GG and SPC/E results give a worse agreement with the reference work of formation in Fig. 2, as they produse even more strongly bound clusters than TIP4P.

Stronger binding suppresses the decay rate of the clusters. This can be seen from Fig. 4, where we show the dimensionless grand canonical probabilities G_1 and D_i of growth and decay of clusters for each model and cluster size. The probabilities have been calculated for clusters in a spherical cavity of 5 nm radius, so that the results for different models are comparable. It is striking to see that the evolution and magnitude of the growth probability is nearly identical in all models. The difference is seen mainly in the decay probability. Again, the dip in the decay probability for a cluster containing four molecules can be clearly seen in all models. It probably results from a favorable geometry of a water cluster containing four molecules compared to adjacent cluster sizes.

For the comparison with experiments, we calculated the full work of formation curves up to the critical cluster sizes at temperatures (218.9, 229.5, 239.6, 249.7, and 259.9 K) and monomer densities corresponding to experiments by Wölk and Strey.[10] Figure 5 shows the simulated critical cluster sizes, along with the experimental predictions obtained by fitting a function $\ln J[T,\ln(S)]$ to the experimental data of Wölk and Strey, and using the nucleation theorem[43]

$$\left(\frac{\partial \ln J}{\partial S}\right)_T = N^* + 1. \tag{48}$$

The form of the chosen fitting function affects the results. We have tried several functional forms loosely based on the classical expression for nucleation rate, and indicate by error bars the uncertainty due to the arbitrary choice. The dramatic variation in the critical cluster sizes gained from different

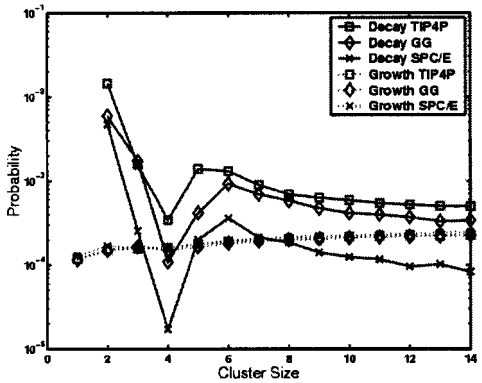

FIG. 4. Grand canonical average growth and decay probabilities as a function of cluster size at 259.9 K and $n_v = 4.61 \times 10^{-4}\ \mathrm{nm}^{-3}$ for a simulation sphere of 5 nm radius. All growth probabilities overlap. The units of probability are arbitrary.

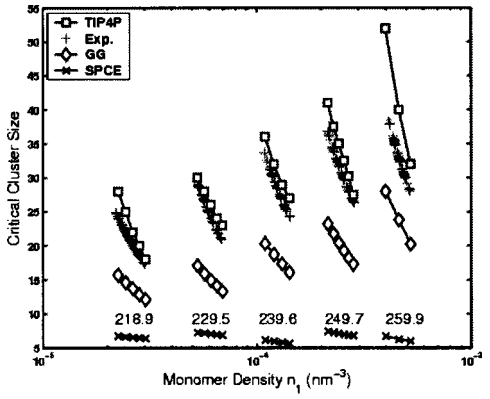

FIG. 5. Critical cluster sizes as a function of monomer density at five different temperatures. We show the simulation results for TIP4P, GG, and SPC/E models. Critical sizes deduced from experiments by Wölk and Strey and classical results are also shown.

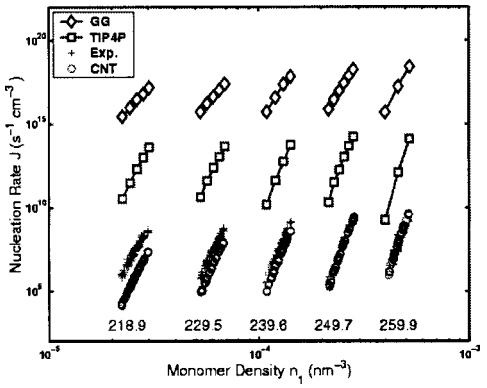

FIG. 7. Nucleation rate as a function of monomer density. We show the simulation results using the TIP4P and GG models at five temperatures. Experimental results of Wölk and Strey and classical theory nucleation rates are also shown.

potential models is clearly seen. The best fit is gained with TIP4P for which the magnitude and monomer density dependency is fairly close to experimental results. The agreement for GG is less satisfactory, and SPC/E produces critical clusters sizes which are clearly too small. Figure 6 shows the classical theory results for critical cluster size calculated from Eq. (43) with the experimental results and TIP4P simulation results at 259.9 and 218.9 K. Classical nucleation theory still gives a better fit with the experiments than the simulations. However, given that TIP4P produces the wrong magnitude for the work of formation, the agreement between the simulated and experimental critical sizes is quite remarkable.

The insertion of the calculated values for the work of formation into Eq. (40) with the classical prefactor [Eq. (41)] gives us the nucleation rates at a given T and n_1. The simulation results for TIP4P and GG are shown in Fig. 7, along with the experimental data and classical nucleation theory results. The classical theory agrees well in magnitude with the experimental data in this temperature region. It also has the right dependency on the monomer density, but a wrong temperature dependency. Due to underestimation of the work of formation, simulations give nucleation rates which are too high. But TIP4P particularly gives a correct dependency on the monomer density and temperature. However, the magnitude is constantly overestimated by a factor of 2.2×10^4. Also GG model performs better than classical theory in terms of temperature dependency, although the magnitude is now constantly overestimated by a factor of 1.3×10^9.

The correct temperature and monomer density dependence is more clearly seen in Fig. 8, where the simulated nucleation rates are divided with factors mentioned above to coincide with the experimental data. TIP4P simulations especially match the experimental values almost perfectly. The fit in the temperature dependency with GG simulation results is also better than the classical predictions, but the monomer density dependency is not quite as good.

Earlier, Hale and DiMattio[20] used a Monte Carlo simulation to determine the Helmholtz free energy differences of small clusters at 260, 280, and 300 K using TIP4P. Based on their simulations, they proposed a scaled form for the nucleation rate that gives a correct temperature dependency for the

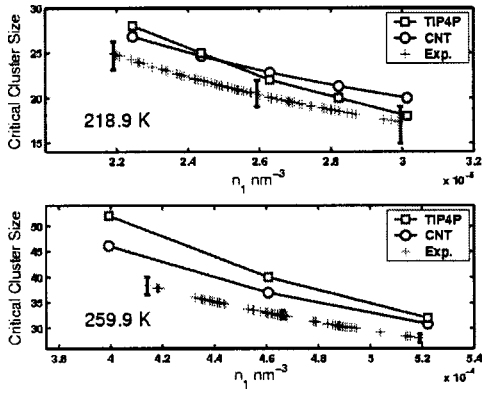

FIG. 6. Critical cluster sizes as a function of monomer density at two different temperatures for TIP4P model with experimental predictions and classical theory results. Error bars indicating the uncertainty in experimental results are also shown.

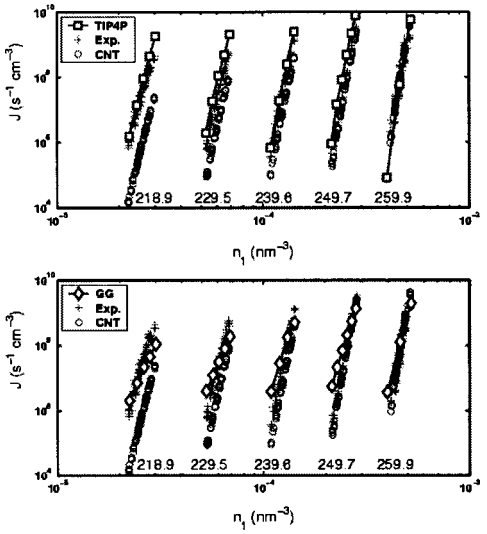

FIG. 8. Nucleation rates from the simulations divided with constant factors. Results of the TIP4P (factor 2.2×10^4) and GG (factor 1.3×10^9) models are shown against monomer density at five different temperatures. Experimental and classical nucleation theory results are also shown.

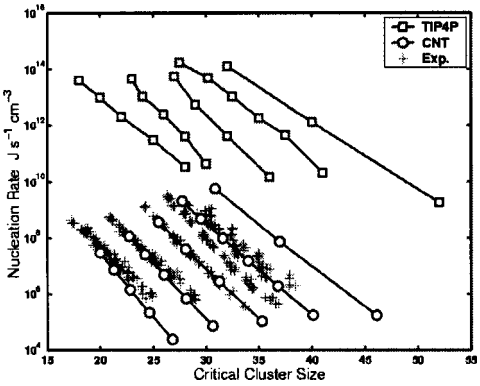

FIG. 9. Nucleation rate as a function of critical cluster size. From left to right the lines correspond to temperatures 218.9, 229.5, 239.6, 249.7, and 259.9 K.

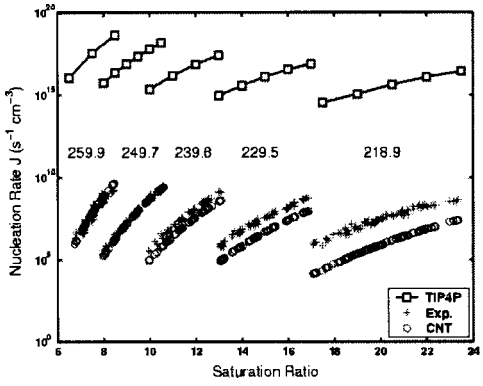

FIG. 10. Nucleation rate as a function of saturation ratio. Note that the temperatures are in a different order than previously.

nucleation rate. They also argued that this scaling also applies to a wider temperature range, but didn't carry out simulations outside the 260–300 K temperature interval. Our results verify that TIP4P produces the correct temperature dependency also in the 220–260 K interval.

Figure 9 shows a plot of critical cluster sizes against the nucleation rate for TIP4P in logarithmic scale. We see near-linear curves that shift gradually toward larger critical cluster sizes with increasing temperature. According to classical theory the curves are not exactly linear, but curve towards a higher nucleation rate with the increasing critical cluster size. The slopes in all the curves are fairly similar, and the separation between the simulated lines is closer to experimental separation than the lines given by the classical theory.

In this paper we have studied nucleation in terms of temperature and monomer density. Another possibility also exists, arising from the incomplete description of real water properties by simple potential models. The models do not give an accurate description of the binodal of water. Thus the saturation pressures for the models are different from real water. Because of this, comparisons of experimental and simulated nucleation rate and critical cluster size at the same temperature and saturation ratio give different results than comparisons at the same temperature and monomer density.

Calculation of the binodal for different water models at the studied temperatures is beyond the scope of this paper. However, Dr. Bin Chen has kindly provided us the saturation pressures for TIP4P close to our simulation temperatures[44] gained from Gibbs Ensemble Monte Carlo simulations: 220 K; 6.2 (0.4) Pa, 230 K; 17.5 (0.9) Pa, 240 K; 55.1 (1.8) Pa and 260 K; 328 (12) Pa. Numbers quoted in the brackets are the standard error of mean analyzed from five independent simulations and each has a length of 200 000 Monte Carlo cycles. 500 molecules were used in these simulations.

Figure 10 shows how results for TIP4P in Fig. 7 change if the input monomer density is calculated from the model saturation vapor pressure. As one can see, the agreement between simulations and experiment does not improve. Simulated nucleation rates become considerably higher, because the saturation pressures and hence the input monomer densities are about 1.4 times higher than in the previous case. The temperature dependence of nucleation rate is not as good either. We note that for SPC/E, the critical cluster sizes would grow and nucleation rates drop since the SPC/E saturation vapor pressures are only fractions of true saturation pressures at the studied temperatures.

VI. CONCLUSION

We have studied homogeneous nucleation of water with a semigrand canonical Monte Carlo method that was originally introduced by Vehkamäki and Ford.[22] This method is based on the calculation of average grand canonical growth and decay probabilities of molecular clusters in a canonical ensemble. In this paper, we have showed how it can be used in calculations of the work of formation of molecular clusters. Furthermore, we have derived a connection between the growth and decay probabilities and kinetic condensation and evaporation rates. This connection opens up new possibilities to study cluster properties. It can also be used in the sensitivity analysis of the nucleation process.

Here we have used the extended method for calculations of critical cluster size, cluster work of formation and the nucleation rate of water. In the simulations clusters were defined by applying the Stillinger cluster definition. We checked that the results are not very sensitive to the chosen nearest neighbor distance of the cluster definition. The simulations were run at temperatures and monomer densities corresponding to experiments by Wölk and Strey.[10] Calculations were carried out with three stiff potential models, namely with the unpolarizable TIP4P[23] and SPC/E[24] models and with a polarizable model developed by Guillot and Guissani,[28] called GG model in this paper. Comparisons with experimental data show that our simulations with TIP4P reproduce the size of the critical cluster fairly accurately. TIP4P also produces a correct temperature and monomer density dependency of nucleation rate, but the magnitude is overestimated by a constant factor of 2.2×10^4. With GG we get nucleation rates which are too high by a factor of 1.3×10^9, but again the temperature and monomer density dependencies of the nucleation rates and critical cluster sizes agree fairly well with the experimental data. The results of the TIP4P and GG models are encouraging since the major drawback of classical nucleation theory is the wrong temperature dependence of the nucleation rate. Simulations with SPC/E resulted in far too small critical cluster sizes and high nucleation rates.

The water models generally do not reproduce the saturation vapor pressure behavior of real water correctly. Because of this, we also compared the TIP4P results with experiments at the same saturation ratios rather than monomer densities. This did not improve the agreement between simulations and experiments.

Overall, the results reflect the high sensitivity of the nucleation process. The resulting critical cluster sizes and nucleation rates are highly affected by the choice of molecular interaction potential model. Small variations in the average potential energies of clusters between the models have a strong effect on the stability of the clusters. A stronger average binding between the molecules in a cluster is reflected as a reduced decay probability, while the growth probability is not significantly affected. Thus, the overall stability of clusters increases with stronger binding. This in turn lowers the critical cluster size and boosts the nucleation rate.

[1] M. Kulmala, L. Pirjola, and J. M. Mäkelä, Nature (London) **404**, 66 (2000).
[2] S. Gao, A. Hegg, G. Frick *et al.*, J. Geophys. Res. **106**, 27619 (2001).
[3] M. Volmer and A. Weber, Z. Phys. Chem., Stoechiom. Verwandtschaftsl. **119**, 277 (1925).
[4] R. Becker and W. Döring, Ann. Phys. (Leipzig) **24**, 719 (1935).
[5] J. Zeldovich, Sov. Phys. JETP **12**, 525 (1942).
[6] C. Hung, M. J. Krasnopoler, and J. L. Katz, J. Chem. Phys. **90**, 1856 (1989).
[7] J. L. Schmitt, G. W. Adams, and R. A. Zalabsky, J. Chem. Phys. **77**, 2089 (1982).
[8] R. Strey, T. Schmelling, and P. E. Wagner, J. Chem. Phys. **85**, 6192 (1986).
[9] A. Kacker and R. H. Heist, J. Chem. Phys. **82**, 2734 (1985).
[10] J. Wölk and R. Strey, J. Phys. Chem. B **105**, 11683 (2001).
[11] J. Lothe and G. M. Pound, J. Chem. Phys. **36**, 2080 (1962).
[12] D. W. Oxtoby and R. Evans, J. Chem. Phys. **89**, 7521 (1988).
[13] A. Dillmann and G. E. A. Meier, J. Chem. Phys. **94**, 3872 (1991).
[14] J. K. Lee, J. A. Barker, and F. F. Abraham, J. Chem. Phys. **58**, 3166 (1973).
[15] N. Garcia and J. M. Soler Torroja, Phys. Rev. Lett. **47**, 186 (1981).
[16] I. Kusaka, Z.-G. Wang, and J. H. Seinfeld, J. Chem. Phys. **108**, 3416 (1998).
[17] K. Oh and X. Zeng, J. Chem. Phys. **110**, 4471 (1999).
[18] B. N. Hale and R. Ward, J. Stat. Phys. **28**, 487 (1982).
[19] B. N. Hale, Aust. J. Phys. **49**, 425 (1996).
[20] B. Hale and D. DiMattio, in *Proceedings of the 15th International Conference on Nucleation and Atmospheric Aerosols*, edited by B. Hale and M. Kulmala (American Institute of Physics, 2000), pp. 31–34.
[21] B. Chen, J. I. Siepmann, K. J. Oh, and M. L. Klein, J. Chem. Phys. **115**, 10903 (2001).
[22] H. Vehkamäki and I. J. Ford, J. Chem. Phys. **112**, 4193 (2000).
[23] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[24] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[25] G. T. Gao, K. J. Oh, and X. C. Zeng, J. Chem. Phys. **110**, 2533 (1998).
[26] K. Yasuoka and M. Matsumoto, J. Chem. Phys. **109**, 8463 (1998).
[27] H. Vehkamäki and I. Ford, Phys. Rev. E **59**, 6483 (1999).
[28] B. Guillot and Y. Guissani, J. Chem. Phys. **114**, 6720 (2001).
[29] L. D. Landau and E. M. Lifshits, *Statistical Physics, Part 1* (Pergamon, Oxford, 1969).
[30] J. Yao, R. Greenkorn, and C. Chao, Mol. Phys. **46**, 587 (1982).
[31] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[32] D. Heermann, *Computer Simulation Methods in Theoretical Physics* (Springer-Verlag, Berlin Heidelberg, 1986).
[33] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. (Academic, New York, 2002).
[34] F. F. Abraham, *Homogeneous Nucleation Theory* (Academic, New York and London, 1974).
[35] J. Frenkel, *Kinetic Theory of Liquids* (Oxford University Press, London, 1946).
[36] S. M. Kathmann, G. K. Schenter, and B. C. Garrett, J. Chem. Phys. **116**, 5046 (2002).
[37] H. Vehkamäki and I. J. Ford, J. Chem. Phys. **113**, 3261 (2000).
[38] K. J. Oh and X. C. Zeng, J. Chem. Phys. **112**, 294 (2000).
[39] P. R. ten Wolde and D. Frenkel, J. Chem. Phys. **109**, 9901 (1998).
[40] F. H. Stillinger, J. Chem. Phys. **38**, 1486 (1963).
[41] O. Preining, P. E. Wagner, F. G. Pohl, and W. Szymanski, *Heterogeneous Nucleation and Droplet Growth* (University of Vienna, Institute of Experimental Physics, Vienna, Austria, 1981).
[42] H. Reiss, A. Tabazadeh, and J. Talbot, J. Chem. Phys. **92**, 1266 (1990).
[43] D. W. Oxtoby and D. Kashchiev, J. Chem. Phys. **100**, 7665 (1994).
[44] B. Chen, private correspondence.