



**QUEEN'S  
UNIVERSITY  
BELFAST**

## The Gibbs free energy of homogeneous nucleation: from atomistic nuclei to the planar limit

Cheng, B., Tribello, G., & Ceriotti, M. (2017). The Gibbs free energy of homogeneous nucleation: from atomistic nuclei to the planar limit. *Journal of Chemical Physics*, 14(10), [104707]. <https://doi.org/10.1063/1.4997180>

### Published in:

Journal of Chemical Physics

### Document Version:

Peer reviewed version

### Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

### Publisher rights

Copyright 2017 AIP.

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

### General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

# The Gibbs free energy of homogeneous nucleation: from atomistic nuclei to the planar limit

Bingqing Cheng,<sup>1,\*</sup> Gareth A. Tribello,<sup>2</sup> and Michele Ceriotti<sup>1</sup>

<sup>1</sup>*Laboratory of Computational Science and Modeling, Institute of Materials,  
École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland*

<sup>2</sup>*Atomistic Simulation Centre, School of Mathematics and Physics,  
Queen's University Belfast, Belfast, BT7 1NN*

(Dated: August 28, 2017)

In this paper we discuss how the information contained in atomistic simulations of homogeneous nucleation should be used when fitting the parameters in macroscopic nucleation models. We show how the number of solid and liquid atoms in such simulations can be determined unambiguously by using a Gibbs dividing surface and how the free energy as a function of the number of solid atoms in the nucleus can thus be extracted. We then show that the parameters (the chemical potential, the interfacial free energy, and a Tolman correction) of a model based on classical nucleation theory can be fit using the information contained in these free-energy profiles but that the parameters in such models are highly correlated. This correlation is unfortunate as it ensures that small errors in the computed free energy surface can give rise to large errors in the extrapolated properties of the fitted model. To resolve this problem we thus propose a method for fitting macroscopic nucleation models that uses simulations of planar interfaces and simulations of three-dimensional nuclei in tandem. We show that when the chemical potentials and the interface energy are pinned to their planar-interface values, more precise estimates for the Tolman length are obtained. Extrapolating the free energy profile obtained from small simulation boxes to larger nuclei is thus more reliable.

## INTRODUCTION

Nucleation is an important component of many technological and natural processes, including metal casting, the assembly of microtubules in cells and the formation of water droplets and ice crystals in the atmosphere [1–3]. In spite of its generic importance, however, our understanding of nucleation kinetics is somewhat unsatisfactory, even for the simplest cases of homogeneous nucleation. The main reason for this is that the experimental determination of nucleation rates is challenging. Many of the available analytical techniques can only be used to perform *ex situ* analysis, which cannot fully capture the dynamical processes that take place during nucleation.

A considerable number of studies in the last two decades have thus used atomistic simulation to study homogeneous nucleation [2, 4–10]. Much of this computational effort has concentrated on investigating how the free energy changes with cluster size  $G(n)$  because the homogeneous nucleation barrier  $G^*$  is the maximum of  $G(n)$  and because the nucleation rate can be determined from  $G^*$  by using transition state theory or the Bennett-Chandler approach [4, 6, 11]. Researchers have thus been able to perform atomistic simulations in which solid critical nuclei containing hundreds of atoms have been observed to form from the melt [4–6, 8, 9].

When analyzing atomistic simulations of nucleation, macroscopic models that provide analytical expressions for the nucleation free energy profile  $G(n)$  can be extremely useful as they can be used to extrapolate the  $G(n)$  obtained for the sub-critical nuclei which form in atomistic simulations so as to obtain the free energies of larger nuclei. This is often essential when predicting  $G^*$ , and when interpolating between results obtained at different temperatures. One simple model which is commonly used to rationalize nucleation is classical

nucleation theory (CNT). This theory assumes that  $G(n)$  can be expressed as the sum of a bulk and a surface term, i.e.

$$G(n) = \mu n + \gamma \Omega v^{\frac{2}{3}} n^{\frac{2}{3}}, \quad (1)$$

Here  $\mu$  is the chemical potential difference between the stable and the metastable phases,  $\gamma$  is the area specific interfacial free energy,  $v$  is the molar volume of the bulk stable phase, and  $\Omega$  is a geometrical constant (e.g.  $(36\pi)^{\frac{1}{3}}$  for a spherical nucleus).

CNT has been shown to give rise to significant systematic errors for all sizes of the nucleus [9]. A large part of problem is that the number of atoms  $n$  in the cluster is not a well-defined quantity [6, 8, 9, 12]. Different definitions of  $n$  inevitably affect the calculated free energy profile  $G(n)$ , which forces one to ask which definition of  $n$  is most appropriate within the CNT framework. Another part of the problem is that the CNT expression in Eqn. (1) is generally used as a fitting model for  $G(n)$ , because the various physical quantities that enter this expression - the chemical potential  $\mu$  and the interfacial free energy  $\gamma$  - cannot be calculated directly from homogeneous nucleation simulations. As a consequence, when the results from simulations deviate from the predictions of CNT, it is not clear which assumption within the model is the main culprit. It is thus difficult to add additional correction terms to the model to compensate for the missing elements.

An alternative method for probing nucleation involves investigating planar interfaces, which resemble the surfaces of a large nucleus. From the simulations of planar interfaces, the chemical potential as well as the interfacial free energies for different crystallographic directions can be evaluated [12–17]. Various techniques have been developed for computing the interfacial free energy or the stiffness of the planar interfaces at the

coexistence temperature [13–18]. Using the values of  $\gamma$  obtained from such simulations within the framework of CNT is problematic though as nucleation occurs in undercooled conditions. Recently, however, the metadynamics method has been extended so that it also works in out-of-equilibrium conditions away from the melting point [12]. This method thus opens the door to a systematic exploration into how the quantities obtained from simulations of planar interfaces can be used to predict the free energy associated with a three-dimensional nucleus under realistic conditions.

In the present study, we apply our recently introduced thermodynamic framework for defining the nucleation free energy profile  $G(n)$  in an unambiguous way [19]. We then fit the computed  $G(n)$  to the expressions that come from macroscopic nucleation models, using the specific interfacial free energy  $\gamma$  and the chemical potential difference  $\mu$  as fitting parameters. We also show that a higher-order correction – generally referred to as the Tolman term – naturally emerges from our formalism. Finally, in the last part of the paper, we compare the predicted values for  $\gamma$  and  $\mu$  that we obtain from simulations of three-dimensional nuclei against those obtained from the simulations of planar interfaces.

## THEORY

### Gibbs dividing surface in atomistic systems

In order to link the atomistic descriptions and the macroscopic representations of a solid-liquid system, we define a Gibbs dividing surface using an arbitrarily-chosen extensive quantity  $\Phi$  that characterises the state of all the atoms within the system. This extensive quantity might be the volume  $V$ , the entropy  $S$ , [20] or a global collective variable (CV)  $\Phi = \sum_i \phi(i)$  that is constructed by summing the order parameter values  $\phi(i)$  for each of the atoms in the microstate [12, 19]. We define the position of the  $\Phi$ -based Gibbs dividing surface by requiring that there be zero surface excess for the extensive quantity  $\Phi$ . In other words, the real system and a reference system in which the presence of the interface has no effect on the properties of the two bulk phases are setup so they have the same values for the extensive quantity [21, 22]. Thermal fluctuations ensure that the instantaneous value of the extensive quantity  $\Phi$  for this reference system is not fixed, but follows a distribution  $\rho_{\text{ref}}(\Phi|n_s(\Phi), N - n_s(\Phi))$  with a finite width that can be calculated by performing simulations of the two bulk phases [19, 23]. To have zero excess for  $\Phi$  the real system and the reference system, both of which have  $n_s(\Phi)$  solid atoms and  $N - n_s(\Phi)$  liquid atoms, should have the same distribution for  $\Phi$  [19], i.e.

$$\rho_{\text{sl}}(\Phi|n_s(\Phi), N - n_s(\Phi)) \equiv \rho_{\text{ref}}(\Phi|n_s(\Phi), N - n_s(\Phi)). \quad (2)$$

In Ref. [19], we argued that Eqn. (2) can be used to link a free energy profile  $\tilde{G}(\Phi)$ , which is expressed

as a function of an extensive quantity  $\Phi$  for the whole system, with the nucleation free energy for a single solid cluster which is expressed as a function of  $n_s(\Phi)$ . This connection can be made under the reasonable assumption that the probability of observing a single, large sub-critical cluster inside an undercooled bulk liquid can be expressed as  $N_s \exp(-\beta G(n_s(\Phi)))$  when  $n_{\text{cut}} \leq n_s(\Phi) \leq n^*$ . In these expressions  $N_s$  is the number of nucleation sites, which in this work we take to be simply the number of atoms, while  $G(n_s(\Phi))$  is the free energy excess associated with the solid cluster. Furthermore, as discussed in Ref. [19], the value of  $n_{\text{cut}}$  is set so that having two clusters with sizes greater than this threshold is extremely unlikely, which necessary for unambiguously mapping each atomistic configuration to a thermodynamically meaningful cluster size. Using the law of total probability, the probability distribution for  $\Phi$  in such systems follows

$$e^{-\beta \tilde{G}(\Phi)} = \int_{n_{\text{cut}}}^{n^*} dn_s(\Phi) \rho_{\text{sl}}(\Phi|n_s(\Phi), N - n_s(\Phi)) N_s e^{-\beta G(n_s(\Phi))}, \quad (3)$$

This expression is valid for values of  $\Phi$  that satisfy  $\rho_{\text{sl}}(\Phi|n_s(\Phi), N - n_s(\Phi)) \approx 0$  for all  $n_s(\Phi) < n_{\text{cut}}$ . To be clear  $\tilde{G}(\Phi)$  can be directly computed from atomistic simulations of nucleation. Values of  $G(n_s(\Phi))$  for  $n_s(\Phi) > n_{\text{cut}}$  can then be found by inverting Eqn.(3) numerically [19].

### A general formulation of classical nucleation theory

With a  $\Phi$ -based dividing surface, the free energy of the solid-liquid system relative to that of a bulk liquid can be naturally decomposed into a bulk and a surface term:

$$G(n_s(\Phi)) = \mu_{\text{sl}} n_s(\Phi) + \gamma_{\text{sl}}^\Phi A(n_s(\Phi)), \quad (4)$$

Here  $\mu_{\text{sl}} = \mu_s - \mu_l$  is the difference between the per-atom chemical potentials of the solid and liquid phases. The free energy excess associated with the  $\Phi$ -based dividing surface is then the product of a specific energy term  $\gamma_{\text{sl}}^\Phi$  and an extensive area term  $A(n_s(\Phi))$ .

For planar interfaces, the surface area  $A(n_s(\Phi))$  in Eqn. (4) is fixed by the boundary conditions so the specific planar interfacial free energy  $\gamma_{\text{sl}}^\Phi = \gamma_\infty^\Phi$  is a constant that depends on the crystallographic direction of the planar interface. Notice that if the extensive quantity  $\Theta$  is used to define the dividing surface instead of  $\Phi$  the composition of the reference system changes. This is important as a change in the composition of the reference state will affect the value obtained for the interfacial free energy. We quantify this change in composition using:

$$\delta n_s^{\Phi, \Theta} = (n_s(\Theta) - n_s(\Phi))/A \quad (5)$$

and further note that this quantity should be constant as replacing  $\Phi$  with  $\Theta$  shifts the location of the dividing

surface by a fixed amount [12]. Furthermore, when  $\Phi$  is replaced by  $\Theta$  the resulting change in the planar interfacial free energy

$$\Delta\gamma_{\infty}^{\Phi,\Theta} = \gamma_{\infty}^{\Theta} - \gamma_{\infty}^{\Phi} = -\mu_{\text{sl}}\delta n_s^{\Phi,\Theta}, \quad (6)$$

is also a constant, because the total free energy in Eqn. (4) should be unaffected by the change in the extensive quantity [12].

For the curved interfaces around a three-dimensional nucleus, the surface area  $A(n_s(\Phi))$  is not equally well-defined. Classical nucleation theory assumes that the nucleus has the same density as the bulk solid, which is what ensures that  $A = \Omega v_s^{\frac{2}{3}} n_s^{\frac{2}{3}}$ . Critically, however, this bulk density assumption is only valid when the equimolar dividing surface is used to define the reference state, which implies that only quantities determined using a dividing surface with no excess volume should be inserted into Eqn. (4). This equation should thus read:

$$G(n_s(V)) = \mu_{\text{sl}}n_s(V) + \gamma_{\text{sl}}^V \Omega v_s^{\frac{2}{3}} n_s^{\frac{2}{3}}(V). \quad (7)$$

Even when the reference state is defined in a manner consistent with this assumption about the bulk density of the nucleus, however, one still has to include a term that incorporates the surface excess free-energies' dependence on curvature into  $\gamma_{\text{sl}}^V$  [22, 24]. This dependence of  $\gamma_{\text{sl}}^V$  on the effective radius of the nucleus  $R = (3/4\pi)^{\frac{1}{3}} v_s^{\frac{1}{3}} n_s^{\frac{1}{3}}(V)$  can be written as

$$\gamma_{\text{sl}}^V(R) = \gamma_{\infty}^V(1 - 2\delta/R + \mathcal{O}(1/R^2)). \quad (8)$$

In this expression  $\delta$  is the Tolman length - a quantity that measures the difference between the locations of the equimolar dividing surface and the surface of tension at the planar limit.  $\gamma_{\infty}^V$ , meanwhile, is the specific interfacial free energy associated with the equimolar surface in the planar limit. Taking only the leading terms in Eqn. (8), Eqn. (7) can be rewritten as

$$G(n_s(V)) \approx \mu_{\text{sl}}n_s(V) + \gamma_{\infty}^V \Omega v_s^{\frac{2}{3}} n_s^{\frac{2}{3}}(V)(1 - \epsilon n_s^{-\frac{1}{3}}(V)), \quad (9)$$

where  $\epsilon = (32\pi/3)^{\frac{1}{3}} v_s^{-\frac{1}{3}} \delta$  is a constant.

At the atomic scale, due to the magnitude of fluctuations, it is difficult to distinguish between the different phases based only on the difference in molar volume. Hence, it is often not convenient to use this quantity to determine the number of solid particles in a simulation. One can use a different extensive quantity to determine the location of the dividing surface as it is possible to convert the free-energy profile obtained using one dividing surface to the result that would have been obtained for a different dividing surface using Eq. (5) [12]. It would be useful, however, to derive a CNT expression based on an arbitrary choice of reference thermodynamic variable  $\Phi$ . To derive such an expression, we first assumed that the difference between the location of the  $\Phi$ -based dividing surface and the location of the equimolar dividing surface is

much smaller than the effective radius of the nucleus  $R$ , as this ensures that the difference between  $n_s(\Phi)$  and  $n_s(V)$  can be approximated using

$$n_s(V) - n_s(\Phi) \approx \delta n_s^{\Phi,V} \Omega v_s^{\frac{2}{3}} n_s^{\frac{2}{3}}(V), \quad (10)$$

where  $\delta n_s^{\Phi,V}$  is the constant that appears in Eqn. (5) and (6) and where  $\Omega v_s^{\frac{2}{3}} n_s^{\frac{2}{3}}(V)$  is a measure of the surface area of the nucleus. We can substitute  $n_s(V)$  for  $n_s(\Phi)$  in Eqn. (9) using the above approximation, and can also use the relation  $\gamma_{\infty}^V - \gamma_{\infty}^{\Phi} = -\mu_{\text{sl}}\delta n_s^{\Phi,V}$ . After dropping the higher order correction terms, we obtain

$$G(n_s(\Phi)) \approx \mu_{\text{sl}}n_s(\Phi) + \gamma_{\infty}^{\Phi} \Omega v_s^{\frac{2}{3}} n_s^{\frac{2}{3}}(\Phi)(1 + \zeta n_s^{-\frac{1}{3}}(\Phi)), \quad (11)$$

where  $\zeta = \frac{2}{3}\delta n_s^{\Phi,V} \Omega v_s^{\frac{2}{3}} - \epsilon$  is a constant. This  $\zeta n_s^{-\frac{1}{3}}(\Phi)$  term in Eqn. (11), which has the same mathematical form as the Tolman correction, has two components: the  $\delta n_s^{\Phi,V}$  term stems from the difference between the  $\Phi$ -based dividing surface and the equimolar surface, while the other  $\epsilon$  term is determined from the Tolman length and has a value that is independent of the choice of the dividing surface. Hereafter, we will use CNT( $\Phi$ )+Tol to denote the expression in Eqn. (11) in order to highlight its dependence on the chosen extensive quantity. An awareness of this dependence on the choice of  $\Phi$  is crucial when comparing studies performed with different protocols. Eqs. (9) and (11) are both valid expressions for the free energy of a nucleus that are consistent with Tolman-corrected CNT. These two expressions differ, however, when they come to defining the size of the nucleus and the value and interpretation of the planar-interface excess free energy and the coefficient for the finite-size correction.

## SIMULATION METHODS

We simulated the processes of homogeneous nucleation for a simple but realistic Lennard-Jones system [15, 25, 26]. The NPT ensemble was employed throughout with the Nose-Hoover thermostat and isotropic barostat. The time step was set equal to 0.004 Lennard-Jones time units and a supercell containing 23328 atoms was used throughout.

Each independent simulation at each temperature was run for approximately  $6 \times 10^6$  steps. To accelerate the sampling so as to obtain reversible formation of a solid nucleus in a viable amount of simulation time, we performed biased sampling using the well-tempered metadynamics protocol with adaptive Gaussians [27, 28], and the collective variable  $\Phi = \sum_i S(\kappa(i))$  that was employed in Ref. [12]. In essence,  $S(\kappa(i))$  is a local structural fingerprint for atom  $i$ , which takes a value of 1 for a perfect *fcc* crystal and 0 for a homogeneous liquid. Fast implementation of this complex simulation setup was made possible by the flexibility of the PLUMED code [29] in combination with LAMMPS [30]. See the Supplemental Material for sample input files. In addition to the simulations

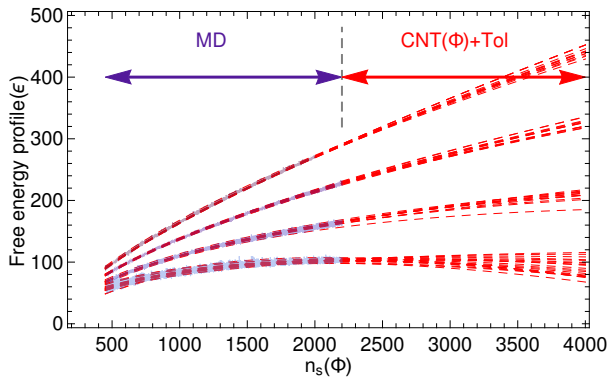


FIG. 1. The four sets of curves from bottom to top correspond to the free energy profiles of  $n_s(\Phi)$  at temperatures of 0.56, 0.58, 0.60 and 0.6185, respectively. Each thin blue curve is computed from a biased molecular dynamics simulation, and each thin red line corresponds to the best CNT( $\Phi$ )+Tol fit of one blue curve.

of homogeneous nucleation, we also simulated planar interfaces along the  $\langle 100 \rangle$ ,  $\langle 111 \rangle$ , and  $\langle 110 \rangle$  crystallographic directions of the *fcc* lattice. For the planar interface simulations, the setups are identical to the ones described in Ref. [12].

## RESULTS AND DISCUSSIONS

At each of the temperatures 0.56, 0.58, 0.60 and 0.6185, a total of 12 independent metadynamics runs were performed. For each simulation run we first computed the free energy profile  $\tilde{G}(\Phi)$  with respect to the extensive quantity  $\Phi$  of the system. We then extracted the nucleation free energy for a single solid cluster as a function of  $n_s(\Phi)$  using the framework introduced above, which is thoroughly described in Ref. [19]. Each free energy profile  $G(n_s(\Phi))$  is plotted as a thin blue curve in Figure 1. For each  $G(n_s(\Phi))$ , we then performed a CNT( $\Phi$ )+Tol fit using Eqn. (11) with  $\mu_{sl}$ ,  $\gamma_\infty^\Phi \Omega$ , and  $\zeta$  as fitting parameters. Each of these fitted curves is shown in red in Figure 1.

The red curves in Figure 1 start to diverge at larger sizes even though they almost overlap at small sizes. In order to understand the origin of this divergence we show the two parameters ( $\mu_{sl}$ ,  $\gamma_\infty^\Phi \Omega$ ) obtained from CNT+Tol( $\Phi$ ) fits for each free energy profile at  $T = 0.60$  in Figure 2. The two parameters  $\mu_{sl}$  and  $\gamma_\infty^\Phi \Omega$  are clustered around a straight regression line in Figure 2, which suggests that there is a very large correlation between them. The consequence of this correlation is that a tiny change in the segment of data used for fitting can make the two parameters vary collectively along the regression line by a significant amount. Thus, small uncertainties in the computed free energy profile for the small sub-critical nuclei that form in simulations can propagate and amplify when these curves are extrapolated to large nuclei.

We wanted to determine whether or not the values obtained for  $\mu_{sl}$  and  $\gamma_\infty^\Phi \Omega$  from the CNT( $\Phi$ )+Tol model

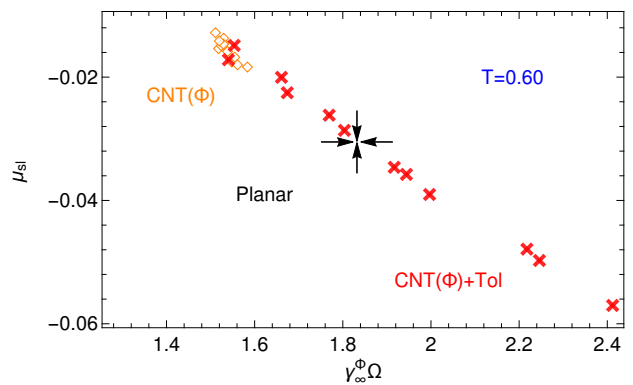


FIG. 2. Each symbol in red or orange represents one pair of fitting parameters ( $\gamma_\infty^\Phi \Omega$  and  $\mu_{sl}$ ) that were obtained by fitting one of the free energy curves shown in figure 1 that were themselves extracted from one of the independent simulations that were performed at  $T = 0.60$ . The black arrows indicate the results obtained from simulations of planar interfaces. All quantities are expressed in Lennard-Jones units.

are physically meaningful. To this end, we also show results obtained from simulations of the planar interfaces in Figure 2. The value of  $\mu_{sl}$  for an interface perpendicular to any lattice direction can be computed by performing biased simulations of that interface [12, 17]. We computed these planar interfacial free energies in our previous work [12] and used a deterministic framework to locate the Gibbs dividing surface. In the Supplementary Material, we show that the values of the interfacial free energies do not change significantly when the probabilistic framework in Eqn. (2) is applied so for this reason we chose to adopt the values for  $\gamma^\Phi$  and  $\gamma^V$  reported in Ref. [12]. To estimate  $\gamma_\infty^\Phi \Omega$ , we first obtained the values of  $\gamma_{100}^\Phi$ ,  $\gamma_{111}^\Phi$  and  $\gamma_{110}^\Phi$  for planar interfaces with normal vectors parallel to the specified lattice directions. We then used a common assumption; namely, that for this Lennard-Jones system the interfacial free energy surface  $\gamma^\Phi(\vec{n})$  can be expanded using cubic harmonics to the third order [31], whose coefficients can be parametrized using the values of  $\gamma_{100}^\Phi$ ,  $\gamma_{111}^\Phi$  and  $\gamma_{110}^\Phi$ . As shown in the Supplementary Material, the shape of the equimolar surface of the nucleus  $R(\vec{n})$  can then be reproduced by performing a Wulff construction using the  $\gamma^V$  values [32]. As the difference between the locations of the  $\Phi$ -based dividing surface and the equimolar dividing surface was assumed to be insignificant compared to the radius of the nucleus, the shape,  $R(\vec{n})$ , that emerges from the Wulff construction can be considered to be very close to the shape of the  $\Phi$ -based dividing surface of the nucleus. From this shape, we can, therefore, estimate the geometrical constant  $\Omega$ , and an effective value of  $\gamma_\infty^\Phi$  for the surface of the whole nucleus by computing the surface integral  $\iint_R \gamma^\Phi(\vec{n}) dA / \iint_R dA$ . At all the temperatures we considered, the shapes  $R(\vec{n})$  that we obtained were close to spherical. In fact, the corresponding  $\Omega$  values were within 0.5% of the geometrical constant for a sphere.

TABLE I. A comparison of the predictions for  $\mu_{s1}$  and  $\gamma_\infty^\Phi$  from different models. For each number, the value in the bracket indicates the statistical uncertainty in the last digit. Each number is the average from the 12 independent runs, and the statistical uncertainty is the standard error for the mean.

model	$\mu_{s1}$				$\gamma_\infty^\Phi$			
	0.56	0.58	0.60	0.6185	0.56	0.58	0.60	0.6185
CNT( $\Phi$ )+Tol	-0.100(9)	-0.061(5)	-0.033(4)	0.001(5)	0.42(4)	0.38(2)	0.39(2)	0.37(2)
Planar	-0.0991(2)	-0.0648(2)	-0.0305(2)	0.0012(2)	0.418(4)	0.393(3)	0.377(2)	0.360(2)
CNT( $\Phi$ )	-0.083(2)	-0.049(1)	-0.016(1)	0.017(1)	0.346(5)	0.330(2)	0.317(1)	0.302(3)

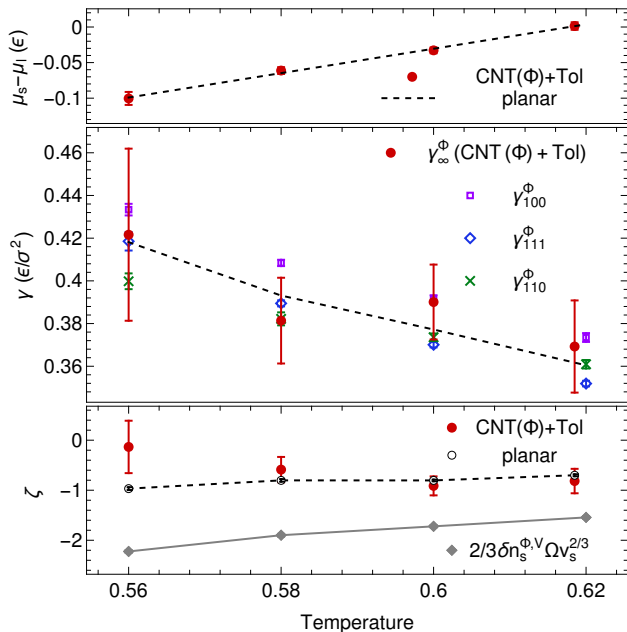


FIG. 3. In the top panel, the red symbols indicate  $\mu_{s1}$  values computed from homogeneous nucleation simulations using the CNT( $\Phi$ )+Tol fit, the dashed line  $\mu_{s1} = 1.714(T - 0.6178)$  describes the fitted values of  $\mu_{s1}$  that was extracted from planar interface calculations using a large simulation box with 20736 atoms. In the middle panel, the purple, blue and green symbols indicate the values of  $\gamma_\infty^\Phi$  for the planar interfaces that are perpendicular to three different lattice directions. The dashed line indicates the effective  $\gamma_\infty^\Phi$  that is obtained when the results for the planar interfaces are averaged over all lattice directions. The red symbols indicate the values of the estimates for the fitting parameters  $\gamma_\infty^\Phi \Omega$  that are obtained from the CNT( $\Phi$ )+Tol model divided by the estimated  $\Omega$  at each temperature. In the bottom panel, the red symbols indicate the Tolman correction constant  $\zeta$  in the CNT( $\Phi$ )+Tol fits, while the black symbols indicate the estimate for  $\zeta$  that is obtained when the planar interface results for  $\mu_{s1}$  and  $\gamma_\infty^\Phi$  are used in Eqn. (11). The gray symbols indicate the equimolar-surface correction to  $\gamma_\infty^\Phi$  from the planar limit,  $\frac{2}{3}\delta n_s^{\Phi,V} \Omega v_s^{\frac{2}{3}}$ . Statistical uncertainties are indicated throughout using error bars.

Figure 2 suggests that the values of  $\mu_{s1}$  and  $\gamma_\infty^\Phi \Omega$  from the CNT( $\Phi$ )+Tol fits are consistent with the planar limit results. However, there is a large spread in the parameter values because of the strong correlation between these two parameters in the fitting. Table I shows all the values of  $\mu_{s1}(T)$  together with the values of  $\gamma_\infty(T)$  that are predicted both from the CNT( $\Phi$ )+Tol model and the planar limit results and further illustrates the very good agreement between the

two models at all temperatures considered. We think this good agreement is a strong indication that the CNT( $\Phi$ )+Tol model is a good macroscopic model for the free energies of atomistic nuclei. Indeed, when a different model is used, the values of the two parameters  $\mu_{s1}$  and  $\gamma_\infty^\Phi \Omega$  may not agree with planar limit results. For example, we also performed a so-called CNT( $\Phi$ ) fit on each computed  $G(n_s(\Phi))$ , using a conventional CNT formulation  $G(n_s(\Phi)) = \mu_{s1}n_s(\Phi) + \gamma_\infty^\Phi \Omega v_s^{\frac{2}{3}} n_s^{\frac{2}{3}}(\Phi)$ . The values of  $\mu_{s1}$  and  $\gamma_\infty^\Phi \Omega$  from the CNT( $\Phi$ ) fits are indicated in Figure 2 and Table I. Once again there are very strong inter-correlations between the parameters. What is even worse, however, is that the fitted values are no longer consistent with the planar interface results. This example highlights the perils associated with using the wrong model to fit the free energy profile for nucleation: although one might obtain a good fit for the data points, the values of the fitted parameters of the model are physically unrealistic and the model will most likely have limited predictive power in scenarios that are outside the range of the existing data set.

The comparison between the values of  $\mu_{s1}$  and  $\gamma_\infty^\Phi \Omega$  from the CNT( $\Phi$ )+Tol and the planar limit results suggests that one should just use the results obtained at the planar interface limit in Eqn. (11) so as to have a model that has  $\zeta$  as the only fitting parameter. This model would have much greater statistical accuracy as there are no correlations between the fitting parameters. We performed such fittings and compared the values of the parameter  $\zeta$  with the values obtained from the regular CNT( $\Phi$ )+Tol fits using all three parameters. The bottom panel of Figure 3, shows that this new approach results in much smaller uncertainties in the values of  $\zeta$ . Therefore, performing simulations of planar interfaces as well as simulations of three dimensional nuclei is worthwhile when studying homogeneous nucleation, especially given how computationally inexpensive such simulations are. We would recommend first computing the values of  $\mu_{s1}$  and  $\gamma_\infty^\Phi \Omega$  from simulations of planar interfaces. This allows one to determine the leading terms  $\mu_{s1}n_s(\Phi) + \gamma_\infty^\Phi \Omega v_s^{\frac{2}{3}} n_s^{\frac{2}{3}}(\Phi)$  in a CNT-type model. Comparing these leading terms with the actual nucleation free energy profiles computed from the simulations of three-dimensional nuclei, then allows one to extract the other higher-order correction terms in the CNT-type model with much higher statistical accuracy.

It is worth discussing the higher-order correction term  $\zeta = \frac{2}{3}\delta n_s^{\Phi,V} \Omega v_s^{\frac{2}{3}} - \epsilon$  in Eqn. (11) a little further.

This factor has two distinct components but we treat it as a single fitting parameter in the present study. To evaluate the first term separately we can exploit the fact that the difference between the number of solid-like atoms per area  $\delta n_s^{\Phi,V}$  for different definitions of the dividing surface can be directly evaluated from simulations of planar interfaces as discussed in Ref. [12]. We evaluated  $\delta n_s^{\Phi,V}$  for planar interfaces with normal vectors parallel to the  $\langle 100 \rangle$ ,  $\langle 111 \rangle$  and  $\langle 110 \rangle$  lattice directions at each temperature, and then approximated  $\delta n_s^{\Phi,V}(\vec{n})$  using a cubic harmonic expansion, and thus estimated its averaged value for a three-dimensional nucleus. These estimated values for  $\frac{2}{3}\delta n_s^{\Phi,V}\Omega v_s^{\frac{2}{3}}$  are shown in the bottom panel of Figure 3. The other term  $\epsilon = (32\pi/3)^{\frac{1}{3}}v_s^{-\frac{1}{3}}\delta$  that enters  $\zeta$  can be determined from the Tolman length  $\delta$ . This quantity can be evaluated from the pressure tensor at each position along the normal to a planar interface for fluid-fluid systems [33–35]. How this quantity should be computed for solid-liquid interfaces is, however, unclear. As such, we chose to subtract the values we obtained for  $\frac{2}{3}\delta n_s^{\Phi,V}\Omega v_s^{\frac{2}{3}}$  from the value we obtained for  $\zeta$  and to extract the Tolman length from  $\epsilon$ . In this way, we predicted the Tolman length to be

$$\delta = \frac{\frac{2}{3}\delta n_s^{\Phi,V}\Omega v_s^{\frac{2}{3}} - \zeta}{(32\pi/3)^{\frac{1}{3}}v_s^{-\frac{1}{3}}}, \quad (12)$$

which is of the order of  $-0.3$  Lennard-Jones length units at the temperatures considered.

Finally, we want to point out that the fact that the CNT( $\Phi$ )+Tol model succeeds in describing small nuclei that contain hundreds of atoms, while being consistent with simulations performed in the limit of a planar solid-liquid interface has theoretical implications. Even though the actual interface between the solid and the liquid is diffuse and fluctuating, the concept of the Gibbs dividing surface can be used to convert the atomistic descriptions of the solid-liquid system into a macroscopic representation. Moreover, finite-size effects are well-captured by simple corrections that take into account the deviation between the chosen dividing surface, the equimolar dividing surface and the surface of tension. In this case, these two corrections are of comparable size.

## CONCLUSIONS

In summary we have studied, by means of atomic-scale simulations, the various terms in classical nucleation theory that contribute to the free energy profile for homogeneous nucleation. When analysing our simulations we define a Gibbs dividing surface from the value of a macroscopic order parameter when calculating the dependence of the free energy on the number of particles in the nucleus. Then, by juxtaposing explicit simulations of a three-dimensional nucleus with simulations of a planar solid-liquid interface, performed in equivalent thermodynamic conditions, we identify the

effects that lead the results on small nuclei to deviate from the predictions of classical nucleation theory. We find that deviations occur because curvature-dependent corrections to the planar-interface surface energy (the so-called Tolman term) are required. These genuine departures from CNT should not be confused with the deviations that occur when an arbitrary order parameter is used to identify the nucleus, however. Such corrections are required simply because we require zero surface excess for the arbitrary extensive quantity  $\Phi$  when calculating the surface excess free energy. As  $\Phi$  is not the volume the value we obtain for this excess free energy differs from the value that would have been obtained had we found the true equimolar dividing surface. The leading order correction for this effect has the same functional form as the Tolman term. It should not, however, be considered as a genuine departure from CNT as it is simply an artifact in the analysis.

In the final parts of the manuscript we discussed the statistical efficiency of different approaches for finding the CNT parameters. We showed that the surface and bulk parameters in the CNT model are strongly correlated, which makes the fitted values of these parameters extraordinarily sensitive to small deviations in the computed free energy surfaces that are used to fit them. We thus concluded that fitting these terms using a free energy surface obtained from a simulation of a 3D nucleus is not optimal and that also using information from simulations of planar-interfaces is thus beneficial. In fact, given that planar-interface models converge faster with respect to both size and simulation time, and that such simulations can be more easily analyzed in terms of different order parameters, we suggest that they should always be used in a preliminary phase. Explicit simulations of 3D nucleation can then be used to identify genuine finite-size effects. We hope that our careful analysis will help resolve some of the ambiguities in atomic-scale studies of both homogeneous and heterogeneous nucleation, and that this work will lay the foundations for the modelling of more complex materials and for the study of different kinds of phase transitions.

## SUPPLEMENTARY MATERIAL

The supplementary material contains a more in-depth discussion of detailed procedures of our approach, together with descriptions of the simulation protocols and a commented sample input file.

---

\* bingqing.cheng@epfl.ch

- [1] D. W. Oxtoby, *Journal of Physics: Condensed Matter* **4**, 7627 (1992).
- [2] P. Yi and G. C. Rutledge, *Annual review of chemical and biomolecular engineering* **3**, 157 (2012).

- [3] G. C. Sosso, J. Chen, S. J. Cox, M. Fitzner, P. Pedevilla, A. Zen, and A. Michaelides, *Chemical reviews* (2016).
- [4] P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, *The Journal of chemical physics* **104**, 9932 (1996).
- [5] P. R. ten Wolde and D. Frenkel, *Physical Chemistry Chemical Physics* **1**, 2191 (1999).
- [6] D. Moroni, P. R. Ten Wolde, and P. G. Bolhuis, *Physical review letters* **94**, 235703 (2005).
- [7] F. Trudu, D. Donadio, and M. Parrinello, *Physical review letters* **97**, 105701 (2006).
- [8] W. Lechner, C. Dellago, and P. G. Bolhuis, *Physical review letters* **106**, 085701 (2011).
- [9] S. Prestipino, A. Laio, and E. Tosatti, *Physical review letters* **108**, 225701 (2012).
- [10] J. McCarty, O. Valsson, and M. Parrinello, *Journal of chemical theory and computation* **12**, 2162 (2016).
- [11] S. Jungblut and C. Dellago, *The European Physical Journal E* **39**, 77 (2016).
- [12] B. Cheng, G. A. Tribello, and M. Ceriotti, *Physical Review B* **92**, 180102 (2015).
- [13] J. Hoyt, M. Asta, and A. Karma, *Physical Review Letters* **86**, 5530 (2001).
- [14] R. L. Davidchack and B. B. Laird, *Physical review letters* **85**, 4751 (2000).
- [15] R. L. Davidchack and B. B. Laird, *The Journal of chemical physics* **118**, 7651 (2003).
- [16] S. Angioletti-Uberti, M. Ceriotti, P. D. Lee, and M. W. Finnis, *Physical Review B* **81**, 125416 (2010).
- [17] U. R. Pedersen, F. Hummel, G. Kresse, G. Kahl, and C. Dellago, *Physical Review B* **88**, 094101 (2013).
- [18] V. G. Baidakov, S. P. Protsenko, and A. O. Tipseev, *Jetp Letters* **98**, 801 (2014).
- [19] B. Cheng and M. Ceriotti, *The Journal of Chemical Physics* **146**, 034106 (2017).
- [20] P. M. Piaggi, O. Valsson, and M. Parrinello, arXiv preprint arXiv:1612.03235 (2016).
- [21] J. W. Gibbs, H. A. Bumstead, W. R. Longley, *et al.*, *The collected works of J. Willard Gibbs*, Vol. 1 (Longmans, Green and Company, 1928).
- [22] R. C. Tolman, *The journal of chemical physics* **16**, 758 (1948).
- [23] V. Yukalov, *Laser Physics Letters* **8**, 485 (2011).
- [24] B. J. Block, S. K. Das, M. Oettel, P. Virnau, and K. Binder, *The Journal of chemical physics* **133**, 154702 (2010).
- [25] S. Angioletti-Uberti, M. Ceriotti, P. D. Lee, and M. W. Finnis, *Phys. Rev. B* **81**, 125416 (2010).
- [26] R. Benjamin and J. Horbach, *The Journal of chemical physics* **141**, 044715 (2014).
- [27] A. Barducci, G. Bussi, and M. Parrinello, *Physical review letters* **100**, 020603 (2008).
- [28] D. Branduardi, G. Bussi, and M. Parrinello, *Journal of Chemical Theory and Computation* **8**, 2247 (2012).
- [29] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, *Computer Physics Communications* **185**, 604 (2014).
- [30] S. Plimpton, *J. Comp. Phys.* **117**, 1 (1995).
- [31] J. Hoyt, M. Asta, and A. Karma, *Phys. Rev. Lett.* **86**, 5530 (2001).
- [32] D. Adams, S. Yalisove, and D. Eaglesham, *Journal of applied physics* **76**, 5190 (1994).
- [33] Y. A. Lei, T. Bykov, S. Yoo, and X. C. Zeng, *Journal of the American Chemical Society* **127**, 15346 (2005).
- [34] E. M. Blokhuis and J. Kuipers, *The journal of chemical physics* **124**, 074701 (2006).
- [35] A. E. van Giessen and E. M. Blokhuis, *The Journal of chemical physics* **131**, 164705 (2009).