Detecting vapour bubbles in simulations of metastable water

Miguel A. González, Georg Menzl, Juan L. Aragones, Philipp Geiger, Frederic Caupin, Jose L. F. Abascal, Christoph Dellago, and Chantal Valeriani

Citation: The Journal of Chemical Physics **141**, 18C511 (2014); doi: 10.1063/1.4896216
View online: http://dx.doi.org/10.1063/1.4896216
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/141/18?ver=pdfcov
Published by the AIP Publishing

Detecting vapour bubbles in simulations of metastable water

Miguel A. González,[1,a)] Georg Menzl,[2,a)] Juan L. Aragones,[1,3] Philipp Geiger,[2] Frederic Caupin,[4] Jose L. F. Abascal,[1] Christoph Dellago,[2,b)] and Chantal Valeriani[1,b)]
[1]*Departamento de Química Fsica, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, 28040 Madrid, Spain*
[2]*Faculty of Physics and Center for Computational Materials Science, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria*
[3]*Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[4]*Laboratoire de Physique de la Matiere Condensee et Nanostructures, Universite Claude Bernard, Lyon 1 et CNRS, Institut Universitaire de France, 43 boulevard du 11 novembre 1918, 69100 Villeurbanne, France*

(Received 17 June 2014; accepted 9 September 2014; published online 26 September 2014)

The investigation of cavitation in metastable liquids with molecular simulations requires an appropriate definition of the volume of the vapour bubble forming within the metastable liquid phase. Commonly used approaches for bubble detection exhibit two significant flaws: first, when applied to water they often identify the voids within the hydrogen bond network as bubbles thus masking the signature of emerging bubbles and, second, they lack thermodynamic consistency. Here, we present two grid-based methods, the M-method and the V-method, to detect bubbles in metastable water specifically designed to address these shortcomings. The M-method incorporates information about neighbouring grid cells to distinguish between liquid- and vapour-like cells, which allows for a very sensitive detection of small bubbles and high spatial resolution of the detected bubbles. The V-method is calibrated such that its estimates for the bubble volume correspond to the average change in system volume and are thus thermodynamically consistent. Both methods are computationally inexpensive such that they can be used in molecular dynamics and Monte Carlo simulations of cavitation. We illustrate them by computing the free energy barrier and the size of the critical bubble for cavitation in water at negative pressure. [http://dx.doi.org/10.1063/1.4896216]

I. INTRODUCTION

By super-heating a liquid beyond the liquid–vapour coexistence temperature or reducing its pressure below the saturated vapour pressure, the liquid becomes metastable and eventually transforms into the thermodynamically stable vapour phase via a bubble nucleation mechanism. Remarkably, in the case of water, strongly negative pressures can be reached experimentally before the liquid "breaks" under the mechanical tension and cavitation occurs, which has important implications for biological processes like water transport in trees[1–3] and the sonocrystallization of ice.[4,5] Since bubble nucleation takes place at the nanoscale, observing it directly in experiments is very difficult. Thus, apart from the nucleation rate, little information can be obtained from experiments. Numerical simulations provide a complementary technique to investigate the microscopic details of bubble nucleation such as the size and shape of the critical bubble and the structure of the liquid around it.

In order to describe the nucleation process, one requires an adequate order parameter to track the phase transition mechanism, i.e., to detect the formation and growth of a bubble. Homogeneous bubble nucleation (or cavitation) from an super-heated fluid has been recently studied in simple liquids (Lennard-Jones) using either the volume of the largest bubble as a local order parameter[6–8] or the density as a global one.[9] In recent work, Meadley and Escobedo[10] demonstrated that when comparing the nucleation free energy barrier obtained with the volume of the largest bubble to the one computed using the global density, the largest bubble volume is a better reaction coordinate than the density; it correlates more strongly with committor probabilities. Even though there is a one-to-one relation between the two order parameters, the authors showed that using a global order parameter can introduce finite size effects as one observes the formation of multiple bubbles instead of capturing the formation of one single critical nucleus.

In general, detecting vapour bubbles in a metastable liquid is a challenging problem, since the vapour phase consists both of "vapour-like" molecules and of "void" spaces. Thus, to compute the volume of the largest bubble and use it as a local order parameter, one needs to be able to identify vapour regions in the metastable liquid, compute their volumes, and select the largest one as the order parameter. This procedure has been applied in Ref. 6, where homogeneous bubble nucleation in a super-heated Lennard-Jones fluid was studied using forward flux sampling, and the largest connected low-density region was used as a local order parameter. The low-density region was defined by means of a "grid-based" procedure: (1) liquid and vapour-like particles were classified using the

[a)]M. A. González and G. Menzl contributed equally to this work.
[b)]Authors to whom correspondence should be addressed. Electronic addresses: christoph.dellago@univie.ac.at and cvaleriani@quim.ucm.es

first-neighbours distribution and, (2) a three-dimensional grid was superimposed on the system, and (3) a cluster analysis was performed on the void space, identifying the bubble with the largest volume as the local order parameter.

Simulations of bubble nucleation in water are scarce. A qualitative picture of bubble formation in strongly superheated water was provided in Ref. 11, and, more recently, a detailed analysis of bubble formation in water under tension was carried out in Ref. 12. In the latter, some of us have recently shown that a valid alternative to a grid-based procedure consists in employing a Voronoi-based analysis to detect the largest bubble. However, the main drawback of this approach is that performing a Voronoi tessellation is a CPU-time consuming task. In addition, most of the available open-source Voronoi packages, as far as we have experienced, are not tailored to target the task of detecting bubbles since the topological distance involves a different set of neighbours than the geometrical one (distant molecules can share a Voronoi-face, thus being Voronoi-neighbours even though they are far away from each other).

In general, an order parameter to detect bubbles in water should meet the following criteria. (1) The order parameter should be local. When studying a nucleation process, the use of a local order parameter allows to detect and follow the growing nucleus instead of the variation of a global property of the system. This is a prerequisite when studying nucleation in a large system, where fluctuations can overshadow the signature of the nucleus when observables are averaged over the entire system, and it allows to obtain structural information about the nucleus and the surrounding liquid. (2) The order parameter should not impose a specific shape on the detected bubble, which is of particular importance when the order parameter is used in a free energy computation method such as umbrella sampling, where the measured free energy barrier height would be altered by excluding certain cluster shapes from the sampling. (3) The order parameter should measure the "true" volume of a bubble. In the same way we can measure the volume of macroscopic objects by the amount of liquid they displace when immersed in a liquid, the volume of a bubble detected by a local order parameter should be equivalent to the average change in volume of the system between the metastable liquid and a system containing such a bubble. An order parameter calibrated in this fashion yields, up to discretization errors, a thermodynamically consistent volume estimate, which allows for direct comparison to experimental data. (4) Since the order parameter has to be evaluated hundreds of thousands of times during a typical simulation run, it should be computationally inexpensive.

In what follows, building on the method introduced in Ref. 6, we propose two novel grid-based methods tailored to work for a network-forming liquid such as water. We will analyze these order parameters with respect to the criteria listed above and present physical features of the nucleation process, in particular the size of the critical cluster and the free energy barrier height for different degrees of metastability.

The rest of the paper is organized as follows. After giving simulation details in Sec. II, we describe the techniques we used to study spontaneous nucleation and the case when nucleation cannot happen spontaneously on timescales accessible to computer simulation because the free energy barrier is too high. In Sec. III we present the two different grid-based methods to identify the volume of the largest bubble in metastable water: the V-method and the M-method. To conclude, in Sec. IV we report a quantitative comparison of the results obtained when computing the order parameter with both methods in the study of spontaneous as well as non-spontaneous nucleation. We discuss our results in Sec. V.

II. SIMULATION DETAILS

A. Molecular dynamics

We simulate metastable water using the rigid non-polarizable TIP4P/2005 model[13] which has been shown to predict a number of water properties with great accuracy.[14] Relevant to this work are the accurate predictions for the vapour–liquid equilibrium[15,16] and the liquid–vapour surface tension.[17] We perform *NpT* molecular dynamics (MD) simulations either using an in-house code, based on the code previously used in Ref. 18, or GROMACS.[19]

When using the home-made code, we simulate a system of $N = 2000$ water molecules. We integrate the equations of motion with a time step of 2 fs using a time-reversible quaternion-based integrator that maintains the rigid geometry of water molecules. In particular, we carry out *NpT*-MD using a slightly modified version of the Verlet integrator proposed by Kamberaj *et al.*,[20] based on the Trotter decomposition schemes applied by Miller *et al.*[21] and Martyna *et al.*[22] In this algorithm, the coupling to the surrounding heat bath is implemented through thermostat chains based on the Nosé–Hoover[23,24] approach with an inverse frequency of 1 ps. Constant pressure is ensured by coupling a barostat based on the Andersen approach[25] to the heat bath with a relaxation time of 3 ps, which is approximately equal to the time a sound wave takes to traverse the simulation box.[26,52] Long-range interactions are treated with Ewald summation.

When using the GROMACS package, we simulate a system of $N = 500$ (when nucleation is spontaneous) or $N = 4000$ (when nucleation is not spontaneous) water molecules in an *NpT* ensemble with a time step of 1 fs. The temperature is kept fixed using the velocity rescaling thermostat[27] with 1 ps (spontaneous nucleation) or 0.2 ps (non-spontaneous nucleation) relaxation time and the pressure is set with the isotropic Parrinello–Rahman barostat[28] with a relaxation time of 2 ps (spontaneous nucleation) or 0.2 ps (non-spontaneous nucleation). We constrain the geometry of the water molecules with the SHAKE algorithm.[29] Long-range electrostatic interactions are treated using the smooth particle mesh-Ewald method.[30]

B. Nucleation rates and barriers

When we over-stretch liquid water, bubbles start appearing and disappearing at random in the system. Only when one of them overcomes a critical threshold, the entire system can cavitate. This corresponds to the system having to cross a nucleation free energy barrier, whose height depends on the

supersaturation (or amount of over-stretching). For low barriers, i.e., high supersaturation, cavitation can be observed spontaneously on the time-scales accessible to molecular dynamics simulations. For high barriers, i.e., low supersaturation, the time until cavitation occurs spontaneously is prohibitively long and thus one has to resort to rare event sampling techniques such as umbrella sampling.

1. When nucleation is spontaneous at the simulation timescale

In order to determine nucleation rates we apply the mean first passage time (MFPT) analysis[31] using the volume of the largest bubble as a local order parameter. If the barrier is high enough to guarantee separation of timescales between relaxation in the metastable basin and the barrier crossing event, the average time $\tau(V_{\text{bubble}})$ it takes until the largest bubble has a volume of V_{bubble} for the first time is[31]

$$\tau(V_{\text{bubble}}) = (\tau_J/2)\{1 + \text{erf}[c(V_{\text{bubble}} - V^*_{\text{bubble}})]\}. \quad (1)$$

Here, erf(x) is the error function, V^*_{bubble} is the size of the critical bubble, c is a constant derived from the local curvature around the top of the free energy barrier and proportional to the Zeldovich factor $Z = c\pi^{-1/2}$, and τ_J is the nucleation time. The average time τ_J the system takes to leave the metastable state and reach the point where the transition to the vapour phase is committed to proceed is related to the nucleation rate J via

$$J = \frac{1}{\tau_J \langle V \rangle}, \quad (2)$$

where $\langle V \rangle$ is the average volume of the system in the metastable liquid state.

The MFPT analysis is performed in the following fashion: by averaging over 200 independent MD trajectories which start from the metastable liquid and subsequently transform into the vapour phase, we compute the average time $\tau(V_{\text{bubble}})$ it takes until the bubble reaches a volume V_{bubble} for the first time. We then fit Eq. (1) to the data to obtain the nucleation rate J, the Zeldovich factor Z, and the size of the critical bubble V^*_{bubble}.

As in Ref. 12, we study nucleation from metastable liquid water above the spinodal line, at 280 K and $p = -2250$ bar (the spinodal at negative pressure has been calculated in Ref. 32 and corresponds to $p_{\text{sp}} \approx -2440$ bar at $T = 280$ K). It has been shown in Ref. 12 that, at these thermodynamic conditions, the volume of the critical cluster is smaller than 1 nm^3 so that a system size of $N = 500$ molecules can easily accommodate a critical bubble.

2. When nucleation is not spontaneous at the simulation timescale

In order to compute the nucleation free energy barrier as a function of bubble volume we use two rare-events numerical techniques:

(1) Umbrella sampling[33] with a hybrid Monte Carlo (HMC) scheme.[34,35] We employ a modified version of the Miller integrator[21] using a Liouville operator decomposition according to Omelyan[36] for the centre of mass equations of motion. During each HMC move the velocities are drawn from the appropriate Maxwell–Boltzmann distribution and the system is propagated according to Newton's equations of motion, where constant pressure is achieved by isotropic volume fluctuations according to the Metropolis criterion.[37] Each HMC step consists of three MD integration steps with a time step of 8 fs. Sampling was enhanced by replica exchange moves[38] between neighbouring windows and the histograms for the individual windows were pieced together using a self consistent histogram method.[39]

(2) A novel molecular dynamics umbrella sampling (HMC-NpT) scheme. Here, the system is propagated via a series of short NpT molecular dynamics trajectories where the resulting configurations are fed to a standard umbrella sampling scheme. More details on this method will be given in Ref. 40.

III. NUMERICAL METHODS

To study homogeneous bubble nucleation from overstretched metastable water we use the volume of the largest bubble, V_{bubble}, as a local order parameter. We propose to detect the largest bubble choosing one of two approaches: the V-method or the M-method. Both methods, inspired by the grid-based analysis of Ref. 6, are tailored to successfully work in a network-forming liquid such as water.

The bubbles emerging in the metastable liquid during cavitation are essentially voids with rare occurrences of vapour-like molecules due to the low density of the vapour phase. Since both methods are based on detecting cubes on a three-dimensional grid which are not occupied by liquid-like molecules, we require a criterion to distinguish between liquid- and vapour-like molecules in order to identify bubbles correctly.

We investigated the suitability of three possible criteria in Appendix A: the ten Wolde–Frenkel (WF) criterion,[41] the hydrogen bond (HB) criterion,[42] and a criterion based on the tetrahedral order parameter of Refs. 43 and 44 (q criterion). Our comparison shows that the WF and the HB criteria perform equally well, while the q criterion is not as well suited to differentiate between liquid and vapour. In this paper, we use the WF criterion in conjunction with the V-method and the HB criterion with the M-method. We would like to stress that one could use either of the two criteria with the M- or the V-method.

In what follows, we will describe each method step by step.

A. V-method for the volume of the largest bubble

Having classified all molecules as liquid or vapour-like, we proceed as follows.

Step 1: We *divide the entire simulation box into small cubes of equal size* by superimposing a three-dimensional grid on the system. Since the total volume fluctuates in a simulation at constant pressure, the cube volume varies with the size of the simulation box. In order to determine "occupied" cubes we assume that each liquid water molecule occupies a sphere with a radius r_S corresponding to the first minimum in the ra-

dial distribution function around its centre of mass. Then we iterate over all cubes in the box and any cube whose center is inside the sphere of a liquid-like water molecule is considered to be occupied, hence, it is not part of a bubble.

Step 2: We *assign unoccupied cubes to clusters* such that a cube and all its face-sharing neighbours are part of the same cluster. Each of the clusters obtained in this fashion constitutes a bubble and its volume v is the total volume of the cubes which belong to the cluster. Even though the method introduces some lattice discretisation errors by construction, we assume these errors to be negligible in view of the approximation of water molecules as spheres and the high spatial resolution of the grid.

Step 3: Even though a method incorporating only the first two steps can be used to describe bubble formation in the metastable liquid by employing the volume of the largest cluster as the order parameter, the definition of the bubble volume is somewhat arbitrary as it depends on the choice of the grid mesh and the radius of the exclusion spheres, which largely determines the "volume" of the water molecules forming the interface. Since the volume of the critical bubble V^*_{bubble} is a property of great interest, which is related to the height of the free energy barrier via the nucleation theorem[45] and thus provides a connection to experimental data,[46] our aim is to develop an order parameter whose estimate of the bubble volume depends as little as possible on the choice of arbitrary properties like r_S or the grid resolution.

In analogy to the macroscopic realm, a reasonable definition of the volume of a nanoscale object[47] is to equate its volume to the volume of the liquid it displaces, i.e., we *equate the volume of a bubble to the increase in system volume* it causes. While for a macroscopic object the effect of density fluctuations of the liquid on such a volume estimate is negligible, for small systems density fluctuations can influence the estimated volume considerably. However, one can still require from the definition of the bubble volume that it corresponds to the increase in system volume on the *average*. In the following we explain how to calibrate the bubble volume such that this requirement is obeyed.

1. Calibration of the bubble volume in the V-method

The thermodynamically consistent bubble volume $V^{\text{V}}_{\text{bubble}}(v)$ is evaluated by computing the partial derivative of the average system volume with respect to the number of bubbles of size v,

$$V^{\text{V}}_{\text{bubble}}(v) = \frac{\partial}{\partial n}\langle V\rangle_{n(v)}. \tag{3}$$

Above, the superscript V indicates that $V^{\text{V}}_{\text{bubble}}$ is the estimate obtained by the V-method. Here, $\langle V\rangle_{n(v)}$ is given by

$$\langle V\rangle_{n(v)} = \frac{1}{Q}\int V e^{-\beta[H(\mathbf{r}^N)+pV]}\delta[n(v,\mathbf{r}^N)-n]\mathrm{d}\mathbf{r}^N\mathrm{d}V,$$

where $n(v,\mathbf{r}^N)$ is the number of bubbles of size v in a given configuration and $Q = \int e^{-\beta[H(\mathbf{r}^N)+pV]}\delta[n(v,\mathbf{r}^N) - n]\mathrm{d}\mathbf{r}^N\mathrm{d}V$ is the reduced partition function. Thus, $\langle V\rangle_{n(v)}$ is the average volume of the system under the condition that the configuration contains exactly n bubbles of size v without any restriction on the number of bubbles with other volumes. The definition (3) quantifies the increase in average system volume if a bubble of size v is added to the system.

For large bubbles, where it is unlikely that multiple such bubbles or any larger bubbles occur in an equilibrium configuration, the partial derivative in Eq. (3) is simply the difference between the average volume $\langle V\rangle_v$ of a system containing a largest bubble of size v and the average volume $\langle V\rangle$ of the metastable liquid (taken at the given thermodynamic state point excluding the stable vapour phase):

$$V^{\text{V}}_{\text{bubble}}(v) = \langle V\rangle_v - \langle V\rangle, \tag{4}$$

where $\langle V\rangle_v = Q^{-1}\int V e^{-\beta[H(\mathbf{r}^N)+pV]}\delta[v(\mathbf{r}^N)-v]\mathrm{d}\mathbf{r}^N\mathrm{d}V$ with the appropriate reduced partition function Q.

Although this order parameter has its roots in the average change of the total system volume due to the presence of the bubble, we stress that it is still a local order parameter which does not impose any particular shape on the bubble, thus fulfilling the requirements 1 and 2 stated in the Introduction.

We compute $V^{\text{V}}_{\text{bubble}}$ in different ways for small and large v. Small bubbles form spontaneously in the metastable liquid. In this regime, we compute the average system volume as a function of the number $n(v)$ of bubbles of size v in a straightforward MD simulation. This calculation yields a linear dependence of system volume on the number of bubbles as shown in Fig. 1. Data points depicted in the inset of Fig. 1 are the partial derivatives $\partial\langle V\rangle_{n(v)}/\partial n$ from Eq. (3), i.e., the average increase in system volume due to the presence of one bubble of size v. In the second regime, i.e., for large bubbles, where it is unlikely that more than one cluster of volume v is present in the system, $\langle V\rangle_v$ is obtained by averaging the system volume within each window, defined as a

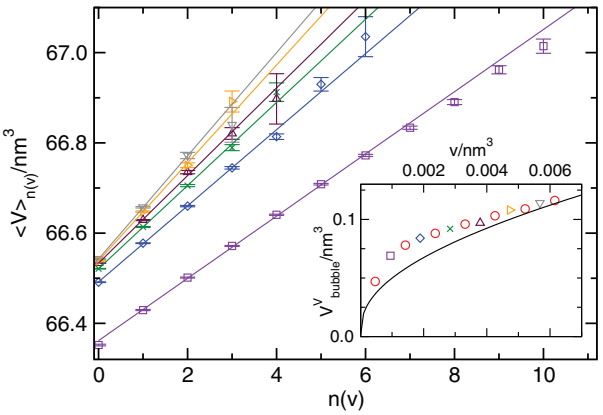

FIG. 1. Average system volume $\langle V\rangle_{n(v)}$ as a function of the number $n(v)$ of bubbles with volume v for spontaneously forming small bubbles. The data sets indicated by different colors are computed for average v-values of 0.95, 1.89, 2.84, 3.77, 4.73, 5.68 Å^3 from bottom to top. Lines are linear fits to the data and error bars indicate $\sigma/\sqrt{N}$, where σ is the standard deviation and N is the number of samples (only points with $N > 50$ were used for fitting). All data were obtained from simulations in the isobaric–isothermal ensemble at $T = 325.0$ K and $p = -1500$ bar. The average volume of the metastable liquid at these conditions is $\langle V\rangle = 66.55\,\text{nm}^3$. *Inset:* $V^{\text{V}}_{\text{bubble}}$ obtained from the slope of the fits in the main plot. Symbols indicate the corresponding data in the main plot, where the data indicated by red circles are omitted for clarity. The v-values shown are averages; fluctuations are smaller than symbol size. The black line represents the fit to $V^{\text{V}}_{\text{bubble}}$ according to Eq. (5) over the entire v-range depicted in Fig. 2.

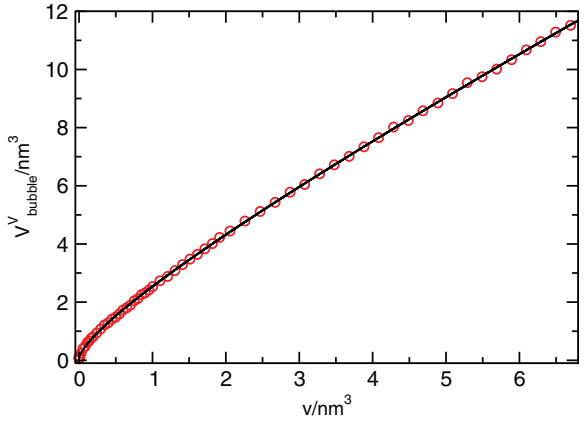

FIG. 2. $V^{\mathrm{V}}_{\mathrm{bubble}}$ as a function of the volume v of vapour-like cubes. Each data point (except for the ones shown in the inset of Fig. 1) represents an average over a window in an umbrella sampling simulation according to Eq. (4). The black line shows the fit to the data by the use of Eq. (5). All data were obtained from simulations in the isobaric–isothermal ensemble at $T = 325.0$ K and $p = -1500$ bar.

range of v-values for the volume of the largest bubble, of a HMC-umbrella sampling calculation. Results of these calculations are shown in Fig. 2.

In order to calibrate the V-method such that the grid-based estimate for the bubble volume is mapped onto the average change in system volume caused by a bubble of volume v we use a fit to the data shown in Fig. 2. Note that the functional form of the fit is arbitrary and does not assume that the bubble has a particular geometrical shape; in cases where one finds no suitable functional form, simple numerical interpolation of the data could be used. Here, we find that a fitting function of the form

$$V^{\mathrm{V}}_{\mathrm{bubble}}(v) \approx v + k_1 v^{2/3} + k_2 v^{1/3}, \tag{5}$$

where $k_1 \approx 1.17$ nm and $k_2 \approx 0.37$ nm^2 at $T = 325.0$ K and $p = -1500$ bar, works very well (indicated by the black line in the figures). This fitting function can be viewed as depending on the cluster "surface" (proportional to $v^{2/3}$) and on a term that takes into account the average "curvature" of the cluster (proportional to $v^{1/3}$).[53]

Therefore, we calibrate our estimate for the bubble volume by mapping the volume v (measured in simulation by summing over the unoccupied cubes forming the cluster) onto the corresponding average change in system volume $V^{\mathrm{V}}_{\mathrm{bubble}}$ caused by a bubble of size v. This calibration depends on the thermodynamic state and as such, it has to be performed for each state point. However, we would like to point out that the procedure is simple and there are no additional simulations necessary to map v onto $V^{\mathrm{V}}_{\mathrm{bubble}}$. Due to this calibration procedure used to approximate the average change in system volume created by the bubble, the V-method is compatible with requirement 3. Moreover, as we show in Appendix C, the V-method is also computationally inexpensive and as such obeys requirement 4.

In order to achieve high spatial resolution, we ensure that the volume of each cube on the grid is less than 0.5 Å^3 for the metastable liquid leading to the choice of 52^3 cubes for a system of $N = 2000$ water molecules. We choose the Stillinger radius $r_{\mathrm{S}} = 3.35$ Å as the radius for the exclusion spheres. This choice is a compromise between the ability to detect the formation of small bubbles and minimising the occurrence of misassigned liquid-like cubes in the metastable liquid.

B. M-method for the volume of the largest bubble

Having classified all molecules as liquid or vapour-like, we proceed as follows.

Step 1: As in the V-method and in Ref. 6, we *superimpose a three-dimensional grid* on our system and assign each cell on the grid to be either *vapour or liquid-like* depending on the molecules occupying the cell. To do this, we consider the oxygen's Lennard-Jones diameter $\sigma = 3.1589$ Å[13] to be each molecule's exclusion sphere: If the center of a cell is under an exclusion sphere, it will be labeled as either liquid- or vapour-like, according to the type of the molecule. When a cell contains more than one type of molecules, for instance a liquid- and a vapour-like molecule, it will be labelled as *liquid*-like.

Once we have classified all cells touched by molecules, we are left with labelling cells which do not clearly belong to any molecule: so called "empty cells." These appear when the size of the grid cells is comparable to the particle's diameter (see Appendix B 2). To classify the empty cells as liquid or vapour-like, we analyse both their first and second neighbour cells. If the number of face-sharing empty/vapour first neighbour cells is at least 7, then we analyze the number of face-sharing empty/vapour second neighbour cells: if they also are at least 7, the empty cell is identified as vapour-like. This procedure allows to avoid considering the typical small "cavities" characterizing a network-forming liquid as vapour-like cells.

Care should be taken when choosing the mesh size of the grid, which is defined as $\Delta = L/\delta$, where L is the box edge and δ the number of cells per edge. δ is a constant, therefore the volume Δ^3 of the grid cells fluctuates with the simulation box in a simulation at constant pressure. For all system sizes studied here, we set Δ to be about half of the oxygen diameter, which is significantly smaller than the value of about 1.5σ chosen in Ref. 6. When choosing the proper value of Δ, we aim to get a good balance between determining the volume of the bubble with good accuracy and avoiding to create small cells not easily assignable as liquid or vapour-like (see Appendices B 1 and B 3): in our study, we set $\Delta = L/19$ for a system of $N = 500$ molecules.

Step 2: Once we have allocated all cells of the grid, *we cluster vapour-like cells into bubbles and identify the bubble with the largest volume as the local order parameter,* $V^{\mathrm{M}}_{\mathrm{bubble}}$, where the superscript "M" indicates the M-method. A typical bubble obtained with the M-method is represented in Figure 3.

By construction, the M-method satisfies the requirements 1 and 2 stated in the Introduction. It is also computationally efficient, thus fulfilling requirement 4 (see Appendix C).

IV. RESULTS

In this section, we first use the two methods introduced ($V^{\mathrm{V}}_{\mathrm{bubble}}$ and $V^{\mathrm{M}}_{\mathrm{bubble}}$) to analyze trajectories of water at ambient

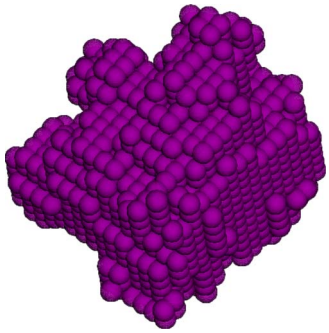

FIG. 3. Typical post-critical bubble obtained with the M-method. The purple spheres represent the vapour-like cells of the grid. We find that the detected bubbles are roughly spherical for the various degrees of metastability investigated in this work, both for the V- and the M-method.

conditions and determine the distribution of bubble volumes. Then, we study trajectories where spontaneous cavitation occurs and compare the estimates for the volume of the largest bubble obtained using both methods. For this case, we compute the nucleation rate and elucidate nucleation properties using the MFPT-formalism. After that, we choose a thermodynamic state where nucleation is a rare event and compute the nucleation free energy barrier as a function of bubble volume evaluated with both the V- and the M-method and correlate both nucleation free energy barriers.

A. Water at ambient conditions

Liquid water exhibits an open structure due to its tendency to form a hydrogen-bond network. The resulting voids in the liquid can pose a challenge to order parameters for bubble nucleation: if the parameters of the method are ill chosen, these voids can be detected as system-spanning bubbles which mask the emergence of "true" bubbles, for instance, transition states from the liquid to the vapour.

In principle, this can be easily avoided by tuning the parameters of the respective order parameters, namely, the mesh point density and the radius of the exclusion spheres, such that the detection of vapour in the liquid phase becomes extremely unlikely. However, unless done carefully, this comes at the price of a decreased spatial resolution (including a significant increase in the size of the smallest bubble that can be detected by the order parameter) and as such our choices for the parameters represent a compromise between avoiding the detection of system-spanning bubbles and obtaining good spatial resolution.

In order to assess how the V- and the M-method perform in liquid water, we analyze molecular dynamics trajectories of liquid water at ambient conditions. Our results are shown in Fig. 4, where the average frequency of occurrence $\langle n(V_{\text{bubble}})\rangle/\langle V\rangle$ is the average number of times a bubble with a volume between V_{bubble} and $V_{\text{bubble}} + \mathrm{d}V$ occurs per configuration per unit volume over the course of a trajectory.

Based upon these data, we conclude that the detected bubbles are both rare and of very small size, which indicates that the voids in the network-forming liquid are not detected as system-spanning bubbles for the chosen parameters of the respective methods. If one divided the systems into parts with

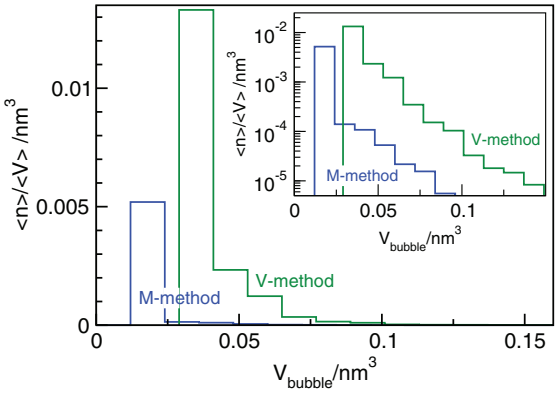

FIG. 4. Frequency of occurrence $\langle n(V_{\text{bubble}})\rangle/\langle V\rangle$ of bubbles of volume V_{bubble} at ambient conditions. The width of each histogram bin is $\mathrm{d}V = 0.012\,\text{nm}^3$. The inset shows $\langle n(V_{\text{bubble}})\rangle/\langle V\rangle$ on a logarithmic scale. The histograms were obtained from an unbiased molecular dynamics simulation at 298 K and 1 bar.

a volume of 1 nm^3 each, at ambient conditions one would find a bubble roughly in one of 180 such cubes when using the M-method (due to its very accurate classification of vapour- and liquid-like cells) and in one of 50 cubes when using the V-method. The typical volume of the smallest bubble detected is 0.017 nm^3 for the M-method and 0.035 nm^3 for the V-method,[54] comparable to the average volume occupied by a water molecule at ambient conditions ($\sim$0.03 nm^3).

B. Comparing the V- and M-methods to detect the volume of the largest bubble

We now consider 200 independent trajectories of spontaneous nucleation in over-stretched water at $T = 280$ K and $p = -2250$ bar and evaluate the volume of the largest bubble with both the V- and M-methods. In Figure 5 we show $V^{\text{M}}_{\text{bubble}}$ versus $V^{\text{V}}_{\text{bubble}}$.

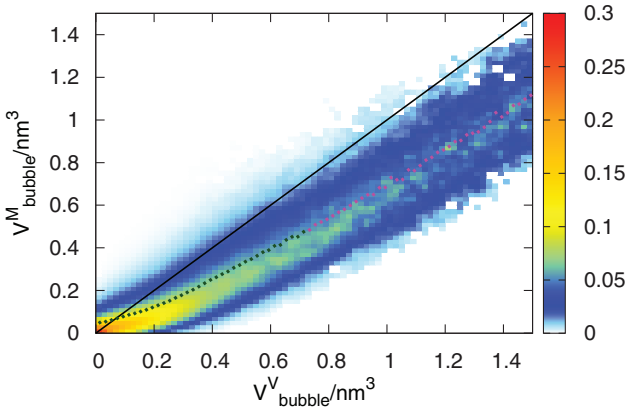

FIG. 5. Largest bubble volume estimates using the M-method (on the y-axis) and the V-method (on the x-axis). The color of each point encodes the probability $P(V^{\text{M}}_{\text{bubble}}|V^{\text{V}}_{\text{bubble}})$ of finding $V^{\text{M}}_{\text{bubble}} \pm 0.01\,\text{nm}^3$ given a value of $V^{\text{V}}_{\text{bubble}} \pm 0.01\,\text{nm}^3$. The dotted line shows the average volume estimate $\langle V^{\text{M}}_{\text{bubble}}(V^{\text{V}}_{\text{bubble}})\rangle$, where green/pink dots correspond to precritical/postcritical bubble volumes (with a critical volume of $\sim$0.7 nm^3, see Table I). The solid black line has a slope of one and is a guide to the eye. The fit parameters for the V-method are $k_1 \approx 0.99$ nm and $k_2 \approx 0.37\,\text{nm}^2$ at these conditions (see Eq. (5)).

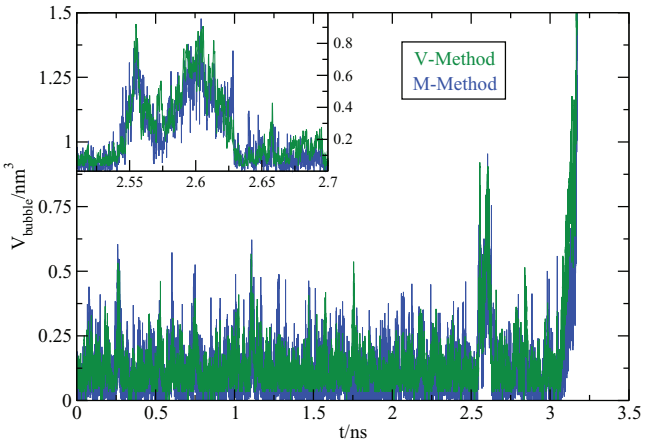

FIG. 6. Comparing the time evolution of the volume of the largest bubble computed with either the M-method (blue) or the V-method (green) for the same cavitating trajectory. The inset shows a zoom around t ∼ 2.60 ns. Notice that the *y*-axis of the inset is on the right-hand side and the units are the same as the main figure.

When comparing the M-method with the V-method, we can distinguish two regimes connected by a crossover region. When the largest bubble is small ($\lesssim$0.3 nm^3), the volume estimates obtained with the two methods are quite different. This is because, by construction, the M-method avoids labelling minute voids in the metastable liquid as bubbles whereas the V-method detects small bubbles with a comparatively higher frequency.

When the volume of the largest bubble exceeds this regime, the two methods give more similar results, even though the V-method yields larger volumes than the M-method on average. The mean value of $V^{\mathrm{M}}_{\mathrm{bubble}}$ at a given value of $V^{\mathrm{V}}_{\mathrm{bubble}}$ (shown as the dotted line in Fig. 5) does not change its shape when passing the critical regime (indicated by the change from green to pink dots) which implies that bubbles on both sides of the free energy barrier have similar structural properties.

For large bubbles ($V^{\mathrm{V}}_{\mathrm{bubble}} \gtrsim 3\,\mathrm{nm}^3$, not shown), the ratio between the average volume estimates obtained by the two methods approaches $\langle V^{\mathrm{M}}_{\mathrm{bubble}}\rangle / V^{\mathrm{V}}_{\mathrm{bubble}} \simeq 0.8$, which allows for an easy conversion between the volume estimates of both methods for large bubbles.

In Fig. 6 we now focus on the time evolution of the volume of the largest bubble during a cavitation trajectory (one of those shown in Fig. 5). While the overall shape of the curves is similar, the volume estimate given by the M-method exhibits fluctuations with a higher frequency than the V-method. As shown in the inset of Fig. 6, both methods are able to detect "almost-critical" V_{bubble} fluctuations, i.e., the growth and shrinking of the largest bubble around its critical size of ∼0.7 nm^3 (Table I). In what follows we will study the effect of the two order parameters on the obtained values of nucleation properties.

C. Spontaneous bubble nucleation in over-stretched water

We first study over-stretched water at thermodynamic conditions where bubble nucleation happens spontaneously in the simulation. In particular, we investigate water bubble nucleation at $T = 280$ K and $p = -2250$ bar as in Ref. 12. For 200 MD trajectories, along which spontaneous cavitation occurs, we compute the volume of the largest bubble and use it as a local order parameter to follow the nucleation mechanism. By means of the MFPT-formalism we compute the nucleation rate, J, and the critical volume, V^*_{bubble}. In order to determine the volume of the largest bubble, we use several approaches: the M- and V-methods proposed in this work as well as the approaches used in Ref. 6 (which uses the WF criterion to distinguish between liquid/vapour molecules) and Ref. 12 (based on the Voronoi tessellation). In the latter algorithm, the bubbles were detected by tracking down interfacial molecules; therefore the volume of those molecules was implicitly included in the final volume of the bubble. Since neither the M-method nor the V-method, by construction, includes the interfacial molecules of the largest bubble as a part of the bubble, when representing the results from Ref. 12 we have re-computed the MFPT removing the volume of the interfacial molecules.

TABLE I. Nucleation time τ_J (ns), Zeldovich factor Z (nm^{-3}), critical volume V^*_{bubble} (nm^3), and nucleation rate J (10^{28} cm^{-3} s^{-1}) at $T = 280$ K and $p = -2250$ bar using different methods to identify the largest bubble. The average volume of the metastable liquid is $\langle V\rangle = 17.23\,\mathrm{nm}^3$.

Method	τ_J	Z	V^*_{bubble}	J
Wang[6]	1.88	4.93	0.11	3.09
Voronoi[12]	1.87	1.48	0.74	3.10
M-method	1.87	1.23	0.74	3.10
V-method	1.85	1.66	0.72	3.13

The calculated values for $\tau(V_{\mathrm{bubble}})$ are presented in Fig. 7. The steepest MFPT curve is obtained with the method of Ref. 6. This method fails to detect small bubbles and severely underestimates the volume of bubbles before the system cavitates, thus underestimating the size of the critical volume. Interestingly, the MFPT curves obtained with the M-method, the Voronoi-based method of Ref. 12 (removing

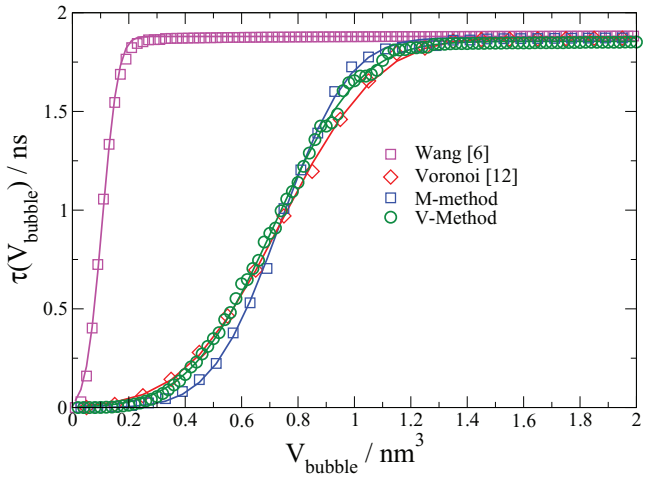

FIG. 7. Mean first-passage time of the largest bubble at $T = 280$ K and $p = -2250$ bar using the order parameters indicated in the legend. The symbols are simulation data and the lines are fits to Eq. (1).

the interfacial molecules) and the V-method have very similar shapes. These three methods allow to identify very small bubbles (as shown by their smooth MFPT curve) and show the inflection points and plateaus at essentially the same values. From the MFPT curves in Fig. 7 we calculate nucleation properties such as τ_J, J, Z, and V^*_{bubble}[31,48] (see Eq. (1)), as reported in Table I.

Notice that the nucleation times τ_J are independent of the method used to identify the largest bubble. Since the nucleation rate only depends on τ_J, the nucleation rate is independent of the method chosen to evaluate the size of the largest bubble.

However, when comparing the Zeldovich factors and the critical bubble volumes V^*_{bubble} provided by the algorithms, we observe some differences because both quantities are sensitive to the method used to evaluate the largest bubble volume. The smallest value of V^*_{bubble} is detected with the method used in Ref. 6, which also gives the largest value for the Zeldovich factor: this implies that the curvature of the free energy barrier is larger than in the other cases. In contrast, the Voronoi-based method from Ref. 12 (removing the interfacial molecules), the M-method and the V-method give very similar values of V^*_{bubble}[55] and Z. This means that not only are the methods able to detect the critical bubble with essentially the same volume, but also that the curvature of the computed free energy barrier is quite similar.

D. Non-spontaneous bubble nucleation in over-stretched water

We now study over-stretched water at thermodynamic conditions ($T = 325$ K and $p = -1500$ bar) where bubble nucleation does not happen spontaneously within the time scales accessible to straightforward molecular dynamics simulation. On the one side, we identify the largest bubble volume with the V-method and use umbrella sampling combined with hybrid Monte Carlo to compute the bubble nucleation free energy barrier. On the other side, we identify the largest bubble volume with the M-method and compute the bubble nucleation free energy barrier using HMC-NpT.[40]

To compute the free energy as a function of the volume V_{bubble} we proceed in the following way. First we carry out a straightforward molecular dynamics simulation at negative pressure and compute $\langle n(V_{\text{bubble}})\rangle$, the average number of bubbles with volume in a narrow interval $[V_{\text{bubble}}, V_{\text{bubble}} + \Delta V_{\text{bubble}}]$. This simulation will yield $\langle n(V_{\text{bubble}})\rangle$ in the range of volumes that are accessible on the time scale of the simulation. To compute the average bubble number $\langle n(V_{\text{bubble}})\rangle$ for larger volumes V_{bubble}, we carry out umbrella sampling simulations with a bias on the volume of the largest bubble. The average bubble numbers obtained in such biased simulations are then conjoined with the result of the straightforward molecular dynamics simulation, yielding the average bubble number $\langle n(V_{\text{bubble}})\rangle$ over a wide range of bubble volumes extending beyond the critical volume.

For large bubble volumes, the probability to find more than one bubble of a given volume at the same time becomes negligible. Accordingly, in this regime, the probability $P(V_{\text{bubble}})\,\Delta V_{\text{bubble}}$ to find a bubble with volume in the interval $[V_{\text{bubble}}, V_{\text{bubble}} + \Delta V_{\text{bubble}}]$ is given by $P(V_{\text{bubble}})\,\Delta V_{\text{bubble}} = \langle n(V_{\text{bubble}})\rangle$. Note that the probability density $P(V_{\text{bubble}})$ defined in this way is independent of the interval width ΔV_{bubble} used for the calculation of the average bubble number $\langle n(V_{\text{bubble}})\rangle$. At negative pressures, the probability density $P(V_{\text{bubble}})$ has a minimum at V^*_{bubble}, the critical bubble volume. According to classical nucleation theory, the rate at which cavitation occurs in a system of total volume $\langle V\rangle$ is proportional to the probability density $P(V^*_{\text{bubble}})$ of finding a bubble of critical volume V^*_{bubble} in the system. Hence, the nucleation rate, which quantifies the number of nucleation events per unit time and unit volume, is proportional to $P(V^*_{\text{bubble}})/\langle V\rangle$. It makes therefore sense to define the free energy in a system of unit volume as

$$F(V_{\text{bubble}}) = -k_B T \ln\left[\frac{V_0^2\langle n(V_{\text{bubble}})\rangle}{\langle V\rangle \Delta V_{\text{bubble}}}\right], \tag{6}$$

where V_0 is a constant determining the unit volume in which the frequency of occurrence of bubbles is measured (in our case, $V_0 = 1\,\text{nm}^3$).

Then, $F(V^*_{\text{bubble}})$ corresponds to the nucleation barrier that needs to be crossed in order to form bubbles of supercritical size.[49] Note that this definition of the nucleation free energy ensures that the nucleation barrier $F(V^*_{\text{bubble}})$ reflects the probability of critical bubbles and does not depend on the probability distribution of small bubbles, which should not affect the nucleation rate. The free energy $F(V_{\text{bubble}})$ obtained using the M-method and the V-method is shown in Fig. 8.

The estimates for the free energy barrier heights obtained using the M-method and the V-method, respectively, differ by about $2\,k_B T$, which constitutes reasonable agreement. At the same time, the estimates for the critical volume at the top of the barrier differ as one would expect: the critical volume obtained using the V-method is $0.36\,\text{nm}^3$(19%) larger than the volume yielded by the M-method. These results are consistent with our findings in the case of lower barriers, namely, that on average the V-method gives higher estimates for the bubble volumes than the M-method (see Fig. 5). Compared to the predictions of classical nucleation theory, the obtained values for the volume of the critical bubble are in reasonable agreement ($V^{*\text{CNT}}_{\text{bubble}} = 2.8\,\text{nm}^3$) while classical nucleation theory overestimates the height of the free energy barrier

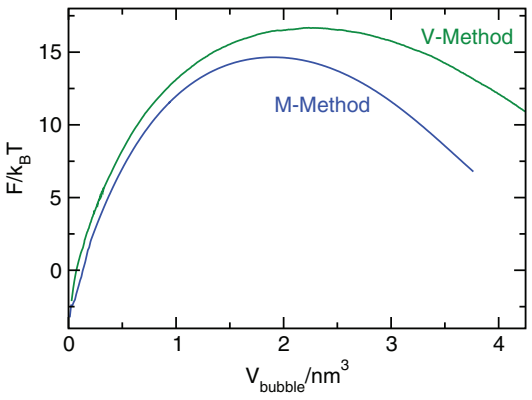

FIG. 8. Free energy F as a function of the bubble volume for the V- and M-methods at $T = 325$ K and $p = -1500$ bar.

($G^{*\,\mathrm{CNT}} = 43.6\,k_\mathrm{B}T$)[50,56] significantly, in agreement with previous simulation studies of nucleation (see, for instance, Ref. 41).

V. DISCUSSION AND CONCLUSIONS

In this paper we present two accurate and efficient grid-based methods to identify bubbles in a network-forming liquid: the M-method and the V-method. While both methods were built upon the grid-based approach in Ref. 6, they strive to improve different aspects of the detection of bubbles in water.

The M-method introduces an accurate approach to identify vapour-like cells in the liquid. In this method, cells are labeled as either liquid- or vapour-like using information about nearest- and next-nearest neighbour cells in an effort to minimize the incorrect detection of small voids always present in network-forming liquids. This implies that the method, by taking into account additional neighbour shells, can be tuned to achieve spatial resolutions that are not accessible to straightforward grid-based approaches without wrongly detecting percolation of voids in the metastable liquid.

The V-method, on the other hand, is focused on obtaining a physically transparent volume estimate that is independent of arbitrary parameter choices such as mesh resolution or the radius of exclusion spheres. This is achieved by a calibration procedure that maps the bubble volume detected in the simulation onto the average change in system volume caused by bubbles of that size. Due to this calibration, the estimated bubble volume is thermodynamically consistent and pV_bubble corresponds to the mechanical work associated with the formation of the bubble. This calibration procedure can also be used to improve other local order parameters when studying nucleation.

Both the M-method and the V-method yield the volume of the largest bubble, which can be used as a local order parameter. This allows one to directly track the evolution of bubbles and analyze their properties in large systems where a global order parameter would not be able to distinguish the emergence of bubbles from fluctuations in the metastable liquid. These order parameters achieve the goal of detecting the volume and shape of bubbles in the liquid with high spatial resolution without constraining the evolving bubbles to a particular shape and are computationally inexpensive (see Appendix C).

When simulating water at ambient conditions, both methods detect bubbles only rarely. Comparing the two methods, owing to the M-method's extremely accurate detection of vapour-like cells, the bubbles detected using the M-method are even less frequent and the volume estimates for these bubbles are lower than in case of the V-method.

When comparing the nucleation properties obtained by using either method, we find that both yield similar results. Under conditions where cavitation occurs spontaneously in a straightforward molecular dynamics simulation, the estimates obtained for the volume of the critical cluster are in excellent agreement between the two methods (see Table I). Under these conditions, the estimates obtained by the methods also agree very well with a Voronoi polyhedra analysis, which is a very precise but computationally expensive method for bubble detection.[12]

Closer to coexistence, when cavitation is extremely unlikely on timescales accessible in simulation, we use umbrella sampling to obtain an estimate for the free energy $F(V_\mathrm{bubble})$ as a function of bubble volume. Here, the estimate for the size of the critical bubble obtained by using the V-method yields, due to the volume calibration procedure, a larger value for the critical cluster and a smaller curvature of the free energy barrier than the M-method. However, the estimate for the height of the free energy barrier, which largely determines the experimentally accessible cavitation rate, obtained by the two methods is in good agreement. In the near future, we plan to use the methods introduced here to perform a thorough study of cavitation of water at negative pressure, evaluating the nucleation rate at several thermodynamic conditions, analysing the nucleation pathways, and characterising the properties of the growing bubble.

ACKNOWLEDGMENTS

The work of M.A.G., J.L.A., J.L.F.A., and C.V. was funded by Grant No. FIS2013/43209-P of the MEC and the Marie Curie Integration Grant No. PCIG-GA-2011-303941 (ANISOKINEQ). C.V. also acknowledges financial support from a Juan de La Cierva Fellowship. The work of G.M., P.G., and C.D. was supported by the Austrian Science Foundation (FWF) under Grant No. P24681-N20. P.G. also acknowledges financial support from FWF Grant No. P22087-N16. Calculations performed at the University of Vienna were carried out on the Vienna Scientific Cluster (VSC). We acknowledge Carlos Vega for helpful discussions and a critical reading of the paper. We thank Michael Grünwald and Jaffar Hasnain for useful discussions.

APPENDIX A: DISTINGUISHING BETWEEN LIQUID AND VAPOUR-LIKE MOLECULES

In our paper, when using the V-method we have classified liquid/vapour molecules with the ten Wolde–Frenkel (WF) criterion,[41] and when using the M-method with the hydrogen bond (HB) criterion.[42] In the following, we will demonstrate that these criteria are suitable for the task of distinguishing between the liquid and the vapour phase of water. As a reference, we also include an alternative criterion based on a tetrahedal order parameter (the q criterion).

The WF criterion has been used by Wang *et al.*[6] to study cavitation in a super-heated Lennard-Jones fluid; it consists of identifying each particle's nearest neighbours using the Stillinger radius r_S (i.e., the first minimum of the radial distribution function (RDF) as a fixed cutoff distance): for water, we will only deal with the oxygen–oxygen RDF. Therefore, a water molecule is defined as vapour-like if it has no neighbours within its Stillinger radius.

The HB criterion is based on the number of donor hydrogen-bonds per molecule.[42]

According to Ref. 42 (see Fig. 9), water molecule 1 donates a hydrogen bond to molecule 2 if (1) the oxygen$_1$–

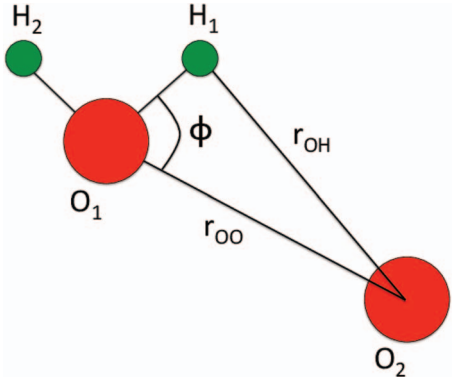

FIG. 9. Sketch of the hydrogen bond criterium from Ref. 42.

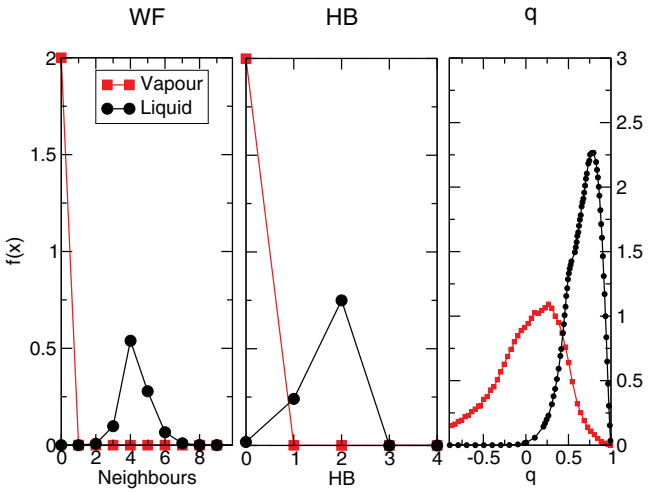

FIG. 10. First-neighbours distribution computed with the WF (left panel) and with the q criterion (right panel) and distribution of donor/acceptor H-bonded molecules computed with the HB criterion (middle panel) for a system consisting of liquid water (black) or vapour (red).

oxygen$_2$ distance (r_{OO}) is smaller than the first minimum of the oxygen–oxygen RDF, (2) the hydrogen$_1$–oxygen$_2$ distance (r_{OH}) between the closest hydrogen of the "donor" molecule and the oxygen of the "acceptor" molecule is smaller than the first minimum of the oxygen–hydrogen RDF, and (3) the angle ϕ between the oxygen$_1$–hydrogen$_1$ vector and the oxygen$_1$–oxygen$_2$ vector is smaller than 30°.[51] Only if all three geometrical conditions are fulfilled, we consider molecule 1 as donating a hydrogen bond to molecule 2.

To distinguish between liquid and vapour molecules in either the M- or the V-method one could also use the tetrahedal order parameter q of Refs. 43 and 44. The q criterion, proposed in Ref. 44, consists of computing for every molecule i the quantity $q_i = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}[cos(\theta_{ijk}) + \frac{1}{3}]^2$, where θ_{ijk} is the angle formed by the oxygens of molecules i and two of its four nearest neighbours j and k (the molecule i being at the vertex of the angle). The q-parameter takes a value of 1 when the four nearest neighbours are in a perfect tetrahedral arrangement around the central one.

We now compute the first-neighbours distribution using either the WF, HB, or q criterion for the liquid and the vapour phase and present our results in Fig. 10. (The thermodynamic conditions at which we have performed these analysis are the following: the liquid, with a density of $\rho = 0.979$ g/cm^3, has been equilibrated at $T = 298$ K and $p = 1$ bar and the vapour, with a density of $\rho = 0.01$ g/cm^3, at $T = 570$ K and $p = 53.45$ bar.)

Using the WF criterion, from the nearest neighbours distribution of the liquid and vapour system, we can identify as liquid-like all particles having a number of neighbours equal to or larger than one and vapour-like otherwise (see Fig. 10, WF).

From the distribution of H-bonded molecules in the liquid and vapour system, we can identify as liquid-like all particles that donate at least one hydrogen bond and vapour molecules are detected by their lack of *donor* hydrogen bonds (see Fig. 10, HB). As expected, the maximum number of donated H-bonds observed is two. In both cases, there is practically no overlap between the liquid and vapour distributions: every vapour molecule lacks hydrogen bonds, whereas most of the liquid molecules have two donor hydrogen bonds, as expected for stable water.

In the q case (see Fig. 10, q), computing the distribution of tetrahedrally ordered molecules in a vapour and a liquid, we observe that the probability distributions overlap considerably.

Given that the distributions are not overlapping, both the HB and the WF criterion are equally well suited to distinguish between vapour and liquid-like molecules, whereas the q criterion seems not to be the best choice due to the clear overlap.

APPENDIX B: ASSIGNING EMPTY CELLS TO BE LIQUID OR VAPOUR-LIKE IN THE M-METHOD

1. Initial labelling of the cells

In order to label each cell we proceed as follows. To start with, we label as liquid-like the cells beneath liquid particles and vapour-like the ones beneath vapour particles. Therefore, depending on the grid mesh L/δ (where L is the simulation box edge), we are left with a number of cells that do not clearly belong either to the liquid or to the vapour. At this initial stage, we make the assumption that empty cells are liquid-like. In what follows, we describe an algorithm to properly label them and correct for the initial guess.

2. The best choice for the threshold of the number of empty/vapour neighbours in the first and second cell

In our study, to identify an empty cell as vapour-like, we check that not only at least 7 of its *first neighbours* are neighbours to each other and vapour-like *but also* 7 of its *second neighbours*. In general, in order to choose the threshold for the first and second neighbour cells, we want to make sure that we avoid identifying empty cells in thermodynamically stable water.

Figure 11 represents the distribution of having the largest bubble of a given volume in the case of thermodynamically stable liquid water at 1 bar and 298 K. The largest bubble has been detected with the HB-criterion, assuming $\delta = 19$, and

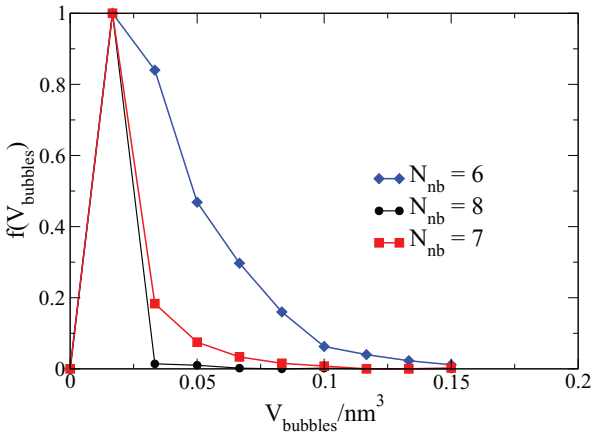

FIG. 11. Distribution of the size of the largest bubble in a thermodynamically stable liquid at 1 bar and 298 K for different numbers of N_{nb} and $\delta = 19$.

different values of the threshold for the number of first and second neighbours N_{nb} (we are always assuming that these two thresholds are the same).

As shown in the figure, all distributions are peaked around 0.017 nm^3, comparable to the volume of a water molecule ($\sim$0.030 nm^3) at these conditions: this means that with the largest probability the largest vapour bubble has the size of a water molecule. However, decreasing the threshold value of N_{nb}, the largest bubble can reach much larger volumes: this corresponds to detecting voids characteristic of the network-forming liquid as vapour bubbles (for instance, in the case $N_{nb} = 6$). Therefore, the best choice for N_{nb} corresponds to the minimum value needed not to observe a distribution of large bubbles (larger than one water molecule in a liquid) in thermodynamically stable water. In our study, this condition is met by the choice of having at least 7 empty cells that are neighbours in the first cell together with 7 that are neighbours in the second one.

3. The best choice for L/δ

As stated in Step 2 of the M-method, when setting the optimal value of the grid's mesh size (L/δ) one has to satisfy two conditions. On the one side, the size of the cell grid must be as small as possible in order to reduce the error coming from the discretization of the bubble's volume. On the other side, the size of the cell grid must be as large as possible to avoid to label the "holes" which exist (even in its thermodynamically stable state) in a network-forming liquid such as water, as vapour cells. Figure 12 represents a two-dimensional sketch of the grid of liquid water molecules (represented by their oxygens) in three possible scenarios characterised by a different grid mesh L/δ: (a) $0.5\sigma < L/\delta$; (b) $0.33\sigma < L/\delta < 0.5\sigma$; and (c) $L/\delta < 0.33\sigma$. The distance between the molecules is slightly larger than σ (corresponding to the first minimum of the oxygen–oxygen pair distribution function).

To explain the criteria for the choice of the threshold of the number of first and second neighbours in Fig. 12 (2D grid), 3 has been used as the minimum number of empty/vapour-like neighbours. When we evaluate one empty cell, we check if this cell has at least 3 first empty/vapour

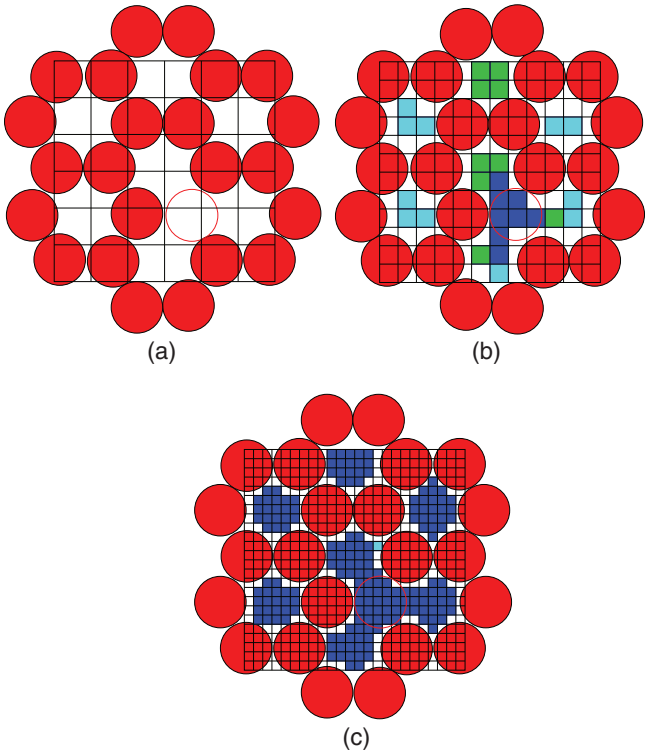

FIG. 12. Two-dimensional grid with usual liquid water-like structure. Solid red circles are oxygens and the empty red circle is one removed particle. (a) $0.5\sigma < L/\delta$; (b) $0.33\sigma < L/\delta < 0.5\sigma$; and (c) $L/\delta < 0.33\sigma$. Blue cells are the final bubble cells, cyan cells and green cells are the empty transformed into liquid cells after applying the first neighbours criteria and second neighbours criteria, respectively. (a) large, (b) medium, and (c) small.

neighbours that are also neighbours to each other and 3 or more second neighbours of the same type. Thus, we can label this cell as vapour-like under first and second neighbour criteria.

As in Ref. 6, in case (a) we determine whether empty cells are liquid- or vapour-like. In this case, it is not necessary to apply our neighbour criteria because the grid size is large enough to allow to define every cell as vapour or liquid cells (there are no empty cells). However, if we find empty cells, we label all cells using only the first neighbour criterion. As shown in the figure, when a void as large as one particle appears in the system, it is difficult to evaluate the bubble volume because we cannot detect empty cells. The drawback is that this choice strongly penalises interfacial molecules, i.e., molecules with a small number of vapour neighbours.

Case (b) allows to better resolve the volume of the bubble. In this case, to assign empty cells we need to recur not only to the first neighbours, but also to the second ones. If we consider only the first-neighbour cells, we will label the empty cells between liquid particles as vapour-like (given that most of their first neighbours are empty or vapour-like). Whereas, considering also the second neighbour cells, we will label the empty cells as liquid-like (given that their second neighbours are not vapour-like).

Compared to the previous cases, the grid in case (c) is the smallest. This situation can give more accurate estimates for the volume, but one would need to recur to the third neighbour cells in order to determine the nature of each empty cell. Even

TABLE II. CPU (real) time (in minutes) for the M-method, the V-method, and the Voronoi-method. Each algorithm was used to analyze the same 10 000 frames of a cavitating trajectory. Note that the presented results are highly implementation dependent and should be considered only as indicative.

Method	CPU-time
M-method	18
V-method	1.7
Voronoi[12]	452

though in this way we improve the accuracy of the definition of the bubble's volume, this method is CPU-time consuming.

Therefore, we conclude that the best compromise in terms of CPU-time and precision in determining the bubbles' volume is case (b). Labeling empty cells according to their type of first and second neighbours corrects for the cells' initial assignation. This allows us to find the real bubble not the natural voids always present in a network-forming liquid. Moreover, having properly detected all empty cells has also the advantage to be able to clearly distinguish between two neighbouring bubbles (without detecting them as a merged single bubble[6]).

APPENDIX C: COMPARING THE COMPUTATIONAL EFFICIENCY OF THE DIFFERENT METHODS

In order to compare the CPU-time needed to compute the volume of the largest bubble using the V-method, the M-method, and the Voronoi-method (as in Ref. 12), we select one of the spontaneously nucleating trajectories (that cavitated in 13 ns) and run each of the algorithms on 10 000 frames of this trajectory.

The benchmark was performed on a computer equipped with an Intel Xeon X5680 @ 3.33 GHz processor. The operating system was 64-bit GNU/Linux and each algorithm was running single-threaded. Results are presented in Table II.

The Voronoi-method is clearly the most computationally expensive algorithm of the three, being two/three orders of magnitude slower than the M-/V-method. When comparing these two methods, the V-method is ten times faster than the M-method. The main reason is that within the M-method, one has to check both the first and the second neighbour shells in order to assign a grid cell as liquid- or vapour-like, whereas the V-method simply checks for molecules within a cutoff distance.

When interpreting the CPU-time estimates given in Table II, one should keep in mind that the estimates include the time it takes to initially read the configurations from the hard drive. The calibration procedure for the V-method is not included in the CPU-time estimate.

In summary, grid-based methods are not only significantly easier to implement than the Voronoi-method but also computationally more efficient, making them a convenient choice when studying bubble nucleation.

[1]H. Cochard, C. R. Phys. **7**, 1018 (2006).
[2]T. D. Wheeler and A. D. Stroock, Nature (London) **455**, 208 (2008).
[3]O. Vincent, P. Marmottant, P. A. Quinto-Su, and C.-D. Ohl, Phys. Rev. Lett. **108**, 184502 (2012).
[4]K. Ohsaka and E. H. Trinh, Appl. Phys. Lett. **73**, 129 (1998).
[5]D. Yu, B. Liu, and B. Wang, Ultrason. Sonochem. **19**, 459 (2012).
[6]Z.-J. Wang, C. Valeriani, and D. Frenkel, J. Phys. Chem. B **113**, 3776 (2009).
[7]K. Torabi and D. S. Corti, J. Phys. Chem. B **117**, 12479 (2013).
[8]K. Torabi and D. S. Corti, J. Phys. Chem. B **117**, 12491 (2013).
[9]V. K. Shen and P. G. Debenedetti, J. Chem. Phys. **111**, 3581 (1999).
[10]S. L. Meadley and F. A. Escobedo, J. Chem. Phys. **137**, 074109 (2012).
[11]D. Zahn, Phys. Rev. Lett. **93**, 227801 (2004).
[12]J. L. F. Abascal, M. A. Gonzalez, J. L. Aragones, and C. Valeriani, J. Chem. Phys. **138**, 084508 (2013).
[13]J. L. F. Abascal and C. Vega, J. Chem. Phys. **123**, 234505 (2005).
[14]C. Vega and J. L. F. Abascal, Phys. Chem. Chem. Phys. **13**, 19663 (2011).
[15]C. Vega, J. L. F. Abascal, and I. Nezbeda, J. Chem. Phys. **125**, 034503 (2006).
[16]J. L. Aragones, L. G. MacDowell, I. J. Siepmann, and C. Vega, Phys. Rev. Lett. **107**, 155702 (2011).
[17]C. Vega and E. de Miguel, J. Chem. Phys. **126**, 154707 (2007).
[18]P. Geiger and C. Dellago, J. Chem. Phys. **139**, 164105 (2013).
[19]D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, J. Comput. Chem. **26**, 1701 (2005).
[20]H. Kamberaj, R. J. Low, and M. P. Neal, J. Chem. Phys. **122**, 224114 (2005).
[21]T. F. Miller, M. Eleftheriou, P. Pattnaik, A. Ndirango, D. Newns, and G. J. Martyna, J. Chem. Phys. **116**, 8649 (2002).
[22]G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, Mol. Phys. **87**, 1117 (1996).
[23]S. Nosé, J. Chem. Phys. **81**, 511 (1984).
[24]W. J. Hoover, Phys. Rev. A **31**, 1695 (1985).
[25]H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).
[26]H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
[27]G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[28]M. Parrinello and A. Rahman, J. Appl. Phys. **52**, 7182 (1981).
[29]J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[30]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[31]J. Wedekind, R. Strey, and D. Reguera, J. Chem. Phys. **126**, 134103 (2007).
[32]M. A. Gonzalez, Master's thesis, Fac. Ciencias Quimicas, Univ. Complutense, Madrid, 2011.
[33]G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[34]S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[35]B. Mehlig, D. W. Heermann, and B. M. Forrest, Phys. Rev. B **45**, 679 (1992).
[36]I. P. Omelyan, I. M. Mryglod, and R. Folk, Phys. Rev. E **65**, 056706 (2002).
[37]M. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. N. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[38]C. J. Geyer and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).
[39]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[40]M. A. Gonzalez, E. Sanz, C. MacBride, J. L. F. Abascal, C. Vega, and C. Valeriani, "Nucleation free-energy barriers using molecular dynamics umbrella sampling," Phys. Chem. Chem. Phys. (in press).
[41]P. R. ten Wolde and D. Frenkel, J. Chem. Phys. **109**, 9901 (1998).
[42]A. Luzar and D. Chandler, J. Chem. Phys. **98**, 8160 (1993).
[43]P. L. Chau and A. J. Hardwick, Mol. Phys. **93**, 511 (1998).
[44]J. R. Errington and P. G. Debenedetti, Nature (London) **409**, 318 (2001).
[45]D. Kashchiev, J. Chem. Phys. **76**, 5098 (1982).
[46]M. E. M. Azouzi, C. Ramboz, J.-F. Lenain, and F. Caupin, Nat. Phys. **9**, 38 (2013).
[47]M. Grünwald and C. Dellago, J. Chem. Phys. **131**, 164116 (2009).
[48]J. Wedekind, G. Chkonia, J. Wölk, R. Strey, and D. Reguera, J. Chem. Phys. **131**, 114506 (2009).
[49]L. Maibaum, Phys. Rev. Lett. **101**, 019601 (2008).
[50]M. Blander and J. L. Katz, AICHE J. **21**, 833 (1975).
[51]E. Guàrdia, J. Martí, L. García-Tarrés, and D. Laria, J. Mol. Liq. **117**, 63 (2005).
[52]The relaxation time of the barostat $\nu^{-1} = 3\,\text{ps}$ is chosen to approximate the length of the simulation box divided by the speed of sound $c = \sqrt{K/\rho}$, where $K = \rho\,\partial P/\partial\rho \approx 2\text{GPa}$ is the bulk modulus of TIP4P/2005 water in the studied pressure range.
[53]The reason for the difference in volume between a cluster of unoccupied grid cubes and the corresponding average change in system volume lies in

the choice of the exclusion sphere radius (in our case, r_S). This radius is chosen sufficiently large as to prevent the erroneous detection of system-spanning bubbles in the homogeneous metastable liquid. At the liquid–bubble interface, this large diameter results in a "gap" between the surface of the liquid and the detected bubble, thereby underestimating the true volume of the bubble.

[54]Note that we used the V-method in the parametrisation obtained from the fit shown in Fig. 2, which was computed at negative pressures. Calibrating the V-method at ambient conditions would result in a (likely small) change in the volume of the detected bubbles but would not affect their frequency of occurrence.

[55]The very similar estimates for V^*_{bubble} appear to be at odds with the data presented in Fig. 5, which show that the V-method gives larger volume estimates than the M-methods on average. However, since the MFPT-analysis is based on the dynamics of the system, the obtained estimate depends not only on the mean of the order parameter used to track bubble nucleation but also on its fluctuations. In particular, a higher frequency of fluctuations in the order parameter leads to a higher estimate for V^*_{bubble}. As Fig. 6 shows, estimates obtained with the M-method exhibit larger fluctuations than estimates by the V-method, which, when V^*_{bubble} is computed from MFPT analysis, cancels the difference in the average bubble volume detected by the two methods, leading to a virtually identical estimate for V^*_{bubble}.

[56]Since classical nucleation theory predicts the reversible work for a single nucleus to grow to critical size, its prediction does not take the volume of the system into account. In order to make the CNT prediction comparable to $F(V^*_{\text{bubble}})$ as defined in Eq. (6), we follow Ref. 50 where the number density ρ in the liquid is chosen as the (somewhat arbitrary) prefactor entering the nucleation rates. Thus, the CNT prediction for the barrier height is given by $G^{*\,\text{CNT}} = 16\pi\gamma^3/3\Delta p^2 - \ln(\rho/\rho_0)$, where $\gamma = 65.7\,\text{mJ/m}^2$ is the surface tension for TIP4P2005 water at 325 K[17] and $\rho_0 = 1/\text{nm}^3$.