Predicting homogeneous nucleation rate from atomistic simulations

Thèse N° 9183

Présentée le 18 janvier 2019

à la Faculté des sciences et techniques de l'ingénieur
Laboratoire de science computationnelle et modélisation
Programme doctoral en science et génie des matériaux

pour l'obtention du grade de Docteur ès Sciences

par

BINGQING CHENG

Acceptée sur proposition du jury

Prof. P. Muralt, président du jury
Prof. M. Ceriotti, directeur de thèse
Prof. M. Parrinello, rapporteur
Dr M. Salvalaglio, rapporteur
Prof. B. Smit, rapporteur

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Acknowledgements

I owe a debt of gratitude to my supervisor for mentorship and encouragement throughout the PhD. I could not have asked for a better supervisor. I want to thank my collaborators for their excellent work and the joy of working together. I am also grateful to some other lab mates, including the ones in the labs that I visited, for their help and the fun time spent together. I thank the people who guided me and gave me enormous inspiration.

I appreciate very much the long-lasting friendships that give me tremendous mental support. I am very grateful to my extremely supportive parents, and my loving husband.

Abstract

Nucleation is ubiquitous, from the formation of clouds to the preparation of pharmaceutical compounds, from metal casting to the tempering of chocolates, and from the growth of beautiful nautilus shells to the assembly of microtubules in cells. The first experiment for observing a nucleation event was performed by Fahrenheit in 1724, and it can be easily replicated in a kitchen: simply place a bottle of purified water in the freezer. The liquid water can be cooled to far below zero degrees Celsius without freezing due to the lack of microscopic nuclei, which are the embryos from which the freezing phase transition can occur. After that, shaking the bottle will induce nucleation which in turn prompts a rapid ice crystallization.

Despite its pivotal importance and long history, we only have a rough idea about the underlying mechanism of nucleation. The classical nucleation theory says that, during the growth of a nucleus inside a bulk phase, an interface that surrounds the nucleus has to be created. This interface is associated with an energy penalty, which the system has to overcome for the nucleus to grow into a critical size which precedes an avalanche of structural transitions. It is analogous to being stuck in a valley on the Alps, so that a great deal of time and energy have to be spent to climb over a peak in order to reach another valley. And yet, we have not reached a quantitative understanding of how high the energy barrier is and how long is the waiting time of nucleation for specific systems. Taking again the example of ice nucleation from bulk liquid water, there has been a long-standing discrepancy by more than 10 orders of magnitude between the measured and the predicted expectation time of nucleation. While the classical nucleation theory is able to paint a physical picture of nucleation, for many systems it is insufficient and thus needs extension.

Despite substantial improvements in recent years, experimental characterization of the dynamical nucleation processes is extremely difficult, which motivates atomistic modelling efforts that use numerical simulation techniques. However, atomistic simulations also faces a number of challenges: firstly the typical time scales accessible to atomistic simulations are confined to below microseconds, while nucleation can take hours or days to occur. We have mitigated the challenge by employing state-of-the-art enhanced sampling methods in the simulation studies of nucleation [1–5]. In a nutshell, instead of naively waiting for a rare event to happen, we place a bias to help the system overcome the energy penalty of nucleation. To return to the earlier analogy with Alpine hiking, we can flatten out the Alps by depositing (a lot of) sand into all the valleys, making the landscape level and easy to explore.

Secondly, only microscopic quantities such as the coordinates and the velocities of each atom can be directly obtained from simulations. On the contrary, macroscopic observables such as heat content can be easily interpreted by human brains, enter analytic expressions that explain the phenomena, and be measured experimentally. Compressing the astronomical amount of information contained in the microscopic quantities into a handful of macroscopic observables can be a daunting task as there are countless ways to do so and not all are meaningful. A major part of the problem may sound semantic: what is one phase and what is another? It is surprisingly hard to tell at the molecular level. Think of an ice nucleus embedded in liquid water: how can we determine which molecules belong to the solid phase, and which ones belong to the liquid? What about the molecules that are close to the interface? Inspired by the concept of Gibbs dividing surface from the founding father of statistical mechanics, we formulated a thermodynamic framework that reconciles the picture emerging from simulations with macroscopic theories of nucleation [1, 2]. Crucially, by defining the interface between two phases and its associated free energy in a rigorous and self-consistent manner, we are able to accurately extract the energy and the entropy costs to form a new phase and to create an interface, enabling stringent validation and extension of the classical nucleation theory [3–5].

The third challenge stems from the shortfalls of atomistic modelling itself. In most cases, the interactions between atoms and molecules are mimicked by so-called empirical force fields, which are analytic functions with tunable fitting parameters that take atomic positions as inputs and return energy and forces. Another option is to compute the interactions from first principles using the tools from quantum chemistry, which is prohibitively expensive and thus not practical for systems with more than hundreds of atoms and for time scales longer than a nanosecond. There is, however, an alternative way to combine the merits of both the first principles and the empirical force field approaches, by exploiting machine learning (ML) techniques to "learn" the atomic interactions from quantum mechanics. In essence, a ML atomic potential "remembers"a pre-computed database of reference structure-property pairs, and makes predictions about a new structure by comparing it to the members of the reference set. Recently, we have constructed a ML potential for water, which is able to reproduce well several thermodynamic properties of water including the melting point [6]. We are currently combining this ML potential with the aforementioned enhanced sampling methods and the thermodynamic framework, in order to arrive at a predictive model of ice nucleation.

The field of nucleation is undergoing a renaissance in recent years. Despite all the progress, we are still in the early stages regarding what we know and what we do not know about nucleation. Perhaps marrying statistical mechanics with machine learning provides the missing ingredients to achieve predictive modelling and quantitative understanding of nucleation.

Key words: nucleation, phase transition, interface, free energy methods, machine learning potential, molecular dynamics

List of publications

Publications directly related to the thesis

1. Bingqing Cheng*, Edgar A Engel, Jörg Behler, Christoph Dellago, Michele Ceriotti. (2018) ab initio thermodynamics of liquid and solid water. Proceedings of the National

Academy of Sciences (In press)

2. Bingqing Cheng*, Christoph Dellago, Michele Ceriotti. (2018) Theoretical prediction of the homogenous ice nucleation rate: disentangling thermodynamics and kinetics. Physical Chemistry Chemical Physics, DOI: 10.1039/C8CP04561E
3. Bingqing Cheng*, Michele Ceriotti. (2018) Computing the Tolman length for solid-liquid interfaces. The Journal of Chemical Physics, 148(23): 231102.
4. Bingqing Cheng*, Michele Ceriotti. (2018) Computing the absolute Gibbs free energy in atomistic simulations: applications to defects in solids. Physical Review B, 97(5): 054102.
5. Bingqing Cheng*, Gareth A Tribello, Michele Ceriotti. (2017) The Gibbs free energy of homogeneous nucleation: from atomistic nuclei to the planar limit. The Journal of Chemical Physics, 147(10): 104707.
6. Bingqing Cheng*, Michele Ceriotti. (2016) Bridging the gap between atomistic and macroscopic models of homogeneous nucleation. The Journal of Chemical Physics, 146(3): 034106.
7. Bingqing Cheng, Gareth A Tribello, Michele Ceriotti. (2015) Solid-liquid interfacial free energy out of equilibrium. Physical Review B, 92(18): 180102.

Other publications

1. Venkat Kapil, Mariana Rossi, Ondrej Marsalek, Riccardo Petraglia, Yair Litman, Thomas Spura, Bingqing Cheng, Alice Cuzzocrea, Robert H. Meißner, David M. Wilkins, Przemys law Juda, Sebastien P.Bienvenue, Wei Fang, Jan Kessler, Igor Poltavsky, Steven Vandenbrande, Jelle Wieme, Clemence Corminboeuf, Thomas D. Kuhne, David E. Manolopoulos, Thomas E. Markland, Jeremy O. Richardson, Alexandre Tkatchenko, Gareth A. Tribello, Veronique Van Speybroeck, Michele Ceriotti. (2018) i-PI V2: A Universal Force Engine for Advanced Molecular Simulations. Accepted in Comp. Phys. Comm.
2. Bingqing Cheng, Anthony T Paxton, Michele Ceriotti. (2018) Hydrogen diffusion and trapping in a-iron: the role of quantum and anharmonic fluctuations. Physical Review Letters, 120(22): 225901.
3. Bingqing Cheng, Joerg Behler, Michele Ceriotti. (2016) Nuclear Quantum Effects in Water at the Triple Point: Using Theory as a Link Between Experiments. Journal of Physical Chemistry Letters, 7(12): 2210-2215.
4. Peggy S S Leung, Hing Shun Leung, Bingqing Cheng, Alfonso H W Ngan. (2015) Size dependence of yield strength simulated by a dislocation-density function dynamics approach. Modelling and Simulation in Materials Science and Engineering, 23(3): 035001.
5. Bingqing Cheng, Michele Ceriotti. (2014) Direct path integral estimators for isotope fractionation ratios. The Journal of Chemical Physics, 141(24): 244112.

6. Bingqing Cheng, Hing Shun Leung, Alfonso H W Ngan. (2014) Strength of metals under vibrations - Dislocation-density-function dynamics simulations. Philosophical Magazine, 95(16-18):1-21.
7. Hing Shun Leung, Peggy S S Leung, Bingqing Cheng, Alfonso H W Ngan. (2014) A New Dislocation-density-function Dynamics Scheme for Computational Crystal Plasticity by Explicit Consideration of Dislocation Elastic Interactions. International Journal of Plasticity, 67: 1-25.
8. Bingqing Cheng*, Alfonso H W Ngan. (2013) Crystal plasticity of Cu nanocrystals during collision. Materials Science and Engineering, 585:326-334.
9. Bingqing Cheng, Alfonso H W Ngan. (2013) The sintering and densification behaviour of many copper nanoparticles: A molecular dynamics study. Computational Materials Science, 74:1-11.
10. Bingqing Cheng*, Alfonso H W Ngan. (2013) Thermally induced solid-solid structural transition of copper nanoparticles through direct geometrical conversion. The Journal of Chemical Physics, 138(16):164314.
11. Bingqing Cheng, Alfonso H W Ngan. (2013) The crystal structures of sintered copper nanoparticles: A molecular dynamics study. International Journal of Plasticity, 47: 65-78.

(* Corresponding author)

Contents

Acknowledgements i

Abstract iii

1 Introduction 1

1.1 Background 1
1.2 Theory of nucleation 2
1.3 Experimental studies of nucleation 6
1.4 Atomistic simulation studies of nucleation 7
1.5 Outline of the thesis/ statement of contributions 11

2 Free energy estimation methods 13

2.1 The fundamentals of free energy 13
2.2 The thermodynamic integration method 17
  2.2.1 A general workflow of thermodynamic integration for solid systems . . . 17
  2.2.2 One example of applying the thermodynamic integration method: vacancy formation energy in BCC iron 24
2.3 Enhanced sampling methods 26
  2.3.1 Free energy estimation from unbiased and biased molecular dynamics . 28
  2.3.2 An example of enhanced sampling methods: metadynamics 28
  2.3.3 Constructing a collective variable for distinguishing solid from liquid . . 29
  2.3.4 An example of local order parameter: FCCUBIC 30

3 The interfacial free energy of solid-liquid interfaces 33

3.1 Gibbs dividing surface in atomistic systems 34
3.2 Simulation details 36
3.3 The planar interfacial free energy γ^{Φ} 37
3.4 Using a different Gibbs dividing surface 38
3.5 Results and discussions 39
3.6 Conclusions 42

4 Bridging the gap between atomistic and macroscopic models of homogeneous nucleation 43

4.1 An idealized multiple-cluster model 43

4.2 A probabilistic definition of Gibbs dividing surfaces 46
4.3 Obtaining cluster-size free energies from an extensive order parameter 48
4.4 Application: solidification of a Lennard-Jones system 49
4.5 Application: nucleation in a two-dimensional Ising Model 52
4.6 Conclusions 53

5 How to use macroscopic nucleation theory to interpret atomistic simulations of homogeneous nucleation? 55
5.1 A recap of the thermodynamic framework 56
5.2 A general formulation of classical nucleation theory 56
5.3 Simulation methods 58
5.4 Results and Discussions 58
5.5 Conclusions 63

6 Computing the Tolman length for solid-liquid interfaces 65
6.1 The Tolman length and the surface of tension 67
6.2 Free energy of surface of tension 67
6.3 Fluctuating Gibbs dividing surface 69
6.4 Simulation methods 71
6.5 Convergence tests 72
6.6 Results 73
6.7 Conclusions 75

7 Theoretical prediction of the homogeneous ice nucleation rate 77
7.1 Motivation 77
7.2 The nucleation free energy of ice Ih 79
7.3 Accounting for stacking disorder 81
7.4 The kinetic factor in homogeneous nucleation 85
7.5 Results 88
7.6 Discussions 89
7.7 Conclusions 91

8 The quest for a predictive model of ice nucleation 93
8.1 The challenges in modelling ice and liquid water 93
8.2 First-principles thermodynamics 94
8.3 A neural network potential energy surface for bulk water 95
8.4 The relative stability of hexagonal and cubic ice 97
8.5 The relative stability of hexagonal ice and liquid water 100
8.6 Comments on the first principles description of water and ML potentials 101

9 Conclusions 103

10 Appendix **105**
10.1 Cluster distributions under the multiple cluster model 105
10.2 A fluctuating Gibbs dividing surface 106

Bibliography **126**

Curriculum Vitae **127**

1 Introduction

1.1 Background

Nucleation is a key step in bulk phase transitions [7–10]. This process plays a crucial role in natural phenomena and in technological applications, from the formation of clouds [10] to self-assembly[11], and from casting to the growth of thin films [12, 13]. An important class of problems in nucleation is solidification (i.e. the formation of solid inside bulk liquid at undercooled conditions). Solidification underlies many natural phenomena such as the freezing of water in clouds and the formation of igneous rock, and is is also crucial for various critical technologies including commercial casting, soldering and additive manufacturing [14–16]. In the present thesis we focus mostly on solidification, because of its pivotal importance, and also because in the case of solid nucleus inside the liquid the Laplace pressure (e.g. is the pressure difference between the inside and the outside of a curved surface that forms the boundary between the two regions) has little influence on the chemical potential of the solid, which makes the problem somehow simpler compared to bubble nucleation or condensation.

Despite the ubiquity of the nucleation phenomena, experimental characterization is difficult due to the highly dynamical process at out-of-equilibrium conditions, and considerable uncertainties involved in the measurements [17, 18]. These difficulties have triggered numerous computational efforts for understanding nucleation, such as evaluating the free energy change associated with nucleation using atomistic simulations [10]. Atomistic modelling of nucleation is plagued with technical difficulties, however, as molecular dynamic simulations have very short time scales and their accuracy is limited by the interatomic potential energy surfaces assumed in the simulations. Notwithstanding the tremendous amount of efforts from both the experimental and the theory sides in the last a few decades, a quantitative interpretation of the nucleation phenomena has not been achieved, and the current understanding goes little beyond the textbook picture of the classical nucleation theory (CNT).

In the remainder of this chapter, we first briefly introduce the classical theory of nucleation, then summarize the state-of-the-art experimental as well as computational methods for studying nucleation, and finally identify the challenges involved in the atomistic simulation

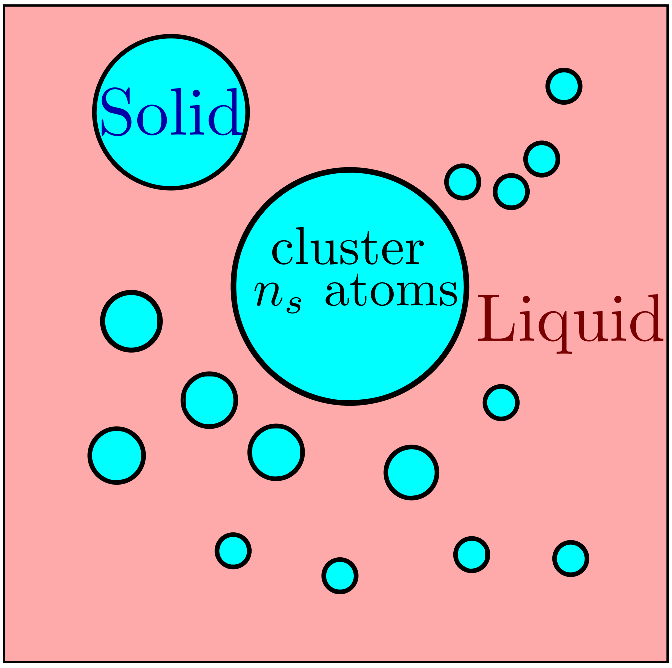

Figure 1.1 – A schematic of an under-cooled liquid system that contains multiple solid sub-critical clusters.

studies of nucleation and outline the motivation behind the current thesis.

1.2 Theory of nucleation

A comprehensive theory of nucleation must have two essential elements: an explanation of the nucleation and phase transition mechanism, and a formulation that can be used to interpret and predict nucleation rates. The starting point of a nucleation theory is that the new structure formed inside the metastable bulk phase can be conceptualized as a separate entity named nucleus whose kinetics and thermodynamic properties can be modelled and analyzed, so that one does not have to consider the rest of the microscopic events occurring inside the remaining metastable bulk phase. This starting point of the nucleation theory comes with a subtle drawback: it is not evident how to identify solid and liquid and draw the boundaries between a crystal nucleus and the surrounding undercooled melt, and as such the very identification of a number of particles that belong to the nucleus can be ambiguous. In the followings we first bypass this drawback and directly introduce the classical nucleation theory and the various extension to it, while bearing in mind that the very foundation of any nucleation theory-the definition of a nucleus-is often the core cause of why a nucleation theory may fail.

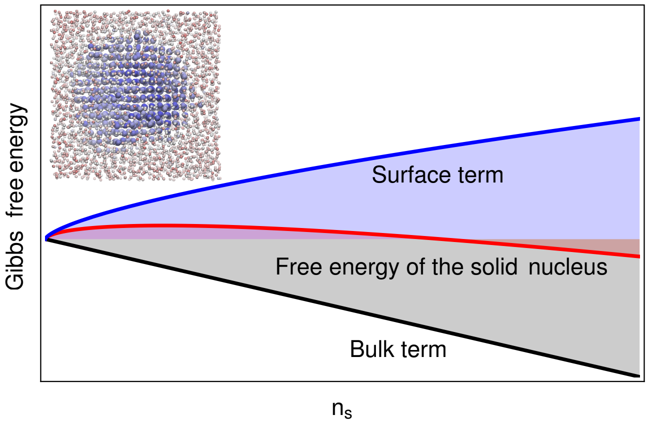

Figure 1.2 – A schematic of the bulk, the surface and the overall Gibbs free energy of a solid cluster as a function of its size n_s.

Classical nucleation theory

The classical nucleation theory (CNT) is the simplest and perhaps the most useful model rationalize nucleation. It assumes that the stable phase forms by growth of nanoscopic nuclei [19]. These clusters are unstable when they are smaller than a critical size $n^\star$, and at any given time the metastable phase contains multiple sub-critical clusters of the stable phase (Figure 4.2).

In the context of the homogeneous nucleation of crystalline solids from bulk liquid, CNT assumes that $G(n_\mathrm{s})$ can be expressed as the sum of a bulk and a surface term [9] i.e.

$$G(n_\mathrm{s}) = \mu_\mathrm{sl} n_\mathrm{s} + \gamma A(n_\mathrm{s}). \tag{1.1}$$

In this expression the first bulk term stems from the difference in chemical potential between the solid and the liquid $\mu_\mathrm{sl} = \mu_\mathrm{s} - \mu_\mathrm{l}$, which is negative below the coexistence temperature T_m. The second term describes the penalty associated with the interface between the two phases, and introduces a kinetic barrier to nucleation. This surface term is the product of a specific free energy excess term γ, and the extensive surface area $A(n_\mathrm{s})$. However, due to the diffuse nature of the interface, there is a degree of ambiguity in the location and area of the dividing surface between phases. In classical nucleation theory (CNT), an infinitesimally-thin dividing surface divides the solid nucleus from the surrounding liquid. These two phases are usually taken to have their bulk densities so the surface area of the solid nucleus can be calculated using $A(n_\mathrm{s}) = \sigma {n_\mathrm{s}}^{2/3}$, where σ is a constant that depends on the shape, e.g. $\sigma = (36\pi)^{1/3} v_\mathrm{s}^{2/3}$ for a spherical nucleus with bulk solid molar volume v_s. Under these assumptions, the free-energy barrier to nucleation can be calculated using $G^\star = \frac{4}{27}\sigma^3 \mu_\mathrm{sl}^{-2} \gamma^3$.

In many cases, nucleation is an activated process, that is, there is usually a very long waiting time before an occurrence of such event, but once it starts to happen the process itself takes a very short amount of time. For such processes with time-scales separation, the Transition State Theory (TST) theory can be employed, and the nucleation rate can be obtained from the standard Arrhenius expression of the transition state theory, i.e.

$$J = J_0 \exp(-G^{\star}/k_B T), \tag{1.2}$$

where J_0 denotes a kinetic prefactor. There are different approaches to estimate the prefactor J_0, which is frequently expressed as

$$J_0 = (1/v_{\mathrm{l}}) Z f^{+} \tag{1.3}$$

where v_{l} is the molar volume of the undercooled liquid, f^{+} is the addition rate of particles to the critical nucleus, and the Zeldovich factor

$$Z = \sqrt{\frac{1}{2\pi k_B T} \frac{d^2 G(n_{\mathrm{s}})}{d n_s^2}} \tag{1.4}$$

can be obtained from the nucleation free energy profile $G(n_{\mathrm{s}})$. Note that the expression for the Zeldovich factor Z was derived by solving the Becker and Döring Master equation for nucleation kinetics [7, 20].

The nucleation rate J is a measure of how frequently nucleation events occur for a unit volume of the system that belongs to the metastable bulk phase. For a specific system with volume V, the expected amount of time for the system to remain in the mestastable phase is characterized by the the induction time $\tau^{\star}$, which is related to the nucleation rate by

$$\tau^{\star} = \frac{1}{JV}. \tag{1.5}$$

The classical nucleation theory has long been criticized for the various simplifying assumptions involved [21]:

- The nucleus embedded inside the metastable substrate is assumed to adopt the same macroscopic properties (e.g. density, enthalpy, free energy) of the stable phase.
- The nucleus is assumed to be spherical.
- The interface between the nucleus and the metastable bulk phase is assumed to be infinitely thin.
- The interfacial free energy of the interface is assumed to be independent from the curvature of the interface, that is, $\gamma(R) = \gamma(\infty)$.

Apart from the danger that these assumptions may lead to oversimplifications, CNT also suffers from the problem that it has limited range of application, e.g. it is unable to explain the vanishing of the nucleation barrier at high undercoolings. Despite these limitations of CNT, it is frequently used as a baseline model due to its general form and physical insights, and the various extensions and developments in the theoretical modelling of nucleation are built upon the foundation of CNT.

Two-step nucleation process

The original formulation of CNT assumes that a crystal nucleus inside the melt grows by the gradual attachment of individual particles, and as such, the number of particles belonging to the cluster is the natural reaction coordinate for describing the nucleation process, and there is a single free energy barrier that the system has to overcome to form a critical nucleus. However, it has been proposed that the formation of crystals from liquids may occur according to a two-step nucleation mechanism: an amorphous cluster is formed first by overcoming a first free energy barrier by means of density fluctuations, and then the cluster begins to adopt crystalline order after surmounting a second free energy barrier.

Ostwald's rule

Another phenomenon that can further complicate the process of nucleation is the polymorphism of crystals, which is the ability of a solid to exist in more than one crystallographic structure. Ostwald's rule states that it is not always that case that the most stable polymorph form during nucleation, but rather may proceed by one or more intermediate polymorphs that are metastable but have free energies closer to the initial phase [22]. It is hard to imagine how polymorphism will affect the validity of the central tenet of CNT (Eqn. (1.1)) provided that the chemical potential and the interfacial free energy terms properly reflect the properties of the polymorph that actually nucleates, but it makes the modelling and interpretation of the nucleation process even more complicated.

Heterogeneous nucleation

Heterogeneous nucleation is nucleation with the nucleus first formed at preferential sites such as an impurity or a surface, is much more common than homogeneous nucleation. In some cases of the heterogeneous nucleation the presence foreign substrate promote nucleation because the phenomenon of wetting changes the equilibrium shape of the critical nucleus. The shape change thereby lowers the effective surface energy and diminishes the free energy barrier. The process is often rationalized such that heterogeneous nucleation free energy is equal to the product of homogeneous nucleation and constant factor less than one, i.e. [10]

$$\Delta G_{heterogeneous} = \Delta G_{homogeneous} \times f(\theta), \tag{1.6}$$

where $f(\theta) = (2 - 3\cos\theta + \cos^3\theta)/4$ is a function of the contact angle θ.

Density functional theory

The motivation of using a continuum model to describe nucleation comes from the realization that physical interfaces are diffuse but are assumed to be infinitesimally thin under the classical nucleation theory. One of the most popular continuum models is density function theory (DFT), which uses the density distribution function of a system to characterize the structures and properties. In the framework of DFT, the mean field Helmholtz free energy of a fluid-fluid system is expressed as a functional of the local density $\rho(\mathbf{r})$ [23]:

$$F(\rho(\mathbf{r})) = \int_V d\mathbf{r} F_h(\rho(\mathbf{r})) - \mu\rho(\mathbf{r}) + \frac{1}{2}\int_V d\mathbf{r} \int_V d\mathbf{r'} \rho(\mathbf{r})\rho(\mathbf{r'}) V(|\mathbf{r}-\mathbf{r'}|), \tag{1.7}$$

where $F_h(\rho(\mathbf{r}))$ denotes the Helmholtz free energy of a uniform fluid with density ρ, μ is the chemical potential, and V is the pair-wise interaction potential between particles. Although DFT methods are in principle exact, approximations are usually made in calculations. One of the most widely used approximations is the Cahn-Hilliard theory of interfaces and nucleation, which approximates the free energy using a local term and a square-gradient term of the density,i.e. [24]

$$F(\rho(\mathbf{r})) = \int_V d\mathbf{r} F_h(\rho(\mathbf{r})) - \mu\rho(\mathbf{r}) + \kappa(\bigtriangledown\rho(\mathbf{r}))^2, \tag{1.8}$$

where κ is a positive constant that is usually determined phenomenologically. The Cahn–Hilliard equation is particularly useful in the regime of spinodal decomposition, where the metastable phase becomes unstable and the free-energy barrier to nucleation vanishes.

The Cahn–Hilliard equation is a good approximation when the term $V(|\mathbf{r}-\mathbf{r'}|)$ in Eqn. (1.7) is short-ranged so the free energy is only slightly non-local, since non-locality of the free energy is only considered by the gradients of the density. For a sharply varying interface, additional correction terms have to be added [25].

Although DFT methods have been widely employed to study simple and multicomponent fluids, the nucleation problem that usually involves sharp gradient in the density is more difficult to treat, particularly when solid-liquid interfaces are involved. It is also worth noting that, although the continuum DFT model does not assume an infinately thin surface, the choice for the definition of the density $\rho(\mathbf{r})$ still affect the location and the width of the interface.

1.3 Experimental studies of nucleation

With the notable exception of high-resolution cryo-TEM [26] which can capture some initial stage of nucleation, crystal nucleation in liquids takes place much too fast to allow for a sequence of snapshots to be taken using experimental instruments. Usually, experimental methods can be used to detect the nuclei inside the bulk but do not provide any microscopic detail. Many experimental probes can be used here, including ultrafast X-ray diffraction [27, 28], optical microscopy [29], high-speed visual or infrared imaging [30], powder X-ray diffraction patterns [31]. The nucleation rate J is not a direct observable from experiments, but instead can be inferred from the induction time $\tau^\star$.

Besides the lack of dynamical information, experimental characterizations have other drawbacks as well. Because even small amounts of impurities induce heterogeneous nucleation that massively accelerates the overall nucleation rate of the system, the droplet technique is often used to address this problem [32]: the metastable liquid is dispersed into a large number of micrometer-sized droplets droplets, such that a significant number of droplets are free from impurities that induce heterogeneous nucleation. This technique also has the additional advantages of reducing the occurrence of multiple nucleation events within the same droplet, as well as examining an ensemble of independent nucleation events simultaneously. However, the droplet size cannot be too small either, in order to eliminate the effect from the surface.

Another problem is that the detectable size of a nucleus is usually larger than the critical size $n^\star$, so some amount of crystal growth must take place after the critical nucleus is formed. However, it is often assumed that the growth rate is normally so much faster than the nucleation rate such that the time it takes for the nucleus to grow into the detectable size is negligible compared with $\tau^\star$.

1.4 Atomistic simulation studies of nucleation

Although experimental techniques for the determination of nucleation rates have seen fast developments in the past decades [10], the measurements are still challenging due to the non-equilibrium conditions, and it is difficult to obtain insights into the detailed nucleation mechanism. As such, employing computational methods for evaluating the free energy change during nucleation as well as investigating the nucleation pathway is an extremely attractive alternative. In the last two decades, a considerable number of studies have thus used atomistic simulation to study homogeneous nucleation, and the key methodologies used in those studies can be roughly classified into three categories with descending computational costs: brute-force molecular dynamics, transition path sampling methods [33] that sample the ensemble of the reactive trajectories from bulk liquids to super-critical nuclei, and quasi-equilibrium free energy methods such as umbrella sampling [34] and metadynamics [35]. In this section we briefly introduce the key aspects and the influential nucleation studies employing these atomistic simulation methods.

Brute-force molecular dynamics

Brute-force molecular dynamics (MD) simulations of homogeneous nucleation are confronted with the fact that nucleation is a rare event, which is associated with a long waiting time far beyond the time scale accessible to simulations. As such, in order to observe the spontaneous formation of a critical nucleus in molecular dynamics simulations, an extreme undercooling of the system has to be imposed. For instance, Brute-force MD simulations employing classical force fields is only feasible when the nucleation rates are above $10^{25} cm^{-3} s^{-1}$, which implies that a liquid water system has to be undercooled by about 80 K in the studies of homogeneous ice nucleation. Besides carefully select the undercooling of the system considering the constraints on the simulation time, one also has to make sure that the system size is significantly larger than the critical nucleus in order to reduce finite size effects, and at the same time collect ample statistics of the stochastic nucleation events. Due to these constraints, Brute-force molecular dynamics have a very limited usage in the nucleation studies.

One method to overcome the nucleation barrier and avoid the long waiting time is called the seeding technique [36], which has gained some popularity during the last a few years. In this approach, the undercooled bulk fluid is seeded with a crystalline cluster, and afterwards the system is allowed to evolve at a certain temperature and pressure. The size of the crystalline nucleus is then closely monitored at different temperatures, in order to find a temperature under which the nucleus is critical and its size remains roughly constant. Finally, this size of the critical nucleus $n^{\star}$ is plugged into the original expression of the classical nucleation theory (Eqn. (1.1)) to get estimates of the nucleation rate, by assuming a constant value for the interfacial free energy at all radius so that $G^{\star} = \mu_{\mathrm{sl}} n^{\star}/2$. The seeding method is simple to implement and enables the estimation of the nucleation rate in a broad range of materials and thermodynamic conditions. However, this is an approximate method as it relies on the validity of both the preconceived structure of the critical nucleus and of classical nucleation theory. It is therefore important to assess and validate the accuracy of this approximate technique for different systems.

Transition path sampling

For many rare events, there is a wide disparity of timescales between fast dynamical transition between two equilibrium states and long waiting time inside an equilibrium state. Transition path sampling (TPS) is a technique that focuses on the sampling of the ensemble of reactive trajectories, which is defined as the paths connecting the two equilibrium states [33]. Typically, a new trajectory is generated by displacing an initial reactive trajectory, and then is collected into the transition path ensemble if it also connects the initial and the final equilibrium states. As such, transition path sampling by itself only needs the definition of what are the regions of the phase space that are associated with the reactant and the product of the reaction, without requiring the identification of a transition state.

An important feature of transition path sampling is that paths harvested in simulations are unbiased dynamical pathways, which in principle can be used to study the mechanism of reactions. However, ergodic sampling of the transition path ensemble is extremely expensive, and the path ensemble by itself does not necessarily contain enough information for computing the reaction rate. As a remedy, the transition interface sampling (TIS) method [37] and forward flux sampling (FFS) method [38] are frequently applied. To use these method, one has to define a set of order parameters which can be used to distinguish not only the product of the reaction but also the transition states. These two methods place a series of interfaces between two equilibrium states employing the order parameters, and measure the effective flux between these interfaces. By so doing, only relatively short trajectories between the neighboring interfaces are collected, rather than the long reactive trajectories that pass through both the initial and final states. The success of the TIS and the FFS methods thus hinges on the quality of the order parameters in characterizing the transition states.

Although being much more efficient and cheaper compared with brute-force MD, transition path sampling is still associated with high costs. As a consequence, TPS-type methods are mostly used to explore a reaction coordinate with a complex form, instead of sampling events that can be represented by a rather simple order parameter. For the problem of nucleation, many times the size of the nucleus is a good order parameter, and the reaction time can be very long due to the relatively high activation barriers. For these reasons, nucleation is not very often studied using transition path sampling, examples include magnetization reversal in the Ising-model [39], the freezing of a Lennard-Jones system [40], and large scale simulations performed on the homogeneous nucleation of stacking disordered ice using a monoatomic water model [41, 42]. Nevertheless, such methods may eventually prove useful in this field.

Quasi-equilibrium free energy methods

A relatively cheap class of free energy methods for studying nucleation is based on the assumption of quasi-equilibrium. Before discussing these methods, let's set the stage by introducing the basics of equilibrium statistical mechanics and collective variables (CVs). For a system in thermodynamic equilibrium at the NPT ensemble, a free energy profile as a function of a chosen collective variable λ can be expressed as

$$G(\lambda) = -k_B T \ln \int d\Omega \exp(-\frac{U(\Omega) + PV(\Omega)}{k_B T})\delta(\lambda(\Omega) - \lambda), \tag{1.9}$$

where the CV ($\lambda(\Omega)$) is constructed to classify each microstate Ω. Here we restrict the discussion within the cases with one CV in canonical ensemble for the simplicity of notation, although the extensions to other ensembles and to include more CVs are straightforward. For most cases, sampling the system according to the Boltzmann distribution is usually not sufficient for purpose of free energy estimation. For better sampling, a bias potential as a function of the CV λ is usually added to the original Hamiltonian,

$$H_{\text{biased}} = H + V(\lambda). \tag{1.10}$$

The bias potential $V(\lambda)$ can take various forms, be it a fixed umbrella potential [34], or an adaptive bias used in metadynamics simulations [35]. After obtaining the biased trajectory based on H_{biased}, a re-weighting procedure has to be performed to retrieve the original Boltzmann distribution. Any observable O as a function of the coordinates of the microstates Ω

can be estimated as

$$\langle O(\Omega)\rangle = \frac{\langle O(\Omega)e^{V(\lambda(\Omega))/k_B T}\rangle_{biased}}{\langle e^{V(\lambda(\Omega))/k_B T}\rangle_{biased}}, \tag{1.11}$$

averaged from configurations generated from the biased simulation.

During the process of the formation of the nuclei belonging to the more stable phase inside the metastable phase, the system is strongly out-of-equilibrium. However, one can make a reasonable assumption, in line with classical nucleation theory, which says the metastable phase is in an equilibrium before forming a critical nucleus and eventually transforming into the stable phase. In this way, the equilibrium free energy methods can be applied to obtain a free energy profile as a function of a chosen collective variable for the system in the metastable phase. Furthermore, in order to compute the nucleation rate constant, one can employ the Bennett-Chandler method, which consists a two-step procedure of first computing the free-energy barrier, and then measuring the kinetic pre-factor from molecular dynamics simulations constrained at the top of the free-energy barrier.

With the help of the quasi-equilibrium assumption and free energy methods, researchers have been able to perform atomistic simulations combined with biased sampling methods in which solid critical nuclei containing hundreds of atoms have been observed to form from the melt [9, 40, 43–53]. A recent comprehensive review article has compiled a list of such atomistic studies for systems including Lennard-Jones, colloids, atomic liquids, water, etc [10]. Much of these computational efforts have concentrated on investigating how the free energy changes with cluster size $G(n)$, and on verifying the accuracy of the CNT model. Some of these studies have found a good agreement between the CNT prediction in Eqn. (1.1) and the free energy profile for a cluster $G(n_\mathrm{s})$ that was computed from simulations [45, 52]. Others, meanwhile, have shown significant systematic differences between the two [49]. Once the free energy barrier of nucleation $G^\star = \max(G(n_\mathrm{s}))$ has been determined, the nucleation rate can be obtained from Eqn. (1.3). The addition rate f^+ can be computed as a diffusion coefficient from the mean square displacement of the cluster size after it is released at the top of the nucleation barrier [45, 54]. However, this approach assumes that $dG(n_\mathrm{s})/dn_\mathrm{s}$ is effectively zero when running multiple trajectories, is influenced by the choice of the initial configuration, and the latent heat created when the size of the nucleus changes also has an impact [55].

The limitations of atomistic studies of nucleation

In general, atomistic simulations of homogeneous nucleation offer tremendous advantages including exploration and characterization of detailed nucleation mechanism, elimination of the influence from impurities, and physical insights to the formulation and the prediction of the nucleation process. However, many challenges and controversies exists, and there is a considerable gap between the experimental measurements of nucleation rates and the theoretical predictions [10].

One challenge stems from the usual problems of atomistic simulations, e.g. reliability of the force fields, the slow dynamics of certain supercooled systems, finite size effects, etc. However, as we anticipated, a major challenge comes from the missing link between atomistic simulations and macroscopic models such as CNT: due to the diffuse nature of the interface, there is a degree of ambiguity in the location and area of the surface between phases and in the determination of the cluster size n_s. Furthermore, in a system of undercooled liquid, there exists a number solid clusters of different sizes as well as local fluctuations. In practice, when analyzing an atomistic model, one typically proceeds by first selecting an arbitrary order parameter that is able to distinguish between the atoms in each of the two phases. The atoms that are thus identified as being part of the more stable phase are then grouped into clusters [40, 45]. These heuristic procedures make the definition of the clusters size n_s, the associated free energy profile $G(n_\mathrm{s})$, and the size of the critical nucleus $n^\star$ ambiguous. If one then combines the values of these quantities with macroscopic nucleation theories such as CNT, in order to predict physical quantities such as the interfacial free energies and nucleation rates, inaccuracies will likely emerge. For example, from the previous computational predictions of homogeneous ice nucleation rates, even when the same water model and the same thermodynamic conditions are assumed, nucleation rates predicted in different studies typically differ by as much as 5-10 orders of magnitude [10]. This discrepancy is due to the fact that it is often necessary to evoke the standard form of classical nucleation theory (CNT) in order to estimate quantities such as the nucleation barrier, as the long time scale of nucleation rules out the option of brute force molecular dynamics simulations. However, a number of approximations within CNT have been shown to be over-simplifications [10], and more importantly, it is highly non-trivial to extract the values of the parameters that enter CNT using the microscopic quantities directly obtained from simulations. For instance, the diffuse nature of solid-liquid interfaces makes it difficult to rationalize and formulate the nucleation free energy in a unique and meaningful way, as the choice of the atomic order parameters used to distinguish ice structures affects the computed free energy profile and the size of the critical nucleus [3, 36, 56].

1.5 Outline of the thesis/ statement of contributions

In the present thesis, we are primarily focused on introducing a rigorous thermodynamic model that links macroscopic theories and atomic-scale simulations, in order to provide a simple and elegant framework for apply and expand classical nucleation theory. The thesis is organized as follows:

- In chapter 2, we cover the concept of free energy under the framework of statistical mechanics, and then describe two key free energy methods that are employed in the thesis. A part of chapter 2 is adapted from Ref. [57].
- A planar interface between two phases resembles the surface of a large nucleus. In chapter 3, we derive a computational method as well as a thermodynamic framework to

rigorously define and characterize the free energy of a solid-liquid interface at out-of-equilibrium conditions. This chapter is adapted from Ref. [1].

- In chapter 4, we address the issue of how to treat multiple nuclei in the metastable phase, by building a thermodynamic framework that reconcile the picture emerging from simulation with macroscopic theories of nucleation. This chapter is adapted from Ref. [2].
- In chapter 5, we discuss how to rigorously extract and analyze the various terms that contribute to the nucleation free energy profile, from atomstic simulations. This chapter is adapted from Ref. [3].
- In chapter 6, we demonstrate how to directly evaluate the Tolman length, which determines the leading order curvature dependence of the interfacial free energy between two phases, from simulations of planar solid-liquid interfaces. This chapter is adapted from Ref. [4].
- In chapter 7, we apply the thermodynamic framework as well as the computational methods that we developed to study the phenomenon of homogenous ice nucleation. This chapter is adapted from Ref. [5].
- Modelling water using empirical force fields is often not accurate enough for quantitatively predicting ice nucleation. To solve this, in chapter 8 we describe a new machine learning (ML) potential for ice and liquid water that is based on density functional theory. We show the performance of this ML potential and discuss how it can be used to model nucleation. This chapter is adapted from Ref. [6].

2 Free energy estimation methods

Free energy estimation is at the core of understanding nucleation, which is an activated process in which the system overcomes a free energy barrier to transition from a metastable state to a more stable one, as detailed in Chapter 1. Furthermore, the free energy of nucleation can be conveniently expressed as the sum of a bulk term stemming from the chemical potential difference between the stable and the metastable phase, and a surface term that is the product of interfacial free energy and surface area. As such, computing the free energy of nucleation as well as the chemical potential difference and surface energy is a key step in rationalizing nucleation phenomena using atomistic simulations. Yet, the implications of free energy estimation extend far beyond that: knowledge of the free energy is crucial in predicting the relative stability of different states of materials and molecules, and underlies a plethora of physical and chemical phenomena including phase diagrams, solubility, equilibrium concentration of defects, and so on. Due to the paramount importance of free energy estimation, this chapter is devoted to giving an overview of the tricks of the trade. This chapter is organized as follow: we first define free energy under the framework of statistical mechanics, then describe two key free energy methods, thermodynamic integration and enhanced sampling methods.

2.1 The fundamentals of free energy

Under the framework of thermodynamics the definition of free energy is somewhat convoluted: in 1873, Willard Gibbs described it as “the greatest amount of mechanical work which can be obtained from a given quantity of a certain substance in a given initial state, without increasing its total volume or allowing heat to pass to or from external bodies, except such as at the close of the processes are left in their initial condition” [58]. To put this in slightly more accessible terms, free energy is a measure of the available work (useful energy) a system can perform when transforming from one state to another. There are two subtle aspects in this statement:

(i) Free energy of a state is always a relative quantity with respect to an explicit or implicit reference state, and the free energy of a state compared to another is only meaningful if the system can be temporarily constrained to the initial state before a transition will happen. (ii) The initial and the final states are inevitably under certain thermodynamic conditions, and depending on these conditions the formulation of the free energy takes a different form. The Helmholtz free energy A is for the systems with fixed temperature and volume, and the Gibbs free energy G is associated with fixed temperature and pressure of the system. In what follows, we will use these notations, and use F to denote the free energy in a generic setting.

The concept of free energy is much clearer under statistical mechanics: The expression for the free energy of a system is closely related to the partition function, which in turn depends on the thermodynamic boundary conditions defining the ensemble. Under the canonical (NVT) ensemble, the partition function of a bulk system that has N *indistinguishable* particles and is contained in a volume V is given by [59, 60]

$$Q(N,V,T) = \frac{V^N}{\Lambda^{3N}N!}\int_{D(V)} \mathrm{d}\mathbf{q}\exp\left[-\frac{U(\mathbf{q})}{k_B T}\right], \tag{2.1}$$

where the potential energy U is a function of the atomic coordinates $\mathbf{q} = \{\mathbf{q}_{1\ldots N}\}$, $D(V)$ denotes the spatial domain defined by the containing volume [60], and $\Lambda = \sqrt{2\pi\hbar^2/mk_B T}$ is the thermal de Broglie wavelength. The expression for the absolute Helmholtz free energy of the system is thus

$$A(N,V,T) = -k_B T\ln Q(N,V,T) = -k_B T\ln\frac{V^N}{\Lambda^{3N}N!} - k_B T\ln\int_{D(V)} \mathrm{d}\mathbf{q}\exp\left[-\frac{U(\mathbf{q})}{k_B T}\right]. \tag{2.2}$$

Always bear in mind that the value of the free energy is always a relative quantity with respect to *something*, and in Eqn. (2.2) the implicit reference is the ideal gas state whose free energy can be expressed as the analytic expression $-k_B T\ln\frac{V^N}{\Lambda^{3N}N!}$.

When the isothermal-isobaric (NPT) ensemble is used instead, the system can be characterized by the absolute Gibbs free energy

$$G(N,P,T) = -k_B T\ln\left(\int dV\exp\left[-\frac{PV}{k_B T}\right]\frac{V^N}{\Lambda^{3N}N!}\int_{D(V)} d\mathbf{q}\exp\left[-\frac{U(\mathbf{q})}{k_B T}\right]\right), \tag{2.3}$$

or starting from the Helmholtz free energy,

$$G(N,P,T) = -k_B T\ln\int dV\exp\left[-\frac{PV}{k_B T}\right]\exp\left[-\frac{A(N,V,T)}{k_B T}\right]. \tag{2.4}$$

Computing directly the partition function for an arbitrary potential is impractical. In addition, only the difference in free energies $\Delta F_{A-B} = F(A) - F(B)$ between two systems or states has physical consequences and thus is related to experimental observables: Not only does ΔF_{A-B} determine the relative probability of finding state A compared to B at thermodynamic equilibrium, i.e.

$$\frac{\rho(A)}{\rho(B)} = e^{-\beta\Delta F_{A-B}}, \tag{2.5}$$

it also governs the transformation rates between A and B under the principle of detailed balance, i.e.

$$\frac{P(B \to A)_t}{P(A \to B)_t} = e^{-\beta\Delta F_{A-B}}, \tag{2.6}$$

where $P(B \to A)_t = P(\Omega_t \in A \mid \Omega_0 \in B)$ is the conditional probability of starting from a microstate belonging to B at time 0, and ending at another microstate belonging to A at time t. Eqn. (2.6) is satisfied under the detailed balance because the forward and back flux between A to B exactly cancels out when the system is in equilibrium, i.e. $\rho(A)P(A \to B)_t = \rho(B)P(B \to A)_t$.

According the Crooks fluctuation theorem [61], Eqn. (2.6) can be further extended to out-of-equilibrium situations involving non-reversible work,

$$\frac{P(B \to A \mid W_{B\to A})_t}{P(A \to B \mid W_{A\to B})_t} = e^{\beta(W_{B\to A}-\Delta F_{A-B})}, \tag{2.7}$$

where $W_{B\to A}$ is the work done on the system during the forward transformation of the system $B \to A$. The work is equal to the energy added to the system described by the original Hamiltonian H_0, through a controllable degree of freedom $\lambda(t)$, with $\lambda = 0$ indicates the system is in B and $\lambda = 1$ indicates the system is in A, i.e.

$$W_{B\to A} = \int_{0,\lambda_0=0}^{t,\lambda_t=1} \frac{d\lambda}{dt}\frac{\partial(H_0 + \Delta H(\lambda(t)))}{\partial\lambda} dt. \tag{2.8}$$

For each forward path $B \to A$, one can always find a reverse path $A \to B$ by simply performing a time reversal and flipping the momenta, and it is easy to see that $W_{A\to B} = -W_{B\to A}$ is the external work for each time-reversed path doing the backward transition from B to A. $P(B \to A \mid W_{B\to A})_t$ is the conditional probability of starting from B at time 0 and ending in A at time t, under the external work $W_{B\to A}$. There are numerous interpretations for the Crooks fluctuation theorem, personally I think of this as a way to state the equivalence of free energy and the potential to do useful work, and as such, applying an amount of external work $W_{A\to B}$ effectively changes the free energy difference ΔF_{A-B} and thereby biases the transformation rates between A and B.

If one applies external work $W_{B\to A}$ to the system such that all the paths that start at B at time 0 end in A at time t, the transformation rates $P(B \to A \mid W_{B\to A})_t$ and $P(A \to B \mid W_{A\to B})_t$ in Eqn. (2.7) will be fixed at unity. Under this condition, Eqn. (2.7) then directly implies the Jarzynski non-equilibrium work relation [62]:

$$\langle e^{-\beta W_{B\to A}}\rangle_{B\to A} = e^{-\beta\Delta F_{A-B}}, \tag{2.9}$$

where the ensemble average is taken from all the simple path (i.e. no re-crossing) for $B \rightarrow A$. The Jarzynski equality in Eqn. (2.9) is probably the most general results of statistical mechanics as the (local) equilibrium assumption is not required during the transition between two states. The only conditions required are: (i) at the beginning of the switching the system is locally-equilibrated in state B; and (ii) the external work $W_{B\rightarrow A}$ applied during the switching is done using the same *protocol*, which mean the controllable degree of freedom $\lambda(t)$ is varied at a pre-defined manner. For example, the most non-restrictive protocol is just imposing that $\lambda_0 = 0$ and $\lambda_t = 1$, and another possible one is requiring that $\lambda(t)$ changes linearly during the switching.

The two limiting cases of the Jarzynski equality are very instructive: (i) when the switching is infinitely fast, meaning that the Hamiltonian of the system is suddenly changed to H_A from H_B, the work done $W_{B\rightarrow A}$ is just the energy difference between the initial and the final microstate under the different Hamiltonians, and Jarzynski becomes the familiar expression of thermodynamic perturbation:

$$\langle e^{-\beta(H_A(\Omega)-H_B(\Omega))}\rangle_B = e^{-\beta\Delta F_{A-B}}, \tag{2.10}$$

where $\langle\ldots\rangle_B$ means the ensemble average over the equilibrium distribution of microstates Ω under the Hamiltonian H_B. On the other hands, at the limit where the driven system is always in equilibrium (i.e. the switching parameter $d\lambda/dt \equiv 0$ and $H_\lambda = (1-\lambda)H_B + \lambda H_A$) and $W_{B\rightarrow A}$ is reversible, the Jarzynski equality reduces to a statement for the second law of thermodynamics

$$\Delta F_{A-B} = \langle W^{\mathrm{rev}}_{B\rightarrow A}\rangle \leq \langle W_{B\rightarrow A}\rangle, \tag{2.11}$$

and the reversible work can be computed as

$$W^{\mathrm{rev}}_{B\rightarrow A} = \int_0^1 \langle\frac{\partial H(\lambda)}{\partial\lambda}\rangle_\lambda d\lambda. = \int_0^1 \langle H_A(\Omega) - H_B(\Omega)\rangle_\lambda d\lambda, \tag{2.12}$$

where $\langle\ldots\rangle_\lambda$ means the ensemble average of the equilibrium distribution under the Hamiltonian $H(\lambda)$.

In practice, the raw forms of the Crooks fluctuation theorem and the Jarzynski equality are seldom used for free energy estimation from atomistic simulations. The reason is, as can be seen from Eqn. 2.9, that computing the free energy difference using Jarzynski requires taking the average of an exponential, which usually converges poorly when the spread of the exponent is much larger than unity [63, 64]. In contrary, when quasi-equilibrium is assumed, Eqn. 2.12 which only involves a linear operation can be applied, resulting in much better statistical efficiency. As such, the commonly used free energy methods usually operate under the quasi-equilibrium assumption. There are two major classes of such methods:

- One class of standard free energy techniques, such as metadynamics, umbrella sampling, and transition path sampling [34, 65, 66], relies on the concept that the phase space of a system can be divided into a number of states using a choice of collective variables (or reaction coordinates), and the free energy difference between two states can then be computed by sampling both states, as well as the transition paths that connect them.

- Another class of free energy methods relies on the thermodynamic integration (TI) method [60, 67, 68]. The TI can be performed along a physical path for example along temperature or pressure, or via an unphysical path between the physical system and a reference system over a switching parameter λ. The TI method does not require a smooth transformation of the atomic coordinates, but just the evaluation of free-energy derivatives as a function of a change in the thermodynamic conditions.

Generally speaking, the thermodynamic integration method is particularly suitable for computing the free energies of bulk phases, and the enhanced sampling methods are better for systems undergoing a transition. In this chapter, we will describe these methods, with a particular emphasis on their application to condensed phase systems and the phenomena of nucleation.

2.2 The thermodynamic integration method

Note: The contents of this section is adapted from Ref. [57].

During a thermodynamic integration (TI), a *reversible* switching is performed along a thermodynamic path between two systems, during which the infinitesimal change in free energy along the path is computed and accumulated. The TI method is very versatile and flexible; it can be performed along a physical path, for example, along temperature or pressure, or via an unphysical path between the two systems with different Hamiltonian. In this section, we discuss on how to compute the absolute free energies of a solid system using the method of thermodynamic integration. We focus on solid systems because in the case of solidification, we are often interested in the free energy difference between different crystalline phases that can nucleate from the melt, as well as the free energy associated with crystallographic defects inside a nucleus. For the case of fluid systems, please refer to Ref. [67, 69].

2.2.1 A general workflow of thermodynamic integration for solid systems

In the case of a solid system, we found it effective to take a harmonic crystal reference, and follow the TI routes illustrated in Figure 2.1. In a nutshell, for evaluating the Helmholtz free energy A of a solid system, we propose to first integrate along λ between the harmonic and the real crystal and then do an integration with respect to the temperature T, which correspond to the yellow and the red arrows in Figure 2.1, respectively. To calculate the Gibbs free energy G of the system, we first obtain the Helmholtz free energy at a low temperature, switch from the NVT to the NPT ensemble, and finally do TI along the temperature T (the yellow, the blue,

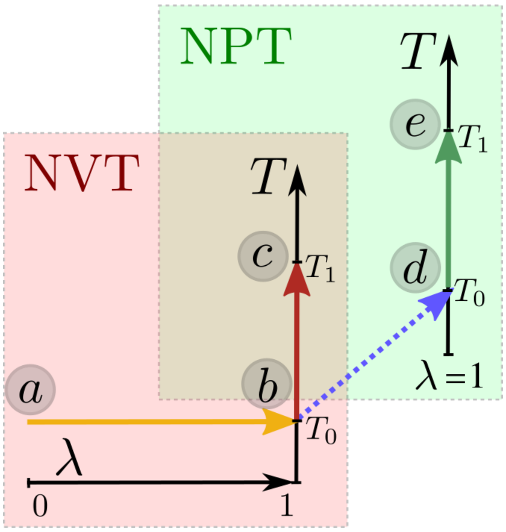

Figure 2.1 – An illustration of the different thermodynamic integration routes employed in the present paper. Under the canonical (NVT) ensemble, the yellow arrow indicates the switching between an harmonic reference system ($\lambda = 0$) and a real system ($\lambda = 1$), and the red arrow illustrates TI with respect to temperature. The dashed blue arrow shows the transformation between the Helmholtz free energy and the Gibbs free energy. The green arrow denotes TI over temperature under the isothermal-isobaric (NPT) ensemble.

and the green arrows in Figure 2.1). In the following, we discuss in detail how each step can be computed conveniently and efficiently.

Before we start the detailed discussion, note that in TI it is advantageous and often necessary to constrain the center of mass (CM) of the system. The Helmholtz free energy difference between the unconstrained and the constrained crystalline system under periodic boundary conditions can be expressed as [70]

$$\Delta A_{cm}(N, V, T) = -k_B T\left(\ln\frac{V}{N} + \frac{3}{2}\ln N + \ln\frac{1}{\Lambda^3}\right), \tag{2.13}$$

which can be considered as a finite size effect. Therefore, when we perform TI we focus solely on systems with fixed CM, and at the end of the calculation the term ΔA_{CM} can be added to retrieve the free energy of the unconstrained system, although at times the influence may be negligible. We will also discuss in more detail other finite-size effects in Section 2.2.1. In the other sections – since we will always work under the constant-number-of-particles framework in TI – we omit N when denoting thermodynamic states.

An absolute reference: the Helmoltz free energy of the Debye crystal

Strictly speaking, only the relative free energy of a system with respect to a reference can be defined without any ambiguity. The “absolute" free energy in this paper refers to the fact that the free energies of the chosen reference systems are analytic, and can be meaningfully compared between distinct reference systems including those that contain different numbers of particles.

A harmonically-coupled crystal of N atoms with a constrained center of mass constitutes a convenient reference system for a solid (point a in Figure 2.1). Taking the phonon frequency for the crystal to be $\{\omega_{i=1\ldots 3N-3}\}$, [1] one can obtain an expression for the classical free energy of such a Debye crystal at the temperature T_0:

$$A_{\mathrm{h}}(V,T_0) = k_B T_0 \sum_{i=1}^{3N-3} \ln \frac{\hbar\omega_i}{k_B T_0}. \tag{2.14}$$

Note that from the standpoint of performing thermodynamic integration, the reference can be any harmonic crystal that has the same number of particles as the real system. For instance, one could even take a reference in which all particles are independently coupled to the lattice sites with a constant spring term (i.e. an Einstein crystal) [68, 71, 72]. However, for better statistical efficiency, it is better to choose a reference harmonic crystal that has the same frequency modes and equilibrium configuration as the real crystal, both of which can be determined for example via local energy minimization followed by a diagonalization of the Hessian matrix [73].

The Helmoltz free energy of an anharmonic crystal

Starting from a reference crystal (a) with a known free energy, one can obtain the Helmholtz free energy of the real crystal (b) using thermodynamic integration in the NVT ensemble, as indicated by the yellow arrow in Figure 2.1. using a parameter λ to perform the switch between the harmonic Hamiltonian $\mathscr{H}_{\mathrm{h}}$ and the actual Hamiltonian $\mathscr{H}$. In practice, one should run multiple simulations with the hamiltonian $\mathscr{H}(\lambda) = (1-\lambda)\mathscr{H}_{\mathrm{h}} + \lambda\mathscr{H}$ at different values of λ, so as to switch between the harmonic Hamiltonian $\mathscr{H}_{\mathrm{h}}$ and the actual Hamiltonian $\mathscr{H}$ [71]. The free energy of the real system with a fixed CM can then be evaluated using

$$A(V,T_0) - A_{\mathrm{h}}(V,T_0) = \int_0^1 d\lambda\, \langle U - U_{\mathrm{h}} \rangle_{V,T_0,\lambda}\,, \tag{2.15}$$

where $\langle \ldots \rangle_{V,T_0,\lambda}$ denotes the ensemble average over NVT simulations using the Hamiltonian $H(\lambda)$.

In practice, to avoid severe statistical inefficiencies and singularities in the integral, one should perform this step at a low temperature T_0 when the system is quasi-harmonic and when diffusive or rotational degrees of freedom are completely frozen. If T_0 is sufficiently low and the real and the reference systems are very similar, one also has the option to evaluate $A(V,T_0) - A_{\mathrm{h}}(V,T_0)$ using the free energy perturbation method, eliminating the integration error altogether. One only needs to run simulations for the reference harmonic crystal, and obtain the free energy of the real system using

---

[1] The zero-frequency translational modes are excluded due to the constraint on CM

$$A(V,T_0) - A_{\mathrm{h}}(V,T_0) = -k_B T_0 \ln \langle \exp\left[-\frac{U-U_{\mathrm{h}}}{k_B T_0}\right]\rangle_{V,T_0,\lambda=0}, \tag{2.16}$$

where $\langle\ldots\rangle_{V,T_0,\lambda=0}$ denotes the ensemble average for the harmonic crystal at T_0 and V, and U and U_{h} denote the real and harmonic potentials, respectively. Note that in order to ensure the statistical efficiency of this perturbative approach, the standard deviation of $U - U_{\mathrm{h}}$ has to be of the order of $k_B T_0$ [63].

The Helmholtz free energy as a function of temperature

Thermodynamic integration from a Debye crystal to the fully-anharmonic potential tends to become very inefficient as the temperature increases. For this reason, it is often useful to perform a thermodynamic integration with respect to temperature from a low to a high temperature under the desired thermodynamic conditions. Let us start by discussing how to perform this step under the NVT ensemble, which is the process indicated by the red arrow (b to c) in Figure 2.1. The Helmholtz free energy of a system that has N atoms and a fixed CM can be expressed by the well-known thermodynamic integration expression

$$\frac{A(V,T_1)}{k_B T_1} = \frac{A(V,T_0)}{k_B T_0} - \int_{T_0}^{T_1} \frac{\langle U\rangle_{V,T} + \langle K\rangle_{V,T}}{k_B T^2} dT, \tag{2.17}$$

where $\langle U\rangle_{V,T}$ and $\langle K\rangle_{V,T}$ are the ensemble averages of the potential energy and the kinetic energy, respectively.

One way to improve the convergence of Eqn. (2.17) is to consider that the ensemble average of the classical kinetic energy of a system that has N atoms and a fixed CM is analytic: $\langle K\rangle_{V,T} = (3N-3)k_B T/2$. Furthermore, one can also consider that if the potential was harmonic, also $\langle U\rangle$ would take the same value. Thus, one can take

$$\langle \delta U\rangle_{V,T} = \langle U\rangle_{V,T} - A(V,0) - (3N-3)\frac{k_B T}{2} \tag{2.18}$$

that measures the temperature-dependent anharmonic part of the potential energy. Note that $A(V,0) = \langle U\rangle_{V,0}$. After performing analytically some of the integrals, Eqn. (2.17) becomes

$$\frac{A(V,T_1)}{k_B T_1} = \frac{A(V,0)}{k_B T_1} + \frac{A(V,T_0) - A(V,0)}{k_B T_0} - (3N-3)\ln\frac{T_1}{T_0} - \int_{T_0}^{T_1} \frac{\langle \delta U\rangle_{V,T}}{k_B T^2} dT. \tag{2.19}$$

In quasi-harmonic systems, one can further reduce the variance of the integrand. One can use again the analytical expression for $\langle K\rangle_{V,T}$, together with the virial theorem, to write

$$(3N-3)\frac{k_B T}{2} = \langle K\rangle = -\frac{1}{2}\sum_{i=1}^{N} \langle \mathbf{F}_i \mathbf{q}_i\rangle. \tag{2.20}$$

Here $\mathbf{q}_i$ and $\mathbf{F}_i$ are the position of atom i and the force vector acting on it. Since the average force $\langle \mathbf{F}_i \rangle$ is zero, one can also add an arbitrary reference position $\hat{\mathbf{q}}_i$, and write [74, 75]

$$\langle \delta U \rangle_{V,T} = \langle U + \frac{1}{2} \sum_{i=1}^{N} \mathbf{F}_i (\mathbf{q}_i - \hat{\mathbf{q}}_i) \rangle_{V,T} - A(V,0) \tag{2.21}$$

If the potential energy surface is perfectly harmonic, and if one takes $\hat{\mathbf{q}}_i$ equal to the equilibrium position of atom i, it is easy to verify that the virial term would cancel completely the fluctuations in the potential energy. Even if the potential is quasi-harmonic, as long as it is not diffusive even at high temperatures, the use of the virial reference and Eqn. (2.21) can substantially improve the statistical efficiency in the estimation of $\langle \delta U \rangle_{V,T}$. However, when the motion of atoms is strongly anharmonic or diffusive, an atom can start vibrating around a different equilibrium position, and in that case the statistical efficiency of the straightforward expression Eqn. (2.18) is better.

Besides improving the convergence of each temperature window, one can try to improve the accuracy and the efficiency of the TI procedure by choosing wisely the discretization points or, equivalently, by performing a change of variables that yields a smoother integrand [76]. In this case, it is convenient to perform a change of variables that ensures that the statistical error in the integrand is roughly constant at all temperatures. Assuming the temperature dependence of the fluctuations in the anharmonic potential energy is similar to its harmonic counterpart, i.e. $\langle \delta U^2 \rangle_{V,T} - \langle \delta U \rangle_{V,T}^2 \sim T$, the required change of variable is $y = \ln(T/T_0)$, which transforms the integral into the form

$$\int_{T_0}^{T_1} \frac{\langle \delta U \rangle_{V,T}}{T^2} dT = \int_0^{\ln(T_1/T_0)} \frac{\langle \delta U \rangle_{V,T_0 e^y}}{T_0 e^y} dy. \tag{2.22}$$

In other words, one should select temperatures that are equally spaced in $\ln(T)$ in simulations. Coincidentally, this selection is also optimal for performing replica exchanges between the systems at different temperatures [77] - which should be done whenever possible as it will greatly benefit statistical convergence.

Another advantage of performing parallel tempering is that it requires sufficient overlap between adjacent replicas at temperatures T_i and T_{i+1}. Under these circumstances, $\langle U \rangle_{V,T}$ for $T_i < T < T_{i+1}$ can be evaluated via re-weighting, such as

$$\langle U \rangle_{V,T} = \frac{\langle U \exp\left[-\frac{U}{k_B} (\frac{1}{T} - \frac{1}{T_i}) \right] \rangle_{V,T_i}}{\langle \exp\left[-\frac{U}{k_B} (\frac{1}{T} - \frac{1}{T_i}) \right] \rangle_{V,T_i}}. \tag{2.23}$$

The integral in Eqn. (2.19) can thus be solved analytically to give an exact (within statistical uncertainty) expression for the contribution to the integral from the $[T_i, T_{i+1}]$ window:

$$\frac{A(V, T_{i+1})}{k_B T_{i+1}} - \frac{A(V, T_i)}{k_B T_i} = -\frac{3N-3}{2} \ln \frac{T_{i+1}}{T_i} - \ln \langle \exp\left[-\frac{U}{k_B} (\frac{1}{T_{i+1}} - \frac{1}{T_i}) \right] \rangle_{V,T_i}, \tag{2.24}$$

which effectively turns the thermodynamic integration formalism into a sequence of free energy perturbations, eliminating completely the integration error.

From the Helmholtz free energy to the Gibbs free energy

More often than not, the isothermal–isobaric ensemble (NPT) provides a more natural framework to describe the thermodynamic conditions of real systems than the NVT ensemble. However, the harmonic crystal (a in Figure 2.1) that was used as the absolute constant-volume reference in previous sections does not extend naturally to the NPT ensemble because its pressure is not well-defined (e.g. the Einstein crystal is a system of independent particles) [78]. As a result, it is not convenient to perform TI with respect to λ from the reference crystal to the real crystal under the NPT ensemble. One way to avoid NPT simulation involves performing multiple simulations at different constant volumes, and then computing the Gibbs free energy and the equilibrium volume of the system by evaluating explicitly the integral (2.4). This is often done using a harmonic expression for the free energy at the different volumes, leading to the so-called quasi-harmonic approximation (QHA) [79, 80]. Alternatively, the equilibrium volume can be computed by performing a single NPT simulation at the desired temperature, and then $A(\langle V\rangle_{P,T}, T)$ is used as a proxy for $G(P,T)$ [78]. In this section, we argue that the transformation between the Helmholtz free energy and the Gibbs free energy, which is the process marked by the dashed blue arrow in Figure 2.1, can be conducted rigorously. This process effectively allows us to convert at will between the two ensembles when performing thermodynamic integrations with respect to ensemble temperature.

The expression for the Gibbs free energy of a system as an integral over the Helmholtz free energy is given by Eqn. (2.4). This expression can be combined with that for the distribution of volume fluctuations for the system under the NPT ensemble

$$\rho\,(V|P,T) = \frac{\exp\left[-\frac{PV}{k_B T}\right]\exp\left[-\frac{A(V,T)}{k_B T}\right]}{\int dV \exp\left[-\frac{PV}{k_B T}\right]\exp\left[-\frac{A(V,T)}{k_B T}\right]},, \tag{2.25}$$

which is just the normalized probability of observing the system to have instantaneous volume V in a simulation under constant P and T. We can then write

$$G(P,T) = A(V,T) + PV + k_B T \ln \rho\,(V|P,T)\,, \tag{2.26}$$

which is valid for arbitrary V.

In practice, one can run NPT simulations for a system and compute $\rho\,(V|P,T)$ just by accumulating the histogram of the instantaneous volume of the system. After that, one can select a volume V, preferably the one that maximizes $\rho\,(V|P,T)$ for the sake of better statistical efficiency in the determination of $\rho\,(V|P,T)$, and compute $A(V,T)$ for the same system at that volume using the route a to b in Figure 2.1. Finally, the Gibbs free energy can be obtained applying Eqn. (2.26).

For a solid system, in order to avoid residual strain and elastic energy, one can vary the shape of the simulation cell instead of using a fixed shape in NPT simulations under a hydrostatic pressure [81]. To account for the degree of freedom associated with the variable cell, in this case Eqn. (2.26) should be modified to read

$$G(P,T) = A(\mathbf{h},T) + P\det(\mathbf{h}) + k_B T \ln \rho\,(\mathbf{h}|P,T)\,, \tag{2.27}$$

where **h** is a matrix that represents the dimensions of a simulation cell, and $A(\mathbf{h},T)$ is the free energy of the system evaluated at constant cell dimensions.

The Gibbs free energy as a function of temperature

Having converted a harmonic-reference Helmoltz free-energy to a constant-pressure Gibbs free energy at a given temperature T_0, one can easily perform a thermodynamic integration over temperature in the NPT ensemble (a path indicated by the green arrow in Figure 2.1). For a system with N atoms and a restricted CM, the expression reads

$$\frac{G(P,T_1)}{k_B T_1} = \frac{G(P,T_0)}{k_B T_0} - \int_{T_0}^{T_1} \frac{\langle U\rangle_{P,T} + (3N-3)\dfrac{k_B T}{2} + P\,\langle V\rangle_{P,T}}{k_B T^2}\,dT, \tag{2.28}$$

where $\langle U\rangle_{P,T} + (3N-3)\dfrac{k_B T}{2} + P\,\langle V\rangle_{P,T}$ is the enthalpy. Starting from this expression, one can apply all the techniques mentioned in Section 2.2.1, for example one can take

$$\langle \delta H\rangle_{P,T} = \langle U\rangle_{P,T} + P\,\langle V\rangle_{P,T} - G(P,0), \tag{2.29}$$

where $G(P,0) = \langle U\rangle_{P,0} + P\,\langle V\rangle_{P,0}$. Doing the integration in Eqn. (2.28) explicitly leaves

$$\frac{G(P,T_1)}{k_B T_1} = \frac{G(P,0)}{k_B T_1} + \frac{G(P,T_0) - G(P,0)}{k_B T_0} \quad - (3N-3)\ln\frac{T_1}{T_0} - \int_{T_0}^{T_1} \frac{\langle \delta H\rangle_{P,T}}{k_B T^2}\,dT. \tag{2.30}$$

In addition, one can also use the virial theorem (Eqn. (2.20)), the change of variable in the integration (Eqn. (2.22)), and parallel tempering to further accelerate the convergence. When performing parallel tempering, one can eliminate the thermodynamic integration error by using a free-energy perturbation to compute the increment of G between two replicas at temperatures T_{i+1} and T_i:

$$\frac{G(P,T_{i+1})}{k_B T_{i+1}} - \frac{G(P,T_i)}{k_B T_i} = -\frac{3N-3}{2}\ln\frac{T_{i+1}}{T_i} - \ln\left\langle \exp\left[-\frac{U+PV}{k_B}\left(\frac{1}{T_{i+1}} - \frac{1}{T_i}\right)\right]\right\rangle_{P,T_i}. \tag{2.31}$$

Finite size effects

Most of the time, one is interested in computing the free energy per atom of a bulk, infinite system, or the excess free energy of a defect in the dilute limit. An atomistic simulation, however, is inevitably restricted to a finite system size, which can result in deviations from the ideal case. Many of these finite-system-size effects have been documented in the literature. First of all, in the limit of small system size, the free energy of the system is not an extensive quantity. Taking the ideal gas part of the Helmholtz free energy $A_{id}(N,V,T) = -k_B T \ln \frac{V^N}{\Lambda^{3N} N!}$ in Eqn. (2.2) for example, one can see that

$$\frac{A_{id}(N,V,T)}{N k_B T} = 1 - \ln \frac{V}{N} + \ln \frac{1}{\Lambda^3} - \frac{\ln N}{2N} + \mathcal{O}\left(\frac{1}{N}\right) \tag{2.32}$$

using Stirling's formula. The leading $\ln N/N$ term is a well-documented finite size effect that reduces to zero in the thermodynamic limit [70, 82]. The Gibbs free energy per atom $G(N,P,T)/N$ as well as $k_B T \ln \rho\,(V|P,T)\,/N$ in Eqn. (2.25) also displays a similar dependence on $\ln N/N$. Constraining the center of mass of the system in simulations also introduces a non-extensive correction to the free energy,

$$\Delta A_{cm}(N,V,T) = A(N,V,T) - A_{cm}(N,V,T), \tag{2.33}$$

where A_{cm} denotes the Helmholtz free energy of the system with fixed center of mass [67, 70]. Fortunately, it is easy to correct for this part, because the expression for ΔA_{cm} in Eqn. (2.13) is analytic and trivial to compute. More subtle sources of finite size effects come from the cutoff of potentials, and from the discretization of the vibrational phonon spectrum due to the size of the supercell in simulations [78]. To help with this issue, there are interpolation techniques that help accelerate the convergence of the computed phonon dispersion relation [83].

It is worth stressing that for system sizes that can be reached easily in simulations using empirical force fields, finite-size effects may not be significant. However, one should always be aware of their presence and check for system-size convergence, particularly in ab initio calculations where the number of atoms that can be simulated is highly restricted. To minimize the impact of finite-size effects one should always compare free energies between systems of similar sizes, to benefit for a (partial) error cancellation.

2.2.2 One example of applying the thermodynamic integration method: vacancy formation energy in BCC iron

Here we provide a concrete example for the application of the thermodynamic integration workflow detailed above, in order to paint a more intuitive picture.

A vacancy is a type of point defects in a crystal, in which an atom is removed from one of the lattice sites. At any given temperature and pressure up to the solid-liquid coexistence line, an equilibrium concentration $\exp[-G_\mathrm{v}/k_B T]$ of vacancies exists, where G_v is the Gibbs free energy of a vacancy. Often, particularly in materials produced by fast quenching, a non-equilibrium concentration of vacancies can persist at low temperature, which can play an important role in technologically relevant solid-state transformations [84].

We studied a BCC iron system using a widely used EAM potential [85, 86]. This potential was fitted with the BCC vacancy formation energy at 0 K but lacks a thermally stable FCC phase [85, 87]. Iron exhibits phase transitions between BCC α-iron, FCC γ -iron and a BCC δ-phase when increasing the temperature at ambient pressure, which are largely due to the magnetism of the material [87]. Since the stabilization of the austenitic phase is due to quantum mechanical effects and this EAM potential does not reproduce it, we neglect the FCC phase in the present study, and performed the simulations considering a perfect BCC crystal (250 atoms) and a BCC crystal with a vacancy (249 atoms). In all the simulations, the centers of mass of the systems were constrained.

At high temperatures, the computation of the free energy of the crystal with a vacancy is particularly problematic using the integration over λ in Eqn. (2.15) due to the onset of diffusion in simulations [88]. This difficulty is circumvented here as we performed the integration from the harmonic crystal to the real crystal at a low temperature T_0 =100 K and at the equilibrium cell size (the yellow arrow in Figure 2.1). This harmonic crystal has the same phonon modes and Hessian matrix as the real system [73, 89]. Note that at this step the Helmholtz free energy of a crystal with a vacancy that sits at a fixed lattice site is computed, as the vacancy does not diffuse during the simulations at T_0 =100 K. After that, we switched to the NPT ensemble (the blue arrow in Figure 2.1), and ran simulations at different temperatures and zero hydrostatic pressure, using stochastic velocity re-scaling for temperature control and the anisotropic Nose-Hoover barostat to vary the dimensions of the orthorhombic periodic supercell [90, 91]. During this step, we obtained the temperature dependence of the free energies using Eqn. (2.28) (the green arrow in Figure 2.1). At high temperatures, the vacancy does diffuse but diffusion does not change the values of $\langle \delta H \rangle_{P,T}$ compared with the case when the vacancy is fixed at one site, due to the translational symmetry of the lattice. As such, the Gibbs free energy of a crystal with a fixed vacancy was obtained after integration using Eqn. (2.28).

The Gibbs free energy of a fixed vacancy in the crystal can be expressed as

$$G_{\mathrm{v}} = G_{\mathrm{vacancy}} - \frac{N_{\mathrm{vacancy}}}{N_{\mathrm{perfect}}} G_{\mathrm{perfect}}. \tag{2.34}$$

We used three different methods to estimate this quantity – namely a minimum-potential energy calculation, a harmonic free-energy estimate and the fully anharmonic TI – and plot the results in Figure 2.2 as a function of temperature. Note that, for a given interatomic potential and within the statistical errors, the thermodynamic integration method gives access to the full Gibbs free energy, and can be considered as the “ground truth”. The other commonly used approximations, on the other hand, rest on different assumptions: the harmonic approximation method assumes that anharmonicity is negligible, while the minimum-potential energy method neglects both the anharmonicity and entropic contributions. By comparing

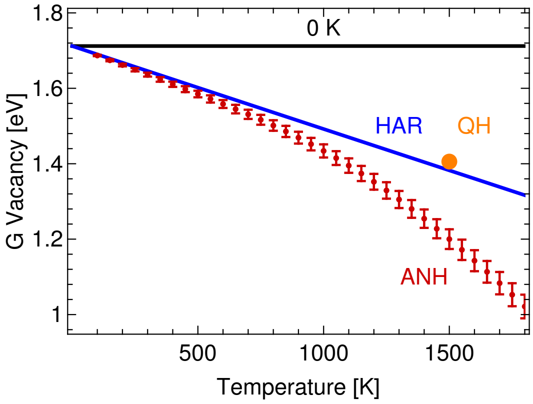

Figure 2.2 – The Gibbs free energy associated with a fixed vacancy in BCC iron estimated using potential energy difference (PE) at 0 K, the harmonic approximation (HAR), and the thermodynamic integration method that considers anharmonicity (ANH). QH indicates the quasi-harmonic approximation using the equilibrium configuration at 1500 K. Statistical uncertainties are indicated by the error bars.

the predictions from these approximation methods to the accurate values computed using thermodynamic integration, we can provide a representative benchmark of the accuracy of these approximations.

The difference between the predictions from the harmonic approximation and the TI is largely negligible at low temperatures, but becomes significant when the temperature approaches the melting point 1772 K for this EAM potential system [85]. To investigate whether this difference stems from a shift in the phonon spectra due to lattice expansion, or from anharmocity, we analyzed the vibrational modes $\{\omega'_{i=1...3N-3}\}$ using the equilibrium configuration of each system at 1500 K, computed the vacancy free energy under this quasi-harmonic approximation, and plotted the result as the orange dot in Figure 2.2. It can be seen that the harmonic contribution at 1500 K cannot explain the difference, and therefore the difference is mainly due to anharmonic effects.

2.3 Enhanced sampling methods

Now we move on to describe the second class of free energy methods, with a common feature that these methods rely on a choice of collective variables (CV) which are smooth functions of the atomic coordinates. We will restrict the discussion within the cases with one CV in canonical ensemble, although the extensions to other ensembles and to include more CVs are straightforward.

For all methods to work, a CV λ is constructed first. Because the momentum distribution of classical systems is analytic (Eqn. 2.2), usually one focuses on the distribution in the configurational space. As such, although λ can be a function of the momenta of the particles

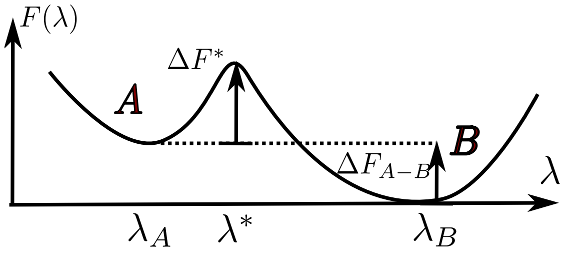

Figure 2.3 – A schematic of a free energy profile as a function of a collective variable λ.

in the system, usually one selects λ to be a function of the atomic coordinates (i.e. $\lambda = \lambda(\mathbf{q})$). Clearly, there are numerous ways to select this function but not all are equally helpful. The construction of CVs is very much of an art, as no clear cut way exists and in most cases it is highly system dependent. As a rule of thumb, for a system that can adopt states (say A and B) that have macroscopically-observable distinctions, a good CV separates the two states as much as possible as well as characterizes the transition states in between. More details on the construction of CVs are postponed to when specific examples are discussed.

For now we assume a good choice of the CV has already been made for a certain system, and the free energy profile (as schematically plotted in Fig. 5.1) can thus be computed as

$$F(\lambda) = -k_B T \ln \frac{\int_{D(V)} \mathrm{d}\mathbf{q} \delta(\lambda(\mathbf{q}) - \lambda) \exp\left[-\dfrac{U(\mathbf{q})}{k_B T}\right]}{\int_{D(V)} \mathrm{d}\mathbf{q} \exp\left[-\dfrac{U(\mathbf{q})}{k_B T}\right]}, \tag{2.35}$$

where $\delta(\ldots)$ denotes a Delta function. For comparing the free energies of state A and B, one first use the CV to identify whether a microstate belongs to state A or B, such as

$$\begin{cases} \lambda < \lambda^*, & \Omega \in A \\ \lambda > \lambda^*, & \Omega \in B \end{cases} \tag{2.36}$$

After that, the free energy difference can be expressed as

$$\Delta F_{A-B} = -k_B T \ln \frac{\int_{-\infty}^{\lambda^*} \mathrm{d}\lambda \exp(-\beta F(\lambda))}{\int_{\lambda^*}^{\infty} \mathrm{d}\lambda \exp(-\beta F(\lambda))} \tag{2.37}$$

In many cases, however, if the wells of A and B are deep and the curvatures at the bottoms of the wells (λ_A and λ_B) are similar, it is sufficient to write

$$\Delta F_{A-B} \approx F(\lambda_A) - F(\lambda_B). \tag{2.38}$$

Finally, it is worth stressing that the free energy profile $F(\lambda)$ is dependent on the arbitrary choice of the function $\lambda(\mathbf{q})$, so it is not a physical observable per se. One can easily verify that if a non-linear transformation $\lambda \rightarrow \lambda'$ is applied, not only the shape of the free energy profile will be changes, the free energy difference computed from Eqn. (2.38) as well as the transition barrier F^* in Fig. 5.1 will change. One thus has to bear in mind what are the actual physical observables and how to obtain these observables from the free energy profile.

2.3.1 Free energy estimation from unbiased and biased molecular dynamics

In a MD simulation, F_{A-B} can be approximated from the histogram of the configurations that belong to the two states

$$\Delta F_{A-B} \doteq -\frac{1}{\beta} \ln \frac{\sum_\Omega \delta(A(\Omega))}{\sum_\Omega \delta(B(\Omega))}, \tag{2.39}$$

provided that the trajectory is long enough to travel back and forth between the two states for a number of times. In conventional MD simulations, however, this estimator is only useful if ΔF_{A-B} and the transition barrier are both on the same order of the thermal fluctuations.

For most cases, sampling the system according to the Boltzmann distribution is usually not sufficient for purpose of free energy estimations. To enhance sampling, a bias potential as a function of the CV λ is usually added to the original Hamiltonian,

$$H_{\text{biased}} = H + V(\lambda). \tag{2.40}$$

If one selects a bias potential that takes the form of a parabolic function, i.e. $V(\lambda) = \kappa(\lambda - \lambda_0)/2$, the biasing scheme is usually called umbrella sampling [34]. If one selects $V(\lambda)$ to be a time-dependent function and is also adaptive, the scheme typically fits into the category of metadynamics [65].

After obtaining the biased trajectory based on H_{biased}, a re-weighting procedure has to be performed to retrieve the original Boltzmann distribution, and any observable O can be estimated as

$$\langle O(\Omega)\rangle = \frac{\langle O(\Omega) e^{\beta V(\lambda(\Omega))}\rangle_{biased}}{\langle e^{\beta V(\lambda(\Omega))}\rangle_{biased}}, \tag{2.41}$$

averaged from configurations generated from the biased simulation. Taking O to be the Delta function $\delta(A(\Omega))$ or $\delta(B(\Omega))$, one arrives at

$$\Delta F_{A-B} \doteq -\frac{1}{\beta} \ln \frac{\sum_\Omega e^{\beta V(\lambda(\Omega))} \delta(A(\Omega))}{\sum_\Omega e^{\beta V(\lambda(\Omega))} \delta(B(\Omega))}. \tag{2.42}$$

2.3.2 An example of enhanced sampling methods: metadynamics

Although the central formulations of the enhanced sampling methods are more or less similar, there are a large spectrum of different designs. Metadynamics is an adaptive method that operates at quasi-equilibrium conditions, making it efficient and easy to implement. It works by adding a time-depend bias $V(\lambda, t)$ to discourage the system from re-visiting the already explored CVs space. During the simulation, the bias is built as a sum of Gaussians of width σ deposited along the trajectory of λ with the deposition rate ω:

$$V(\lambda, t) = \int_0^t dt\, \omega \exp\left(-\frac{(\lambda - \lambda(t))^2}{2\sigma^2}\right). \tag{2.43}$$

In the well-tempered variant, the bias deposition rate reduces with time such as

$$\omega = \omega_0 e^{-\frac{V(\lambda,t)}{k_B \Delta T}}, \tag{2.44}$$

where $\Delta T + T$ is the effective temperature at which the CV is sampled. In the adaptive version, the Gaussian width σ also varies according to diffusional or geometrical considerations [92].

Most of the time, metadynamics is regarded as a free energy method based on the equilibrium distribution of the states. The effect of the external bias on any system observables can be removed by applying Eq. (2.41). On the other hand, it can also be interpreted as a transition path sampling method; the bias $V(\lambda, t)$ provides an acceleration factor to help the system travel to the transition states from the energy basin. Taking state B as the equilibrium state and state A as the transition state, Bennett-Chandler rate theory assumes,

$$P(B \to A, t) = \kappa(t)\phi(\lambda_A)e^{-\beta F_{A-B}}, \tag{2.45}$$

where $\phi(\lambda_A)$ is the effective velocity of the CV on the top of the transition state

$$\phi = \frac{\langle \delta(\lambda(q) - \lambda_A)\, |\mathbf{m}^{-\frac{1}{2}} \dfrac{\partial \lambda(\mathbf{q})}{\partial \mathbf{q}}| \rangle}{\sqrt{2\pi\beta}\langle \delta(\lambda(q) - \lambda_A)\rangle}, \tag{2.46}$$

where $\mathbf{q}$ and $\mathbf{m}$ indicate the coordinates and the mass matrix. When $W_{B\to A} < F_{A-B}$ and the process is still thermally activated, the above equation can be extended as

$$P(B \to A \mid W_{B\to A})_t = \kappa(t)\phi(\lambda_A)e^{-\beta F_{A-B}} e^{\beta W_{B\to A}} = P(B \to A)_t e^{-\beta(V(\lambda_A,t) - V(\lambda_B,0))} \tag{2.47}$$

and the last steps follows the definition of work in Eq. (2.8). This equation implies that the real transition rate is related to the rate in metadynamics by a simple scaling factor.

Because of the above mentioned versatility of metadynamics, it will be the workhorse in the present project. Fast setups of metadynamics simulations together with other biasing protocols were made possible by the PLUMED software [93].

2.3.3 Constructing a collective variable for distinguishing solid from liquid

As mentioned before, a good CV is highly system dependent. For the case of solid clusters nucleating from bulk liquid, a good CV should be highly correlated with the proportion of

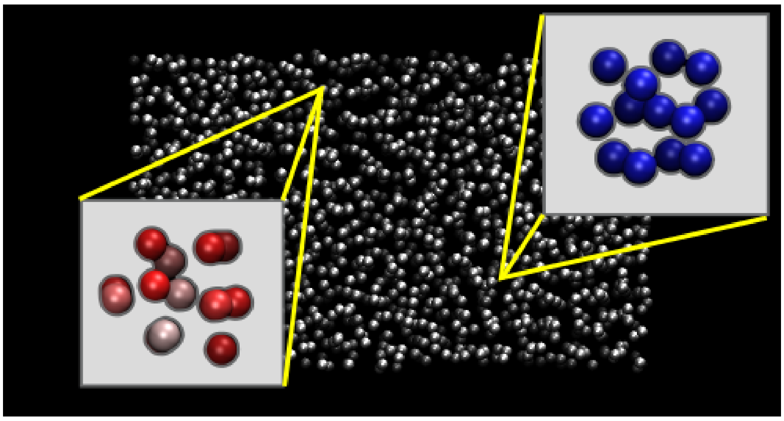

Figure 2.4 – A schematic for selecting and distinguishing local atom environments in a microstate.

the solid phase inside the whole system. It can be a thermodynamic variable such as energy, volume, proxy of entropy [94], etc. In many cases, it is advantageous to build a CV from local order parameters (Fig. 2.4): (i) Firstly, take each environment $\{r_{ij}\}$ centered at each atom i, (ii) For each atom-centered environment, compute an atomic order parameter $\phi_i = \phi(\{r_{ij}\})$. At the atomistic level, several families of order parameter ϕ_i can be used to determine whether the environment around atom i is solid-like or liquid-like [95]. In the next section, we will introduce a modified version of the cubic harmonic order parameter. (iii) Finally, for each microstate, we construct a global CV $\Phi = \sum_i \phi(i)$ by summing over all the atoms.

The PLUMED [93] plug-in provides a simple and efficient way to compute CVs based on local order parameters: At each step of the simulation, the PLUMED software receives the atomic configuration from LAMMPS, computes local order parameters and global collective variables, and returns an external bias potential to the MD code.

2.3.4 An example of local order parameter: FCCUBIC

Here we described a cubic-harmonic-based order parameter ϕ, which discriminates between the liquid, the solid, and the various different crystal orientations. This order parameter was first introduced in Ref. [96], and we latter refined it by re-parameterization and applying a hyperbolic switching function.

This local order parameter ϕ for each atom i depends on the position of its nearest neighbors. To single out the first-shell neighbors, we introduce a radial cutoff function between each pair of atoms i and j:

$$c_r(r_{ij}) = \begin{cases} 1 & r_{ij} \le r_1 \\ 0 & r_{ij} \ge r_0 \\ (y-1)^2(1+2y) & r_1 < r_{ij} < r_0 \end{cases} \tag{2.48}$$

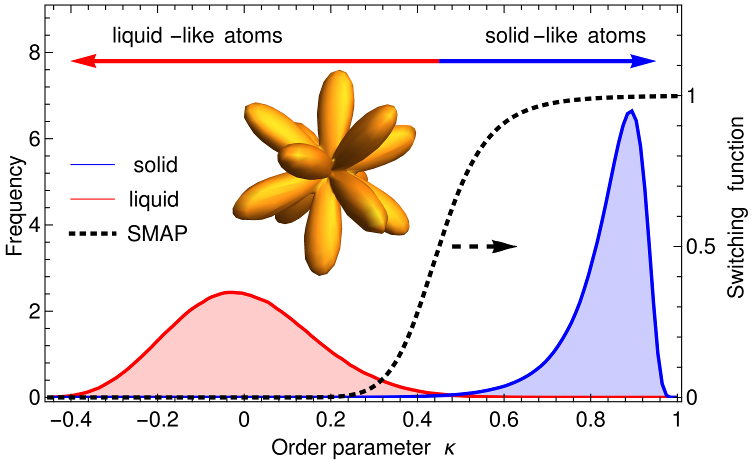

Figure 2.5 – The distributions of κ for a bulk *fcc* crystal oriented with the ⟨100⟩ direction parallel the axes of the simulation cell, and for the bulk liquid. All distributions have been computed at $T = T_\mathrm{m}$. SMAP indicates the sketch-map switching function [97] that we used to define $\phi = S(\kappa)$.

where $y = (r_{ij} - r_1)/(r_0 - r_1)$. We then compute an angular term $c_\alpha(r_{ij})$ to identify whether atom j sits on an *fcc* lattice point relative to i. This angular term has a functional form inspired by the cubic harmonics

$$c_\alpha(\mathbf{r}_{ij}) = (x^4y^4 + y^4z^4 + x^4z^4)/r_{ij}^8 - ax^4y^4z^4/r_{ij}^{12}, \tag{2.49}$$

Fig. 2.5 shows the shape of the above cubic harmonic function function. The parameter $\alpha = 27$ has been changed from the original value used in Ref. [96]. This new value allows us to differentiate more clearly between different orientations of the *fcc* environment. This is useful as c_α is not rotationally invariant. It can thus be used in conjunction with a rotation matrix to ensure that an *fcc* crystal is only detected when it has a particular orientation, that can be specified using a set of Euler angles (ϕ, ψ, θ), i.e.,

$$c_\alpha^{(\phi,\psi,\theta)}(\mathbf{r}_{ij}) = c_\alpha(\mathbf{R}^{(\phi,\psi,\theta)} \cdot \mathbf{r}_{ij}), \tag{2.50}$$

where

$$\mathbf{R}^{(\phi,\psi,\theta)} = \begin{vmatrix} \cos\phi\cos\psi - \cos\theta\sin\psi\sin\phi & \cos\phi\sin\psi + \cos\theta\cos\psi\sin\phi & \sin\phi\sin\theta \\ -\sin\phi\cos\psi - \cos\theta\sin\psi\cos\phi & -\sin\phi\sin\psi + \cos\theta\cos\psi\cos\phi & \cos\phi\sin\theta \\ \sin\theta\sin\psi & -\sin\theta\cos\psi & \cos\theta \end{vmatrix}. \tag{2.51}$$

The un-scaled local order parameter κ^0 is then defined as

$$\kappa^0(i) = \frac{\sum\limits_{i \neq j} c_r(r_{ij}) c_\alpha(r_{ij})}{\sum\limits_{i \neq j} c_r(r_{ij})}. \tag{2.52}$$

This quantity is then linearly scaled using $\kappa(i) = (\kappa^0(i) - \kappa^0_l)/(\kappa^0_s - \kappa^0_l)$ to ensure that the perfect *fcc* lattice corresponds to $\kappa = 1$, while the average κ value for atoms in bulk liquid is zero. The value of $\kappa^0(i)$ in the perfect *fcc* crystal with the desired orientation $\kappa^0_s = 1/16$ can be readily computed by summing over the positions of the ideal positions of nearest neighbors. It is also possible to analytically integrate the value of $\kappa^0(i)$ in the bulk liquid, as one can safely assume that the distribution of $\mathbf{r}_{ij}$ is uniform in the polar directions, i.e.,

$$\kappa^0_l = \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \, c_\alpha(r_{ij}(\phi, \theta, r = 1)) = \frac{143 - a}{5005}. \tag{2.53}$$

For a given choice of a, $\kappa(i) = \frac{80080}{2717+16a}\kappa^0(i) + \frac{16(a-143)}{2717+16a}$.

Figure 2.5 shows that the atoms in bulk *fcc* crystal and bulk liquid have a minimal amount of overlap in the distribution of the local order parameter κ. Consequently, a tunable sigmoid switching function [98] can be used to map each atom onto the solid-like or the liquid-like regime, i.e.,

$$\phi(i) = S(\kappa(i)) = 1 - (1 + (2^{a/b} - 1)(\kappa(i)/R_0)^a)^{-b/a}, \tag{2.54}$$

The switching function used in this work is shown in figure 2.5. It has $a = 8$, $b = 8$, and $R_0 = 0.45$. These values were selected so that $\phi = 0.5$ corresponds to the crossover point at which the κ distributions for the atoms belonging to the bulk crystal and the bulk liquid cross.

3 The interfacial free energy of solid-liquid interfaces

As detailed in Chapter 1, CNT assumes that the free energy of a solid cluster $G(n_\mathrm{s})$ in homogeneous nucleation can be expressed as the sum of a bulk and a surface term [9] i.e.

$$G(n_\mathrm{s}) = \mu_\mathrm{sl} n_\mathrm{s} + \gamma A(n_\mathrm{s}), \tag{3.1}$$

where $\mu_\mathrm{sl} = \mu_\mathrm{s} - \mu_\mathrm{l}$ is the difference in chemical potential between the solid and the liquid, and the second surface term is the product of a specific free energy excess term γ, and the extensive surface area $A(n_\mathrm{s})$.

However, in an atomistic simulation of homogeneous nucleation, it is difficult to extract the various terms that enter Eqn. (3.1) for the following reasons: (i) In a finite undercooled liquid system, there are many fluctuating solid clusters and it is non-trivial to single out an individual one and compute the free energy associated with it. (ii) Due to the diffuse nature of the interface, there is a degree of ambiguity in the location and area of the dividing surface between the two phases, making the determination of γ and A problematic. (iii) In general, the interfacial free energy γ exhibits a curvature dependence, so it cannot be treated as a constant when fitting the computed nucleation free energy profile to the CNT expression Eqn. (3.1).

A great way to simply the problem and circumvent most of the aforementioned issues is to evaluate the interfacial free energy γ for an idealized planar interface, which resembles the surface of a mesoscopic nucleus. In the planar limit the value of the surface area $A(n_\mathrm{s})$ is unequivocal. In addition, different crystallographic orientations can be treated separately, so γ and its anisotropy can be defined more rigorously. At the coexistence temperature of the system T_m, a few techniques have been developed for computing $\gamma(T_\mathrm{m})$, including the capillary fluctuation method, which relies on an analysis of the height profile for the atomically-rough solid-liquid interfaces [99, 100] and the cleaving method, which computes the reversible work required to cleave the bulk solid and liquid and to join the pieces to generate an interface [101–103]. However, to investigate the solid-liquid interface at conditions that mirror those in experiments, a method that works for planar interfaces and that operates under realistic levels of undercooling ($T < T_\mathrm{m}$) is required. In this chapter, we derive just such a method, as well as a thermodynamic framework to rigorously define and characterize the free energy of a solid-liquid interface at out-of-equilibrium conditions. We then use this method to compute the temperature dependence of γ for a realistic interatomic potential.

3.1 Gibbs dividing surface in atomistic systems

In order to relate atomic-scale descriptions with mesoscopic models and experiments, it is necessary to introduce a consistent framework for quantifying the number of solid atoms n_s. At the atomistic level, several families of structural fingerprints $\phi(i)$ can be used to determine whether the environment around atom i is solid-like or liquid-like [95]. In our simulations, for example, we used a cubic harmonic order parameter introduced in chapter 2.3.4. Instead of defining n_s by introducing an ad-hoc threshold on ϕ, which is quite common in simulations of this kind, we adopted an alternative approach, that is closely related to the classical definition of a Gibbs dividing surface.

A Gibbs dividing surface is defined to be an infinitely thin geometrical surface that is sensibly coincident with the physical surface of discontinuity [58, 104]. The surface is meant to be an idealization of the transition region between the two phases, and one should choose, whenever possible, a geometry that is consistent with the boundary conditions and the symmetry of the problem. The precise position and shape of the Gibbs dividing surface are important, for instance when one needs to determine its area [58, 104]. When instead one only needs to define the extent of the two bulk phases, what matters most is that the surface divides the system into a solid part that has n_s atoms and a liquid part that has n_l atoms, with no atom assigned to the interface. It is then useful to construct a reference system, in which the solid part and the liquid part maintain their bulk properties up to the dividing surface. With a dividing surface in place, the free energy and the properties of any two-phase system can be naturally decomposed into a term corresponding to the reference bulk system and an excess term associated with the interface.

To remove the degree of freedom associated with the choice of a dividing surface, it is customary to select a surface such that there is no surface excess of a certain extensive quantity Φ, i.e. such that the real system and the reference system exhibit the same value for the chosen extensive quantity. This extensive quantity could be the volume occupied by that region, its internal energy or its entropy for example. More often than not, it is also convenient to use an extensive order parameter $\Phi = \sum_i \phi_i$, where the atomic order parameter ϕ_i is calculated based on the local environment of each of the particles in the system. The zero surface excess condition can be schematically expressed as

$$\Phi(n_\text{s}, n_\text{l}) \equiv \Phi_\text{ref}(n_\text{s}, n_\text{l}), \tag{3.2}$$

where $\Phi(n_\text{s}, n_\text{l})$ and $\Phi_\text{ref}(n_\text{s}, n_\text{l})$ stands for the values of the extensive quantity Φ of the real solid-liquid system and the reference bulk system that both comprise n_s solid atoms and n_l liquid atoms, subject to the conservation of the total number of atoms $n_\text{s} + n_\text{l} = N$ for a solid-liquid system comprising a total of N atoms. In turn, in the reference bulk system we have $\Phi_\text{ref}(n_\text{s}, n_\text{l}) = \Phi_\text{s}(n_\text{s}) + \Phi_\text{l}(n_\text{l})$

Assuming that the instantaneous value of the extensive quantity Φ_s of the reference bulk solid system that comprises n_s bulk solid atoms can be characterized by the average value for the order parameter of each atom ϕ_s, and Φ_l of the bulk liquid system can be characterized by the average value ϕ_l, one can have the equation $\Phi_\text{ref}(n_\text{s}, n_\text{l}) = \phi_\text{s} n_\text{s} + \phi_\text{l}(N - n_\text{s})$. One can then define the number of atoms of solid using a deterministic mapping

$$n_\text{s}(\Phi) = \left(\Phi - N\phi_\text{l}\right) / \left(\phi_\text{s} - \phi_\text{l}\right). \tag{3.3}$$

This mapping corresponds to a Gibbs dividing surface between the two phases that has zero excess for the extensive variable Φ. Within this approach, the choice of the order parameter implicitly determines the position of the solid-liquid dividing surface while automatically averaging over roughness and thermal fluctuations.

The free energy of the reference bulk system composed of n_s solid atoms and $n_\text{l} = N - n_\text{s}$ liquid atoms is equal to $N\mu_\text{l} + \mu_\text{sl} n_\text{s}$, where μ_l is the chemical potential of the liquid and μ_sl is the difference between the chemical potentials of the solid and liquid. Setting the free energy of the liquid as the reference zero, the free energy of the solid-liquid system with an interface can be expressed

$$G(\Phi) = \mu_\text{sl} n_\text{s}(\Phi) + G_\text{surf}(\Phi), \tag{3.4}$$

where $G_\text{surf}(\Phi)$ is the excess term due to the solid-liquid interface. It is customary to express the surface free energy excess as the product of a specific surface energy term γ^Φ and the area of the interface, i.e. $G_\text{surf}(\Phi) = \gamma^\Phi A$. For a solid-liquid system the surface area is fixed by the boundary condition of the system, $A = 2\Delta y \Delta z$, where $\Delta y \Delta z$ is the cross-section area of the simulation cell and the interface is perpendicular to x axis.

It is worth pointing out that, the choice of the Gibbs dividing surface is not unique, however, regardless of which choice is made the Gibbs free energy of the given solid-liquid system should not change. For instance, if an extensive quantity Θ instead of Φ is used to define the dividing surface, we have

$$G = \mu_\text{sl} n_\text{s}(\Phi) + \gamma^\Phi A = \mu_\text{sl} n_\text{s}(\Theta) + \gamma^\Theta A. \tag{3.5}$$

In out-of-equilibrium conditions $\mu_\text{sl} \neq 0$, Eqn. (3.5) thus implies that the subdivision of G

into bulk and interface terms depends on the choice of the dividing surface for a system containing a planar interface. The specific free energy excesses obtained using two different order parameters ϕ and θ to define the reference state are related by:

$$\gamma^{\Theta} - \gamma^{\Phi} = -\mu_{\mathrm{sl}} \frac{n_{\mathrm{s}}(\Theta) - n_{\mathrm{s}}(\Phi)}{A}. \tag{3.6}$$

This change in the interfacial free energy could been seen to be coming from a shift in the position of the dividing surface. This shift means that different numbers of atoms are assigned to the solid and liquid parts of the system.

Although many choices of the Gibbs dividing surface can be made and they are all equally valid provided that they are used consistently, a few choices have certain merits and as a result are more commonly used: The surface of tension (σ-surface), corresponding to the position at which the mechanical definition of tension applies, is regarded as "special" because the associated interfacial free energy γ^{σ} is curvature-independent. On the other hand, the equimolar dividing surface with interfacial free energy γ^{V} that has no surface excess of volume (V) is commonly used when analyzing nucleation, because this surface encloses a nucleus that has the same density as the bulk, which streamlines the formulation of nucleation free energy profiles [105, 106].

3.2 Simulation details

We simulated the solid-liquid interface for a simple but realistic Lennard-Jones system [103, 107, 108]. To accelerate the sampling so as to obtain reversible formation of a solid-liquid interface in a viable amount of simulation time, we performed well-tempered metadynamics simulations with adaptive Gaussians [92, 109]. using $\Phi = \sum_i S(\kappa(i))$ (see chapter 2.3.4) as the CV. A simulation supercell with dimensions commensurate to the equilibrated lattice at the temperature T, was employed. This cell was elongated along the x axis which was chosen to be parallel to the crystallographic orientation of the solid-liquid interface. Simulations were run at constant temperature and pressure, with only the x direction left free to fluctuate. Under these conditions, the planar interfaces, whenever present, are always perpendicular to the chosen lattice direction and always have a surface area $A = 2\Delta y \Delta z$ equal to twice the supercell cross section. The sampling has to be restricted to the relevant regions of phase space: the melt, the coexistence state with correctly-oriented solid-liquid interfaces, and the defect-free fcc crystal. Choosing an appropriate CV is critical for achieving this end and, in addition, several judiciously tuned restraints are required to prevent the formation of grain boundaries and twin defects. The presence of these defects was already observed in biased simulations at T_{m} [110] and they become very likely at undercooled conditions. Fast implementation of this complex simulation setup was made possible by the flexibility of the PLUMED code [93] in combination with LAMMPS [91].

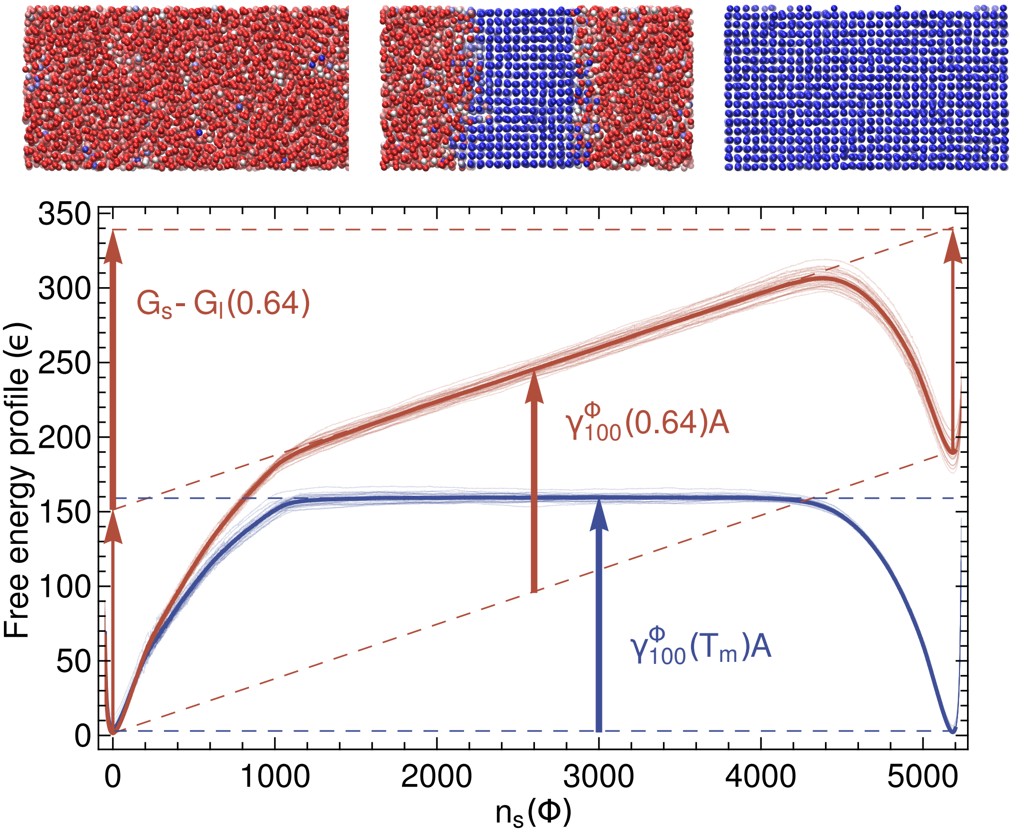

Figure 3.1 – The free energy profile averaged from 36 independent metadynamics simulation runs, and shown as a function of $n_s(\Phi)$, for an interface perpendicular to $\langle 100\rangle$ and a simulation box composed of $16 \times 9 \times 9$ *fcc* unit cells. Representative atomic configurations for different values of $n_s(\Phi)$ are shown, with atoms colored according to the local order parameter ϕ. The two sets of curves correspond to the free energy profiles at the melting temperature ($\gamma(T_m) = 0.371(2)$) and at a moderate degree of over-heating ($\gamma(0.64) = 0.365(3)$). All quantities are expressed in Lennard-Jones units.

3.3 The planar interfacial free energy γ^Φ

Figure 3.1 shows the free energy G as a function of $n_s(\Phi)$ at two different temperatures. These curves were reconstructed from the metadynamics simulations using on-the-fly re-weighting [111, 112]. At $T = T_\mathrm{m}$, $G(n_s(\Phi))$ displays a broad horizontal plateau that corresponds to the progressive movement of the solid-liquid interface. As discussed in Ref. [107], $\gamma(T_\mathrm{m})A$ can be unambiguously taken as the difference between the free energy of the plateau and the free energy at the bottom of the solid and liquid minima. The bias potential allows the interface to form even when μ_sl is non-zero, at temperatures away from T_m. Figure 3.1 thus also reports the free energy profile computed for $T > T_m$. The free energies of the minima corresponding to the bulk solid and liquid phases now take different values. The slope of the line that joins these minima is equal to $\mu_\mathrm{sl}(T)$. The region that was a plateau at $T = T_m$ has become an oblique line, with a slope that is also equal to μ_{sl} as moving the dividing surface now has an energetic cost that is associated with forcing atoms to undergo the phase transition [113]. The vertical red arrow in Figure 3.1 indicates the free energy difference between the simulated system with a solid-liquid interface and a reference state composed of $n_{\mathrm{s}}(\Phi)$ atoms of bulk solid and $N - n_{\mathrm{s}}(\Phi)$ atoms of bulk liquid. This quantity is equal to $\gamma(T)A$.

3.4 Using a different Gibbs dividing surface

At this point it is important to note that at $T \neq T_{\mathrm{m}}$ a change in the definition of the order parameter ϕ affects how the free energy of a given state is partitioned between the bulk and surface terms. We will demonstrate later that the value we obtain for $\gamma(T)$ thus depends on ϕ. As expressed in Eqn. (3.5), this does not change the total free energy of the solid-liquid system, provided that a consistent framework is used to define the nucleus' size and surface energy. It can, however, lead to confusion when comparing simulations, phenomenological models and interpretations of experiments as these may have been analyzed using different definitions. In particular, CNT-based models generally assume that the nucleus has the bulk density. This is equivalent to taking a reference state based on the molar volume, or to using a zero-excess-volume dividing surface. A V-based value of γ is therefore desirable as one can then establish a direct link between simulations and CNT models. In addition, simulations run with different order parameters can be compared straightforwardly if the volume-based interfacial free energy γ^V is also computed.

In principle it is possible to compute γ using a different local order parameter θ by on-the-fly reweighing the biased trajectory. In practice, however, any results obtained are only reliable when the sampling of $\Theta = \sum_i \theta(i)$ is as thorough as the sampling of the CV used in the metadynamics simulations. In addition, in order to extract γ, the free energy profile as a function of $n_{\mathrm{s}}(\Theta)$ must exhibit a clear plateau region that corresponds to solid-liquid coexistence. Unfortunately, the molar volume does not fulfill these two criteria. To illustrate this problem, we performed 2D metadynamics simulations in which $\Phi = \sum_i S(\kappa(i))$ (see chapter 2.3.4) and the total volume V were used as CVs. In Figure 2 the resulting free energy surface is displayed as a function of the solid atom counts obtained when these two extensive quantities are inserted into Eq. (3.3). The fluctuations in the molar volume for the two bulk phases are so large that they overlap with the plateau region. The molar volume, by itself, is thus not an effective fingerprint for distinguishing the solid and the liquid phases at the atomic scale.

Even though one cannot immediately obtain γ^V from the free energy profile as a function of $n_{\mathrm{s}}(V)$, it can still be determined indirectly. To do so, one simply computes the change in atom count that occurs when V rather than Φ is used to define the reference state:

$$\Delta n_{\mathrm{s}}(\Phi, V) = \langle n_{\mathrm{s}}(V)\delta\left[n_{\mathrm{s}}(\Phi) - n_{\mathrm{s}}(\Phi')\right]\rangle - n_{\mathrm{s}}(\Phi). \tag{3.7}$$

In the above, the Dirac δ ensures that only states with a certain value of $n_{\mathrm{s}}(\Phi)$ are selected, and

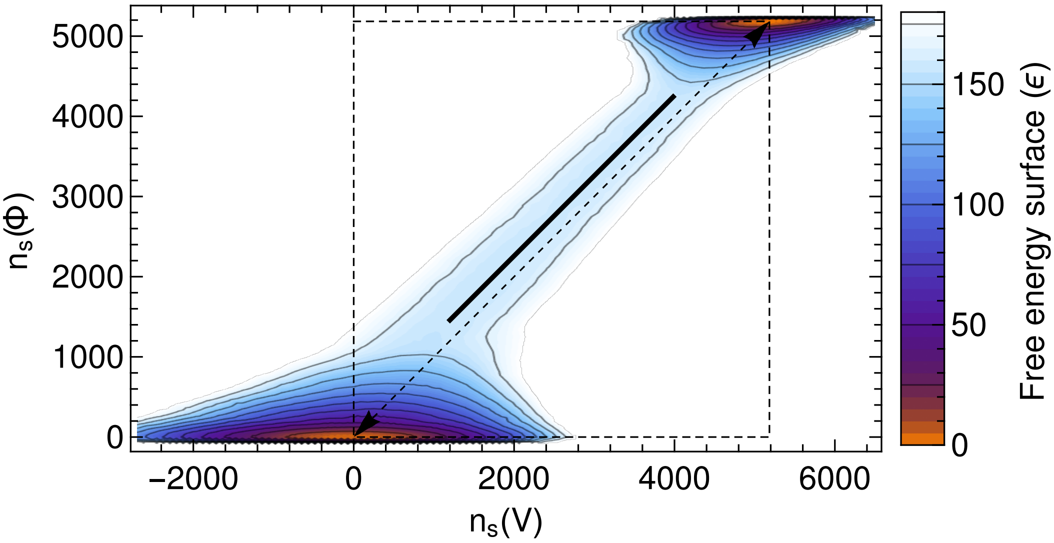

Figure 3.2 – The free energy as a function of $n_s(\Phi)$ and $n_s(V)$ for a simulation at $T_m = 0.6185$ computed using a supercell 16 × 9 × 9 times the *fcc* unit cell. The dashed line indicates the $n_s(\Phi) = n_s(V)$ ideal line, whereas the full line indicates the mean value of $n_s(V)$ for a given value of $n_s(\Phi)$. The horizontal offset between the two curves corresponds to the surface excess for the atom count associated with the Φ-based dividing surface, $\Delta n_s(\Phi, V)$, which for this particular calculation is equal to 259 ± 2 atoms.

$\langle\cdot\rangle$ represents a NPT ensemble average, which can be obtained by appropriately re-weighting the biased simulation. $\Delta n_s(\Psi, V)$ can be determined to a very high statistical accuracy as long as one of the two order parameters can identify the coexistence region. Graphically, this quantity corresponds to the horizontal offset between the solid black line and the dashed diagonal in Figure 3.2, which is constant over the entire plateau region. This implies that, when a solid-liquid interface is present, a change in the definition of the Gibbs dividing surface leads to a constant shift in n_s. The interfacial free-energy based on the volume, $\gamma^V(T)$, can thus be inferred from the quantity $\gamma^\Phi(T)$ using $\gamma^V = \gamma^\Phi - \mu_{sl}\Delta n_s(\Phi, V)/A$.

3.5 Results and discussions

Figure 3.3 shows the temperature dependence for μ_{sl} and γ computed for the $\langle 100\rangle$ and $\langle 111\rangle$ crystal lattice orientations. μ_{sl} is a bulk property so its value should not depend on the orientation of the surface as is observed in the top panel. The magnitude and anisotropy of the interface free energy at $T = T_m$ match the values reported in the literature Ref. [103, 107, 108]. Its temperature dependence, however, depends strongly on the choice of reference state. The lower panel shows that γ^Φ has a near-constant negative slope for both orientations which indicates that the excess entropy associated with the Φ-based dividing surface is positive. The importance of using a consistent reference state when studying the solid-liquid interface in out-of-equilibrium conditions is highlighted by the dramatically different behavior observed

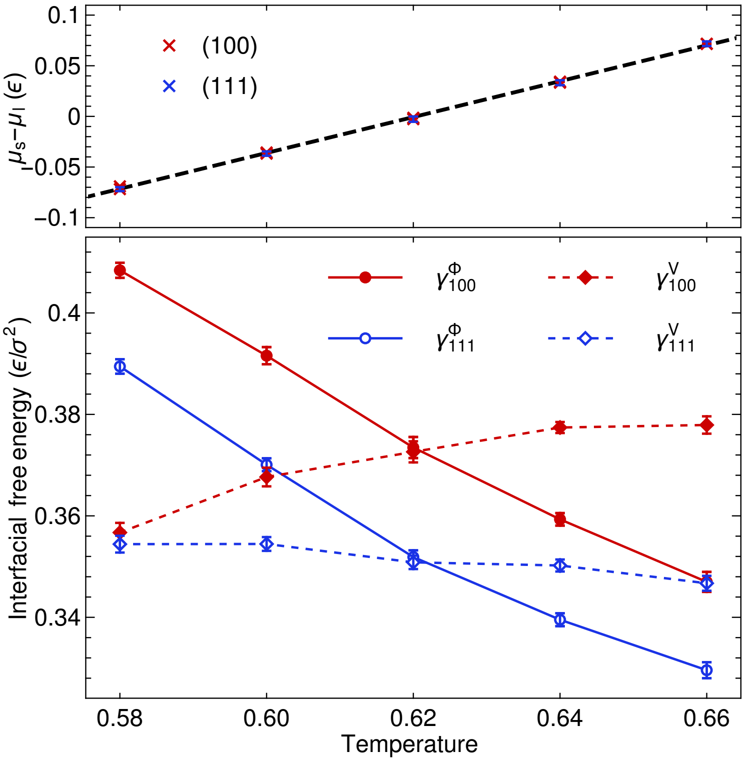

Figure 3.3 – Chemical potential difference $\mu_{\mathrm{sl}} = \mu_{\mathrm{s}} - \mu_{\mathrm{l}}$ (upper panel) and interfacial free energy γ (lower panel) for two different crystal orientations, as a function of temperature. γ^{Φ} and γ^{V} were computed using a Gibbs dividing surface that has zero surface excess for Φ and V, respectively (see equation 3.3). These two different ways of partitioning the free energy of the system between bulk and surface terms ensure that there are significant differences between γ^{Φ} and γ^{V} when $T \neq T_{\mathrm{m}}$. Each data point represents an average over 24 independent metadynamics runs and statistical uncertainties are indicated by the error bars. The uncertainties in γ^{Φ} and $\mu_{\mathrm{sl}}\Delta n_{\mathrm{s}}(\Phi, V)$ were assumed to be independent during the error estimation. The data points were obtained by restricting the sampling to prevent complete melting. The simulation boxes comprised about 1200 atoms, which implies a small finite-size effect that shifts T_{m} to 0.62. At T_{m} the interfacial free energies, for the two orientations we considered are $\gamma_{100}(0.62) = 0.373(2)$ and $\gamma_{111}(0.62) = 0.352(1)$. All quantities are expressed in Lennard-Jones units.

when the volume is used to define the reference state. γ^V_{100} *increases* with temperature as $\Delta n_\mathrm{s}(\Phi, V)$ is large and negative. When the molar volume is used to define the reference state a larger fraction of the interface region is seen as an ordered liquid rather than a disordered solid. The interface for the $\langle 100 \rangle$ crystal orientation thus has a negative excess surface entropy. For the $\langle 111 \rangle$ orientation, meanwhile, $\Delta n_\mathrm{s}(\Phi, V)$ is smaller and the slope of γ^V_{111} remains negative.

It is important to remember that the total free energy of the nucleus is the only quantity that affects its thermodynamic stability. However, the dependence of γ on the dividing surface will have consequences on predictions and data analysis if the same reference state is not used consistently. Usually, CNT models are built assuming that the nucleus' density is constant and equal to its bulk value, and the equimolar Gibbs dividing surface and the associated value of γ^V is most consistent with this assumption.

The observation of a positive slope for $\gamma^V(T)$ is consistent with a simple phenomenological model for a hard-sphere system [114], and with the effective γ values that were determined by fitting the values of $G^\star(T)$ obtained from simulations run at deeply undercooled conditions using a CNT expression [46]. Our results paint a somewhat more nuanced picture, however. The behavior of γ^V_{100} deviates significantly from linearity even at small undercoolings. In addition, the temperature dependence of the solid-liquid interface energy computed relative to a zero-excess-volume reference depends strongly on orientation, with the $\langle 111 \rangle$ direction showing a small negative slope. As a consequence, the anisotropy in γ^V between the two orientations we considered decreases down to $T = 0.58$, suggesting that the shape of the critical nucleus may depend on the degree of undercooling. Fully determining the size and shape of the critical nucleus would require one to determine γ and $\partial\gamma/\partial T$ for other high-symmetry orientations and for small mis-orientations. A direct comparison between planar-interface calculations and 3D nucleation simulations at intermediate undercoolings would also allow one to test the validity of the various assumptions within CNT. For example, such simulations would allow one to determine whether or not the Laplace pressure caused by the curvature of the interface has a significant effect on the chemical potentials of the two phases. These calculations, as well as the investigation of more realistic inter-atomic potentials, will be the subject of future work, which will provide more accurate parameters for mesoscale phase-field models of nucleation and growth [115, 116].

Using a thermodynamic definition for the size of the solid region allows one to treat atomic-scale fingerprints and macroscopic observables on the same footing, thereby providing a practical procedure for converting the value of γ obtained with a computationally-effective order parameter into the value consistent with the assumptions of mesoscopic theories of nucleation. The combination of this framework and accelerated molecular dynamics makes it possible to generate and stabilize planar interfaces at $T \neq T_\mathrm{m}$. Consequently, simulations can be performed with conditions that more closely resemble those in experiments. Such simulations could be used to shed light on the changes in morphology of the critical nucleus, to monitor capillary fluctuations at $T \neq T_\mathrm{m}$ [99], and to give atomistic insight on the evolution of interfaces away from equilibrium[117]. In addition, the methodology and the thermodynamic considerations we make here, whilst examining solidification, would apply with minor modifications to the study of nucleation in other contexts, be it melting [118], precipitation [119], condensation, order-disorder transitions, or other situations in which a phase transition is hindered by a surface energy term[120].

3.6 Conclusions

The properties of the interface between solid and melt are key to solidification and melting, as the interfacial free energy introduces a kinetic barrier to phase transitions. This makes solidification happen below the melting temperature, in out-of-equilibrium conditions at which the interfacial free energy is ill-defined. Here we draw a connection between the atomistic description of a diffuse solid-liquid interface and its thermodynamic characterization. This framework resolves the ambiguities in defining the solid-liquid interfacial free energy above and below the melting temperature. In addition, we introduce a simulation protocol that allows solid-liquid interfaces to be reversibly created and destroyed at conditions relevant for experiments. We directly evaluate the value of the interfacial free energy away from the melting point for a simple but realistic atomic potential, and find a more complex temperature dependence than the constant positive slope that has been generally assumed based on phenomenological considerations and that has been used to interpret experiments.

4 Bridging the gap between atomistic and macroscopic models of homogeneous nucleation

In the previous chapter, we discussed how to define a Gibbs dividing surface and compute the free energy associated with a planar solid-liquid interface, we also hint that the situation of homogeneous nucleation is more complicated because of the presence of multiple nuclei in the undercooled bulk liquid system. Moreover, there is a conceptual gap in assuming that fluctuations in the metastable phase involving a few atoms should be regarded as a nucleus of a stable phase that is only defined in the thermodynamic limit. In this chapter we address these issues by building a thermodynamic framework that reconcile the picture emerging from simulation with macroscopic theories of nucleation. To achieve this, we first investigate a multiple cluster model that we use as a proxy for an idealized atomistic system. Then, we develop a thermodynamic framework that is consistent with the multiple cluster model, requires fewer assumptions, and is fully applicable to the atomistic systems simulated in molecular dynamics or Monte Carlo studies. For the sake of clarity, we will develop our theoretical framework making reference to the case of solidification from the melt (Figure 4.2), but our results are general, and we present an application to a two-dimensional Ising model (Figure 4.4) to demonstrate that it can be applied to all sorts of activated phase transition processes.

4.1 An idealized multiple-cluster model

We start by taking an idealized model of a metastable bulk liquid phase, in which all the solid clusters can be identified unambiguously. We then further assume that the interactions between clusters are insignificant (e.g. negligible volume exclusion). We use the symbol p_n to denote the number of solid clusters containing n atoms. The total number of solid atoms in the system is thus just $n_{\text{tot}} = \sum_{n=1}^{\infty} n p_n$. We assume the average population of cluster sizes is given by the CNT expression:

$$\langle p_n \rangle \propto \exp(-\beta G(n)), \tag{4.1}$$

where $\beta = 1/k_B T$, and $G(n)$ is the free energy excess associated with a single cluster of size n. The probability distribution for the cluster populations $P(n, p_n)$ can be approximated using a Poisson distribution, i.e.

$$\begin{aligned} P(n, p_n) =& \lambda(n)^{p_n} e^{-\lambda(n)} / p_n!\,, \\ \lambda(n) = \langle p_n \rangle =& N_\text{s} e^{-\beta G(n)}, \end{aligned} \tag{4.2}$$

where $G(n)$ is the free energy of a single cluster of size n, and N_s is the number of nucleation sites. N_s is proportional to the total number of particles in the system, and guarantees the appropriate scaling with system size. $G(n)$ can take any form as long as it is a monotonically increasing function for n smaller than the size of the critical nucleus, but for this idealized system we use $G(n) = \mu n + \sigma v^{\frac{2}{3}} n^{\frac{2}{3}}$ and the effective interfacial free energy σ is equal to the specific interfacial free energy γ times a geometrical constant Ω. For our toy system, we selected parameters $k_B = 1$, $T = 0.58$, $\mu = -0.0562\epsilon$, $\gamma = 0.354\epsilon/\sigma^2$, $\Omega = 4.836$, $v^{\frac{2}{3}} = 1.035\sigma^2$, and ϵ and σ are the Lennard-Jones energy and length units. We also assumed $N_\text{s} = 700$. These parameters were selected to mimic those obtained in the calculations we performed for the Lennard-Jones system described in chapter 2. The free energy profile for this equation with these parameters is shown in red in Figure 4.1.

As we have assumed that the cluster size distribution follows Eqn. (4.2), we can derive the following expression for the free energy profile of the whole system as a function of the *total* number of solid atoms, $\tilde{G}(n_\text{tot})$,

$$e^{-\beta \tilde{G}(n_\text{tot})} = \sum_{p_1=0}^{\infty} \sum_{p_2=0}^{\infty} \cdots \sum_{p_{n^\star}=0}^{\infty} \delta\left(\sum_{n=1}^{\infty} n p_n - n_\text{tot} \right) \prod_{n=1}^{n^\star} P(n, p_n), \tag{4.3}$$

by explicitly enumerating all the possible combinations of cluster sizes that result in the same n_tot. Here $n^\star$ is the size of the critical nucleus, which is taken as an upper bound for the cluster size, in order to restrict the analysis to the range of metastability of the liquid [47]. We computed $\tilde{G}(n_\text{tot})$ analytically using Eqn. (4.3), and plotted the result in blue in Figure 4.1. It is important to note that $\tilde{G}(n_\text{tot})$ does not explicitly depend on the size composition of all the clusters. Defining this quantity is thus an important step towards the formulation of a macroscopic view of nucleation that is reliant on extensive quantities calculated over the whole system.

Figure 4.1 shows that the liquid contains $\langle n_\text{tot} \rangle$ solid atoms on average. By writing out explicitly the cluster size composition, we noticed that when $n_\text{tot} < \langle n_\text{tot} \rangle$ the most probable configuration for the system was composed of several small clusters. However, as n_tot gets larger, it typically contained one large solid cluster accompanied by many smaller ones. One can thus define a cutoff size n_cut, such that for $n_\text{tot} \gg n_\text{cut}$ it is orders of magnitude more

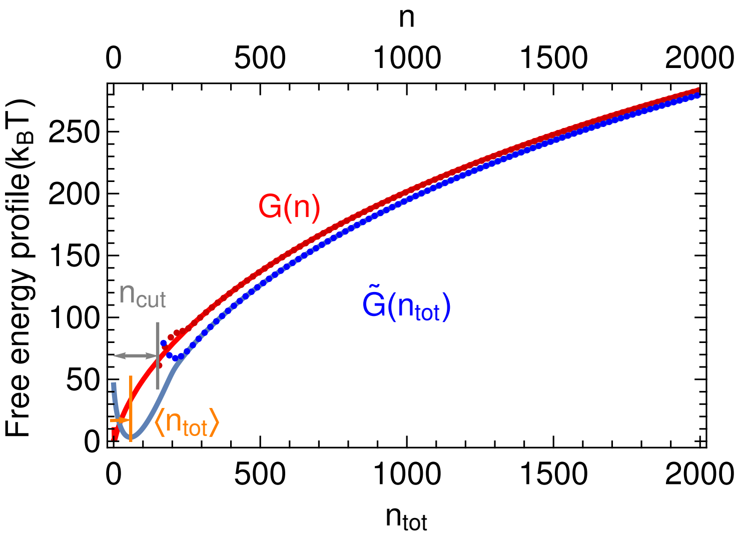

Figure 4.1 – The red line represents the free energy profile $G(n)$ of a single cluster. The blue line shows the exact $\tilde{G}(n_{\mathrm{tot}})$ of the atomistic system with multiple clusters. The red dots and the blue dots indicate $G(n)$ and $\tilde{G}(n_{\mathrm{tot}})$ that are approximated using Eq. (4.4), respectively. The grey and yellow vertical lines indicate n_{cut} and $\langle n_{\mathrm{tot}}\rangle$, respectively.

likely to have precisely one cluster with size $n > n_{\mathrm{cut}}$ than to have several or none of such large clusters. At $n_{\mathrm{tot}} \gg n_{\mathrm{cut}}$, the largest cluster can be interpreted as a standalone solid cluster with n atoms, associated with a probability $P(n,1)$. The rest of clusters in the background can be treated as a separate bulk liquid system that follows the same distributions (Eqn. (4.2)) as the original whole system. Under such treatment, at $n_{\mathrm{tot}} \gg n_{\mathrm{cut}}$ the expression for $\tilde{G}(n_{\mathrm{tot}})$ can be simplified tremendously as

$$\exp(-\beta\tilde{G}(n_{\mathrm{tot}})) = \sum_{n=n_{\mathrm{cut}}}^{n_{\mathrm{tot}}} N_{\mathrm{s}} e^{-\beta G(n)} e^{-\beta\tilde{G}(n_{\mathrm{tot}}-n)}, \tag{4.4}$$

considering also that $P(n,1) \approx N_{\mathrm{s}}\lambda(n)$ for large n. The blue dots in Figure 4.1 correspond to the approximate $\tilde{G}(n_{\mathrm{tot}})$ that can be computed using Eqn. (4.4). These points overlap perfectly with the exact values at $n_{\mathrm{tot}} \gtrsim n_{\mathrm{cut}} + \langle n_{\mathrm{tot}}\rangle$. What is more, Eqn. (4.4) suggests that there is a one-to-one mapping between the $\tilde{G}(n_{\mathrm{tot}})$ for the whole system and the $G(n)$ for a single cluster with $n > n_{\mathrm{cut}}$. As such, Eq. (4.4) allows us to calculate $G(n)$ from a knowledge of $\tilde{G}(n_{\mathrm{tot}})$, without any information on the sizes of individual clusters in each snapshot. The $G(n)$ that is reconstructed from Eq. (4.4) using a fixed-point numerical scheme is indicated using red dots in Figure 4.1, and overlaps perfectly with the exact $G(n)$.

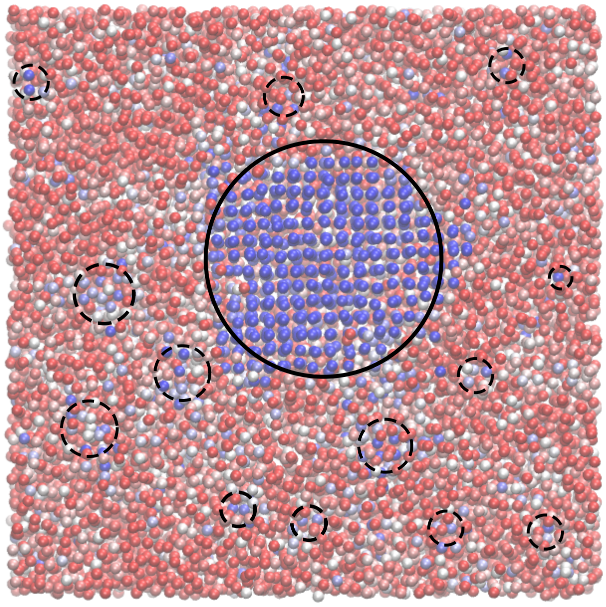

Figure 4.2 – A snapshot of an under-cooled liquid system that has a large solid cluster (solid circle), and many smaller clusters (dashed circles). Atoms are colored based on the value of a local order parameter so solid-like atoms are colored in blue, while liquid-like atoms are colored in red.

4.2 A probabilistic definition of Gibbs dividing surfaces

Our discussion thus far demonstrates that the average distribution of cluster sizes can be extracted from the distribution of the total number of atoms assigned to the stable phase. Unfortunately, in actual atomistic simulations it is impossible to assign individual atoms or molecules to either of the two phases without additional empirical assumptions. In what follows, we will therefore use the concept of Gibbs dividing surface, and introduce a thermodynamic approach that draws a connection between an atomistic and a macroscopic description of homogeneous nucleation.

The starting point of the framework is again the concept of the Gibbs dividing surface [58, 104]. In the previous chapter, we have described in detail about the definition of the dividing surface. In essence, a Gibbs dividing surface defined by an extensive quantity Φ of the solid-liquid system comprising a total of N atoms fulfills the zero surface excess condition

$$\Phi_{\mathrm{sl}}(n_{\mathrm{s}}, n_{\mathrm{l}}) \equiv \Phi_{\mathrm{ref}}(n_{\mathrm{s}}, n_{\mathrm{l}}), \tag{4.5}$$

where $\Phi_{\mathrm{sl}}(n_{\mathrm{s}}, n_{\mathrm{l}})$ and $\Phi_{\mathrm{ref}}(n_{\mathrm{s}}, n_{\mathrm{l}})$ stands for the values of the extensive quantity Φ of the real solid-liquid system and the reference system that both comprise n_{s} solid atoms and $n_{\mathrm{l}} = N - n_{\mathrm{s}}$ liquid atoms. In the previous chapter, we assumed that the value of n_{s} can be found by applying the deterministic mapping $\Phi_{\mathrm{ref}}(n_{\mathrm{s}}, n_{\mathrm{l}}) = \phi_{\mathrm{s}} n_{\mathrm{s}} + \phi_{\mathrm{l}}(N - n_{\mathrm{s}})$ where ϕ_{s} and ϕ_{l} are the average value for the order parameter of each atom in the solid and liquid respectively. However, it is important to realize that an instantaneous extensive quantity of a finite piece of bulk solid or bulk liquid can fluctuate even at fixed thermodynamic conditions. As a consequence, the extensive quantity Φ of a reference system that has a bulk solid part with n_s atoms and a bulk liquid part with $N - n_\mathrm{s}$ atoms also have fluctuations. In what follows, we will describe a probabilistic framework that takes into account the fluctuations when determining the dividing surface.

First, consider an unbiased simulation of the bulk solid phase. Then, select contiguous portions of the solid of varying size, and determine for each case the number of atoms n contained in the region and the value of the extensive quantity Φ. By computing the histogram of these quantities one can estimate

$$\rho_s(\Phi|n) = \int \delta(\Phi(\Omega) - \Phi)\mathrm{d}\Omega / \int \mathrm{d}\Omega, \tag{4.6}$$

where Ω denotes a possible microstate for the n atoms of that region, distributed with a probability consistent with the thermodynamic conditions, and $\Phi(\Omega)$ is the value of Φ for that microstate. This distribution, $\rho_s(\Phi|n)$, can be regarded as the conditional probability for observing Φ in a system consists of a bulk solid region containing n atoms. Since this solid region should mimic the bulk solid part in the reference system, strictly speaking its shape should be delimited by the the Gibbs dividing surface and the boundaries of the reference system. It is worth noting, however, that in the cases we considered in this study the shape of the solid region has little impact on $\rho_s(\Phi|n)$, as long as a compact shape is chosen. An analogous distribution $\rho_l(\Phi|n)$ can be derived for a bulk liquid region that contains n atoms and also has a shape determined by the dividing surface and the boundaries of the reference system.

Contrary to the multiple cluster model discussed above, “solid” and “liquid” in this case indicate well-defined thermodynamic states. The bulk solid state encompasses all the possible configurations for a system of solid, which can contain point defects, other crystal defects, and even small molten pools. Similarly, a bulk liquid comprises local crystalline orderings and sub-critical solid clusters.

The finite width of the distributions for $\rho_s(\Phi|n)$ and $\rho_l(\Phi|n)$ in Eqn. (4.6) ensures that the value of Φ for a given microstate cannot be used to determine the composition of a reference system with absolute certainty. Instead, we can compute the distribution of Φ for a reference system composed of n_s solid atoms and n_l liquid atoms using

$$\rho_\mathrm{ref}(\Phi|n_\mathrm{s}, n_\mathrm{l}) = \int \mathrm{d}\varphi \rho_s\left(\varphi\middle|n_\mathrm{s}\right)\rho_l\left(\Phi - \varphi\middle|n_\mathrm{l}\right). \tag{4.7}$$

According to the concept of the Gibbs dividing surface, if the reference and the actual system both have n_s solid atoms and n_l liquid atoms, they should also both have the same distribution for Φ. In other words, the zero-excess condition in Eqn. (4.5) takes a probabilistic form,

$$\rho_{\mathrm{sl}}\left(\Phi|n_{\mathrm{s}},n_{\mathrm{l}}\right) \equiv \rho_{\mathrm{ref}}\left(\Phi|n_{\mathrm{s}},n_{\mathrm{l}}\right). \tag{4.8}$$

Only when $\rho_s\left(\Phi|n\right)$ and $\rho_l\left(\Phi|n\right)$ are both δ functions at any given n, Eqn. (4.7) is reduced to a deterministic mapping between Φ and n_{s} analogous to that introduced in chapter 3.1.

4.3 Obtaining cluster-size free energies from an extensive order parameter

Let us now describe how to extract the free energy profile for a solid cluster from atomistic simulations of undercooled liquid. In simulations, the values of Φ can be easily computed for every microstate, so the associated free energy $\tilde{G}(\Phi)$ can be directly obtained from biased or unbiased molecular dynamics simulations. Since the atomistic simulations are constructed so that they sufficiently sample all configurations in the undercooled liquid, the computed free energy profile $\tilde{G}(\Phi)$ directly characterizes the distribution of Φ in the liquid, i.e. $\rho_l\left(\Phi|N\right) = \exp(-\beta\tilde{G}(\Phi))$. On the other hand, the bulk liquid sampled in simulations can have configurations that contain sub-critical nuclei of large sizes, such as the one illustrated in Figure 4.2. Those configurations can have a value of Φ approaching those typically encountered for a solid sample. As we have discussed in the multiple cluster model, configurations that contain a large number of solid-like atoms are overwhelmingly likely to comprise one and only one cluster of size larger than n_{cut} and a liquid-like background. By a similar logic, a configuration with a value of Φ that has enough solid-like characteristics can be interpreted as a single solid cluster larger than n_{cut} and the surrounding liquid.

Consider a single solid cluster that has $n_{\mathrm{s}} > n_{\mathrm{cut}}$ atoms together with a liquid background of $N - n_{\mathrm{s}}$ atoms. The values of the extensive quantity for this combination of phases follow the distribution $\rho_{\mathrm{sl}}\left(\Phi|n_{\mathrm{s}}, N-n_{\mathrm{s}}\right)$. Inside the undercooled bulk liquid, the average population for solid clusters of size n_{s} can be expressed as $\langle p_{n_s}\rangle = N_s \exp(-\beta G(n_{\mathrm{s}}))$, where N_s is the number of nucleation sites that can often be considered as the number of atoms or molecules or lattice sites in homogeneous nucleation, and $\exp(-\beta G(n_{\mathrm{s}}))$ the probability that a nucleus of size n_{s} has grown around a nucleation site in the metastable liquid. In other words, $G(n_{\mathrm{s}})$ represents the free energy excess associated with a solid cluster that has n_{s} atoms relative to the bulk liquid. Notice also that for $n_{\mathrm{s}} > n_{\mathrm{cut}}$, the average population $N_s \exp(-\beta G(n_{\mathrm{s}}))$ is also the probability of observing a solid cluster of size n_{s} in the bulk liquid system. Based on these considerations, and using the law of total probability, the probability distribution for Φ in such systems follows

$$e^{-\beta\tilde{G}(\Phi)} = \int_{n_{\mathrm{cut}}}^{n^{\star}} \mathrm{d}n_{\mathrm{s}}\rho_{\mathrm{sl}}\left(\Phi|n_{\mathrm{s}}, N-n_{\mathrm{s}}\right) N_s e^{-\beta G(n_{\mathrm{s}})}. \tag{4.9}$$

This expression is valid for values of Φ that satisfy $\rho_{\mathrm{sl}}\left(\Phi|n, N-n\right) \approx 0$ for all $n < n_{\mathrm{cut}}$, so that the system can be considered to have a single cluster of size larger than n_{cut}.

In principle, $G(n_\mathrm{s})$ can be determined from Eqn. (4.9), as both $\tilde{G}(\Phi))$ and $\rho_\mathrm{sl}\,(\Phi|n_\mathrm{s}, N-n_\mathrm{s})$ can be computed from simulations. However, to avoid the numerical instabilities in the direct deconvolution process, we cast the problem as a fixed-point iteration. The average Φ value for a system containing n_s solid atoms, $\bar{\Phi}(n_\mathrm{s}) = \int \mathrm{d}\Phi \rho_\mathrm{sl}\,(\Phi|n_\mathrm{s}, N-n_\mathrm{s})\,\Phi$, follows a monotonic relation with n_s, as shown in the Supplementary Material. One can invert this relation and obtain a value $\bar{n}_\mathrm{s}(\Phi)$ at each Φ such that $\Phi = \bar{\Phi}(\bar{n}_\mathrm{s}(\Phi))$. More generally, after some simple manipulations, we can rewrite Eqn. (4.9) as:

$$\tilde{G}(\Phi) = G(\bar{n}_\mathrm{s}(\Phi)) - \frac{1}{\beta}\log N_s - \frac{1}{\beta}\log\int_{n_\mathrm{cut}}^{n^\star} \mathrm{d}n \rho_\mathrm{sl}\,(\Phi|n, N-n)\, e^{-\beta[G(n)-G(\bar{n}_\mathrm{s}(\Phi))]}. \tag{4.10}$$

This equation can be rearranged, exploiting the inversion between $\bar{n}_\mathrm{s}$ and Φ, into a self-consistency condition on $G(n_\mathrm{s})$

$$G(n_\mathrm{s}) = \tilde{G}(\bar{\Phi}(n_\mathrm{s})) + \frac{1}{\beta}\log N_s + \frac{1}{\beta}\log\int_{n_\mathrm{cut}}^{n^\star} \mathrm{d}n \rho_\mathrm{sl}\left(\bar{\Phi}(n_\mathrm{s})\big|n, N-n\right) e^{-\beta[G(n)-G(n_\mathrm{s})]}, \tag{4.11}$$

which can be solved iteratively starting from the initial guess $G_0(n_\mathrm{s}) = \tilde{G}(\bar{\Phi}(n_\mathrm{s}))$, and plugging the old guess onto the right-hand side to obtain a new estimate at each iteration. Upon convergence, $G(n_\mathrm{s})$ is an estimate of the free energy for a solid cluster containing n_s atoms relative to the bulk liquid.

It is worth stressing that the cluster size n_s and the associated free energy $G(n_\mathrm{s})$ in Eqn. (4.9) are still dependent on the choice of Φ, because the reference system is defined based on a Gibbs dividing surface that has zero excess for the extensive quantity Φ. Due to the diffuse nature of the physical interface, a different choice for the extensive variable can result in a different location of the Gibbs dividing surface and a different reference system.

4.4 Application: solidification of a Lennard-Jones system

To demonstrate how this thermodynamic framework can be applied to an atomistic simulation of a phase transition, we simulated the processes of homogeneous solidification for a Lennard-Jones system of 23328 atoms at $T = 0.58$ [103, 107, 108] – corresponding to a moderate undercooling relative to the melting temperature of $T_m = 0.6185$. [1, 103, 107] We performed 12 independent biased sampling runs using the well-tempered metadynamics protocol with adaptive Gaussians (see chapter 2.3.2) [91–93, 109]. We used a collective variable $\Phi = \sum_i S(\kappa(i))$ in the biased simulations (see chapter 2.3.4). The solid blue line in Figure 4.3 indicates the free energy profile $\tilde{G}(\Phi)$ that was obtained by re-weighting the trajectories. Assuming the Gibbs dividing surface separating the solid cluster and the bulk liquid has a spherical shape, we also computed the two probability distributions $\rho_s\,(\Phi|n)$ and $\rho_l\,(\Phi|n)$ from unbiased simulations of the bulk phases.

The snapshot in Figure 4.2 is taken from one of the biased runs, with each atom colored according to the value of $S(\kappa(i))$. Analyzing the population of cluster sizes in this snapshot

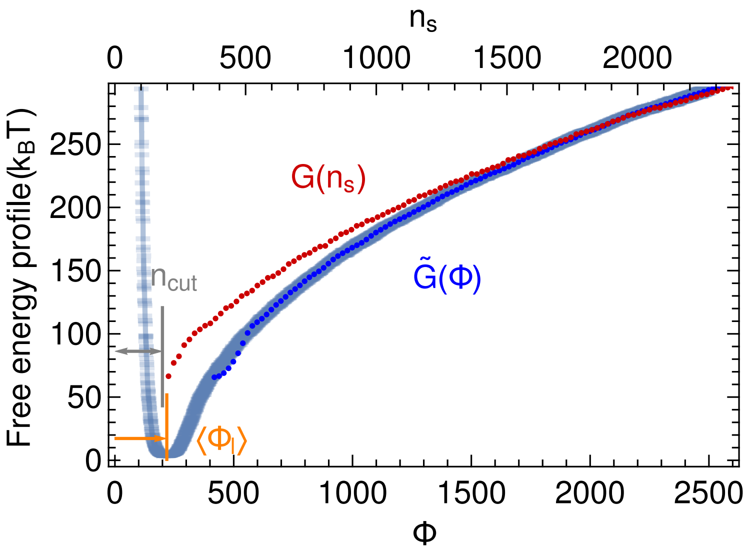

Figure 4.3 – The solid blue line is the free energy profile $\tilde{G}(\Phi)$, with statistical errors indicated by the error bars. The red and the blue dots indicate the reconstructed curves for $G(n_s)$ and $\tilde{G}(\Phi)$, respectively. The grey and yellow vertical lines indicate n_{cut} and the average extensive quantity $\langle \Phi_l \rangle$, respectively.

requires a man-made choice of a cutoff value for $S(\kappa)$, and a complex procedure to identify adjacent groups of solid atoms. Instead, by applying the thermodynamic model introduced above, we can simply characterize the behavior of Φ in the solid and the liquid phases, and use that knowledge to convert $\tilde{G}(\Phi)$ of the whole system into $G(n_s)$, using the iterative expression in Eqn. (4.11). Here we also assume the total number of atoms in the system N to be the number of nucleation sites N_s, although any other choice would simply amount to a vertical shift of the free-energy curve. This curve of $G(n_s)$ plotted as the red dots in Figure 4.3, corresponds to the free energy for a single cluster relative to the bulk liquid. In order to demonstrate the convergence of the conversion, we used the computed $G(n_s)$ to reconstruct the free energy profile $\tilde{G}(\Phi)$ using Eqn. (4.9). As shown in Figure 4.3, the reconstructed $\tilde{G}(\Phi)$ is indistinguishable from that obtained directly from the simulation.

As suggested by the many similarities between Figure 4.3 and Figure 4.1, the multiple cluster model and the thermodynamic model are very closely related. In the Supplementary Material we show that under a few additional assumptions Eqn. (4.9) is exactly the same as Eqn. (4.4), with Φ taking the role of n_{tot}. Eqn. (4.9) serves a dual purpose: it converts the extensive quantity Φ into an estimate for the overall solid fraction and it singles out the free-energy excess for the largest cluster from the fluctuations of the background liquid.

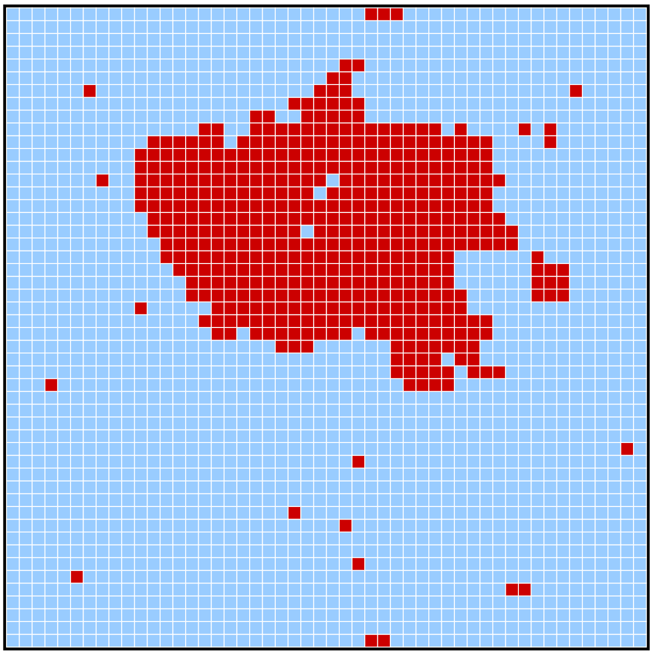

Figure 4.4 – A snapshot of a two-dimensional square-lattice Ising model undergoing a transition between "up" (blue) and "down" (red) ferromagnetic phases.

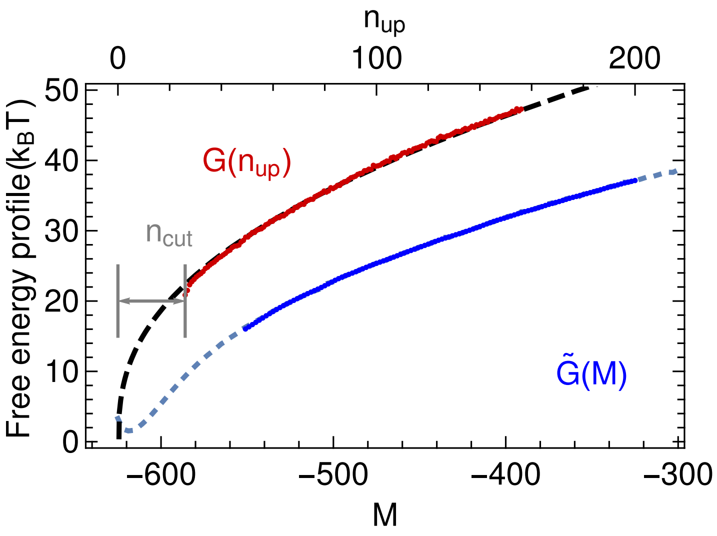

Figure 4.5 – The dotted blue line is the free energy profile $\tilde{G}(M)$ as a function of the magnetization, for a square-lattice Ising model with size $L = 25$ and periodic boundary conditions, simulated at $T = 1.5$. The solid red and the blue lines indicate the reconstructed curves for $G(n_{up})$ and $\tilde{G}(M)$, respectively. The dashed black line is the prediction from Eqn. (4.13). Note that $G(n_{up})$ and $\tilde{G}(M)$ are discrete functions, as n_{up} can only have integer values and M can only be odd integers in this system. The grey vertical lines indicate the choice of n_{cut} for this system.

4.5 Application: nucleation in a two-dimensional Ising Model

In order to demonstrate the general applicability of our thermodynamic framework, we discuss in this section its application to the homogeneous nucleation of a two-dimensional Ising model, in the absence of external magnetic field. The model is described by the usual first-neighbour Heisenberg Hamiltonian

$$H = -J \sum_{\langle i,j \rangle} s_i s_j, \tag{4.12}$$

where $J = 1$ is the coupling constant, the spin s_i at site i is either up (+1) or down (-1), and the sum extends over all its nearest neighbors in the lattice. We used a periodic square lattice with side $L = 25$, and performed a Monte Carlo simulation with biased sampling [111] at the temperature $T = 1.5$, well below the critical temperature $T_c = 2.269$. We started the simulation with all spins down, and also restricted the sampling to the states with negative total magnetization.

The snapshot in Figure 4.4 (from a simulation with $L = 50$) shows a large cluster with positive magnetization embedded in the phase with spins down. Note that one can observe spontaneous fluctuations of opposite spins not only in the negatively-magnetized background, but also inside the large nucleating cluster, underscoring the ambiguity in defining cluster sizes by counting the number of contiguous spins with the same orientation. [47] In contrast, our thermodynamic model does not rely on any clustering algorithm to identify nuclei of the different phases, but instead only focuses on the total magnetization $M = \sum_i s_i$ as a macroscopic order parameter to characterize the overall state of the system. The dotted blue line in Figure 4.5 indicates the free energy profile $\tilde{G}(M)$. Taking a reference state with a circular Gibbs dividing surface, we computed the probability distribution $\rho\left(M \middle| n_{up}, n_{down}\right)$ from unbiased simulations of the bulk phases as shown in the Supplementary Material. Using the iterative expression in Eqn. (4.11), and assuming the number of nucleation sites N_s is the total size of the lattice, L^2, we obtained $G(n_{up})$, which corresponds to the free energy for a single positively-magnetized cluster relative to the bulk negatively-magnetized phase. $G(n_{up})$ is plotted as the solid red line in Figure 4.5, together with the free energy profile $\tilde{G}(M)$ reconstructed using Eqn. (4.9), that matches perfectly the directly computed free-energy curve, signaling the convergence of self-consistent iterations.

In Ref. [121], the computed nucleation free energy profile of the 2D Ising model was found to agree well with the expression

$$G(n) = 2\sqrt{\pi n}\sigma + \tau k_B T \ln n + d, \tag{4.13}$$

where $\sigma = 1.20585$ is the temperature-dependent interfacial free energy for this Ising model that can be computed analytically [122], $\tau k_B T \ln n$ accounts for the shape fluctuations of the cluster ($\tau = \frac{5}{4}$ for the 2D Ising model), and the term $d = 8 - 2\sqrt{\pi}\sigma$ ensures that the free energy of a isolated spin is correctly captured. In Figure 4.5, we plotted the exact prediction of Eqn. (4.13) as the black dashed line. This prediction matches perfectly the $G(n_{up})$ obtained from our thermodynamic framework, without using any fitting parameters and without explicitly performing a cluster analysis of the simulation.

4.6 Conclusions

The thermodynamic framework introduced here provides a link between the molecular and the macroscopic scales. Any extensive quantity can be chosen to discriminate between the solid and the liquid, be it built upon local descriptors, or a traditional thermodynamic quantity such as the total volume, the energy or the magnetization. By characterizing the fluctuations of this extensive quantity we can rigorously define, in a probabilistic manner (Eqn. 4.7), a reference state consistent with a zero-excess Gibbs dividing surface that encloses a single cluster of the stable phase.

Our method is applicable to all types of phase transitions – from solidification, to precipitation or condensation – and it can be combined with any sampling method one chooses to accelerate nucleation [9, 40, 43–52]. From such simulations the free-energy for the overall system as a function of any extensive quantity can be computed, and then converted into the free energy of a single cluster relative to the metastable bulk. By avoiding the need of singling out atom-size clusters that are inherently ill-defined, our approach is both practically simple and conceptually elegant. Since no assumption is made on the functional form of the computed free energy profile for nucleation, our approach can be used to test the limits of classical nucleation theory, and extended so that it also describes heterogeneous nucleation, and thus further advances our understanding of bulk and interface-driven phase transitions.

5 How to use macroscopic nucleation theory to interpret atomistic simulations of homogeneous nucleation?

When analyzing atomistic simulations of nucleation, macroscopic models that provide analytic 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 the nucleation barrier $G^{\star}$, and when interpolating between results obtained at different temperatures. Classical nucleation theory has often been used for this purpose, but it has been shown to give rise to significant systematic errors for all sizes of the nucleus [49]. Although there can be genuine deficiencies of the CNT, *how* the CNT has been applied can also cause inaccuracies: a large part of problem is that the number of atoms n in the cluster is not a well-defined quantity [1, 40, 48, 49]. 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 is generally used as a fitting model for $G(n)$, because the various physical quantities that enter this expression - the chemical potential μ and the interfacial free energy γ - cannot be calculated directly from homogeneous nucleation simulations. In addition, the multiple clusters encountered in the metastable bulk system further complicate the analysis. 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.

In this chapter, we apply the previously developed thermodynamic framework for defining the nucleation free energy profile $G(n)$ in an unambiguous way (Chapter 3 and 4). We then fit the computed $G(n)$ to the expressions that come from macroscopic nucleation models, using the specific interfacial free energy γ and the chemical potential difference μ as fitting parameters. We also show that a higher-order correction – generally referred to as the Tolman term – naturally emerges from our formalism. Finally, we compare the predicted values for γ

and μ that we obtain from simulations of three-dimensional nuclei against those obtained from the simulations of planar interfaces.

5.1 A recap of the thermodynamic framework

In order to link the atomistic descriptions and the macroscopic representations of a solid-liquid system, we apply the thermodynamic framework developed in chapter 3 and 4. Using the Gibbs dividing surface framework introduced in chapter 3, in chapter 4 we provided a link between $\tilde{G}(\Phi)$, which is the free energy profile of the whole system as a function of the global extensive quantity Φ, with the nucleation free energy $G(n_\mathrm{s}(\Phi))$ of a single solid cluster. We also introduced the expression

$$e^{-\beta\tilde{G}(\Phi)} = \int_{n_\mathrm{cut}}^{n^\star} \mathrm{d}n_\mathrm{s}(\Phi)\rho_\mathrm{sl}\left(\Phi|n_\mathrm{s}(\Phi), N-n_\mathrm{s}(\Phi)\right) N_s e^{-\beta G(n_\mathrm{s}(\Phi))}. \tag{5.1}$$

To be clear, $\tilde{G}(\Phi)$ can be directly computed from atomistic simulations of nucleation, and $G(n_\mathrm{s}(\Phi))$ then be found by inverting Eqn.(5.1) numerically, as described in Chapter 4.

5.2 A general formulation of classical nucleation theory

With a Φ-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:

$$G(n_\mathrm{s}(\Phi)) = \mu_\mathrm{sl} n_\mathrm{s}(\Phi) + \gamma^\Phi A(n_\mathrm{s}(\Phi)), \tag{5.2}$$

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

For planar interfaces, the surface area $A(n_\mathrm{s}(\Phi))$ in Eqn. (5.2) is fixed by the boundary conditions so the specific planar interfacial free energy $\gamma^\Phi = \gamma^\Phi_\infty$ is a constant that depends on the crystallographic direction of the planar interface. Notice that if the extensive quantity Θ is used to define the dividing surface instead of Φ 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:

$$\delta {n_\mathrm{s}}^{\Phi,\Theta} = (n_\mathrm{s}(\Theta) - n_\mathrm{s}(\Phi))/A \tag{5.3}$$

and further note that this quantity should be constant as replacing Φ with Θ shifts the location of the dividing surface by a fixed amount (see chapter 3). Furthermore, when Φ is replaced by Θ the resulting change in the planar interfacial free energy

$$\Delta\gamma^{\Phi,\Theta}_\infty = \gamma^\Theta_\infty - \gamma^\Phi_\infty = -\mu_\mathrm{sl}\delta {n_\mathrm{s}}^{\Phi,\Theta}, \tag{5.4}$$

is also a constant, because the total free energy in Eqn. (5.2) should be unaffected by the change in the extensive quantity.

For the curved interfaces around a three-dimensional nucleus, the surface area $A(n_{\mathrm{s}}(\Phi))$ 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 $A = \Omega v_{\mathrm{s}}^{\frac{2}{3}} n_{\mathrm{s}}^{\frac{2}{3}}$. 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. (5.2). This equation should thus read:

$$G(n_{\mathrm{s}}(V)) = \mu_{\mathrm{sl}} n_{\mathrm{s}}(V) + \gamma^{V} \Omega v_{\mathrm{s}}^{\frac{2}{3}} n_{\mathrm{s}}^{\frac{2}{3}}(V). \tag{5.5}$$

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 γ^{V} [104, 123]. This dependence of γ^{V} on the effective radius of the nucleus $R = (3/4\pi)^{\frac{1}{3}} v_{\mathrm{s}}^{\frac{1}{3}} n_{\mathrm{s}}^{\frac{1}{3}}(V)$ can be written as

$$\gamma^{V}(R) = \gamma^{V}_{\infty}(1 - 2\delta/R + \mathcal{O}(1/R^{2})). \tag{5.6}$$

In this expression δ 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. γ^{V}_{∞}, meanwhile, is the specific interfacial free energy associated with the equimolar surface in the planar limit. Taking only the leading terms in Eqn. (5.6), Eqn. (5.5) can be rewritten as

$$G(n_{\mathrm{s}}(V)) \eqsim \mu_{\mathrm{sl}} n_{\mathrm{s}}(V) + \gamma^{V}_{\infty} \Omega v_{\mathrm{s}}^{\frac{2}{3}} n_{\mathrm{s}}^{\frac{2}{3}}(V)(1 - \epsilon n_{\mathrm{s}}^{-\frac{1}{3}}(V)), \tag{5.7}$$

where $\epsilon = (32\pi/3)^{\frac{1}{3}} v_{\mathrm{s}}^{-\frac{1}{3}} \delta$ 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 diving 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. (5.3). It would be useful, however, to derive a CNT expression based on an arbitrary choice of reference thermodynamic variable Φ. To derive such an expression, we first assumed that the difference between the location of the Φ-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 $n_{\mathrm{s}}(\Phi)$ and $n_{\mathrm{s}}(V)$ can be approximated using

$$n_{\mathrm{s}}(V) - n_{\mathrm{s}}(\Phi) \eqsim \delta n_{\mathrm{s}}^{\Phi,V} \Omega v_{\mathrm{s}}^{\frac{2}{3}} n_{\mathrm{s}}^{\frac{2}{3}}(V), \tag{5.8}$$

where $\delta n_{\mathrm{s}}{}^{\Phi,V}$ is the constant that appears in Eqn. (5.3) and (5.4) and where $\Omega v_{\mathrm{s}}^{\frac{2}{3}} n_{\mathrm{s}}{}^{\frac{2}{3}}(V)$ is a measure of the surface area of the nucleus. We can substitute $n_{\mathrm{s}}(V)$ for $n_{\mathrm{s}}(\Phi)$ in Eqn. (5.7) using the above approximation, and can also use the relation $\gamma_{\infty}^{V} - \gamma_{\infty}^{\Phi} = -\mu_{\mathrm{sl}}\delta n_{\mathrm{s}}{}^{\Phi,V}$. After dropping the higher order correction terms, we obtain

$$G(n_{\mathrm{s}}(\Phi)) \approx \mu_{\mathrm{sl}} n_{\mathrm{s}}(\Phi) + \gamma_{\infty}^{\Phi}\Omega v_{\mathrm{s}}^{\frac{2}{3}} n_{\mathrm{s}}{}^{\frac{2}{3}}(\Phi)(1 + \zeta n_{\mathrm{s}}{}^{-\frac{1}{3}}(\Phi)), \tag{5.9}$$

where $\zeta = \frac{2}{3}\delta n_{\mathrm{s}}{}^{\Phi,V}\Omega v_{\mathrm{s}}^{\frac{2}{3}} - \epsilon$ is a constant. This $\zeta n_{\mathrm{s}}{}^{-\frac{1}{3}}(\Phi)$ term in Eqn. (5.9), which has the same mathematical form as the Tolman correction, has two components: the $\delta n_{\mathrm{s}}{}^{\Phi,V}$ term stems from the difference between the Φ-based dividing surface and the equimolar surface, while the other ϵ 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(Φ)+Tol to denote the expression in Eqn. (5.9) in order to highlight its dependence on the chosen extensive quantity. An awareness of this dependence on the choice of Φ is crucial when comparing studies performed with different protocols. Eqs. (5.7) and (5.9) 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.

5.3 Simulation methods

We simulated the processes of homogeneous nucleation for a simple but realistic Lennard-Jones system [103, 107, 108]. 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 6×10^6 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 [92, 109], and the collective variable $\Phi = \sum_i S(\kappa(i))$ that was described in chapter 2.3.4.

5.4 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 Φ of the system. We then extracted the nucleation free energy for a single solid cluster as a function of $n_{\mathrm{s}}(\Phi)$ using the framework introduced above, which is thoroughly described in Ref. [124]. Each free energy profile $G(n_{\mathrm{s}}(\Phi))$ is plotted as a thin blue curve in Figure 5.1. For each $G(n_{\mathrm{s}}(\Phi))$, we then performed a CNT(Φ)+Tol fit using Eqn. (5.9) with μ_{sl}, $\gamma_{\infty}^{\Phi}\Omega$, and ζ as fitting parameters. Each of these fitted curves is shown in red in Figure 5.1.

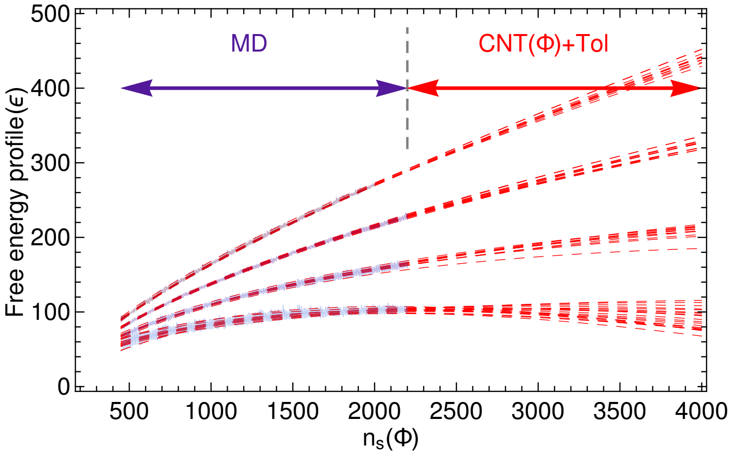

Figure 5.1 – The four sets of curves from bottom to top correspond to the free energy profiles of $n_\mathrm{s}(\Phi)$ at temperatures of 0.56, 0.58, 0.60 and 0.6185, respectively. Each thin blue curve is computed from a biased molecular dynamics simulation, and each thin red line corresponds to the best CNT(Φ)+Tol fit of one blue curve.

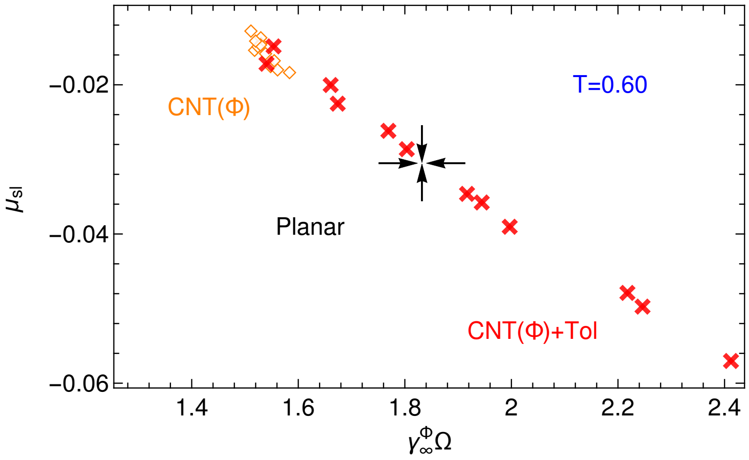

Figure 5.2 – Each symbol in red or orange represents one pair of fitting parameters ($\gamma_\infty^\Phi \Omega$ and μ_sl) that were obtained by fitting one of the free energy curves shown in figure 5.1 that were themselves extracted from one of the independent simulations that were performed at $T = 0.60$. The black arrows indicate the results obtained from simulations of planar interfaces. All quantities are expressed in Lennard-Jones units.

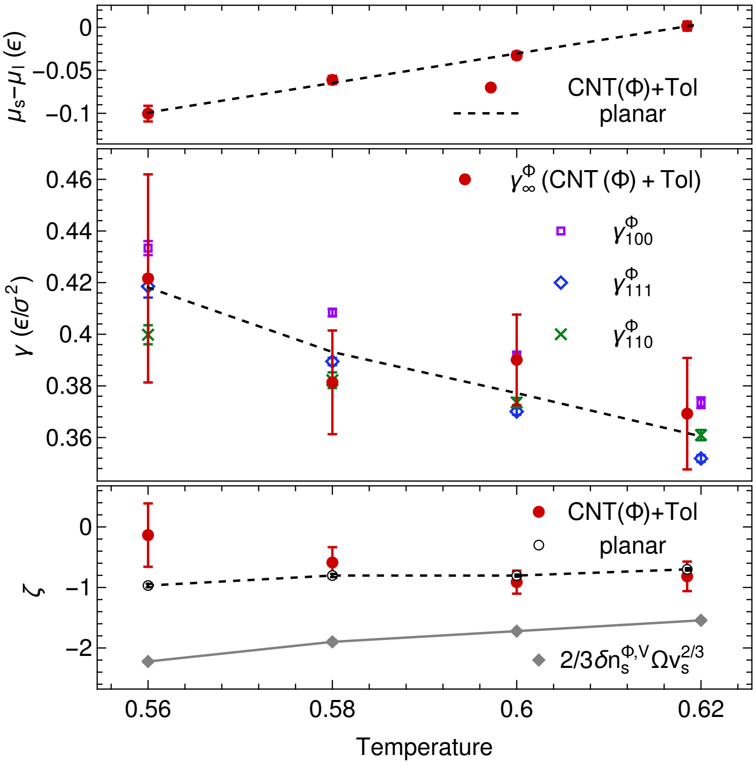

Figure 5.3 – In the top panel, the red symbols indicate μ_{sl} values computed from homogeneous nucleation simulations using the CNT(Φ)+Tol fit, the dashed line $\mu_{\mathrm{sl}} = 1.714(T - 0.6178)$ describes the fitted values of μ_{sl} that was extracted from planar interface calculations using a large simulation box with 20736 atoms. In the middle panel, the purple, blue and green symbols indicate the values of γ_∞^Φ for the planar interfaces that are perpendicular to three different lattice directions. The dashed line indicates the effective γ_∞^Φ 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(Φ)+Tol model divided by the estimated Ω at each temperature. In the bottom panel, the red symbols indicate the Tolman correction constant ζ in the CNT(Φ)+Tol fits, while the black symbols indicate the estimate for ζ that is obtained when the planar interface results for μ_{sl} and γ_∞^Φ are used in Eqn. (5.9). The gray symbols indicate the equimolar-surface correction to γ_∞^Φ 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.

Table 5.1 – A comparison of the predictions for μ_{sl} and γ_{∞}^{Φ} 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.

	μ_{sl}				γ_{∞}^{Φ}			
model	0.56	0.58	0.60	0.6185	0.56	0.58	0.60	0.6185
CNT(Φ)+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(Φ)	-0.083(2)	-0.049(1)	-0.016(1)	0.017(1)	0.346(5)	0.330(2)	0.317(1)	0.302(3)

The red curves in Figure 5.1 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 (μ_{sl}, $\gamma_{\infty}\Omega$) obtained from CNT+Tol(Φ) fits for each free energy profile at $T = 0.60$ in Figure 5.2. The two parameters μ_{sl} and $\gamma_{\infty}\Omega$ are clustered around a straight regression line in Figure 5.2, 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 μ_{sl} and $\gamma_{\infty}\Omega$ from the CNT(Φ)+Tol model are physically meaningful. To this end, we also show results obtained from simulations of the planar interfaces (see chapter 3) in Figure 5.2. The value of μ_{sl} for an interface perpendicular to any lattice direction can be computed by performing biased simulations of that interface [1, 113]. We computed these planar interfacial free energies in our previous work [1] and used a deterministic framework to locate the Gibbs dividing surface. We observed that the values of the interfacial free energies do not change significantly when the probabilistic framework in Eqn. (**??**) is applied, so we chose to adopt the values for γ^{Φ} and γ^{V} reported in chapter 3. To estimate $\gamma_{\infty}^{\Phi}\Omega$, we first obtained the values of γ_{100}^{Φ}, γ_{111}^{Φ} and γ_{110}^{Φ} 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 [99], whose coefficients can be parametrized using the values of γ_{100}^{Φ}, γ_{111}^{Φ} and γ_{110}^{Φ}. On the other hand, the shape of the equimolar surface of the nucleus $R(\vec{n})$ can then be reproduced by performing a Wulff construction using the γ^{V} values [125]. As the difference between the locations of the Φ-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 Φ-based dividing surface of the nucleus. From this shape, we can, therefore, estimate the geometrical constant Ω, and an effective value of γ_{∞}^{Φ} 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 Ω values were within 0.5% of the geometrical constant for a sphere.

Figure 5.2 suggests that the values of μ_{sl} and $\gamma_{\infty}^{\Phi}\Omega$ from the CNT(Φ)+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 7.1 shows all the values of $\mu_{\mathrm{sl}}(T)$ together with the values of $\gamma_{\infty}(T)$ that are predicted both from the CNT(Φ)+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(Φ)+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 μ_{sl} and $\gamma_{\infty}^{\Phi}\Omega$ may not agree with planar limit results. For example, we also performed a so-called CNT(Φ) fit on each computed $G(n_{\mathrm{s}}(\Phi))$, using a conventional CNT formulation $G(n_{\mathrm{s}}(\Phi)) = \mu_{\mathrm{sl}} n_{\mathrm{s}}(\Phi) + \gamma_{\infty}^{\Phi}\Omega v_{\mathrm{s}}^{\frac{2}{3}} n_{\mathrm{s}}^{\frac{2}{3}}(\Phi)$. The values of μ_{sl} and $\gamma_{\infty}^{\Phi}\Omega$ from the CNT(Φ) fits are indicated in Figure 5.2 and Table 7.1. 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 μ_{sl} and $\gamma_{\infty}^{\Phi}\Omega$ from the CNT(Φ)+Tol and the planar limit results suggests that one should just use the results obtained at the planar interface limit in Eqn. (5.9) so as to have a model that has ζ 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 ζ with the values obtained from the regular CNT(Φ)+Tol fits using all three parameters. The bottom panel of Figure 6.4, shows that this new approach results in much smaller uncertainties in the values of ζ. 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 μ_{sl} and $\gamma_{\infty}^{\Phi}\Omega$ from simulations of planar interfaces. This allows one to determine the leading terms $\mu_{\mathrm{sl}} n_{\mathrm{s}}(\Phi) + \gamma_{\infty}^{\Phi}\Omega v_{\mathrm{s}}^{\frac{2}{3}} n_{\mathrm{s}}^{\frac{2}{3}}(\Phi)$ 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 $\zeta = \frac{2}{3}\delta n_{\mathrm{s}}{}^{\Phi,V}\Omega v_{\mathrm{s}}^{\frac{2}{3}} - \epsilon$ in Eqn. (5.9) 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 $\delta n_{\mathrm{s}}{}^{\Phi,V}$ for different definitions of the dividing surface can be directly evaluated from simulations of planar interfaces as discussed in chapter 3. We evaluated $\delta n_{\mathrm{s}}{}^{\Phi,V}$ for planar interfaces with normal vectors parallel to the ⟨100⟩, ⟨111⟩ and ⟨110⟩ lattice directions at each temperature, and then approximated ${\delta n_\mathrm{s}}^{\Phi,V}(\vec{n})$ using a cubic harmonic expansion, and thus estimated its averaged value for a three-dimensional nucleus. These estimated values for $\frac{2}{3}{\delta n_\mathrm{s}}^{\Phi,V}\Omega v_\mathrm{s}^{\frac{2}{3}}$ are shown in the bottom panel of Figure 6.4. The other term $\epsilon = (32\pi/3)^{\frac{1}{3}}\, {v_\mathrm{s}}^{-\frac{1}{3}}\delta$ that enters ζ can be determined from the Tolman length δ. This quantity can be evaluated from the pressure tensor at each position along the normal to a planar interface for fluid-fluid systems [126–128]. How this quantity should be computed for solid-liquid interfaces is, however, unclear. As such, we chose to subtract the values we obtained for $\frac{2}{3}{\delta n_\mathrm{s}}^{\Phi,V}\Omega v_\mathrm{s}^{\frac{2}{3}}$ from the value we obtained for ζ and to extract the Tolman length from ϵ. In this way, we predicted the Tolman length to be

$$\delta = \frac{\frac{2}{3}{\delta n_\mathrm{s}}^{\Phi,V}\Omega v_\mathrm{s}^{\frac{2}{3}} - \zeta}{(32\pi/3)^{\frac{1}{3}}\, {v_\mathrm{s}}^{-\frac{1}{3}}}, \tag{5.10}$$

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(Φ)+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.

5.5 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 Φ when calculating the surface excess free energy. As Φ 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.

Finally, 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.

6 Computing the Tolman length for solid-liquid interfaces

In the previous chapter, we demonstrated that, after the size of the nucleus is determined by a Φ-based Gibbs dividing surface, the nucleation free energy profile as a function of the number of solid atoms $n_\mathrm{s}(\Phi)$ can be expressed as

$$G(n_\mathrm{s}(\Phi)) \eqsim \mu_\mathrm{sl} n_\mathrm{s}(\Phi) + \gamma_\infty^\Phi \Omega v_\mathrm{s}^{\frac{2}{3}} n_\mathrm{s}{}^{\frac{2}{3}}(\Phi)(1 + \zeta n_\mathrm{s}{}^{-\frac{1}{3}}(\Phi)), \tag{6.1}$$

where $\zeta = \frac{2}{3}\delta n_\mathrm{s}{}^{\Phi,V} \Omega v_\mathrm{s}^{\frac{2}{3}} - (32\pi/3)^{\frac{1}{3}} v_\mathrm{s}{}^{-\frac{1}{3}} \delta$ is the Tolman correction, which has two components: the $\delta n_\mathrm{s}{}^{\Phi,V}$ term stems from the difference between the Φ-based dividing surface and the equimolar surface, while the other term $-(32\pi/3)^{\frac{1}{3}} v_\mathrm{s}{}^{-\frac{1}{3}} \delta$ is determined by the Tolman length δ. Equally validly, when the volume V is used to define the dividing surface, i.e. the equimolar dividing surface, the free energy profile of nucleation can be expressed as:

$$G(n_\mathrm{s}(V)) = \mu_\mathrm{sl} n_\mathrm{s}(V) + \gamma^V \Omega v_\mathrm{s}^{\frac{2}{3}} n_\mathrm{s}{}^{\frac{2}{3}}(V)(1 - \epsilon n_\mathrm{s}{}^{-\frac{1}{3}}(V)), \tag{6.2}$$

where $\epsilon = (32\pi/3)^{\frac{1}{3}} v_\mathrm{s}{}^{-\frac{1}{3}} \delta$. One can treat μ_sl, γ^V and δ as three fitting parameters, or obtain μ_sl and γ^V independently from other metadynamics simulations of planar interfaces [1, 113] and just use δ as the only fitting parameter.

Although the Tolman length δ can be estimated from fitting the free energy profile to the CNT expression, it is highly desirable to directly compute it from atomistic simulations, because (i) there is no guarantee that the CNT is valid for a certain system, (ii) obtaining δ from the fitting is extraordinarily sensitive to small deviations in the computed free energy profiles, because all three fitting parameters in the CNT expression are strongly correlated, as demonstrated in the last chapter, and (iii) the values of the Tolman length can be different for solid-liquid interfaces along different crystallographic directions, and only an effective value average over all orientations can be obtained from simulations of homogeneous nucleation.

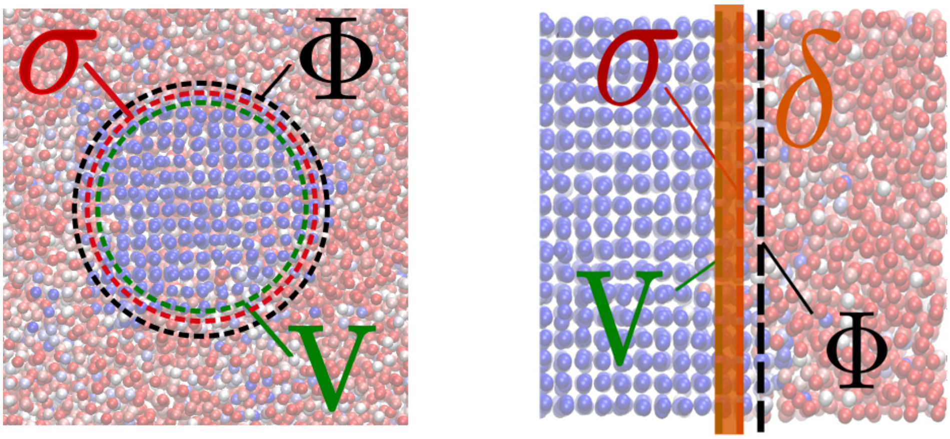

Figure 6.1 – A schematic showing the location of different dividing surfaces (equimolar V, surface of tension σ, and Φ-based dividing surface) for a 3-dimensional nucleus (left) and a planar interface (right). The distance between the surface of tension and the equimolar dividing surface at the planar limit is the Tolman length δ.

The Tolman length δ determines the leading order curvature dependence of the interfacial free energy between two phases [104, 129]. As such, besides nucleation, it also plays a important role in other phenomena that involve curved interfaces such as Ostwald ripening [130]. For many decades, the sign and magnitude of the Tolman length has remained a source of considerable controversy [126, 131–133], due to the conceptual and practical difficulties involved in measuring or computing it.

For fluid-fluid interfaces, the Tolman length can be computed from molecular dynamics simulations by summing over the local pressure tensor across the phase boundary [128, 132, 134–136]. This same procedure is not directly applicable for solid-liquid interfaces, however, because the local pressure tensor is ill-defined for systems that are not uniform fluids [134, 137, 138], and also because the elastic energy stored inside a solid interferes with the determination of the stress components that are associated with the surface tension. Due to these theoretical and technical challenges, the value δ for a solid-liquid interface has never been evaluated directly. Instead it has only ever been used as a curvature correction fitting parameter that enters the free energy expression for nucleation [49, 106]. Furthermore, in many nucleation studies [10] this curvature correction term in interfacial free energies is altogether omitted because of the complexities and the statistical error involved in simultaneously determining multiple parameters in a numerical fit of nucleation free energies [2]. As such, the lack of the reference values for δ for solid-liquid interfaces significantly hampers the efforts to quantitatively model and predict nucleation barrier and rate.

Here we introduces a methodology that can be used to determine directly and accurately the Tolman length δ for solid-liquid interfaces using atomistic simulations. To this aim, we first revisit the key thermodynamic concepts that are related to δ, and then reformulate the problem in an alternative manner at out-of-equilibrium conditions, while refining the capillary fluctuation method for computing interfacial free energies.

6.1 The Tolman length and the surface of tension

The definition of the Tolman length is closely related to the concept of the Gibbs dividing surface, which is an infinitely thin geometrical surface whose position sensibly coincides with the discontinuity between the two phases. The most common way to determine its precise location is to consider a reference system in which the bulk phases extend up to the interface, and has the same value of a chosen extensive quantity (e.g. volume, energy, the sum of order parameters of all atoms) as the actual system, with no excess term associated with the interface [58]. Among all the possible choices for the diving surface, the surface of tension (σ-surface), corresponding to the position at which the mechanical definition of tension applies, is regarded as “special" because the associated interfacial free energy γ^{σ} is curvature-independent. On the other hand, the equimolar dividing surface (V-surface) that has no surface excess of volume is commonly used when analyzing nucleation, because this surface encloses a nucleus that has the same density as the bulk, which streamlines the formulation of nucleation free energy profiles [105, 106]. The Tolman length characterizes the specific free energy of a spherical V-surface with radius R, such that $\gamma^{V}(R) = \gamma^{V}(1 - 2\delta/R + \mathcal{O}(1/R^{2}))$. Furthermore, in the limit of a planar interface, the Tolman length is just the difference between the location of the surface of tension and of the equimolar dividing surface, i.e. [104]

$$\delta = h^{V} - h^{\sigma} \tag{6.3}$$

where h^{V} and h^{σ} indicate the height of the two dividing surfaces for the planar interface, and $h = -\infty$ is inside the bulk solid. Eqn. (6.3) also implies that the Tolman length is related to the interfacial free energy difference between the V and the σ-surfaces by [129]

$$\delta = v_{\mathrm{s}} \frac{\gamma^{\sigma} - \gamma^{V}}{\mu_{\mathrm{sl}}}, \tag{6.4}$$

where $\mu_{\mathrm{sl}} = \mu_{\mathrm{s}} - \mu_{\mathrm{l}}$ is the chemical potential difference between the solid and liquid phases, and v_{s} is the molar volume of the bulk solid. δ / v_{s} can be interpreted as the adsorption of solid atoms per unit area at the surface of tension, and as such Eqn. (6.4) directly stems from Gibbs adsorption isotherm [1, 104]. Notice that Eqn. (6.4) is only useful away from the coexistence point so that $\mu_{\mathrm{sl}} \neq 0$.

6.2 Free energy of surface of tension

Eqn. (6.4) provides a recipe for computing the Tolman length using the values of γ^{V} and γ^{σ} at out-of-equilibrium conditions. A method that employs metadynamics for computing γ^{V} away from the coexistence temperature T_{m} has only recently become available [1, 65, 96], so in the present study we focus on obtaining γ^{σ} by applying a capillary wave model.

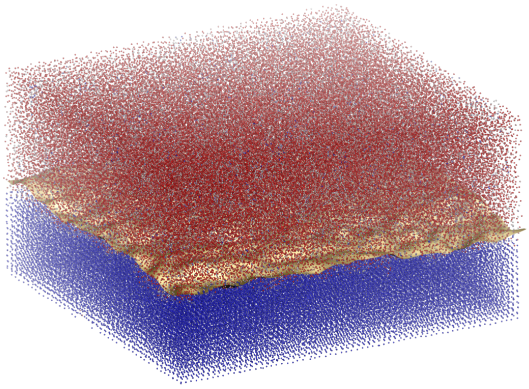

Figure 6.2 – A schematic showing the capillary waves for a solid-liquid planar interface.

Even at the macroscopic scale, a planar solid-liquid interface is not completely flat due to the long wavelength thermal distortions of the interface, which are generally referred to as capillary waves (Figure 6.2) [99, 139]. When the extra surface area δs is generated due to the capillary waves, the lattice spacing of the solid is conserved and its elastic energy is unchanged. Following a mechanical definition of the surface energy, without heat transfer, change in system size or change in composition of the two phases, the variation in free energy is entirely captured by the change in surface energy $\delta E = \gamma^\sigma \delta s$. According to the capillary wave model, the surface energy of the fluctuating interface with a height function $h(x, y)$ can be approximated by integrating the orientation-dependent surface tension over the interface [99, 140], i.e.

$$E_{surf} = \int_s ds \gamma^\sigma(\vec{n}) \tag{6.5}$$

where $\vec{n}$ is the interface normal vector, and $ds \approx dxdy(1 + (dh/dx)^2/2 + (dh/dy)^2/2)$ which is the surface element. The local curvature of the interface does not enter the equation above as γ^σ is curvature independent by construction. If each capillary fluctuation mode is in thermal equilibrium, the interfacial stiffness and the ensemble average of the long wavelength Fourier components are related via

$$\langle |A_h(k_x, k_y)|^2 \rangle = \frac{k_B T}{l_x l_y (k_x^2 \widetilde{\gamma}_{11} + 2k_x k_y \widetilde{\gamma}_{12} + k_y^2 \widetilde{\gamma}_{22})} \tag{6.6}$$

The $\widetilde{\gamma}_{ij} = \gamma^\sigma + \dfrac{\partial^2 \gamma^\sigma}{\partial \theta_i \theta_j}|_{\theta_{i,j}=0}$ are the components of stiffness tensor.

Eqn. (6.6) is a standard expression for the capillary fluctuation method (CFM) [99, 100, 141], which is formally valid at thermal equilibrium and so far has only been employed at the coexistence temperature T_m when the planar interface is not under the driving force to migrate. [1] It was shown in previous work that estimations of the interfacial free energy at T_m for various types of solid-liquid systems using CFM [99, 100, 141] are consistent with those obtained using other free energy methods such as metadynamics and thermodynamic integration [96, 102, 143, 144], which suggests that the central tenet of CFM – the mechanical definition of the surface energy – is valid for solid-liquid interfaces that exhibit capillary fluctuations.

However, for our purposes of using Eqn. (6.4), γ^σ has to be computed at $T \neq T_\mathrm{m}$ so the capillary fluctuation method has to be extended to out-of-equilibrium conditions. At such conditions, the planar interface is under a driving force to migrate, because of the chemical potential imbalance between the metastable and the stable phases. However, an umbrella potential can be introduced to counter-balance this chemical potential difference and to pin the interface away from the coexistence temperature T_m [113]: taking the actual Hamiltonian of the system to be $\mathscr{H}(\mathbf{q})$, the biased Hamiltonian can be expressed as

$$\mathscr{H}_{biased}(\mathbf{q}) = \mathscr{H}(\mathbf{q}) + \frac{\alpha}{2}\left(\Phi - \bar{\Phi}\right)^2, \tag{6.7}$$

where $\Phi = \sum_{i=1}^{N} \phi_i$ is an extensive quantity of the whole system that is constructed by summing up order parameter of each atom, and ϕ_i is an atom-centered order parameter that is able to discriminate between the atoms that belong to each of the two different phases.

With the biased Hamiltonian in Eqn. (6.7), which introduced a constrained ensemble of states at out-of-equilibrium conditions, CFM can be extended to conditions that are away from the coexistence temperature T_m. To elucidate why this is the case, as well as to evaluate $\langle |A_h(k_x, k_y)|^2 \rangle$ in Eqn. (6.11) in a more efficient and accurate manner, we propose a new and efficient method for locating the fluctuating surface, which is an extension of the original formulation of the Gibbs dividing surface.

6.3 Fluctuating Gibbs dividing surface

Let us consider a solid-liquid system that has N atoms, a box size $\{l_x, l_y, l_z\}$, and a planar interface perpendicular to the z axis. We first introduce an instantaneous order parameter density field

$$\widetilde{\phi}(x, y, z) = \sum_{i=1}^{N} \phi_i g(x - x_i) g(y - y_i) g(z - z_i), \tag{6.8}$$

where (x_i, y_i, z_i) and ϕ_i denote the coordinate and the order parameter of the i-th atom, respectively, and g is a normalized kernel function which is chosen to be a Dirac delta function in this case. The zero-excess condition that defines the height of the interface $h(x, y)$ at any point can then be written in terms of a line integral of the phase field (Eqn. (6.8)) along the z

---

[1] CFM has been employed for the water-ice interface formed at undercooled conditions due to premelting [142].

axis over a range that contains the interface

$$\int_0^{l_z} \mathrm{d}z\,\widetilde{\phi}(x,y,z) = \int_0^{h(x,y)} \mathrm{d}z\,\widetilde{\phi}_\mathrm{s} + \int_{h(x,y)}^{l_z} \mathrm{d}z\,\widetilde{\phi}_\mathrm{l}, \tag{6.9}$$

Here $\widetilde{\phi}_\mathrm{s}$ and $\widetilde{\phi}_\mathrm{l}$ indicate the density field inside the bulk solid and liquid phases, respectively. Eqn. 6.9 can be seen as an extension of the planar Gibbs dividing surface for the whole system [1, 124] that is restricted to an infinitesimally thin domain centered around (x,y). Combining Eqn. (6.8) and (6.9), one can write

$$\int_0^{h(x,y)} \mathrm{d}z\,\widetilde{\phi}_\mathrm{s} + \int_{h(x,y)}^{l_z} \mathrm{d}z\,\widetilde{\phi}_\mathrm{l} = \sum_{i=1}^{N} \phi_i g(x-x_i) g(y-y_i). \tag{6.10}$$

One can take a 2-D Fourier expansion for both sides of Eqn. (6.10) over the whole cross section of the simulation box $\{l_x, l_y\}$ and perform an ensemble average for the amplitude of each Fourier mode. By making the assumption that the bulk fluctuations and the surface fluctuations are mutually independent, one obtains

$$\begin{aligned}(\langle\widetilde{\phi}_\mathrm{s}\rangle - \langle\widetilde{\phi}_\mathrm{l}\rangle)^2 \langle|A_h(k_x,k_y)|^2\rangle + \langle|A_s(\langle h\rangle;k_x,k_y)|^2\rangle + \langle|A_l(l_z-\langle h\rangle;k_x,k_y)|^2\rangle \\ = \langle \left[\frac{1}{l_x l_y} \sum_{i=1}^{N} \phi_i \exp(-\mathrm{i}k_x x_i - \mathrm{i}k_y y_i) \right]^2 \rangle.\end{aligned} \tag{6.11}$$

In Eqn. (6.11), $\langle\widetilde{\phi}_\mathrm{s}\rangle$ and $\langle\widetilde{\phi}_\mathrm{l}\rangle$ are the averaged values from the bulk fields. $A_s(\langle h\rangle;k_x,k_y)$ and $A_l(l_z-\langle h\rangle;k_x,k_y)$ are the Fourier coefficients characterizing the bulk fluctuations, respectively for a slab of bulk solid that has a cross section $\{l_x,l_y\}$ and thickness $\langle h\rangle$, and for a bulk liquid with the dimensions $\{l_x,l_y,l_z-\langle h\rangle\}$. The average amplitudes of these bulk quantities can all be evaluated separately from simulations of the bulk phases using a simulation box of the same cross section as the solid-liquid system, using expressions such as

$$\langle|A_s(\langle h\rangle;k_x,k_y)|^2\rangle = \langle \left[\frac{1}{l_x l_y} \sum_{i=1}^{N_s} \phi_i \exp(-\mathrm{i}k_x x_i - \mathrm{i}k_y y_i) H(\langle h\rangle - z_i) H(z_i) \right]^2 \rangle, \tag{6.12}$$

where $H(\dots)$ is the Heaviside function and N_s is the number of atoms in the bulk solid system. In the Appendix we also discuss additional and optional approximations that one can make to simplify the evaluation of $\langle|A_s(\langle h\rangle;k_x,k_y)|^2\rangle$ and $\langle|A_l(\langle h\rangle;k_x,k_y)|^2\rangle$. As the right hand side of Eqn. (6.11) can also be evaluated directly from the snapshots of atomic coordinates for the solid-liquid system in molecular dynamics simulations, the only remaining term $\langle|A_h(k_x,k_y)|^2\rangle$ that enters the CFM expression (Eqn. (6.6)) can be determined. Note that when there are two parallel planar interfaces in the system, one can simply modify Eqn. (6.11) by adding a factor of two to the term $\langle|A_h(k_x,k_y)|^2\rangle$.

The zeroth Fourier mode of the height function $A_h(0,0)$, corresponds to the average height of the fluctuating interface. $\langle h\rangle = \int_0^{l_x} \mathrm{d}x \int_0^{l_y} \mathrm{d}y\, h(x,y)/l_x l_y$ corresponds to the position of the conventional planar Gibbs dividing surface of the whole system that has zero surface excess for the extensive quantity $\Phi = \sum_{i=1}^{N} \phi_i$, since

$$\Phi = \int_0^{l_x} \mathrm{d}x \int_0^{l_y} \mathrm{d}y \left[\int_0^{\langle h \rangle} \widetilde{\phi}_\mathrm{s} \mathrm{d}z + \int_{\langle h \rangle}^{l_z} \widetilde{\phi}_\mathrm{l} \mathrm{d}z \right]. \tag{6.13}$$

As extensively discussed in Ref. [1, 104], the location of this planar dividing surface $\langle h \rangle$ determines its surface absorption and thus affects its interfacial free energy. Meanwhile, the magnitudes of all other Fourier modes with non-zero frequencies ($A_h(k_x, k_y)$) do not change the proportion between the solid and the liquid atoms in the system, and thus do not affect the surface absorption.

The average height $\langle h \rangle$ of the interface is special also because it only depends on the global extensive quantity Φ. As the umbrella bias potential in Eqn. (6.7) is only a function of Φ, the bias does not act on any capillary fluctuation modes other than the zeroth Fourier mode $\langle h \rangle$. Consequently, the effect of the bias balances the chemical potential difference between the two phases, without altering the energy distributions for other non-zero frequencies. Because the system under the biased Hamiltonian (Eqn. (6.7)) is in thermodynamic equilibrium, the equipartition theorem holds for capillary fluctuation modes with non-zero frequencies, and Eqn. (6.6) is thus still valid.

Another subtle point is that, at a reasonably large length scale, the height function $h(x, y)$ of the fluctuating interfaces defined by different order parameters are parallel to each other, and the magnitudes for long wavelength Fourier modes are identical. As extensively discussed in the Appendix, for each (k_x, k_y) wave vector that is smaller than a certain cutoff, the Fourier amplitude and the corresponding value of $\widetilde{\gamma}$ are independent from the choice of the order parameter that is used to define the fluctuating dividing surface. Previous studies that employ different order parameters and even distinct criteria for locating the interface also arrived at consistent estimations for $\widetilde{\gamma}$ and γ^σ at T_m [99, 100, 141, 145].

6.4 Simulation methods

We simulated the solid-liquid planar interfaces for a simple but realistic Lennard-Jones system [96, 103, 108]. The NPT ensemble was employed throughout with the stochastic velocity rescaling thermostat [90]. A Nose-Hoover barostat was used along the z axis which was set up to be perpendicular to the interface, The dimensions of the supercell along x and y are commensurate with the equilibrium lattice parameters. We considered planar interfaces along the $\langle 100 \rangle$ and $\langle 110 \rangle$ crystallographic directions of the *fcc* lattice, using supercells containing 256000 and 315392 atoms, respectively. For each orientation and at each temperature (0.56, 0.58, 0.60 and $T_\mathrm{m} = 0.6178$), 10 independent umbrella sampling [34] simulations with the biased Hamiltonian (Eqn. (6.7)) were performed, using the collective variable $\Phi = \sum_i \phi_i = \sum_i S(\kappa(i))$ (see chapter 2.3.4). Fast implementation of this simulation setup was made possible by the flexibility of the PLUMED code [93] in combination with LAMMPS [91].

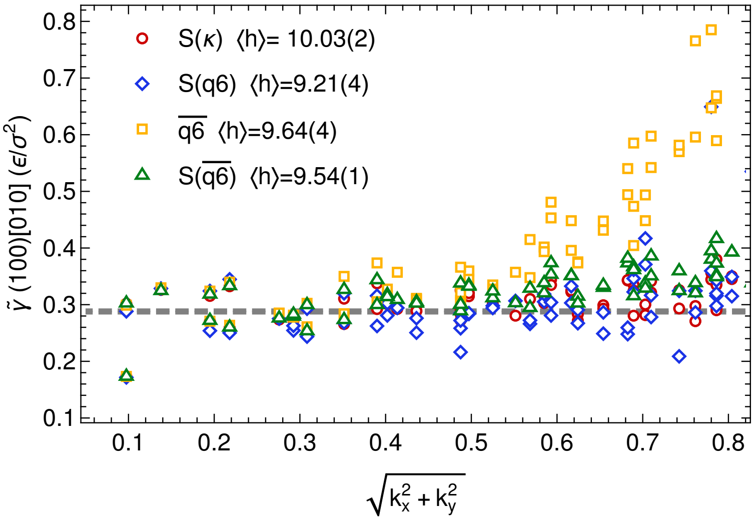

Figure 6.3 – The interfacial stiffness $\tilde{\gamma}(100)[010]$ estimated for each pair of (k_x, k_y) wave vector using dividing surfaces defined using different order parameters at $T = 0.58$, for a single set of trajectories. The FCCUBIC κ [96], Q6 and locally averaged Q6 [146] atomic order parameters were employed, and S indicates a switching function [147]. The converged value of the stiffness is indicated using a dashed gray band. The average height $\langle h \rangle$ of each dividing surface is also specified in the Lennard-Jones length unit, and the numbers in bracket indicate the error of the last digit.

By combining Eqn. (6.6) and Eqn. (6.11), the interfacial stiffness $\tilde{\gamma}$ can be extracted from each thermal capillary fluctuation mode during post processing. For the ⟨100⟩ or ⟨110⟩ interfaces, we first extracted the interfacial stiffness tensors $\tilde{\gamma}_{11}$ and $\tilde{\gamma}_{22}$ at each temperature using the atomic order parameter $\phi_i = S(\kappa(i))$. After computing the stiffness tensors, we expanded it in cubic harmonics [100], which are consistent with the symmetry of the *fcc* crystal, so as to extract the orientation-dependent interfacial free energy $\gamma^\sigma(\vec{n})$.

6.5 Convergence tests

Before presenting the results from the production runs, we want to demonstrate the validity of employing Eqn. (6.6) and (6.11) to extract $\tilde{\gamma}$ at undercooled conditions using different order parameters. We thus selected a few trajectories generated from umbrella simulations for the ⟨100⟩ solid-liquid interface at $T = 0.58$, and analyzed them using distinct order parameters (e.g. FCCUBIC κ, Q6 and locally averaged Q6). Note that these order parameters are based on distinct constructions, and for the solid-liquid system the average heights of the dividing surfaces $\langle h \rangle$ defined using these order parameters are significantly apart. For each (k_x, k_y) wave vector, and for each choice of the atom-centred order parameter, the Fourier components $\langle |A_h(k_x, k_y)|^2 \rangle$ for the instantaneous interface height were first computed using Eqn. (6.11), and the interfacial stiffness tensors ($\tilde{\gamma}_{11} = \tilde{\gamma}_{22} = \tilde{\gamma}(100)[010]$ and $\tilde{\gamma}_{12} = 0$ for the ⟨100⟩ interface) for that wave vector were obtained from Eqn. (6.6). For each (k_x, k_y) wave vector that is small, the values of $\widetilde{\gamma}(100)[010]$ obtained using different order parameters are in a perfect agreement, even when ergodic sampling for that Fourier mode has not been achieved. This consistency is a strong indication that the fluctuating Gibbs dividing surfaces defined by different order parameters are parallel to each other at the length scale of $2\pi/k_{x,y}$. At large $k_x, k_y \to 1$, which corresponds to a wavelength of just a few lattice spacing, the $\langle|A_h(k_x, k_y)|^2\rangle$ that enters Eqn. (6.6) start to depend on the specific choice of the order parameter, implying that these fluctuating Gibbs dividing surfaces cease to be parallel to the surface of tension at such length scale. In other words, $\widetilde{\gamma}$ can only be extracted at the constant regime in where each Fourier mode (k_x, k_y) fulfills the equi-partition theorem and the effective stiffness of that mode is independent from the wave vector.

Although the same value of stiffness can be extracted using dividing surfaces defined by different order parameters, the statistical efficiency associated with each order parameter can be quite different for two reasons: 1) the values of stiffness can only be obtained from capillary waves at small (k_x, k_y) and one should set a cutoff k_{cut} such that the effective stiffness extracted from each low k mode is constant. For diving surfaces defined by different order parameters, the correlation length along the cross section is different. For example, the order parameters that were constructed based on local averaging tend to be associated with longer correlation length, and thus k_{cut} has to be set smaller, which hinders the statistical efficiency. 2) the fluctuations in the bulk phases also contribute to the Fourier amplitudes (Eqn. (6.11)), and the order parameter that gives smaller $\langle|A_s(\langle h\rangle; k_x, k_y)|^2\rangle$ and $\langle|A_l(l_z - \langle h\rangle; k_x, k_y)|^2\rangle$ tends to give better efficiency.

6.6 Results

In Figure 6.4 we plot the interfacial free energy of the surface of tension γ^σ for the $\langle 100\rangle$, $\langle 111\rangle$ and $\langle 110\rangle$ interfaces, together with free energies of planar equimolar dividing surfaces γ^V that were obtained from our previous metadynamics simulations [1]. The interfacial free energies at T_m for all three crystalline orientations agree very well with previous results using different methods including thermodynamic integration, metadynamics, and CFM [1, 100, 103, 141, 145]. Unsurprisingly, the value, temperature dependence, and anisotropy among the principal orientations of γ vary considerably with the choice of the Gibbs dividing surface. In theory, at T_m, $\gamma^\sigma = \gamma^V$, but in reality we notice a small difference between them. This is due to the fact that γ^V was computed using metadynamics simulations with small supercells comprising about 1200 atoms, which implies a small finite-size effect that increases interfacial free energies [96], and also shifts T_m to 0.62 [1]. In order to compensate for this small finite-size effect, we shifted vertically the values of γ^V so that $\gamma^\sigma(T_m) = \gamma^V(T_m)$ before computing the Tolman length using Eqn. (6.4).

The purple, blue and green symbols in the top panel of Figure 6.4 indicate the estimates of the Tolman length for the $\langle 100\rangle$, $\langle 111\rangle$ and $\langle 110\rangle$ interfaces. There is significant anisotropy

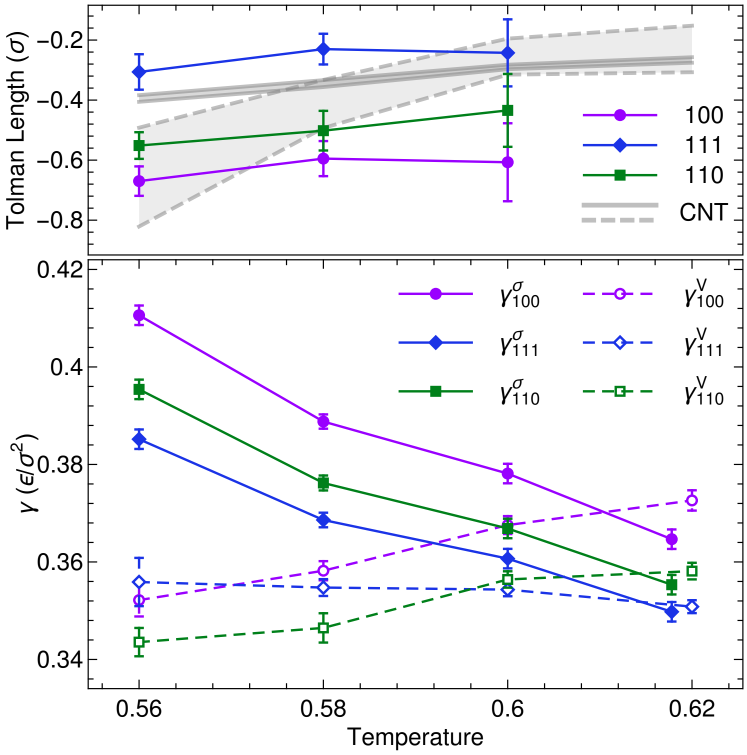

Figure 6.4 – The bottom panel shows the the planar interfacial free energies of the surface of tension (γ^{σ}) and the equimolar dividing surface (γ^{V}) for three different lattice directions at different temperatures ($T_{\mathrm{m}} = 0.6178$). The data for γ^{V} were obtained from Ref. [1]. The purple, blue and green symbols in the top panel indicate the predictions of the Tolman length for the ⟨100⟩, ⟨111⟩ and ⟨110⟩ solid-liquid interfaces. The dashed gray band shows indirect estimates of the orientation-averaged Tolman length of a solid nucleus from homogeneous nucleation simulations, where the Tolman length, the chemical potential difference and the interfacial free energy were treated as three fitting parameters in a classical nucleation theory expression [106]. The solid gray band indicates another set of results from fitting nucleation free energy profiles, where δ is used as the sole parameter, and the interfacial free energies and chemical potentials are obtained elsewhere from independent planar interface simulations [1, 113]. Statistical errors of the mean estimations are indicated using either error bars or band widths. All quantities are in Lenard Jones units.

in δ among the three crystal orientations, and little temperature dependence. The negative value of δ suggest that the equimolar dividing surface is closer to the bulk solid than the surface of tension, and that the interfacial energy of a curved equimolar surface $\gamma^V(R)$ has a positive curvature dependence. For comparison, in Figure 6.4 we also plot the estimate of the orientation-averaged Tolman length for three dimensional solid nucleus from previous homogeneous nucleation simulations [106]. These values were obtained by fitting nucleation free energy profiles with a classical nucleation theory expression with a Tolman correction term in Eqn. (6.1) and Eqn. (6.2). One can treat μ_{sl}, γ and δ as three fitting parameters, and the dashed gray band in Figure 6.4 shows the indirect estimation of the Tolman length δ using this approach. One can also obtain μ_{sl} and γ independently from other metadynamics simulations of planar interfaces [1, 113] and just use δ as the only fitting parameter. δ estimated in such a way is plotted with the solid gray band in Figure 6.4. Overall, the estimations of the Tolman length from all these three different methods agree. This agreement not only corroborates the present framework and method, but also indicates that a three-parameters CNT (including the chemical potential difference μ_{sl}, the interfacial free energy γ^V and the Tolman length δ) can accurately describe the free energy profile for homogeneous nucleation of the Lennard-Jones crystal from its melt.

6.7 Conclusions

In summary, we presented a thermodynamic framework that enables a direct evaluation of the Tolman length for planar solid-liquid interfaces in atomistic simulations. This framework is the coronation of an effort to streamline the study of nucleation by means of atomistic simulations, and relies on (i) the rigorous definition of the Gibbs dividing surface based on an atomic-scale order parameter and its fluctuations [1, 124], which we have used here to formulate an elegant and efficient version of the capillary fluctuation method that does not require the explicit geometric location of the dividing surface; (ii) the use of metadynamics to compute the planar-interface surface energy for the equimolar dividing surface γ^V in out-of-equilibrium conditions [1, 96]; (iii) the calculation of the mechanical surface tension γ^σ based on a capillary fluctuation analysis at $T \neq T_m$. Based on these theoretical advances, we computed γ^σ and the Tolman lengths δ of the solid-liquid planar interfaces of three principal crystal directions for a model system. In this case, we find that the values of δ that we obtained by evaluating it directly are in good agreement with the ones obtained by fitting homogeneous nucleation free energy profiles using CNT expressions. The framework we presented opens the door to a rigorous determination of the Tolman length δ and the free energy of the surface of tension γ^σ for various physical systems, and for different classes of homogeneous and heterogeneous phase transitions. The ability to compute these quantities is crucial both to verify the consistency of classical nucleation theory for a given problem, and to obtain quantitative predictions of nucleation rates by means of atomistic modelling.

7 Theoretical prediction of the homogeneous ice nucleation rate

In the previous chapters, we have developed a rigorous and comprehensive thermodynamic model that bridges the gap between macroscopic model and atomistic simulation of homogeneous nucleation. We removed the ambiguities in macroscopic classical nucleation theory, and demonstrated how the information contained in atomistic simulations of homogeneous nucleation should be used when fitting the parameters in macroscopic models. In this chapter, we will apply the thermodynamic framework to the case of homogeneous ice nucleation from undercooled liquid water. It is not only an ubiquitous process in nature that influences global phenomena such as climate change, but also has many practical implications in refrigeration, anti-freezing, solidification and melting of solutions, as well as many other technological applications [7, 9, 10]. Along with the application, we will also develop additional techniques that bring us an efficient and rigorous estimate of the ice nucleation rate.

7.1 Motivation

Despite its pivotal importance, our understanding of homogeneous ice nucleation and homogeneous nucleation in general is far from complete, which is partly due to that fact the experimental investigation of dynamical nucleation processes is very difficult and often costly [26]. An alternative is to rely on atomistic simulations to study nucleation, which has gained a lot of popularity in the last two decades [10]. However, predicting ice nucleation rates using atomistic simulations is plagued by difficulties. One major challenge arises from the inaccuracies in modelling the unique properties of water using either empirical potentials or *ab initio* methods [148–153]. On the other hand, as illustrated in Figure 7.1, even when the same water model and the same thermodynamic conditions are assumed, nucleation rates predicted in different studies typically differ by as much as 5-10 orders of magnitude [10]. This discrepancy is due to the fact that it is often necessary to evoke the standard form of classical nucleation theory (CNT) in order to estimate quantities such as the nucleation barrier, as the long time scale of nucleation rules out the option of brute force molecular dynamics simulations. However, a number of approximations within CNT have been shown to be over-simplifications [10], and more importantly, it is highly non-trivial to extract the values of the parameters that enter CNT

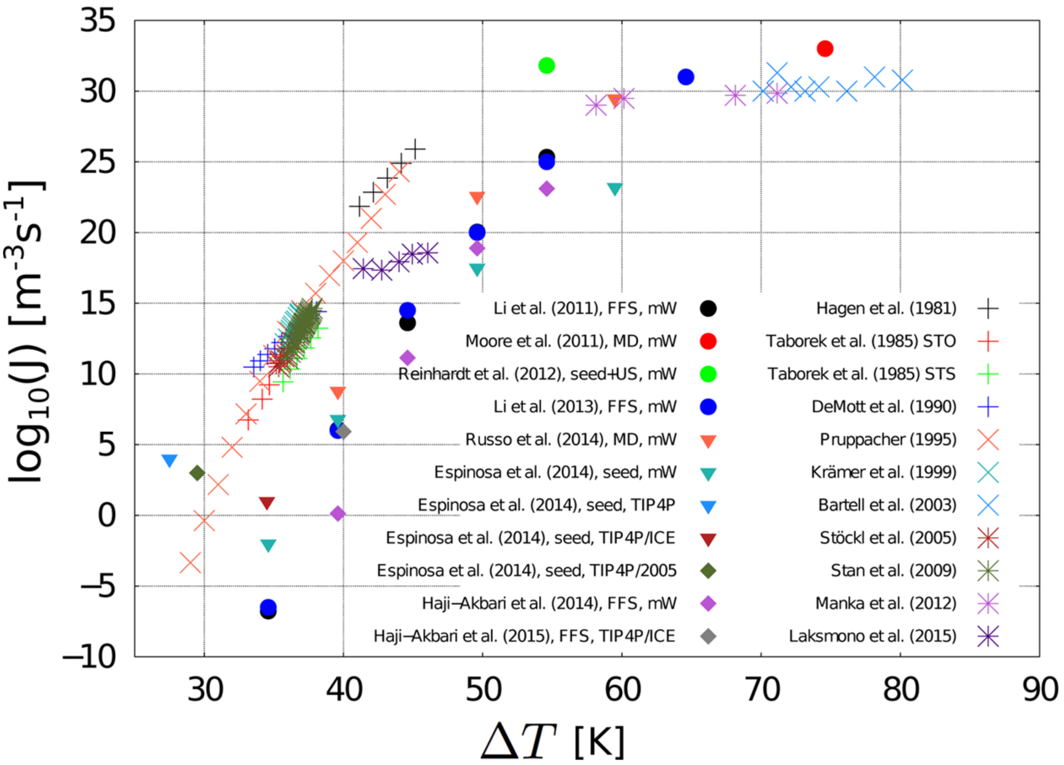

Figure 7.1 – A comparison between experimental measurements (crosses) and theoretical predictions (filled symbols) of the ice nucleation rate at different temperatures. Figure adopted from Ref. [10].

using the microscopic quantities directly obtained from simulations. For instance, the diffuse nature of solid-liquid interfaces makes it difficult to rationalize and formulate the nucleation free energy in a unique and meaningful way, as the choice of the atomic order parameters used to distinguish ice structures affects the computed free energy profile and the size of the critical nucleus [3, 36, 56]. Furthermore, an ice nucleus with complex stacking-disordered structures is usually formed during the homogeneous nucleation process [41, 42, 154], but it is computationally very expensive to exhaustively sample the numerous corresponding nucleation pathways even for nuclei that are within a small window of sizes [41, 42]. These multiple nucleation pathways make the theoretical analysis of the nucleation process even more difficult.

The present chapter describes a rigorous and efficient framework to estimate the absolute homogeneous nucleation rate at realistic thermodynamic conditions, and investigate how different physical quantities, including the interfacial free energy, entropic contributions due to stacking disorder, and the kinetic prefactor contribute to the overall rate. To demonstrate this framework on an important system with complex nucleation pathways, we employ a monoatomic water (mW) model [155], which has proved very successful in reproducing many thermodynamic and structural properties of water including the melting point and the relative stability of different ice phases [156] and as a result has been widely used to study ice nucleation [10]. We focus on a temperature $T = 240$ K and pressure $P = 1$ bar, which are thermodynamic conditions commonly encountered in clouds [157], freezers, and glaciers.

7.2 The nucleation free energy of ice Ih

Rather than attempting to sample multiple nucleation pathways, we take an alternative route in which we first compute the nucleation rate for perfect ice Ih nuclei and then add correction terms to account for stacking disorder. This approach does not only allow us to fully converge the nucleation rate, but also disentangles contributions of different physical origins. In order to restrict the sampling to the part of configuration space where only ice Ih nuclei form inside the liquid, we opt for a combination of the umbrella sampling method and the seeding technique: at the beginning of the simulation, the fluid is seeded with an initial pure Ih ice cluster of a certain size. At the same time, an umbrella potential is added to the Hamiltonian $\mathcal{H}(\mathbf{q})$ of the system [34], so the biased Hamiltonian used in simulations is

$$\mathcal{H}_{biased}(\mathbf{q}) = \mathcal{H}(\mathbf{q}) + \frac{\kappa}{2}\left(\Phi - \bar{\Phi}\right)^2, \tag{7.1}$$

where κ denotes the spring constant of a rather stiff umbrella potential, and the global collective variable (CV) $\Phi = \sum_i \phi(i)$ is constructed by summing the order parameter values $\phi(i)$ for each of the atoms in the system. Here, $\phi = S(\overline{Q_6})$ is the locally-averaged bond order parameter [146], which we transform with a hyperbolic switching function to enhance its resolving power between solid and liquid-like atomic environments. The parameter $\bar{\Phi}$ is set to a value that ensures the stabilization of the initial ice Ih cluster at out-of-equilibrium conditions. Repeating this procedure many times within a relevant range of sizes of the initial nuclei and different values of $\bar{\Phi}$, one can reconstruct a free energy profile $\tilde{G}(\Phi)$ by using the WHAM method [158].

This combination of the umbrella sampling method and the seeding technique is the cornerstone of the efficient framework for estimating the nucleation rate: (i) seeding the nucleus means that little time is spent in growing the nucleus and waiting for the possible defects formed during the kinetic growth process to annihilate; (ii) equilibrating the system under the presence of an umbrella potential means that local quasi-equilibrium can be achieved, and that system is not prone to kinetic factors such as the latent heat; (iii) considering just one nucleation pathway associated with the pure nucleus ensures easy convergence of the computed free energy profile.

For the umbrella sampling simulations, the NPT ensemble was employed throughout with the stochastic velocity rescaling thermostat [90] and an isotropic Nose-Hoover barostat. A total of 8 sets of simulation runs with the biased Hamiltonian (Eqn. (7.1)) were performed using a system size of 8192 molecules. For each set, about 50 umbrella sampling windows were used, and each trajectory lasted about 0.5 ns. Fast implementation of this simulation setup was made possible by the flexibility of the PLUMED code [93] in combination with

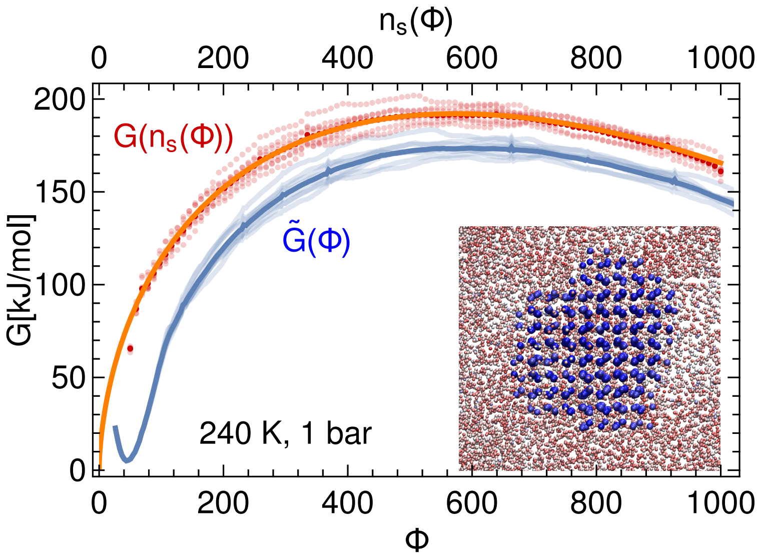

Figure 7.2 – The light blue curves are the free energy profiles as a function of the collective variable Φ for 8 sets of umbrella sampling simulations, and the dark blue curve is the averaged $\tilde{G}(\Phi)$ from these runs. Each dotted light red curve is the free energy profile of a perfect Ih nucleus extracted from each set of simulations, the dotted dark red curve is the averaged result, and the orange curve indicates a CNT fit using Eqn. (7.2). The inset shows a snapshot of an ice Ih nucleus embedded in liquid water.

LAMMPS [91].

Note that the atomic order parameter $\overline{Q_6}$ does not distinguish ice Ic and ice Ih phases but one can monitor the time evolution of another order parameter such as locally-average bond order parameter $\overline{Q_4}$ that is able to differentiate the two ice phases. One can also place a constraint on the number of particles that have Ic stacking. That said, even without such explicit constraint very few stacking faults form during the umbrella sampling simulations, and almost all of them are in very small nuclei with fewer than 200 atoms. The very low occurrence of the stacking faults here is probably due to that the umbrella potential places a strong constraint on the nucleus size in each simulation, making it difficult for a basal bi-layer to dissolve and re-crystallize, which is needed for the formation of a stacking disorder.

In Figure 7.2 we plotted the free energy $\tilde{G}(\Phi)$ that is associated with the formation of pure ice Ih nuclei. From $\tilde{G}(\Phi)$, one can extract the free energy profile $G_{\mathrm{Ih}}(n_s(\Phi))$ for the ice Ih nucleus as a function of the cluster size by using the thermodynamic framework introduced in chapter 3-5. Any extensive quantity, which is chosen to be the collective variable, Φ in this case, can be used to unambiguously define a Gibbs dividing surface, which surrounds a nucleus and determines its size. Chapter 4 describes in detail the conversion between the free energy profile $\tilde{G}(\Phi)$ and the nucleation free energy as a function of the size of the solid nucleus $n_s(\Phi)$. We obtained $G_{\mathrm{Ih}}(n_s(\Phi))$ for the ice Ih nucleus with sizes larger than the cutoff value $n_{\mathrm{cut}} = 50$, and plotted the result in Figure 7.2.

For a single reaction channel, the size $n_{\mathrm{s}}(\Phi)$ alone is sufficient to characterize the nucleation free energy profile. In such a case, CNT is commonly used to rationalize nucleation. With a Φ-based dividing surface, the free energy of the ice Ih nucleus relative to the bulk liquid can be naturally decomposed into a bulk and a surface term,

$$G_{\mathrm{Ih}}(n_{\mathrm{s}}(\Phi)) \eqsim \mu_{\mathrm{Ih}} n_{\mathrm{s}}(\Phi) + \gamma_{\infty}^{\Phi} \Omega v_{\mathrm{s}}^{\frac{2}{3}} n_{\mathrm{s}}{}^{\frac{2}{3}}(\Phi)(1+\zeta n_{\mathrm{s}}{}^{-\frac{1}{3}}(\Phi)). \tag{7.2}$$

Here, μ_{Ih} is the difference between the chemical potentials of ice Ih and the liquid phase, γ_{∞}^{Φ} is the ice Ih-liquid interfacial free energy of the Φ-based dividing surface at the planar limit, and $\Omega = (36\pi)^{1/3}$ is the geometrical constant for a spherical nucleus. The $\zeta n_{\mathrm{s}}{}^{-\frac{1}{3}}(\Phi)$ term, which is determined by the distance between the Φ-based dividing surface and the surface of tension in the planar limit, captures the curvature dependence of the interfacial free energy γ^{Φ} of a curved interface.

The orange curve in Figure 7.2 shows a CNT fit to the nucleation free energy profile of the ice Ih nucleus using Eqn. (7.2). The chemical potential $\mu_{\mathrm{Ih}} = -0.649$ kJ/mol at 240 K was obtained from previous calculations [159], and the other two parameters $\gamma_{\infty}^{\Phi} = 28.2(2)\,\mathrm{mJ/m^2}$ and $\zeta = 0.2(1)$ were determined from the fit. From this value of ζ, we estimate the distance between the Φ-based dividing surface to the surface of tension [3, 4] to be $d = -\zeta(3v_{\mathrm{s}}/32\pi)^{1/3} = 0.2\,\text{Å}$. For comparison, if all three parameters are used in the fit, the predicted values are $\mu_{\mathrm{Ih}} = -0.69(3)$ kJ/mol, $\gamma_{\infty}^{\Phi} = 30(2)\,\mathrm{mJ/m^2}$ and $\zeta = 0.0(1)$.

The surface energy for the ice-liquid interface was computed to be around $35\,\mathrm{mJ/m^2}$ for the mW model at its melting point $T_{\mathrm{m}} = 274.6$ K [143, 144]. In general, the interfacial free energy exhibits a temperature dependence, and is also dependent on the specific choice of the extensive quantity used to define the Gibbs dividing surface away from the melting point [1]. However, as extensively discussed in chapter 5, regardless of which dividing surface is used, as long as it is used consistently and the curvature correction term ζ is included in the formulation of the nucleation free energy profile, no discrepancy will emerge. In this case, the correction term ζ is relatively small such that omitting it altogether from the CNT fit does not lead to significant changes in the estimation of the interfacial free energy, however, one should bear in mind that this may not be the case when using a different Gibbs dividing surface or when studying a different system. In general, employing an interpolation or extrapolation using the naive version of the CNT without the curvature correction, as is often the case in computational studies of this kind [36, 160], may result in a systematic error in the prediction of the interfacial free energy as well as the nucleation barrier [3].

7.3 Accounting for stacking disorder

Above, we have considered a single nucleation pathway to form a pure ice Ih nucleus in liquid water. We now formulate the difference in free energy of stacking-disordered and hexagonal ice crystallites. Because the ice-water interfacial free energies are indistinguishable for ice Ih and

Ic phases [144], the free energy difference due to the formation of stacking disorders can only be associated with the bulk term. Assuming the nucleus has a certain shape (e.g. spherical), the number of bi-layers k_{max} along the basal plane can be easily calculated for a nucleus of a given size (see the inset of Figure 7.3). The area of each plane A_k at each bi-layer can also be determined analytically as the bi-layers are equally spaced along the diameter of the spherical nucleus. We use an analytic model similar to the one-dimension Ising-like system discussed in Ref. [161]. We assume that the energy of each bi-layer only depends on its neighboring bi-layers, thus on each layer the stacking order is independent, and the probability ρ_k of forming an ice Ic stacking order at the plane k follows the Boltzmann distribution, i.e.,

$$\rho_k = \frac{1}{Z_k} \exp\left(-\frac{A_k \gamma_{\mathrm{sf}}(T)}{k_B T} \right), \tag{7.3}$$

with the partition function for the k-th bi-layer

$$Z_k = 1 + \exp\left(-\frac{A_k \gamma_{\mathrm{sf}}(T)}{k_B T} \right), \tag{7.4}$$

where $\gamma_{\mathrm{sf}}(T)$ denotes the temperature dependent stacking fault free energy per unit area. Note that the term stacking fault here refers to a stacking disorder with respect to the standard Ih stacking, which means a pure Ic nucleus can be considered as having a stacking fault on every bi-layer. The free energy difference between a pure ice Ih nucleus and one that has the same size and a stacking disordered structure can be expressed as

$$\Delta G_{\mathrm{sf}} = -k_B T \ln \prod_k Z_k. \tag{7.5}$$

Using Eqn. (7.5), the problem of computing the free energy difference between a pure Ih stacking nucleus and a mixed one has been reduced to characterizing the stacking fault free energy $\gamma_{\mathrm{sf}}(T)$, which we determine next. To do that, we have carefully selected a combination of multiple thermodynamic integration routes [57], in order to take into account vibrational entropy and anharmonicity, and make it possible to disentangle the different contributions to the stacking fault free energy.

The Gibbs free energy for a perfect bulk ice Ih structure and of a Ih bulk crystal with two stacking fault layers were computed separately using a sequence of thermodynamic integration routes, following the general strategy outlined in chapter 2.2. We started by computing the Helmoltz free energy of a real system at a low temperature (90 K) by thermodynamic integration starting from a reference harmonic crystal. Afterwards, we switched from the NVT to the NPT ensemble and obtained the Gibbs free energy. Finally, independent molecular dynamics simulations were performed in the NPT ensemble at temperatures ranging from 90 K to 300 K, in order to obtain the temperature dependence of the Gibbs free energy for each system by thermodynamic integration with respect to T.

Figure 7.3 shows that, considering only the potential energy difference at 0 K or under

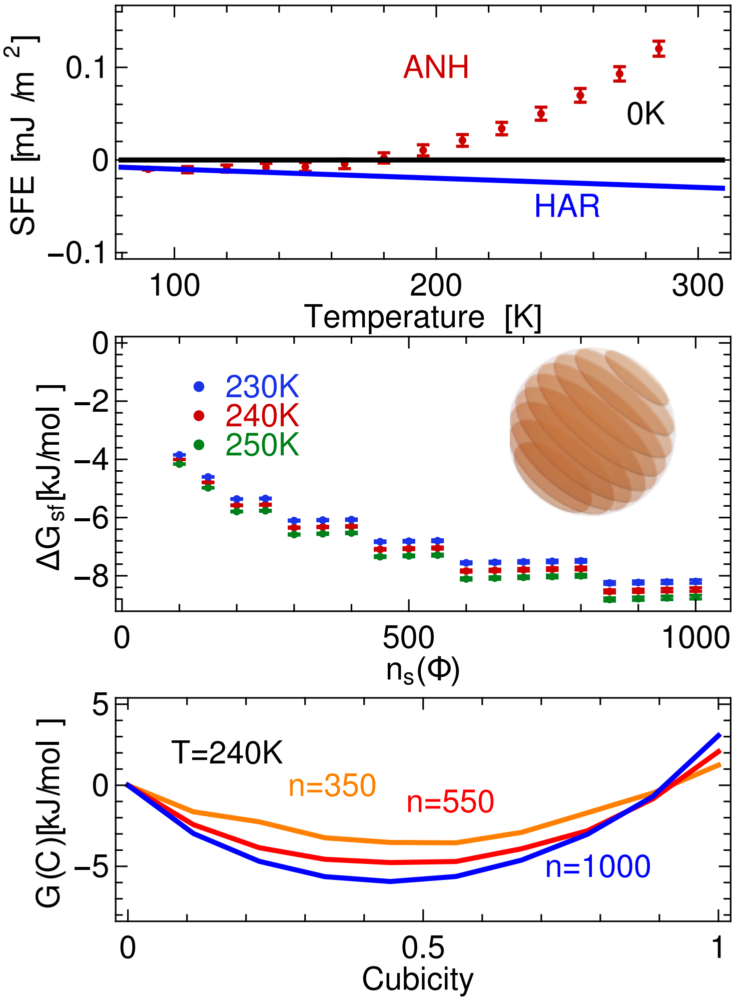

Figure 7.3 – Upper panel: The stacking fault free energy per area as a function of temperature. The black line is the estimate from the potential energy difference at 0 K, the blue curve represents the harmonic approximation (HAR), and the red dots show the results from thermodynamic integration that considers anharmonicity (ANH). Statistical uncertainties are indicated by the error bars. Middle panel: the free energy difference ΔG_{sf} between a pure ice Ih nucleus and a one that has the same size and a mixed stacking disorder at temperatures 230 K, 240 K and 250 K as predicted by the analytic model illustrated in the inset. Lower panel: The free energy profile as a function of cubicity for nuclei of three different sizes at 240 K.

the harmonic approximation, the stacking fault free energy per unit area $\gamma_{\mathrm{sf}}(T)$ is estimated to be zero or negative. Only when anharmonicity is taken into consideration, $\gamma_{\mathrm{sf}}(T)$ is predicted to be positive, indicating that the ice Ih phase is more stable than Ic. This observation highlights the importance of accounting for anharmonic effects in studies of this kind. In Reference [156], $\gamma_{\mathrm{sf}}(T_{\mathrm{m}})$ was predicted to be 0.11(2) mJ/m^2 by the method of growing ice bi-layers in simulations at near coexistence conditions, which agrees well with our results using thermodynamic integration. Notice also that the magnitude of $\gamma_{\mathrm{sf}}(T)$ is very small at all the temperatures considered here, which means that for small ice nuclei the term in the exponent in the partition function Z_k (Eqn. (7.4)) is very close to zero.

Using the values of $\gamma_{\mathrm{sf}}(T)$, ΔG_{sf} obtained from Eqn. (7.5) as a function of the size of the nucleus at three different temperatures is also shown in Figure 7.3. The magnitude of the correction term ΔG_{sf} is larger at high temperature because it mostly stems from the entropic gain of forming stacking disorders, but it is relatively insensitive with respect to temperature.

Because the analytic model introduced above accounts for the distribution of the stacking faults at each layer using the accurate values of $\gamma_{\mathrm{sf}}(T)$, one can enumerate all the possible combinations of the stacking disorder sequences, and thereby compute the free energy as function of cubicity for a nucleus of a given size. The cubicity here is defined as the fraction of ice molecules that are in local Ic environments. The lower panel of Fig. 7.3 shows that for the three sizes considered here it is more favorable for the ice nucleus to adopt a cubicity close to 0.5, which is fully consistent with the observation in Ref. [160]. For the three nucleus sizes considered here, the free energy gain associated with the mixed stacking disorder is larger for the larger nuclei. The most favorable degree of cubicity decreases with cluster size, consistent with the fact that for ice nuclei approaching macroscopic size, the term $A_k\gamma_{\mathrm{sf}}(T)$ in Eqn. (7.4) will dominate, and pure Ih stacking should become favorable. Note also that cubicity= 1 corresponds to a ice Ic nucleus, and thus the analytic model here predicts that the chemical potential difference μ_{Ic-Ih} between Ic and Ih is 0.0026(4) kJ/mol at 240 K. This value agrees well with $\mu_{Ic-Ih}(240) = 0.0031(2)$ kJ/mol that we computed independently using the thermodynamic integration method. This agreement validates our assumption that the each basal plane can be considered independently when calculating the free energy difference associated with stacking disorders. Note that the analytic model here neglects the possibility of intersecting stacking disorders: for ice Ih lattice stacking disorder can only occur along the two basal faces, but in the uncommon cases when the nucleus consists a large enough domain of ice Ic, stacking can occur along the four (111) planes of the Ic lattice. It is also possible to construct a 2D Ising-like model to mimic the intersecting stacking disorders [42], although one has to make assumptions on the free energy penalty of a grain boundary-like structure in the intersection.

The cubicity= 1 (i.e. pure ice Ic nucleus) case is illuminating also because one can perform a set of seeding and umbrella sampling calculations described in the previous section for the nucleation pathway of pure Ic nuclei. In Figure 7.4 we report the free energy profile G_{Ic} as a function of the nucleus size. The difference in the nucleation barriers $G^{\star}_{\mathrm{Ic}} - G^{\star}_{\mathrm{Ih}}$

between the Ic and the Ih nucleus is estimated to be 2 ± 2 kJ/mol, and the sizes of the critical nuclei (about 570 molecules) are similar in both cases. This difference in nucleation barriers can be entirely explained by the chemical potential difference between the phases of ice $\mu_{Ic-Ih}(240\,\mathrm{K})$. This result thus supports our assumption as well as the conclusions in previous calculations [144] that the ice-water interfacial free energies are indistinguishable for ice Ih and Ic phases. In Ref. [42] $G^{\star}_{\mathrm{Ic}} - G^{\star}_{\mathrm{Ih}}$ was estimated to be about 6 kJ/mol at 230 K using the method of transition path sampling over ice nuclei whose sizes are close to that of the critical nucleus. The difference between their result and ours is still comparable to the magnitude of the statistical error, and might be related to the fact that the Ic nuclei in our case were initially seeded in the undercooled liquid and therefore almost defect-free, but the ones in their study were generated by sampling dynamical trajectories so may contain a higher concentration of defects.

Further comparisons with Ref. [42] can be made regarding the free energy profile as a function of cubicity (lower panel in of Figure 7.3). Our analytic model, although simple, is able to capture the overall trend in the dependence of the free energy on cubicity. Our estimation for the free energy difference associated with staking disorders $\Delta G_{\mathrm{sf}} = 8$ kJ/mol of the critical nucleus is smaller compared to the prediction 14 kJ/mol in Ref. [42]. The discrepancy may be due to that our analytic model neglects intersecting stacking disorders, grain boundaries, and domains of other phases of ice, which may further lower the free energy of ice nuclei. In any case, this difference in ΔG_{sf} here would only accounts for about one order of magnitude change in the estimated nucleation rate J. It is also worth pointing out that it has been debated that the mW model underestimates the free energy penalty associated with the Ic structures [161], and one advantage of using an analytic model is the possibility to employ the experimental values or *ab initio* results for γ_{sf} when estimating ΔG_{sf} as well as the cubicity of nuclei of different sizes [161]. We also want to point out that one has the option to sample the near critical ice nucleus with stacking disorder and other defects, using transition path sampling [42, 160] or Monte Carlo methods [161], in order to compute ΔG_{sf}. Regardless how one chooses to evaluate ΔG_{sf}, it can later be directly added on top of G_{Ih} for estimating the free energy profile of ice nucleus with defects, i.e. $G(n_{\mathrm{s}}(\Phi)) = G_{\mathrm{Ih}}(n_{\mathrm{s}}(\Phi)) + \Delta G_{\mathrm{sf}}(n_{\mathrm{s}}(\Phi))$.

7.4 The kinetic factor in homogeneous nucleation

Once the free energy barrier of nucleation $G^{\star} = \max(G(n_{\mathrm{s}}))$ has been determined, the nucleation rate can be obtained from [45, 54]

$$J = (1/v_{\mathrm{l}}) Z f^{+} \exp(-G^{\star}/k_B T) \tag{7.6}$$

where v_{l} is the molar volume of the undercooled liquid, f^{+} is the addition rate of particles to the critical nucleus, and the Zeldovich factor

$$Z = \sqrt{\frac{1}{2\pi k_B T} \frac{d^2 G(n_{\mathrm{s}})}{d n_s^2}} \tag{7.7}$$

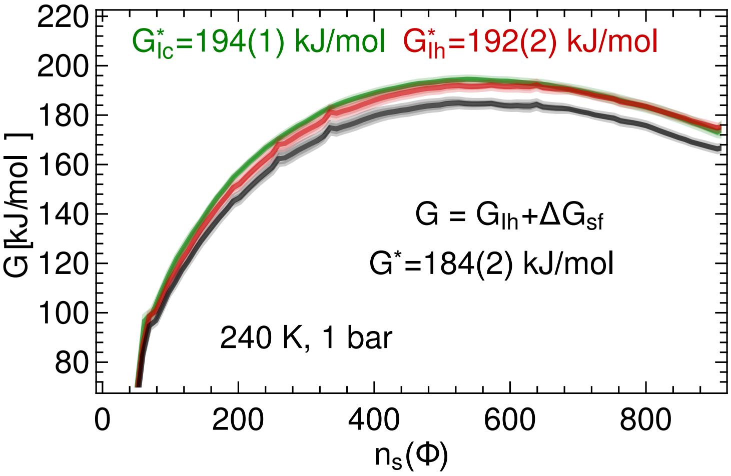

Figure 7.4 – The red curve is the free energy profile G_{Ih} of a pure Ih nucleus, the green curve is G_{Ic} of a pure Ic nucleus, and the black curve indicates the free energy profile of an ice nucleus that can have a mixed stacking order. The width of the curves indicates the statistical error in the free energy estimation, computed from the error of the mean from independent simulation runs.

can be obtained numerically from the nucleation free energy profile $G(n_\mathrm{s}(\Phi))$ in Figure 7.4. The addition rate f^+ can be computed as a diffusion coefficient from the mean square displacement of the cluster size after it is released at the top of the nucleation barrier [45, 54]. However, this approach assumes that $dG(n_\mathrm{s})/dn_\mathrm{s}$ is effectively zero when running multiple trajectories, is influenced by the choice of the initial configuration, and by the latent heat created when the solid nucleus changes size [55]. In order to overcome these shortcomings, we computed f^+ accurately and directly from the umbrella sampling trajectories with the biased Hamiltonian Eqn. (7.1) by applying a stochastic model originally proposed to mimic the kinetics of planar interfaces [162].

In the stochastic model, the time evolution of an extensive quantity Φ is viewed as resulting both from molecules changing from one phase to the other as well as from fluctuations in the bulk phases,

$$\Phi(t) = (\phi_\mathrm{s} - \phi_\mathrm{l})n_\mathrm{s}(t) + \phi_\mathrm{l} N + f(t). \tag{7.8}$$

Here, ϕ_s and ϕ_l are the averages of the atomic order parameter in the bulk solid and the bulk liquid phases, respectively. The first term in the above equation stems from the temporal change of the size $n_\mathrm{s}(t)$ of the solid cluster evolving under the action of the biased Hamiltonian. The term $f(t)$, on the other hand, takes into account fluctuations that do not change the composition of the solid-liquid system, but are due to changes of the extensive quantity Φ within the bulk phases, for instance caused by phonons propagating through the system.

In general, the time evolution of these two terms in Eqn. (7.8) occurs on distinct time scales, as the change in $n_s(t)$ is determined by the relatively slow growth rate of the solid-liquid interface, and the dynamics of $f(t)$ happens on the time scale of lattice vibrations. These different time scales are reflected in the power spectrum $S(\omega)$ of $\Phi(t)$, related to the time autocorrelation function $\langle\Phi(0)\Phi(t)\rangle$ by

$$S(\omega) = \int_{-\infty}^{\infty} \langle\Phi(0)\Phi(t)\rangle\, e^{-i\omega t} dt. \tag{7.9}$$

In Figure 7.5 we plot $\omega S(\omega)$ obtained for a solid-liquid system that contains a critical solid nucleus (blue curve). For comparison, we also show the results for the bulk solid as well as bulk the liquid under the same thermodynamic conditions. It can be seen that for all three systems there is a peak at high frequency, which corresponds to the fast fluctuations. Only for the solid-liquid system with a critical nucleus there is another well separated peak at a low frequency, which stems from the growth of the crystalline nucleus embedded in the liquid.

To rationalize the power spectrum $S(\omega)$ further and extract quantitative information on the growth process from it, we now postulate that the time evolution of the collective variable $\Phi(t)$ can be modeled using a pair of coupled Langevin equations as described in Ref. [162]:

$$\begin{aligned} \gamma\dot{q} &= -\kappa(f+q-\bar{\Phi})+\eta(t) & (7.10)\\ m_f\ddot{f} &= -\kappa_f f-\kappa(f+q-\bar{\Phi})-\gamma_f\dot{f}+\eta_f(t), & (7.11) \end{aligned}$$

where the variable q, representing the slowly evolving part of Φ, is defined as $q = (\phi_s - \phi_l)n_s(t) + \phi_l N$. In the above equation, γ and γ_f are friction constants associated with q and f, respectively, $\eta(t)$ and $\eta_f(t)$ are Gaussian random forces. While the dynamics of $q(t)$ is assumed to be over-damped, inertial effects are included for the variable $f(t)$, which is assigned an effective mass of m_f. The force constant κ is the sum of the umbrella spring constant of value 0.005 kJ/mol (Eqn. (7.1)) and the curvature of the free energy for the critical nucleus, $\kappa' = 2d^2G/dn_s^2 = -0.0003$ kJ/mol, which is negligible in this case. For this model, the power spectrum $S(\omega)$ of $\Phi(t)$ can be determined analytically,

$$S(\omega) = \frac{2k_BT}{\omega^2}\mathrm{Re}\left[\left[\frac{1}{\gamma}+\left[\gamma_f+i\left(\omega m_f-\frac{\kappa_f}{\omega}\right)\right]^{-1}\right]^{-1}-\frac{i\kappa}{\omega}\right]^{-1}. \tag{7.12}$$

We have fitted this expression to the power spectrum obtained from the umbrella sampling simulation and the result is shown in Figure 7.5. As can be inferred from the figure, the simple Langevin model captures both peaks of the power spectrum, although the high-frequency peak is reproduced less accurately than the low-frequency peak, most likely because $f(t)$ is sensitive to the details of the order parameter in use. However, we are primarily interested in the slow mode $q(t)$ and, particularly, in the value of the friction constant γ because this parameter is related to the addition rate f^+ is by $f^+ = k_BT/\gamma$. Due to the separation of time scales, the parameters m_f and κ_f associated with the fast mode $f(t)$ have little bearing on the dynamics of the slow mode. From the fit, we obtained $\gamma = 0.06$ ps kJ/mol, yielding an

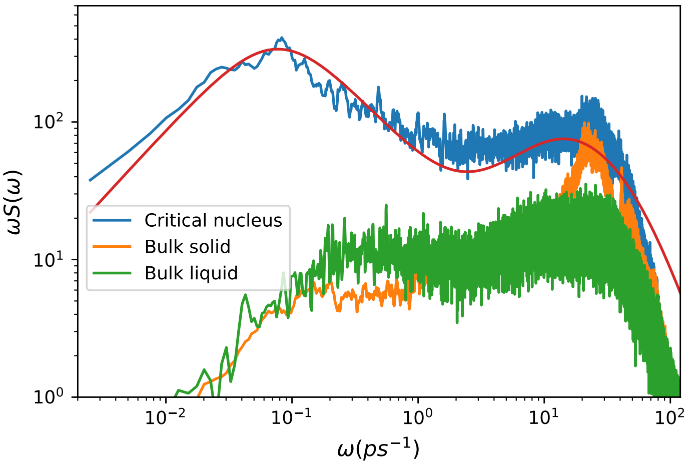

Figure 7.5 – The orange, green and blue curves are the spectra $\omega S(\omega)$ for bulk solid, bulk liquid, and a solid-liquid system that contains a solid critical nucleus ($n^\star = 550$) at 240 K and 1 bar, respectively. The red curve is the fitting curve using Eqn. (7.12) with parameters $m_f = 3\times10^{-5}$ ps^2 kJ/mol, $\kappa_f = 0.03$ kJ/mol, $\gamma_f = 0.0025$ ps kJ/mol, and $\gamma = 0.06$ ps kJ/mol.

Table 7.1 – Key parameters for the estimation of the homogeneous ice nucleation rate at $T = 240$ K and $P = 1$ bar. The value for the chemical potential $\mu_{\text{Ih}} = -0.649 kJ/mol$ is from Ref. [159].

type	μ_{sl} [kJ/mol]	v_{s} [Å^3]	v_{l} [Å^3]	$G^\star$ [kJ/mol]	Z	f^+ [s^{-1}]	J [m^{-3}s^{-1}]	$\log_{10}(J\text{m}^3\text{s})$
Ih	-0.649	30.48	29.79	192(2)	0.003	3×10^{13}	5×10^{-3}	-2.3
Ic	-0.646	30.48	29.79	194(1)	0.004	3×10^{13}	2×10^{-3}	-2.8
mixed	-	30.48	29.79	184(2)	0.004	3×10^{13}	0.3	-0.5

addition rate of particles to the critical nucleus of $f^+ = 3\times10^{13}\,\text{s}^{-1}$, which is smaller compared with previous results ($7\times10^{13}\,\text{s}^{-1}$) [36] but of the same order of magnitude. In addition, we estimated $f^+ = 2.5\times10^{13}\,\text{s}^{-1}$ and $5\times10^{13}\,\text{s}^{-1}$ for nuclei of about 340 and 900 atoms, respectively, confirming that the addition rate increases with nucleus size.

7.5 Results

Combining the free energy profile of stacking-disordered ice nucleus and the kinetic prefactor using Eqn. (7.6), we estimate the nucleation rate to be $J = 0.3\,\text{m}^{-3}\text{s}^{-1}$ at 240 K and 1 bar for the mW model. The key data for this estimation are tabulated in Table 1. For our choice of order parameter, disregarding the influence from the curvature dependence of the interfacial free energy changes the predicted nucleation rate by 3 orders of magnitude in this case. The effects from the curvature dependence, however, may be much larger for other systems or when using a different choice of the Gibbs dividing surface [3]. Finally, we can estimate that the free energy gain associated with stacking disorders accelerates the nucleation rate by about 2 orders of magnitude.

For comparison, for the same water model and at the same thermodynamics conditions (Figure 7.1), a previous estimate using the seeding approach yields $J = 10^{-9}\ \mathrm{m}^{-3}\mathrm{s}^{-1}$ [36], and the forward flux sampling method gives $J = 2 \times 10^{-7}\ \mathrm{m}^{-3}\mathrm{s}^{-1}$ [160]. In most experimental measurements (Figure 7.1), the predicted homogeneous ice nucleation rate is around $J = 10^{7}\ \mathrm{m}^{-3}\mathrm{s}^{-1}$ at an undercooling of 35 K [163–165]. To compare, our estimate is faster by about 9 orders of magnitude compared with previous literature values, but 7 orders of magnitude slower relative to experiments.

7.6 Discussions

Theoretical prediction of the homogeneous ice nucleation rate is an extremely challenging problem not only because the high nucleation barrier and complex nucleation pathways impose computational challenges, but also because the predicted rate is highly sensitive on the fine details of the simulation methods as well as the water model employed (e.g. a change of just 10 kJ/mol in nucleation barrier affects the rate by more than 2 orders of magnitude). In order to systematically improve the existing computational methods and ultimately achieve quantitative accuracy in the predictions, it is therefore important to understand where the discrepancies between different computational predictions and between computation and experiment may come from. Our method helps to build such understanding, as it not only provides a reference value, but also breaks down the problem into segments and considers each quantity that affects the nucleation rate in a careful and rigorous manner.

Our estimate of the homogeneous nucleation rate of the mW model is about 9 orders of magnitude higher than the previous estimates using the seeding approach [36] (Figure 7.1) . A large part of the difference is probably due to a different definition for the size of the critical nucleus $n^\star$, which is crucial in approximating the nucleation barrier if one employs the original expression of the classical nucleation theory for which $G^\star = n^\star|\mu_{\mathrm{sl}}|/2$ [36]. It is common practice to set a threshold on the atomic order parameter in order to distinguish solid and liquid-like atoms and to determine the nucleus sizes [36, 42, 54, 56, 160, 166], however, the nucleus size metric in use has a very strong influence on the estimated size [42, 56] and can thus affect the estimated rate by many orders of magnitude [166]. For example, with the Φ-based Gibbs dividing surface we estimated $n^\star \approx 550$ molecules, and in Ref. [54] the estimate is 688 molecules, which implies that the estimated rate would differ by about 8 orders of magnitude using the original CNT expression. It is worth pointing out that our framework eliminates this ambiguity in the determination of the critical nucleus and the nucleation rate, making the classical nucleation theory much more rigorous. In addition, the seeding approach neglects the affects from stacking disorders, which may lower the rate by a few orders of magnitude.

Our estimated rate is also 6 orders of magnitude faster than the one from forward flux sampling [160] (Figure 7.1). A similar discrepancy has been reported at 220 K, where the ice nucleation rate of mW water computed with umbrella sampling [56, 167] was found to be about 5 orders of magnitude higher than the corresponding forward flux sampling result [160]. In addition, for the case of a hard sphere system, the nucleation rates predicted by umbrella sampling and the ones predicted by forward flux sampling are more consistent though the former are still marginally faster at low supersaturations [168]. Although the path-based techniques have a number of advantages including sampling real dynamics without the presence of bias potentials, but there has been some worries that it may not proceed down the correct pathway that allows for sufficient local equilibration [167]. From the practical side, we want to point out that the umbrella sampling method used here is computationally much cheaper: we performed a total of about 200 nanosecond simulations to obtain a well-converged free energy profile of nucleation with uncertainty estimations, compared to tens or hundreds microseconds in the previous studies using path-based techniques [41, 42].

Finally, our estimated rate is 7 orders of magnitude slower compared with experiments (Figure 7.1), which were measured often with micro-droplets to avoid heterogeneous nucleation due to foreign particles [163–165], although there have been concerns that residual aerosol particles or surfactants may still affect the measured rates [165]. In general, at deep undercoolings ($T_\mathrm{m} - T > 50\mathrm{K}$), previous computational predictions and experimental measurements agree rather well, but at moderate undercoolings ($T_\mathrm{m} - T < 50\mathrm{K}$) the experimental rates are usually faster by 10-15 orders of magnitudes compared with simulations using different water models including mW, TIP4P and TIP4P/Ice [10]. Our underestimation of the nucleation rate is thus in line with this systematic trend.

To further elucidate the origin of this theoretical underestimation of the rate using the mW model, in what follows we will examine the various relevant terms. First of all, the mW model overestimates the kinetic prefactor in Eqn. (7.6), as its diffusion coefficient is 2 orders of magnitude higher compared with experiments [159]. Stacking disorders accelerate the rate by two orders of magnitude according to our analytic model, and by 3 orders of magnitude according to previous transition path sampling simulations [42]. As such, the underestimation of the rate can only come from the overestimation of the nucleation free energy barrier. Recall that the nucleation free energy profile of the ice Ih nucleus (Figure 7.2) can be well described by the CNT expression in Eqn. (7.2), suggesting that CNT is accurate for this case and that the nucleation barrier can be expressed as $G^\star \approx \dfrac{4(\gamma_\infty^\Phi \Omega v_\mathrm{s}^{2/3})^3}{27|\mu|^2} + \zeta \dfrac{2(\gamma_\infty^\Phi \Omega v_\mathrm{s}^{2/3})^2}{3|\mu|}$. The Tolman correction $\zeta = 0.2(1)$ is small in this case, and thus unlikely to play a significant role. On the other hand, the chemical potential $\mu_\mathrm{Ih} = -0.649$ kJ/mol of mW water at 240 K [159] compares well with the experimental value −0.6756 kJ/mol, which we calculated from the heat capacities reported in Ref. [41]. The molar volume v_s of ice predicted by the mW model is about 7% smaller compared with experiments (density 0.983g/cm^2 for mW compared with the actual value 0.92g/cm^2), which goes in the direction of lowering the nucleation barrier compared with experiments. Hence, we infer that our underestimation is probably due to the fact that the mW model overestimates the interfacial free energy γ_∞^Φ, which is the only remaining key factor. Away from the melting point, the values of γ_∞^Φ depend on the choice of the Gibbs dividing surface [1], but one can unequivocally compare the values at T_m: the mW model predicts 35 mJ/m^2 [143, 144], compared with the range 24 – 33 J/m^2 measured experimentally [165, 169, 170], three TIP4P-type water models (TIP4P, TIP4P/2005, TIP4P/Ice) all predict values between 28 J/m^2 and 30 J/m^2 [143]. Indeed, if one adjust the value for γ_∞^Φ of the mW model (decrease by 10 – 15 %) in the CNT expression (Eqn. (7.2)), and at the same time correct for the molar volume (increase by 7%) as well as the kinetic factor (decrease by 2 orders of magnitude), a rate of $J = 10^4 - 10^{10}$ m^{-3}s^{-1} would be obtained, in a much better agreement with experiments.

The analysis above highlights the sensitivity of the predicted ice nucleation rate on the underlying water model, and thus suggests that accurately predicting the rate may require a high-level water model. In general, empirical water models suffer from various limitations. As extensively discussed in Ref. [159], TIP4P models usually underestimate the chemical potential difference μ_sl between the undercooled liquid and the ice phases by as much as 20-30%, the TIP4P models and the mW model also underestimate the heat of fusion, some TIP4P models such as TIP4P/Ice underestimates the diffusion coefficient, etc. In the next chapter, we will explore the possibility of using *ab-initio*-accuracy potential energy surfaces to predict the rate. The theoretical framework and the method presented in this paper allow us to go beyond inexpensive empirical water models, because of the significant reduction in the computational cost compared with previous path-based methods. Furthermore, disentangling all the terms that contribute to the overall rate means one can choose to evaluate these terms independently.

7.7 Conclusions

In order to provide an accurate determination of the absolute nucleation rate for a monatomic water model, and to decompose it in physically-meaningful terms, we followed three key steps: Firstly, we computed and characterized the free energy profile for pure ice Ih nuclei at a very affordable computational cost. Then we took into account multiple nucleation pathways due to the possibility of forming stacking disordered Ic layers in ice nuclei, by adding an analytic free energy correction term. We then calculated the kinetic prefactor using a stochastic model, and finally obtained the homogeneous ice nucleation rate using a general formulation of classical nucleation theory.

Our framework removed the ambiguity in defining the size of the nucleus, and did not rely on many commonly-adopted approximations, such as neglecting the curvature dependence in interfacial free energy and the effect from stacking disorders, both of which may have contributed to the long-standing controversy on the predicted nucleation rates, that varies by many orders of magnitude among previous studies.

The presented protocol for the study of homogeneous nucleation involving multiple nucleation pathway makes fast and efficient computation of nucleation rates possible: cost is lowered by about three orders of magnitude compared with the previous path-based studies [41, 42], and further reduction is possible if one uses the CNT expression in Eqn. (7.2) to perform interpolation or extrapolation of the computed free energy profile of nucleation. More importantly, our approach disentangles all the terms that contribute to the overall rate, including the difference in chemical potentials, the interfacial free energy, the stacking fault free energy, and the kinetic prefactor. Separating those terms not only greatly facilitates the theoretical understanding and formulation of the homogeneous nucleation process, but also enables one to evaluate these terms independently. For instance, one can choose to compute the relatively important terms such as the chemical potential and interfacial free energy employing an *ab initio* potential energy surface [153, 171], perhaps even considering nuclear quantum effects which have been found to play in role in stabilizing the ice Ih phase [172], and then combine them to estimate the rate. Our framework thus opens the door to the prediction of homogeneous ice nucleation rates using *ab initio* potential energy surfaces, allowing for stringent cross-validations between theoretical and experimental estimates of this important quantity. In addition, it also provides a recipe for how to tackle even more complex nucleation phenomena, such as the crystallization of molecular crystals, the precipitation of halide salts, and the aggregation of hydrates.

8 The quest for a predictive model of ice nucleation

In the previous chapter, we computed the homogeneous ice nucleation rate using an empirical water model, and then discussed the possible cause of discrepancy with experiments. In general, empirical water models suffer from various limitations. Most notably, the predicted chemical potential differences μ_{sl} using commonly used TIP4P-type models are usually lower than the actual values by as much as 20%, and μ_{sl} heavily influences the nucleation barrier. The shortfalls of empirical models thus suggests that that accurately predicting the ice nucleation rate may require a high-level water model.

In this chapter, we first model the systems of ice and liquid water from first principles using the tools of quantum chemistry. We compute the chemical potential difference between the ice Ic and Ih phases of water, and between ice Ih and liquid water, at a hybrid density-functional-theory (DFT) level of theory, taking into account NQEs, proton disorder, and anharmonicity. This is made possible by exploiting advances in machine learning (ML) techniques to avoid the prohibitively large computational expenses otherwise incurred by extensively sampling phase space using first principles methods, and employing state-of-the-art thermodynamic integration techniques in order to accurately and rigorously compute the Gibbs free energies of the different phases of water. After that, we explore the possibility of using this *ab-initio*-accuracy potential energy surfaces predict the rate.

8.1 The challenges in modelling ice and liquid water

Liquid water and ice are ubiquitous on Earth and are intricately related to its climate system [173]. The low-density hexagonal form of ice, Ih, is the thermodynamically stable phase that occurs in abundance. The cubic form (Ic), a metastable ambient-pressure ice phase, derives its importance from its relation to the stacking-disordered ice nucleus that occurs as a precursor of freezing. The relative stability of Ic with respect to ice Ih plays a central role in ice cloud formation in the Earth's atmosphere [42, 174, 175], but is extremely difficult to measure experimentally [173].

Theoretical prediction of the thermodynamic properties of liquid water and ice is challenging, because of (i) the shortcomings of common water models including conventional force-fields [149] and (semi-) local DFT approaches [176–178], (ii) proton-disorder in ice, and (iii) the importance of nuclear quantum effects (NQEs) [179]. In particular, calculating the chemical potential difference $\Delta\mu^{\mathrm{Ih}\rightarrow\mathrm{Ic}} = \mu^{\mathrm{Ic}} - \mu^{\mathrm{Ih}}$ between proton-disordered Ic and proton-disordered Ih, which characterizes the relative stability, has been a severe computational challenge because the zero-point configurational entropies [180] and the configurational [181] and harmonic vibrational energies of ice Ih and Ic [172] differ by less than one meV/H_2O, so that anharmonic quantum nuclear motion plays a decisive role.

Water and ice have been described with varying success invoking approximations of differing severity, including simple electrostatic dipole models for the energetics of proton-ordering [182], force-field based path-integral molecular dynamics (PIMD) studies [171, 183–185], first principles quasi-harmonic (QHA) [185, 186], and VSCF [172, 187] studies which provide an approximate upper bound for $\Delta\mu^{\mathrm{Ih}\rightarrow\mathrm{Ic}}$. These have greatly advanced our understanding of the nature of liquid water and ice, but also highlight the harsh trade-offs between the accuracy of the description of the potential energy surface governing nuclear motion and the associated cost of sampling relevant atomistic configurations are sampled.

8.2 First-principles thermodynamics

As the underlying electronic-structure description, we employ the hybrid revPBE0 [188–190] functional with a Grimme D3 dispersion correction [191, 192], which has been demonstrated to accurately predict the structure, dynamics, and spectroscopy of liquid water in molecular dynamics (MD) and PIMD simulations [193]. revPBE0-D3 predicts that the difference in lattice energy between the most stable proton-ordered forms of ice Ic and Ih is $U^{\mathrm{Ic}} - U^{\mathrm{Ih}} = -0.3\,\mathrm{meV/H_2O}$ [194]. For the same quantity, diffusion Monte Carlo predicts $U^{\mathrm{Ic}} - U^{\mathrm{Ih}} = -0.4 \pm 2.9\,\mathrm{meV/H_2O}$ [181], two random phase approximation methods predict $-0.2\,\mathrm{meV/H_2O}$ and $0.7\,\mathrm{meV/H_2O}$ [195].

Since thorough sampling of the phase space of water at the revPBE0-D3 level of theory is prohibitively expensive, we sample the phase space using a surrogate ML potential energy surface (PES), U_{ML} and exploit the fact that the Gibbs free energy of the surrogate systems, G_{ML}, can be promoted to the revPBE0-D3 level of theory using free energy perturbation

$$G(p,T) - G_{\mathrm{ML}}(p,T) = -k_B T \ln \langle \exp\left[-\frac{U - U_{\mathrm{ML}}}{k_B T}\right]\rangle_{p,T,\mathscr{H}_{\mathrm{ML}}}, \tag{8.1}$$

where $\langle \ldots \rangle_{p,T,\mathscr{H}_{\mathrm{ML}}}$ denotes the ensemble average for the system sampled at temperature T and pressure p using the surrogate Hamiltonian $\mathscr{H}_{\mathrm{ML}}$. Evaluation of Eqn. (8.1) is rendered particularly affordable and robust by the high fidelity of our surrogate machine-learning PES, which substantially exceeds that obtained from empirical force fields or local DFT calculations, which were previously used as implicit surrogates [196, 197].

8.3 A neural network potential energy surface for bulk water

We constructed a flexible and fully dissociable neural network (NN) potential for bulk liquid water and ice following the framework of Behler and Parrinello [153, 198, 199] using the RuNNer code [200], which was trained on the basis of revPBE0-D3 energies and forces for 1,593 diverse reference structures of 64 molecules of liquid water computed using the CP2K package [201].

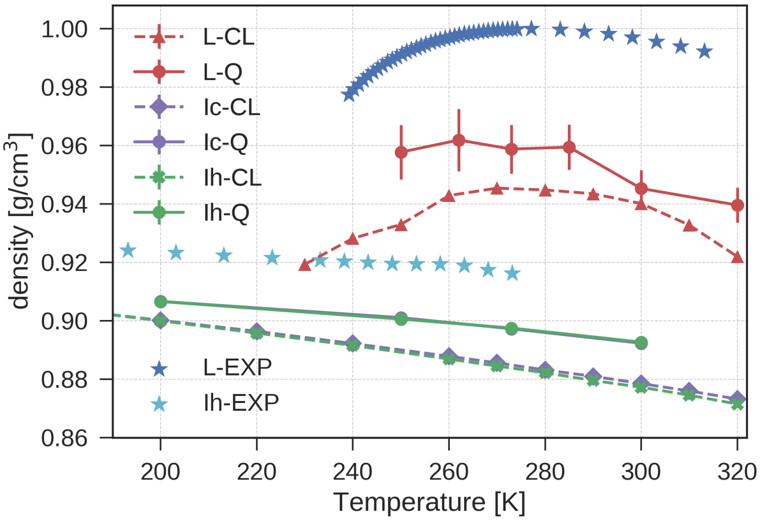

Figure 8.1 – Classical (CL) and quantum (Q) density isobars for ice Ic, ice Ih, and liquid water (L) at $P = 1$ bar computed via (PI)MD simulations using the NN potential. The predicted densities of ice Ic and Ih almost overlap both at the quantum and the classical level. The experimental results for undercooled water are taken from Ref. 202.

The revPBE0-D3-based NN potential describes the density (Fig. 8.1) and structural properties of water (Fig. 8.2) in very good agreement with experiments. Fig. 8.1 shows density isobars computed using both classical MD and PIMD simulations in the NPT ensemble for 64-molecule ice Ic, ice Ih, and liquid water systems [1]. All (PI)MD simulations use 56 beads and are performed employing the i-PI code [89] in conjunction with LAMMPS [204] with a NN potential implementation [205]. Fig. 8.1 highlights that (i) the predicted densities of liquid water and ice Ih and Ic agree with experiment to within 3%, (ii) the predicted thermal expansion coefficients show excellent agreement with experimental data, and (iii) the temperature of maximum density for liquid water matches the experimental determination of 3.98 C. It also shows that NQEs lead to an increase of around 1% in the density of the three phases of water. This anomalous increase for the ice phase has been observed in previous QHA calculations employing a number of different DFT functionals [185]. Experimentally, the suppression of NQEs can be partially achieved by deuteration, and it has been observed that the molar volume of D_2O is 0.4% [206] larger compared with H_2O for liquid water at

---

[1] We have confirmed that the equilibrium density computed with 64 water molecules in classical molecular dynamics simulations is consistent with the values obtained for systems with about 2,000 molecules.

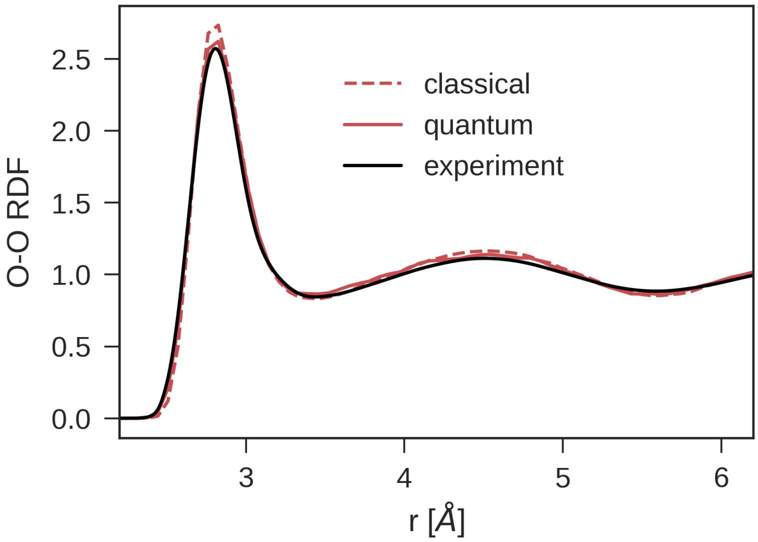

Figure 8.2 – Oxygen-oxygen radial distribution functions (RDF) at 300 K and experimental density computed via (PI)MD simulations in the NVT ensemble using the NN potential. The experimental RDF was obtained from Ref 203.

the ambient temperature, and about 0.3% larger for hexagonal ice at 250 K [207]. Fig. 8.2 shows that NQEs have a minor effect on the oxygen-oxygen radial distribution functions in Fig. 8.2, as one would expect because of competing quantum effects [184, 208]. The slight de-structuring in the simulations with quantum nuclei brings the RDF in excellent agreement with experiment [203], as also seen in the *ab initio* (PI)MD calculations with revPBE0-D3 [193].

Despite these excellent agreements, the fitting strategy, the finite cut-off radii applied to the description of atomic environments, and possible “holes” in the training set inevitably lead to small residual errors between NN predictions and the underlying first-principles reference. To assess their significance, we have trained a collection of NN potentials using different training sets and/or initial random seeds, which demonstrates that predictions of the chemical potential difference between ice Ic and Ih from two different NN potentials can be as large as 1 meV/H_2O. Promoting the results to the DFT level eliminates these residual errors and any dependence on the specific NN potential employed. This allows us to achieve sub-meV accuracy in free energies (as required to resolve the greater stability of ice Ih compared to Ic) and to make unbiased properties predictions at the reference *ab initio* level of theory in general. The temperature-dependent DFT corrections to the NN chemical potentials of different phases of water, $\Delta\mu_{NN} = \mu - \mu_{\mathrm{NN}} = (G - G_{\mathrm{NN}})/N$, as obtained from free energy perturbations (Eqn. (8.1)) performed on 64-molecule systems, are shown in Fig. 8.3. For each ice phase (Ic and Ih) 16 different proton-disordered initial configurations with zero net polarization were generated using the Hydrogen-Disordered Ice Generator [209]. The standard deviation of the potential energy for the 16 proton-disordered ice Ic configurations is 0.3 meV/H_2O (0.25 meV/H_2O) using the NN potential (DFT), respectively, and for ice Ih

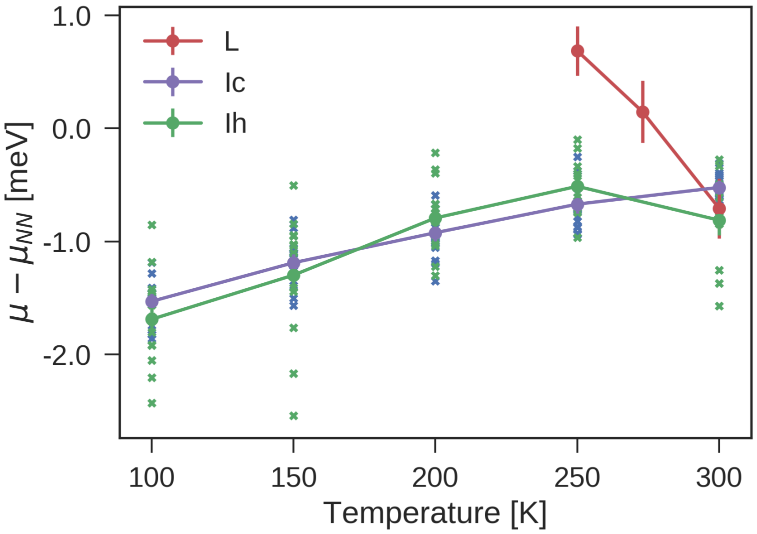

Figure 8.3 – The difference in the chemical potential $\Delta\mu_{\mathrm{NN}} \equiv \mu - \mu_{\mathrm{NN}}$ between revPBE0-D3 and NN-based MD simulations at $P = 1$ bar. Standard errors of the mean are indicated by the error bars. The violet (green) crosses indicate the results from 16 different 64-molecule proton-orderings of Ic (Ih). The violet (green) line shows the average $\Delta\mu_{\mathrm{NN}}$ across proton-orderings.

is 0.4 meV/H_2O (0.25 meV/H_2O) using the NN potential (DFT). Starting from these different initial configurations is crucial here, because (i) the proton order is effectively "frozen-in" at the timescales available to simulation [210] and (ii) there are significant differences between $\Delta\mu_{NN}$ of different proton-disordered states (see Fig. 8.3). For liquid water, 1000 single-point revPBE0-D3 calculations for un-correlated configurations generated from NN-based NPT simulations suffice to converge the value of the calibration term $\Delta\mu^{\mathrm{L}}_{NN}$ to about 0.2 meV/H_2O. For each proton-disordered ice structure, 200 such single-point calculations are enough to converge $\Delta\mu^{\mathrm{Ic}}_{NN}$ and $\Delta\mu^{\mathrm{Ih}}_{NN}$ to 0.1 meV/H_2O.

8.4 The relative stability of hexagonal and cubic ice

To evaluate the absolute quantum-mechanical Gibbs free energies of ice Ih and Ic (and thereby the chemical potential difference $\Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}}$), we first compute the classical free energies of the two phases at the NN level, using a sequence of TI steps. The speed and linear scaling of the NN potential allows us to simulate systems containing as many as 768 water molecules, which is essential to represent the wide spectrum of possible local arrangements realized in proton-disordered ice. During the subsequent step of promoting the NN free energies to the revPBE0-D3 level using free energy perturbation performed over 64-molecule ice systems, averaging over different proton disordered structures is also important, as demonstrated by the spread of $\Delta\mu_{NN}$ between different structures in Fig. 8.3.

We closely follow the TI methods described in Ref. [57]: we first integrate from a Debye crystal to classical ice at 25 K in the NVT ensemble, then transition to the NPT ensemble, and finally evaluate the temperature dependence of the Gibbs free energy using MD simulations in the NPT ensemble at temperatures between 25 K and 300 K. At this point the classical chemical potential difference between ice Ih and Ic with revPBE0-D3 can be evaluated as $\Delta\mu_{\mathrm{cl}}^{\mathrm{Ih}\to\mathrm{Ic}} = \Delta\mu_{\mathrm{cl,NN}}^{\mathrm{Ih}\to\mathrm{Ic}} + \Delta\mu_{NN}^{Ic} - \Delta\mu_{NN}^{Ih}$.

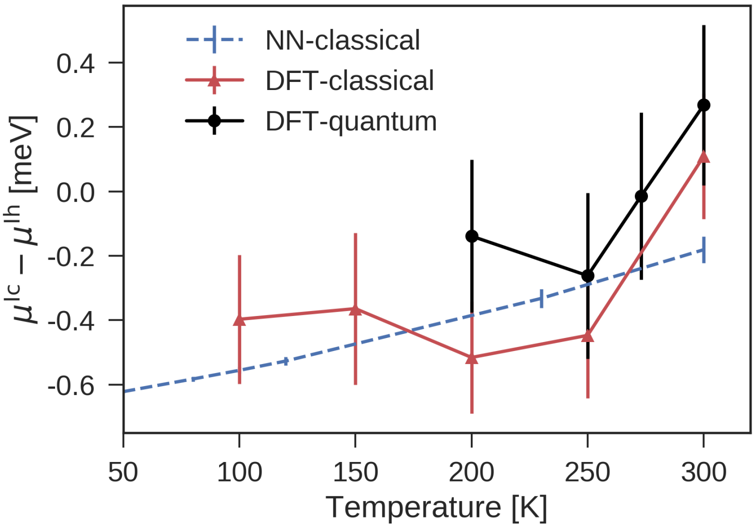

Figure 8.4 – Temperature dependence of the chemical potential difference between ice Ih and Ic at 1 bar. The errors associated with the classical and quantum-mechanical revPBE0-D3 values arise predominantly from the differences in $\Delta\mu_{\mathrm{NN}}$ between different proton-orderings.

NQEs can be taken into account by integrating the quantum centroid virial kinetic energy $\langle T_{CV}\rangle$ with respect to the fictitious "atomic" mass $\tilde{m}$ from the classical mass (i.e. infinity) to the physical masses of oxygen and hydrogen atoms [64, 171, 211, 212]. , i.e.

$$\Delta\mu_{\mathrm{NN}}^{\mathrm{Ih}\to\mathrm{Ic}} - \Delta\mu_{\mathrm{cl,NN}}^{\mathrm{Ih}\to\mathrm{Ic}} = \int_m^\infty dm \frac{\langle T_{CV}^{Ic}(\tilde{m})\rangle - \langle T_{CV}^{Ih}(\tilde{m})\rangle}{\tilde{m}} \tag{8.2}$$

where m are the physical masses of the elements. In practice, a change of variable $y = \sqrt{m/\tilde{m}}$ is applied to reduce the discretisation error in the evaluation of the integral [211], and the integrand is evaluated using PIMD simulations for $y = 1/4, 1/2\sqrt{2}, 1/2, 1/\sqrt{2}, 1$, i.e.

$$\Delta\mu_{\mathrm{NN}}^{\mathrm{Ih}\to\mathrm{Ic}} - \Delta\mu_{\mathrm{cl,NN}}^{\mathrm{Ih}\to\mathrm{Ic}} = 2\int_0^1 \frac{\langle T_{CV}^{Ic}(1/y^2)\rangle - \langle T_{CV}^{Ih}(1/y^2)\rangle}{y} dy. \tag{8.3}$$

To evaluate the integral in Eqn. (8.3), we perform NN-based PIMD simulations that use 56 beads at the NPT ensemble for systems containing 64 molecules, and assess the impact of NQEs on the chemical potential at the NN level using $\Delta\mu_{\mathrm{NN}}^{\mathrm{Ih}\to\mathrm{Ic}} - \Delta\mu_{\mathrm{cl,NN}}^{\mathrm{Ih}\to\mathrm{Ic}}$. We note that the NN potential is not "biased" towards Ic or Ih as the NN to revPBE0-D3 calibration terms $\Delta\mu_{NN}^{\mathrm{Ic}}$ and $\Delta\mu_{NN}^{\mathrm{Ih}}$ are similar (Fig. 8.3), and that the difference in $\langle T_{CV}\rangle$ of difference water phases

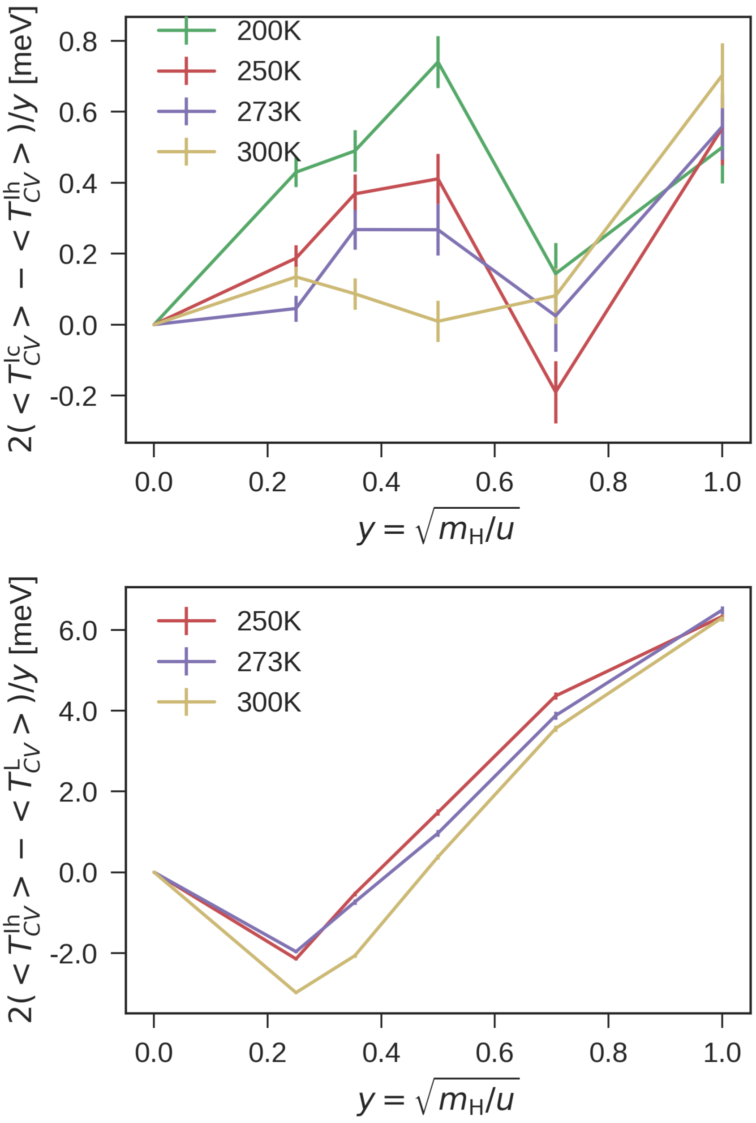

Figure 8.5 – The integral from the classical limit to the full quantum treatment (Eqn. (8.3)), for the case of ice Ic and Ih (upper panel), and ice Ih and liquid water (lower panel).

was previously found to be very similar for three completely different atomic potentials [171]. As such, finally we arrive at the result $\Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}} = \Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}}_{\mathrm{cl}} + \Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}}_{\mathrm{NN}} - \Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}}_{\mathrm{cl,NN}}$. [2]

Fig. 8.4 shows that the NN predictions of $\Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}}$ and the revPBE0-D3 results are statistically indistinguishable. At the classical level $\Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}}_{\mathrm{cl}}$ is negative, especially at low temperatures. Consistent with the VSCF results of Ref. 172, proton disorder introduces substantial variations in the chemical potential of ice Ic and Ih. More importantly, nuclear quantum motion is crucial to stabilize ice Ih. At the quantum-mechanical level $\Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}}$ is close to zero at 200-250 K and increases to $0.2 \pm 0.2\,\mathrm{meV/H_2O}$ at 300 K, suggesting ice Ih is more stable after all. For comparison, at the classical level, the monoatomic water model [155] – which omits hydrogen atoms – predicts a negligible difference ($\Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}}(240\ K) = 0.032 \pm 0.002\,\mathrm{meV}$ [5]), while the MB-pol forcefield [152], which includes many-body terms fitted to the coupled-clusters level of theory, predicts a small negative value ($-0.4\,\mathrm{meV/H_2O}$) [3].

8.5 The relative stability of hexagonal ice and liquid water

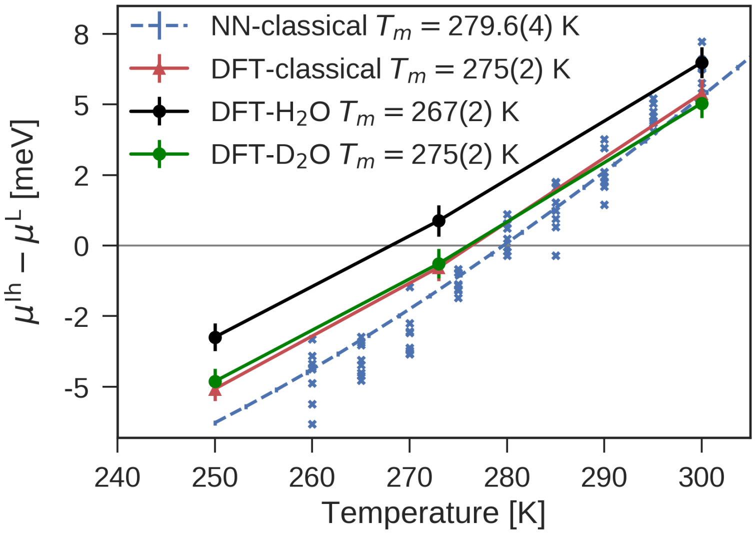

Figure 8.6 – Temperature dependence of the chemical potential difference between liquid water and ice Ih at 1 bar. Blue crosses indicate $\Delta\mu^{\mathrm{L}\to\mathrm{Ih}}_{\mathrm{cl,NN}}$ from independent interface pinning simulations, and the blue dashed line indicates the best fit of these results to the TI expression in Eqn (8.4).

To evaluate the difference in chemical potential $\Delta\mu^{\mathrm{L}\to\mathrm{Ih}} = \mu^{Ih} - \mu^{L}$ between the proton-disordered ice Ih and liquid water, we first perform interface pinning simulations [162] using the PLUMED code [93] on an ice-liquid system containing 5760 molecules at temperatures ranging from 250 K to 300 K and pressure 1 bar, employing the NN potential. Then we fit the results $\Delta\mu^{\mathrm{L}\to\mathrm{Ih}}_{\mathrm{cl,NN}}$ from independent simulations to the TI expression

---

[2] One can also choose to re-weight the whole ring-polymer using Eqn. (8.1), but this is more costly.

[3] We computed $\Delta\mu^{\mathrm{Ih}\to\mathrm{Ic}}(240\ K)$ for MB-Pol by performing a thermodynamic integration in the NPT ensemble, using a parameter λ to perform the switching between the NN Hamiltonian $\mathscr{H}_{\mathrm{NN}}$ and the MB-Pol Hamiltonian $\mathscr{H}_{\mathrm{MB}}$.

$$\Delta\mu^{\mathrm{L}\to\mathrm{Ih}}_{\mathrm{cl,NN}}(T) = -k_B T \int_{T_m}^{T} \frac{\langle H^{\mathrm{Ih}}_{\mathrm{cl,NN}} \rangle_{P,T} - \langle H^{\mathrm{L}}_{\mathrm{cl,NN}} \rangle_{P,T}}{k_B T^2} dT, \tag{8.4}$$

where $H_{cl,NN}$ is the enthalpy of the classical system described by the NN potential, whose value has been computed from separate NPT simulations [194]. The calibration terms for chemical potentials $\Delta\mu^{\mathrm{L}}_{\mathrm{NN}}$ and $\Delta\mu^{\mathrm{Ih}}_{\mathrm{NN}}$ (Fig. 8.3) are added in order to obtain the revPBE0-D3 predictions, which is shown as the red line in Fig. 8.6. NQEs in $\mathrm{H_2O}$ water and $\mathrm{D_2O}$ water are considered by running a series of PIMD simulations at different fictitious masses using the NN potential [64, 171, 211, 212].

Fig. 8.6 shows that both the NN potential and revPBE0-D3 give excellent predictions on the melting point of water (experimental result $T_\mathrm{m} = 273.15\,\mathrm{K}$). When performing TI from physical masses (m_H for H) to the classical limit, NQEs first stabilize water more compared to ice from m_H to about $6m_\mathrm{H}$, and afterwards stabilize ice more from $6m_\mathrm{H}$ to ∞ [194]. Similar phenomena have been observed before for q-TIP4P/F water [183] and for stacked polyglutamine [213], and have been interpreted as a manifestation of competing quantum effects. NQEs lower the melting point of $\mathrm{H_2O}$ by about 8 K compared with classical water. The difference in T_m between the $\mathrm{D_2O}$ and $\mathrm{H_2O}$ is predicted to be 8 ± 2 K, consistent with the result predicted using the q-TIP4P/F water model [183], and in rough agreement with experiment (3.82 K) [206]. The T_m of $\mathrm{D_2O}$ is about the same as the classical water, thanks to a cancellation of NQEs when performing TI from the atomic mass of deuterium to the classical mass. The heat of fusion ($H_f = H^{\mathrm{L}}(T_\mathrm{m}) - H^{\mathrm{Ih}}(T_\mathrm{m})$) is estimated to be 52 meV/$\mathrm{H_2O}$ for $\mathrm{H_2O}$, 54 meV/$\mathrm{H_2O}$ for $\mathrm{D_2O}$, and 58 meV/$\mathrm{H_2O}$ for classical water at the revPBE0-D3 level, which agree rather well with the experimental values of $H_f = 62.3$ meV/$\mathrm{H_2O}$ for $\mathrm{H_2O}$, and 64.5 meV/$\mathrm{H_2O}$ for $\mathrm{D_2O}$.

8.6 Comments on the first principles description of water and ML potentials

We demonstrate that it is possible to achieve a sub-meV level of statistical accuracy in predicting the thermodynamic properties of a complex system such as water at a hybrid DFT level of theory, using a machine-learning potential as an intermediate surrogate model. We show that a revPBE0-D3 description of the electronic structure predicts properties for ice Ih, ice Ic and liquid water that are in excellent quantitative agreement with experiment. In addition, we rigorously assess and disentangle the individual contributions from NQEs, proton disorder, and anharmonicity. The ideas behind our approach, and the free energy methods that we apply, are generally applicable to other materials, and enable the accurate and efficient prediction of the thermodynamic properties of physical systems using machine-learning potentials as a ladder to achieve first-principles accuracy.

On the other hand, the NN potential for bulk water has a number of desirable properties on its own, including a good description of the density isobar, the relative stabilities of ice Ic and Ih, and the melting point. As such, it is interesting to see what this NN potential will predict regarding the homogeneous ice nucleation rate, in comparison with the predictions from empirical water models. The framework and the methods presented in chapter 7 are directly applicable for such simulations. Including NQEs and promoting to DFT results are also possible as a post-processing step.

9 Conclusions

Predictive modelling and quantitative understanding of nucleation is essential for predicting phase transformation processes in nature and precisely controlling material synthesis and processing. Atomistic modeling is a powerful tool for capturing the dynamical processes and investigating the underlying mechanism of nucleation, but it faces several key challenges.

In the present thesis, we tackled the problem of nucleation by attacking it on several fronts. First of all we employed as well as devised enhanced sampling strategies to make the computation affordable. In the case of homogeneous ice nucleation, we computed the free energy profile associated with a single nucleation pathway, and then accounted for the free energy gain from the possibility of exhibiting stacking disorders in the nucleus.

We then formulated a thermodynamic framework to bridge the gap between the microscopic and macroscopic pictures of nucleation, and thus provides a simple and elegant framework to verify and extend classical nucleation theory. Using this framework, we accurately and rigorously extracted different physical quantities that affect nucleation, including the chemical potential, the interfacial free energy, and the Tolman length. By comparing the results that we obtained from simulations of homogeneous nucleation to the ones computed at the planar limit, we verified our thermodynamic framework, as well as benchmarked the accuracy of the classical nucleation theory.

Finally, we constructed a machine learning potential based on hybrid DFT data, in order to better model the interatomic interactions in water systems. We predicted the thermodynamic properties of liquid water as well as hexagonal and cubic ice, rigorously taking into account quantum nuclear motion, anharmonic fluctuations and proton disorder. The ab initio description not only leads to structural properties, density isobar, and melting point in excellent agreement with experiments, but also provides insights on how nuclear quantum effects modulate the stabilities of different phases of water. In addition, this ab initio modelling of water opens up many possibilities of future work, including a first principle description of ice nucleation.

To sum up, the present thesis provides the key instruments for investigating nucleation using atomistic simulations, and represents a substantial development in the quantitative understanding of the nucleation phenomenon. The work present here also opens the door for a number of future research opportunities that involve nucleation. For example, the precipitation of salt from solution, the cubic to orthorhombic phase transformation in cesium lead halide perovskites, and gas bubble formation from liquid water. More in general, the conceptual approach to reconcile atomistic and macroscopic modeling, and the combination of state-of-the-art sampling and machine-learning techniques, provide a robust, generally applicable framework that can be applied to model and understand other important open problems in materials science, such as computing contact angles, modelling emulsion instabilities, and predicting the thermal stabilities of polymorphic molecular crystals.

10 Appendix

10.1 Cluster distributions under the multiple cluster model

The free energy profile $\tilde{G}(n_{\text{tot}})$ as a function of total number of atoms in all the clusters can be computed by enumerating all the possible combinations using Eqn. (4.3). If a cutoff size n_{cut} is selected, one can also compute the distribution of n_{tot} under the constraint that no cluster in the system is of size larger than n_{cut}, i.e.

$$\exp\left(-\beta\tilde{G}^{(0)}(n_{\text{tot}})\right) = \sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\dots\sum_{k_{n_{\text{cut}}}=0}^{\infty}\delta\left(\sum_{n=1}^{\infty} nk_n - n_{\text{tot}}\right)\prod_{n=1}^{n_{\text{cut}}} P(n,k_n), \tag{10.1}$$

On the other hand, if one restricts the system to have exactly one cluster larger than n_{cut}, the distribution of n_{tot} can be written as

$$\exp\left(-\beta\tilde{G}^{(1)}(n_{\text{tot}})\right) = \sum_{n_b=n_{\text{cut}}}^{n_{\text{tot}}} P(n_b,1)\exp(-\beta\tilde{G}^{(0)}(n_{\text{tot}}-n_b)) \tag{10.2}$$

Furthermore, if one restricts the system to have two clusters that have sizes larger than n_{cut}, the distribution of n_{tot} can be written as

$$\exp\left(-\beta\tilde{G}^{(2)}(n_{\text{tot}})\right) = \sum_{n_a=n_{\text{cut}}}^{n_{\text{tot}}}\sum_{n_b=n_{\text{cut}}}^{n_a} P(n_a,1)P(n_b,1)\exp(-\beta\tilde{G}^{(0)}(n_{\text{tot}}-n_a-n_b)) \tag{10.3}$$

We selected $n_{\text{cut}} = 150$, computed $\tilde{G}(n_{\text{tot}})$, $\tilde{G}^{(0)}(n_{\text{tot}})$, $\tilde{G}^{(1)}(n_{\text{tot}})$, and $\tilde{G}^{(2)}(n_{\text{tot}})$, and plotted them in Figure 10.1. It can be seen that at $n_{\text{tot}} > n_{\text{cut}} + \langle n_{\text{tot}}\rangle$, $\tilde{G}^{(1)}(n_{\text{tot}})$ overlaps with $\tilde{G}(n_{\text{tot}})$. In addition, at $n_{\text{tot}} > n_{\text{cut}} + \langle n_{\text{tot}}\rangle$ it is many orders of magnitude less likely to have zero or two clusters of size larger than n_{cut} compared with only having one such cluster. Therefore, at $n_{\text{tot}} > n_{\text{cut}} + \langle n_{\text{tot}}\rangle$ the system effectively only contains one large cluster with size larger than n_{cut}, and $\tilde{G}^{(1)}(n_{\text{tot}})$ is a good approximation of $\tilde{G}(n_{\text{tot}})$. Note that n_{cut} does not need to be decided exactly. As long as n_{cut} is reasonably large, the conclusions above apply.

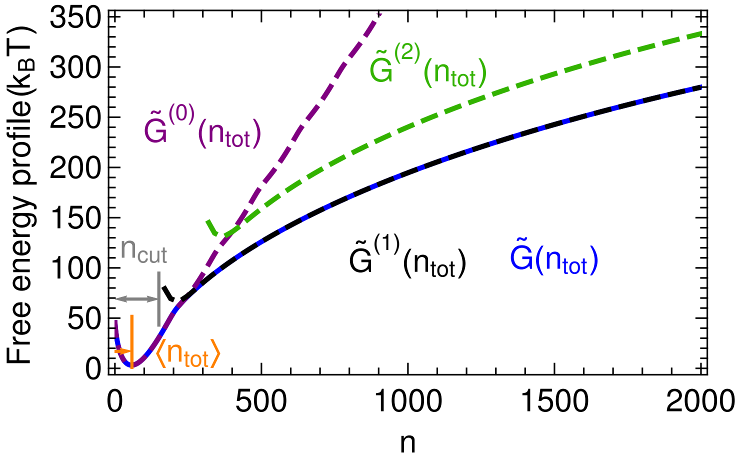

Figure 10.1 – Free energy profiles $\tilde{G}(n_{\text{tot}})$, $\tilde{G}^{(0)}(n_{\text{tot}})$, $\tilde{G}^{(1)}(n_{\text{tot}})$, and $\tilde{G}^{(2)}(n_{\text{tot}})$ of the atomistic system of multiple clusters. At $n_{\text{tot}} < n_{\text{cut}}$, $\tilde{G}^{(0)}(n_{\text{tot}})$ is exactly equal to $\tilde{G}(n_{\text{tot}})$. At $n_{\text{tot}} > n_{\text{cut}} + \langle n_{\text{tot}}\rangle$, $\tilde{G}^{(1)}(n_{\text{tot}})$ lies on the top of $\tilde{G}(n_{\text{tot}})$.

10.2 A fluctuating Gibbs dividing surface

For a solid-liquid system that has N atoms, a box size $\{l_x, l_y, l_z\}$, and a planar interface perpendicular to the z axis, an instantaneous order parameter density field can be introduced as

$$\widetilde{\phi}(x,y,z) = \sum_{i=1}^{N} \phi_i g(x-x_i)g(y-y_i)g(z-z_i), \tag{10.4}$$

where ϕ_i and (x_i, y_i, z_i) denote the atomic order parameter and coordinates of each atom, respectively. The 3-dimensional normalized kernel $g(x-x_i)g(y-y_i)g(z-z_i)$ can be chosen to be a Gaussian function, e.g.

$$g(x-x_i)g(y-y_i)g(z-z_i) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^3 \exp -\frac{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{2\sigma^2}. \tag{10.5}$$

Take a thin domain that has an infinitesimal cross-section centered at (x, y), the extensive quantity based on the atomic order parameter for this needle-shaped domain is $\Phi(x, y)$, and the instantaneous height $h(x, y)$ of the Gibbs dividing surface that gives zero access of the extensive quantity Φ for this domain fulfills

$$\Phi(x,y) = \int_0^{l_z} dz\widetilde{\phi}(x,y,z) = \int_0^{h(x,y)} dz\widetilde{\phi}_{\text{s}} + \int_{h(x,y)}^{l_z} dz\widetilde{\phi}_{\text{l}}, \tag{10.6}$$

where $\widetilde{\phi}_{\text{s}}$ and $\widetilde{\phi}_{\text{l}}$ indicate the density field that contains small fluctuations inside the bulk solid and the bulk liquid. Combining Eqn. (10.4) and (10.5), one can write

$$\int_0^{h(x,y)} dz\widetilde{\phi}_\mathrm{s} + \int_{h(x,y)}^{l_z} dz\widetilde{\phi}_\mathrm{l} = \sum_{i=1}^{N} \phi_i \frac{1}{2\pi\sigma^2} \exp - \frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}. \tag{10.7}$$

Take a 2-D Fourier expansion for the left hand side of Eqn. (10.7) over the whole cross section of the simulation box $\{l_x, l_y\}$, i.e.

$$FT(\Phi(x,y)) = \frac{1}{l_x l_y} \int_0^{l_x} dx \int_0^{l_y} dy \left(\int_0^{h(x,y)} dz\widetilde{\phi}_\mathrm{s} + \int_{h(x,y)}^{l_z} dz\widetilde{\phi}_\mathrm{l} \right) \exp\left(-\mathrm{i}k_x x - \mathrm{i}k_y y\right), \tag{10.8}$$

where $k_x = 2\pi m/l_x, m = 0, \pm 1, \pm 2, \ldots$, and similarly for k_y. To simplify the evaluation of the above Fourier expansion, one can split the bulk density field into an averaged term and a fluctuating term, e.g. $\widetilde{\phi}_\mathrm{s} = \langle \widetilde{\phi}_\mathrm{s} \rangle + \Delta\widetilde{\phi}_\mathrm{s}$. At the meantime, $h(x,y) = \langle h \rangle + \Delta h(x,y)$, with the average height of the interface $\langle h \rangle = \int_0^{l_x} dx \int_0^{l_y} dy h(x,y)/l_x l_y$. As such, the left hand side of the Eqn. (10.7) can be rewritten as

$$\langle \widetilde{\phi}_\mathrm{s} \rangle\, h(x,y) + \langle \widetilde{\phi}_\mathrm{l} \rangle\, (l_z - h(x,y)) + \int_0^{\langle h \rangle} dz\Delta\widetilde{\phi}_\mathrm{s} + \int_{\langle h \rangle}^{l_z} dz\Delta\widetilde{\phi}_\mathrm{l} + \int_{\langle h \rangle}^{h(x,y)} dz\Delta\widetilde{\phi}_\mathrm{s} + \int_{h(x,y)}^{\langle h \rangle} dz\Delta\widetilde{\phi}_\mathrm{l}. \tag{10.9}$$

Now let us consider the Fourier coefficients of the non-zero frequency modes of Eqn. (10.9). For the first two terms in Eqn. (10.9),

$$FT\left(\langle \widetilde{\phi}_\mathrm{s} \rangle\, h(x,y) + \langle \widetilde{\phi}_\mathrm{l} \rangle\, (l_z - h(x,y)) \right) = \left(\langle \widetilde{\phi}_\mathrm{s} \rangle - \langle \widetilde{\phi}_\mathrm{l} \rangle \right) A_h(k_x, k_y), \tag{10.10}$$

where

$$A_h(k_x, k_y) = \frac{1}{l_x l_y} \int_0^{l_x} dx \int_0^{l_y} dy h(x,y) \exp\left(-\mathrm{i}k_x x - \mathrm{i}k_y y\right). \tag{10.11}$$

is just the Fourier coefficients of the height function

$$h(x,y) = \sum_{k_x, k_y} A_h(k_x, k_y) \exp\left(\mathrm{i}k_x x + \mathrm{i}k_y y\right). \tag{10.12}$$

As for the third in Eqn. (10.9), the Fourier coefficients characterizing the fluctuations $\Delta\widetilde{\phi}_\mathrm{s}$ in a slab of bulk solid that has a cross section $\{l_x, l_y\}$ and thickness $\langle h \rangle$ are

$$A_s(\langle h \rangle; k_x, k_y) = \frac{1}{l_x l_z} \int_0^{l_x} dx \int_0^{l_y} dy \int_0^{\langle h \rangle} dz\widetilde{\phi}_\mathrm{s}(x,y,z) \exp\left(-\mathrm{i}k_x x - \mathrm{i}k_y y\right). \tag{10.13}$$

Similarly, for the fourth term in Eqn. (10.9), the Fourier coefficients are

$$A_l(l_z - \langle h \rangle; k_x, k_y) = \frac{1}{l_x l_z} \int_0^{l_x} dx \int_0^{l_y} dy \int_0^{l_z - \langle h \rangle} dz\widetilde{\phi}_\mathrm{l}(x,y,z) \exp\left(-\mathrm{i}k_x x - \mathrm{i}k_y y\right). \tag{10.14}$$

Note that $A_s(\langle h \rangle; k_x, k_y)$ and $A_l(l_z - \langle h \rangle; k_x, k_y)$ can be computed independently from molecular dynamics simulations of bulk solid and bulk liquid using a simulation box of the same cross section $\{l_x, l_y\}$. In addition, if one assumes that the bulk fluctuations and the surface fluctuations are mutually independent, the last two terms in Eqn. (10.9) do not contribute to the Fourier components. To wrap up, the Fourier expansion of the left hand side of Eqn. (10.9) yields

$$(\langle\widetilde{\phi}_{\mathrm{s}}\rangle - \langle\widetilde{\phi}_{\mathrm{l}}\rangle)A_h(k_x,k_y) + A_s(\langle h\rangle\,;k_x,k_y) + A_l(l_z - \langle h\rangle\,;k_x,k_y), \tag{10.15}$$

and as the three terms above are all mutually independent, the ensemble average of the Fourier amplitudes can be expressed as

$$\langle|FT(\Phi(x,y))|^2\rangle = (\langle\widetilde{\phi}_{\mathrm{s}}\rangle - \langle\widetilde{\phi}_{\mathrm{l}}\rangle)^2\,\langle|A_h(k_x,k_y)|^2\rangle + \langle|A_s(\langle h\rangle\,;k_x,k_y)|^2\rangle + \langle|A_l(l_z - \langle h\rangle\,;k_x,k_y)|^2\rangle \tag{10.16}$$

Furthermore, additional approximations can be made to facilitate the computation of $\langle|A_s(\langle h\rangle\,;k_x,k_y)|^2\rangle$ and $\langle|A_l(l_z - \langle h\rangle\,;k_x,k_y)|^2\rangle$: If one assumes that the correlations between the fluctuations of different layers in the bulk phases along the z axis are short-ranged, it is possible to select a length scale λ that is much larger than this spatial correlation length but much smaller than the simulation cell height l_z, and one can make the approximations such as

$$\langle|A_s(\langle h\rangle\,;k_x,k_y)|^2\rangle = \frac{\langle h\rangle}{\lambda}\,\langle|A_s(\lambda;k_x,k_y)|^2\rangle\,, \tag{10.17}$$

and

$$\langle|A_s(\lambda;k_x,k_y)|^2\rangle = \langle\left[\frac{1}{l_x l_y}\sum_{i=1}^{N_s}\phi_i \exp\left(-\mathrm{i}k_x x_i - \mathrm{i}k_y y_i\right)H(\lambda - z_i)H(z_i)\right]^2\rangle\,, \tag{10.18}$$

where $H(\ldots)$ is the Heaviside function and N_s is the number of atoms in the bulk solid system confined in a simulation box of the cross section $\{l_x, l_y\}$.

Now consider the 2-D Fourier expansion for the right hand side of Eqn. (10.7) over the whole cross section of the simulation box $\{l_x, l_y\}$:

$$\begin{aligned}
&\frac{1}{l_x l_y}\int_0^{l_x} dx \int_0^{l_y} dy \sum_{i=1}^{N}\phi_i \frac{1}{2\pi\sigma^2}\exp - \frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}\exp\left(-\mathrm{i}k_x x - \mathrm{i}k_y y\right)\\
&= \frac{1}{l_x l_y}\sum_{i=1}^{N}\phi_i \int_{-\infty}^{\infty} dx' \int_{-\infty}^{\infty} dy' \frac{1}{2\pi\sigma^2}\exp - \frac{x'^2 + y'^2}{2\sigma^2}\exp\left(-\mathrm{i}k_x x' - \mathrm{i}k_y y'\right)\exp\left(-\mathrm{i}k_x x_i - \mathrm{i}k_y y_i\right)\\
&= \frac{1}{l_x l_y}\sum_{i=1}^{N}\phi_i \exp - \frac{\sigma^2(k_x^2 + k_y^2)}{2}\exp\left(-\mathrm{i}k_x x_i - \mathrm{i}k_y y_i\right).
\end{aligned} \tag{10.19}$$

Because the wavelength associated with k has to be much larger than the atomic spacing in order for the capillary wave treatment to be meaningful (i.e. $2\pi/k_x \gg \sigma$), one can simply take

$\exp(-\sigma^2(k_x^2+k_y^2)/2) \equiv 1$ and obtain

$$\langle |FT(\Phi(x,y))|^2\rangle = \langle \left[\frac{1}{l_x l_y} \sum_{i=1}^{N} \phi_i \exp(-\mathrm{i}k_x x - \mathrm{i}k_y y) \right]^2 \rangle . \tag{10.20}$$

To sum up, the ensemble average of the non-zero frequency Fourier components follows

$$\begin{aligned} (\langle \widetilde{\phi}_{\mathrm{s}} \rangle - \langle \widetilde{\phi}_{\mathrm{l}} \rangle)^2 \langle |A_h(k_x,k_y)|^2\rangle + \langle |A_s(\langle h\rangle ; k_x,k_y)|^2\rangle + \langle |A_l(l_z - \langle h\rangle ; k_x,k_y)|^2\rangle \\ = \langle \left[\frac{1}{l_x l_y} \sum_{i=1}^{N} \phi_i \exp(-\mathrm{i}k_x x_j - \mathrm{i}k_y y_j) \right]^2 \rangle . \end{aligned} \tag{10.21}$$

Furthermore, when more than one planar interface are present in the system, the average amplitudes of their capillary waves are also additive to the above expression. For instance, if there are two parallel planar interfaces h_1 and h_2, the left hand side of Eqn. (10.21) becomes

$$(\langle \widetilde{\phi}_{\mathrm{s}} \rangle - \langle \widetilde{\phi}_{\mathrm{l}} \rangle)^2 \left(\langle |A_{h1}(k_x,k_y)|^2\rangle + \langle |A_{h2}(k_x,k_y)|^2\rangle \right) + \langle |A_s(\langle h\rangle ; k_x,k_y)|^2\rangle + \langle |A_l(l_z - \langle h\rangle ; k_x,k_y)|^2\rangle . \tag{10.22}$$

If the two interfaces are not interacting, the two amplitudes $\langle |A_{h1}(k_x,k_y)|^2\rangle$ and $\langle |A_{h2}(k_x,k_y)|^2\rangle$ are independent so the sum of the two average amplitudes is just two times the value for one interface, i.e.

$$\langle |A_{h1}(k_x,k_y)|^2\rangle + \langle |A_{h2}(k_x,k_y)|^2\rangle = 2\langle |A_h(k_x,k_y)|^2\rangle . \tag{10.23}$$

Bibliography

[1] Bingqing Cheng, Gareth A Tribello, and Michele Ceriotti. Solid-liquid interfacial free energy out of equilibrium. *Physical Review B*, 92(18):180102, 2015.

[2] Bingqing Cheng and Michele Ceriotti. Bridging the gap between atomistic and macroscopic models of homogeneous nucleation. *The Journal of chemical physics*, 146(3):034106, 2017.

[3] Bingqing Cheng, Gareth A Tribello, and Michele Ceriotti. The gibbs free energy of homogeneous nucleation: From atomistic nuclei to the planar limit. *The Journal of Chemical Physics*, 147(10):104707, 2017.

[4] Bingqing Cheng and Michele Ceriotti. Communication: Computing the tolman length for solid-liquid interfaces. *The Journal of Chemical Physics*, 148(23):231102, 2018.

[5] Bingqing Cheng, Christoph Dellago, and Michele Ceriotti. Theoretical prediction of the homogeneous ice nucleation rate: disentangling thermodynamics and kinetics. *Physical Chemistry Chemical Physics*, 2018.

[6] Bingqing Cheng, Edgar A Engel, J örg Behler, Christoph Dellago, and Michele Ceriotti. ab initio thermodynamics of liquid and solid water. *Proceedings of the National Academy of Sciences*, In press.

[7] David W Oxtoby. Homogeneous nucleation: theory and experiment. *Journal of Physics: Condensed Matter*, 4(38):7627, 1992.

[8] Jürn Schmelzer, Gerd Röpke, and Vyatcheslav B Priezzhev. *Nucleation theory and applications.* Wiley Online Library, 2005.

[9] Peng Yi and Gregory C Rutledge. Molecular origins of homogeneous crystal nucleation. *Annual review of chemical and biomolecular engineering*, 3:157–182, 2012.

[10] Gabriele C Sosso, Ji Chen, Stephen J Cox, Martin Fitzner, Philipp Pedevilla, Andrea Zen, and Angelos Michaelides. Crystal nucleation in liquids: Open questions and future challenges in molecular dynamics simulations. *Chemical reviews*, 2016.

[11] Pascal Jonkheijm, Paul van der Schoot, Albertus PHJ Schenning, and EW Meijer. Probing the solvent-assisted nucleation pathway in chemical self-assembly. *Science*, 313(5783):80–83, 2006.

[12] W.J. Boettinger, S.R. Coriell, A.L. Greer, A. Karma, W. Kurz, M. Rappaz, and R. Trivedi. Solidification microstructures: recent developments, future directions. *Acta Materialia*, 48(1):43 – 70, 2000.

[13] JA Venables, GDT Spiller, and M Hanbucken. Nucleation and growth of thin films. *Reports on Progress in Physics*, 47(4):399, 1984.

[14] Jonathan A Dantzig and Michel Rappaz. *Solidification*. EPFL press, 2009.

[15] Dianzhong Li, Xing-Qiu Chen, Paixian Fu, Xiaoping Ma, Hongwei Liu, Yun Chen, Yanfei Cao, Yikun Luan, and Yiyi Li. Inclusion flotation-driven channel segregation in solidifying steels. *Nature Comm.*, 5:5572, 2014.

[16] Sean V Murphy and Anthony Atala. 3d bioprinting of tissues and organs. *Nature biotechnology*, 32(8):773–785, 2014.

[17] JJ Hoyt, Mark Asta, and Alain Karma. Atomistic and continuum modeling of dendritic solidification. *Materials Science and Engineering: R: Reports*, 41(6):121–163, 2003.

[18] David T Wu, László Gránásy, and Frans Spaepen. Nucleation and the solid–liquid interfacial free energy. *MRS bulletin*, 29(12):945–950, 2004.

[19] Martin Volmer and A Weber. Keimbildung in übersättigten gebilden. *Zeitschrift für physikalische Chemie*, 119(1):277–301, 1926.

[20] Richard Becker and Werner Döring. Kinetische behandlung der keimbildung in übersättigten dämpfen. *Annalen der Physik*, 416(8):719–752, 1935.

[21] S Karthika, TK Radhakrishnan, and P Kalaichelvi. A review of classical and nonclassical nucleation theories. *Crystal Growth & Design*, 16(11):6663–6681, 2016.

[22] Wilhelm Ostwald. Studien über die bildung und umwandlung fester körper. *Zeitschrift für physikalische Chemie*, 22(1):289–330, 1897.

[23] David W Oxtoby and R Evans. Nonclassical nucleation theory for the gas–liquid transition. *The Journal of chemical physics*, 89(12):7521–7530, 1988.

[24] John W Cahn and John E Hilliard. Free energy of a nonuniform system. i. interfacial free energy. *The Journal of chemical physics*, 28(2):258–267, 1958.

[25] XC Zeng and David W Oxtoby. Gas–liquid nucleation in lennard-jones fluids. *The Journal of chemical physics*, 94(6):4472–4478, 1991.

[26] Emilie M Pouget, Paul HH Bomans, Jeroen ACM Goos, Peter M Frederik, Nico AJM Sommerdijk, et al. The initial stages of template-controlled caco3 formation revealed by cryo-tem. *Science*, 323(5920):1455–1458, 2009.

[27] Jonas A Sellberg, C Huang, Trevor A McQueen, ND Loh, H Laksmono, Daniel Schlesinger, RG Sierra, D Nordlund, CY Hampton, Dmitri Starodub, et al. Ultrafast x-ray probing of water structure below the homogeneous ice nucleation temperature. *Nature*, 510(7505):381, 2014.

[28] Hartawan Laksmono, Trevor A McQueen, Jonas A Sellberg, N Duane Loh, Congcong Huang, Daniel Schlesinger, Raymond G Sierra, Christina Y Hampton, Dennis Nordlund, Martin Beye, et al. Anomalous behavior of the homogeneous ice nucleation rate in “no-man’s land”. *The journal of physical chemistry letters*, 6(14):2826–2832, 2015.

[29] James M Campbell, Fiona C Meldrum, and Hugo K Christenson. Is ice nucleation from supercooled water insensitive to surface roughness? *The Journal of Physical Chemistry C*, 119(2):1164–1169, 2015.

[30] Sigurd Bauerecker, Peter Ulbig, Victoria Buch, Luboš Vrbka, and Pavel Jungwirth. Monitoring ice nucleation in pure and salty water via high-speed imaging and computer simulations. *The Journal of Physical Chemistry C*, 112(20):7631–7636, 2008.

[31] David Ehre, Etay Lavert, Meir Lahav, and Igor Lubomirsky. Water freezes differently on positively and negatively charged surfaces of pyroelectric materials. *Science*, 327(5966):672–675, 2010.

[32] Bernard Vonnegut. Variation with temperature of the nucleation rate of supercooled liquid tin and water drops. *Journal of colloid science*, 3(6):563–569, 1948.

[33] Christoph Dellago, Peter G Bolhuis, and Phillip L Geissler. Transition path sampling. *Advances in chemical physics*, 123:1–78, 2002.

[34] Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[35] A Laio and M Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*, 99:12562–12566, 2002.

[36] Jorge R Espinosa, Carlos Vega, Chantal Valeriani, and Eduardo Sanz. Seeding approach to crystal nucleation. *The Journal of Chemical Physics*, 144(3):034501, 2016.

[37] Titus S Van Erp and Peter G Bolhuis. Elaborating transition interface sampling methods. *Journal of computational Physics*, 205(1):157–181, 2005.

[38] Rosalind J Allen, Chantal Valeriani, and Pieter Rein ten Wolde. Forward flux sampling for rare event simulations. *Journal of physics: Condensed matter*, 21(46):463102, 2009.

[39] Albert C Pan and David Chandler. Dynamics of nucleation in the ising model. *The Journal of Physical Chemistry B*, 108(51):19681–19686, 2004.

[40] Daniele Moroni, Pieter Rein Ten Wolde, and Peter G Bolhuis. Interplay between structure and size in a critical crystal nucleus. *Physical review letters*, 94(23):235703, 2005.

[41] Amir Haji-Akbari and Pablo G Debenedetti. Direct calculation of ice homogeneous nucleation rate for a molecular model of water. *Proceedings of the National Academy of Sciences*, 112(34):10582–10588, 2015.

[42] Laura Lupi, Arpa Hudait, Baron Peters, Michael Grünwald, Ryan Gotchy Mullen, Andrew H Nguyen, and Valeria Molinero. Role of stacking disorder in ice nucleation. *Nature*, 551(7679):218, 2017.

[43] Pieter Rein ten Wolde, Maria J Ruiz-Montero, and Daan Frenkel. Numerical calculation of the rate of crystal nucleation in a lennard-jones system at moderate undercooling. *The Journal of chemical physics*, 104(24):9932–9947, 1996.

[44] Pieter Rein ten Wolde and Daan Frenkel. Homogeneous nucleation and the ostwald step rule. *Physical Chemistry Chemical Physics*, 1(9):2191–2196, 1999.

[45] Stefan Auer and Daan Frenkel. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature*, 409(6823):1020, 2001.

[46] Federica Trudu, Davide Donadio, and Michele Parrinello. Freezing of a lennard-jones fluid: From nucleation to spinodal regime. *Physical review letters*, 97(10):105701, 2006.

[47] Lutz Maibaum. Phase transformation near the classical limit of stability. *Physical review letters*, 101(25):256102, 2008.

[48] Wolfgang Lechner, Christoph Dellago, and Peter G Bolhuis. Role of the prestructured surface cloud in crystal nucleation. *Physical review letters*, 106(8):085701, 2011.

[49] Santi Prestipino, Alessandro Laio, and Erio Tosatti. Systematic improvement of classical nucleation theory. *Physical review letters*, 108(22):225701, 2012.

[50] Matteo Salvalaglio, Claudio Perego, Federico Giberti, Marco Mazzotti, and Michele Parrinello. Molecular-dynamics simulations of urea nucleation from aqueous solution. *Proceedings of the National Academy of Sciences*, 112(1):E6–E14, 2015.

[51] James McCarty, Omar Valsson, and Michele Parrinello. Bespoke bias for obtaining free energy differences within variationally enhanced sampling. *Journal of chemical theory and computation*, 12(5):2162–2169, 2016.

[52] Pablo Miguel Piaggi, Omar Valsson, and Michele Parrinello. Fdreact16: A variational approach to nucleation simulation. *Faraday Discussions*, 2016.

[53] Yuri Lifanov, Bart Vorselaars, and David Quigley. Nucleation barrier reconstruction via the seeding method in a lattice model with competing nucleation pathways. *The Journal of Chemical Physics*, 145(21), 2016.

[54] JR Espinosa, C Navarro, E Sanz, C Valeriani, and C Vega. On the time required to freeze water. *The Journal of Chemical Physics*, 145(21):211922, 2016.

[55] Dmitri Rozmanov and Peter G Kusalik. Isoconfigurational molecular dynamics study of the kinetics of ice crystal growth. *Physical Chemistry Chemical Physics*, 14(37):13010–13018, 2012.

[56] Santi Prestipino. The barrier to ice nucleation in monatomic water. *The Journal of chemical physics*, 148(12):124505, 2018.

[57] B. Cheng and M. Ceriotti. Computing the absolute Gibbs free energy in atomistic simulations: Applications to defects in solids. *Physical Review B*, 97:054102, 2018.

[58] Josiah Willard Gibbs, Henry Andrews Bumstead, William Raymond Longley, et al. *The collected works of J. Willard Gibbs*, volume 1. Longmans, Green and Company, 1928.

[59] Michael P Allen and Dominic J Tildesley. *Computer simulation in chemical physics*, volume 397. Springer Science & Business Media, 2012.

[60] Mark Tuckerman. *Statistical mechanics: theory and molecular simulation*. Oxford University Press, 2010.

[61] Gavin E Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, 60(3):2721, 1999.

[62] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

[63] Michele Ceriotti, Guy AR Brain, Oliver Riordan, and David E Manolopoulos. The inefficiency of re-weighted sampling and the curse of system size in high-order path integration. *Proc. R. Soc. A*, page rspa20110413, 2011.

[64] Bingqing Cheng and Michele Ceriotti. Direct path integral estimators for isotope fractionation ratios. *The Journal of chemical physics*, 141(24):244112, 2014.

[65] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[66] Peter G Bolhuis, David Chandler, Christoph Dellago, and Phillip L Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual review of physical chemistry*, 53(1):291–318, 2002.

[67] C Vega, E Sanz, JLF Abascal, and EG Noya. Determination of phase diagrams via computer simulation: methodology and applications to water, electrolytes and proteins. *Journal of Physics: Condensed Matter*, 20(15):153101, 2008.

[68] Luca M Ghiringhelli, Jan H Los, Evert Jan Meijer, A Fasolino, and Daan Frenkel. Modeling the phase diagram of carbon. *Physical review letters*, 94(14):145701, 2005.

[69] E Sanz, C Vega, JLF Abascal, and LG MacDowell. Phase diagram of water from computer simulation. *Physical review letters*, 92(25):255701, 2004.

[70] James M Polson, E Trizac, S Pronk, and D Frenkel. Finite-size corrections to the free energies of crystalline solids. *The Journal of Chemical Physics*, 112(12):5339–5342, 2000.

[71] James M Polson and Daan Frenkel. Calculation of solid-fluid phase equilibria for systems of chain molecules. *The Journal of chemical physics*, 109(1):318–328, 1998.

[72] Eva G Noya, MM Conde, and Carlos Vega. Computing the free energy of molecular solids by the einstein molecule approach: Ices xiii and xiv, hard-dumbbells and a patchy model of proteins. *The Journal of chemical physics*, 129(10):104704, 2008.

[73] Philippe Y Ayala and H Bernhard Schlegel. Identification and treatment of internal rotation in normal mode vibrational analysis. *The Journal of chemical physics*, 108(6):2314–2325, 1998.

[74] Sabry G Moustafa, Andrew J Schultz, and David A Kofke. Very fast averaging of thermal properties of crystals by molecular simulation. *Physical Review E*, 92(4):043303, 2015.

[75] Andrew J Schultz, Sabry G Moustafa, Weisong Lin, Steven J Weinstein, and David A Kofke. Reformulation of ensemble averages via coordinate mapping. *Journal of chemical theory and computation*, 12(4):1491–1498, 2016.

[76] Michele Ceriotti and Thomas E Markland. Efficient methods and practical guidelines for simulating isotope effects. *J. Chem. Phys.*, 138:014112, 2013.

[77] David J Earl and Michael W Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005.

[78] Rodrigo Freitas, Mark Asta, and Maurice de Koning. Nonequilibrium free-energy calculation of solids using lammps. *Computational Materials Science*, 112:333–341, 2016.

[79] Jianjun Xie, Stefano de Gironcoli, Stefano Baroni, and Matthias Scheffler. First-principles calculation of the thermal properties of silver. *Physical Review B*, 59(2):965, 1999.

[80] R Ramírez, N Neuerburg, M-V Fernández-Serra, and CP Herrero. Quasi-harmonic approximation of thermodynamic properties of ice ih, ii, and iii. *The Journal of chemical physics*, 137(4):044502, 2012.

[81] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

[82] William G Hoover. Entropy for small classical crystals. *The Journal of Chemical Physics*, 49(4):1981–1982, 1968.

[83] Paolo Giannozzi, Stefano De Gironcoli, Pasquale Pavone, and Stefano Baroni. Ab initio calculation of phonon dispersions in semiconductors. *Physical Review B*, 43(9):7231, 1991.

[84] S. Pogatscher, H. Antrekowitsch, M. Werinos, F. Moszner, S. S. A. Gerstl, M. F. Francis, W. A. Curtin, J. F. Löffler, and P. J. Uggowitzer. Diffusion on Demand to Control Precipitation Aging: Application to Al-Mg-Si Alloys. *Phys. Rev. Lett.*, 112:225701, 2014.

[85] MI Mendelev, S Han, DJ Srolovitz, GJ Ackland, DY Sun, and M Asta. Development of new interatomic potentials appropriate for crystalline and liquid iron. *Philosophical magazine*, 83(35):3977–3994, 2003.

[86] Giovanni Bonny, Roberto C Pasianot, Nicolas Castin, and Lorenzo Malerba. Ternary fe–cu–ni many-body potential to model reactor pressure vessel steels: First validation by simulated thermal annealing. *Philosophical magazine*, 89(34-36):3531–3546, 2009.

[87] Cemal Engin, Luis Sandoval, and Herbert M Urbassek. Characterization of fe potentials with respect to the stability of the bcc and fcc phase. *Modelling and Simulation in Materials Science and Engineering*, 16(3):035005, 2008.

[88] Stephen M Foiles. Evaluation of harmonic methods for calculating the free energy of defects in solids. *Physical Review B*, 49(21):14930, 1994.

[89] Michele Ceriotti, Joshua More, and David E Manolopoulos. i-pi: A python interface for ab initio path integral molecular dynamics simulations. *Computer Physics Communications*, 185(3):1019–1026, 2014.

[90] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[91] Steve Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comp. Phys.*, 117:1–19, 1995.

[92] Davide Branduardi, Giovanni Bussi, and Michele Parrinello. Metadynamics with adaptive gaussians. *Journal of Chemical Theory and Computation*, 8(7):2247–2254, 2012.

[93] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, 2014.

[94] Pablo M Piaggi, Omar Valsson, and Michele Parrinello. Enhancing entropy and enthalpy fluctuations to drive crystallization in atomistic simulations. *Physical Review Letters*, 119(1):015701, 2017.

[95] Paul J Steinhardt, David R Nelson, and Marco Ronchetti. Bond-orientational order in liquids and glasses. *Phys. Rev. B*, 28:784–805, 1983.

[96] Stefano Angioletti-Uberti, Michele Ceriotti, Peter D Lee, and Mike W Finnis. Solid-liquid interface free energy through metadynamics simulations. *Physical Review B*, 81(12):125416, 2010.

[97] Michele Ceriotti, Gareth A Tribello, and Michele Parrinello. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences*, 108(32):13023–13028, 2011.

[98] Michele Ceriotti, Gareth A Tribello, and Michele Parrinello. From the Cover: Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. USA*, 108:13023–8, 2011.

[99] JJ Hoyt, Mark Asta, and Alain Karma. Method for computing the anisotropy of the solid-liquid interfacial free energy. *Physical review letters*, 86(24):5530, 2001.

[100] Chandler A Becker, DL Olmsted, M Asta, JJ Hoyt, and SM Foiles. Atomistic simulations of crystal-melt interfaces in a model binary alloy: Interfacial free energies, adsorption coefficients, and excess entropy. *Physical Review B*, 79(5):054109, 2009.

[101] Jeremy Q Broughton and George H Gilmer. Molecular dynamics investigation of the crystal–fluid interface. vi. excess surface free energies of crystal–liquid systems. *The Journal of chemical physics*, 84(10):5759–5768, 1986.

[102] Ruslan L Davidchack and Brian B Laird. Direct calculation of the hard-sphere crystal/melt interfacial free energy. *Physical review letters*, 85(22):4751, 2000.

[103] Ruslan L Davidchack and Brian B Laird. Direct calculation of the crystal–melt interfacial free energies for continuous potentials: Application to the lennard-jones system. *The Journal of chemical physics*, 118(16):7651–7657, 2003.

[104] Richard C Tolman. Consideration of the gibbs theory of surface tension. *The journal of chemical physics*, 16(8):758–774, 1948.

[105] A Cacciuto, S Auer, and D Frenkel. Solid–liquid interfacial free energy of small colloidal hard-sphere crystals. *The Journal of chemical physics*, 119(14):7467–7470, 2003.

[106] Bingqing Cheng, Gareth A. Tribello, and Michele Ceriotti. The gibbs free energy of homogeneous nucleation: From atomistic nuclei to the planar limit. *The Journal of Chemical Physics*, 147(10):104707, 2017.

[107] Stefano Angioletti-Uberti, Michele Ceriotti, Peter D Lee, and Mike W Finnis. Solid-liquid interface free energy through metadynamics simulations. *Phys. Rev. B*, 81:125416, 2010.

[108] Ronald Benjamin and Jürgen Horbach. Crystal-liquid interfacial free energy via thermodynamic integration. *The Journal of chemical physics*, 141(4):044715, 2014.

[109] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical review letters*, 100(2):020603, 2008.

[110] S Angioletti-Uberti. The solid-liquid interface free-energy of Pb: comparison of theory and experiments. *J. Phys. Cond. Matt.*, 23:435008, 2011.

[111] G.M. M Torrie and J.P. P Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comp. Phys.*, 23:187–199, 1977.

[112] Pratyush Tiwary and Michele Parrinello. A Time-Independent Free Energy Estimator for Metadynamics. *J. Phys. Chem.. B*, 2014.

[113] Ulf R Pedersen, Felix Hummel, Georg Kresse, Gerhard Kahl, and Christoph Dellago. Computing gibbs free energy differences by interface pinning. *Physical Review B*, 88(9):094101, 2013.

[114] Frans Spaepen. Homogeneous nucleation and the temperature dependence of the crystal-melt interfacial tension. *Solid State Phys.*, 47:1–1, 1994.

[115] WA Curtin. Density-functional theory of the solid-liquid interface. *Physical review letters*, 59(11):1228, 1987.

[116] Long-Qing Chen. Phase-field models for microstructure evolution. *Annual review of materials research*, 32(1):113–140, 2002.

[117] M. E. Gurtin and P. W. Voorhees. The thermodynamics of evolving interfaces far from equilibrium. *Acta Materialia*, 44:235–247, 1996.

[118] Amit Samanta, Mark E Tuckerman, T.-Q. Yu, and W. E. Microscopic mechanisms of equilibrium melting of a solid. *Science*, 346:729–732, 2014.

[119] Matteo Salvalaglio, Thomas Vetter, Federico Giberti, Marco Mazzotti, and Michele Parrinello. Uncovering Molecular Details of Urea Crystal Growth in the Presence of Additives. *J. Chem. Am. Soc.*, 134:17221–17233, 2012.

[120] Rustam Z Khaliullin, Hagai Eshet, Thomas D Kühne, Jörg Behler, and Michele Parrinello. Nucleation mechanism for the direct graphite-to-diamond phase transition. *Nature materials*, 10:693–7, 2011.

[121] Seunghwa Ryu and Wei Cai. Validity of classical nucleation theory for ising models. *Physical Review E*, 81(3):030601, 2010.

[122] VA Shneidman, KA Jackson, and KM Beatty. On the applicability of the classical nucleation theory in an ising system. *The Journal of chemical physics*, 111(15):6932–6941, 1999.

[123] Benjamin J Block, Subir K Das, Martin Oettel, Peter Virnau, and Kurt Binder. Curvature dependence of surface free energy of liquid drops and bubbles: A simulation study. *The Journal of chemical physics*, 133(15):154702, 2010.

[124] Bingqing Cheng and Michele Ceriotti. Bridging the gap between atomistic and macroscopic models of homogeneous nucleation. *The Journal of Chemical Physics*, 146(3):034106, 2017.

[125] DP Adams, SM Yalisove, and DJ Eaglesham. Interfacial and surface energetics of cosi2. *Journal of applied physics*, 76(9):5190–5194, 1994.

[126] Yi An Lei, Tikhon Bykov, Soohaeng Yoo, and Xiao Cheng Zeng. The tolman length: Is it positive or negative? *Journal of the American Chemical Society*, 127(44):15346–15347, 2005.

[127] Edgar M Blokhuis and Joris Kuipers. Thermodynamic expressions for the tolman length. *The journal of chemical physics*, 124(7):074701, 2006.

[128] Alan E van Giessen and Edgar M Blokhuis. Direct determination of the tolman length from the bulk pressures of liquid drops via molecular dynamics simulations. *The Journal of chemical physics*, 131(16):164705, 2009.

[129] Richard C Tolman. The effect of droplet size on surface tension. *The journal of chemical physics*, 17(3):333–337, 1949.

[130] Peter W Voorhees. The theory of ostwald ripening. *Journal of Statistical Physics*, 38(1-2):231–252, 1985.

[131] Frank P Buff. The spherical interface. i. thermodynamics. *The Journal of Chemical Physics*, 19(12):1591–1594, 1951.

[132] AE Van Giessen and EM Blokhuis. Determination of curvature corrections to the surface tension of a liquid–vapor interface through molecular dynamics simulations. *The Journal of chemical physics*, 116(1):302–310, 2002.

[133] MA Anisimov. Divergence of tolman's length for a droplet near the critical point. *Physical review letters*, 98(3):035702, 2007.

[134] JH Irving and John G Kirkwood. The statistical mechanical theory of transport processes. iv. the equations of hydrodynamics. *The Journal of chemical physics*, 18(6):817–829, 1950.

[135] John G Kirkwood and Frank P Buff. The statistical mechanical theory of surface tension. *The Journal of Chemical Physics*, 17(3):338–343, 1949.

[136] Tatyana Zykova-Timan, Davide Ceresoli, Ugo Tartaglino, and Erio Tosatti. Why are alkali halide surfaces not wetted by their own melt? *Physical review letters*, 94(17):176105, 2005.

[137] JS Rowlinson. Symmetry of the pressure tensor in a nonuniform fluid. *Physical review letters*, 67(3):406, 1991.

[138] Edgar M Blokhuis and Dick Bedeaux. Pressure tensor of a spherical interface. *The Journal of chemical physics*, 97(5):3576–3586, 1992.

[139] Dick Bedeaux and John D Weeks. Correlation functions in the capillary wave model of the liquid–vapor interface. *The Journal of chemical physics*, 82(2):972–979, 1985.

[140] Vladimir Privman. Fluctuating interfaces, surface tension, and capillary waves: an introduction. *International Journal of Modern Physics C*, 3(05):857–877, 1992.

[141] Ruslan L Davidchack, James R Morris, and Brian B Laird. The anisotropic hard-sphere crystal-melt interfacial free energy from fluctuations. *The Journal of chemical physics*, 125(9):094710, 2006.

[142] David T Limmer and David Chandler. Premelting, fluctuations, and coarse-graining of water-ice interfaces. *The Journal of chemical physics*, 141(18):18C505, 2014.

[143] Jorge R Espinosa, Carlos Vega, and Eduardo Sanz. Ice–water interfacial free energy for the tip4p, tip4p/2005, tip4p/ice, and mw models as obtained from the mold integration technique. *The Journal of Physical Chemistry C*, 120(15):8068–8075, 2016.

[144] Michael Ambler, Bart Vorselaars, Michael P Allen, and David Quigley. Solid–liquid interfacial free energy of ice ih, ice ic, and ice 0 within a mono-atomic model of water via the capillary wave method. *The Journal of Chemical Physics*, 146(7):074701, 2017.

[145] Edoardo Baldi, Michele Ceriotti, and Gareth A Tribello. Extracting the interfacial free energy and anisotropy from a smooth fluctuating dividing surface. *Journal of Physics: Condensed Matter*, 29(44):445001, 2017.

[146] Wolfgang Lechner and Christoph Dellago. Accurate determination of crystal structures based on averaged local bond order parameters. *The Journal of chemical physics*, 129(11):114707, 2008.

[147] Gareth A Tribello, Michele Ceriotti, and Michele Parrinello. Using sketch-map coordinates to analyze and bias molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 109(14):5196–5201, 2012.

[148] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[149] Carlos Vega, Jose LF Abascal, MM Conde, and JL Aragones. What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday discussions*, 141:251–276, 2009.

[150] Carlos Vega and Jose LF Abascal. Simulating water with rigid non-polarizable models: a general perspective. *Physical Chemistry Chemical Physics*, 13(44):19663–19688, 2011.

[151] Francesco Paesani and Gregory A Voth. The properties of water: Insights from quantum simulations. *The Journal of Physical Chemistry B*, 113(17):5702–5719, 2009.

[152] Sandeep K Reddy, Shelby C Straight, Pushp Bajaj, C Huy Pham, Marc Riera, Daniel R Moberg, Miguel A Morales, Chris Knight, Andreas W Götz, and Francesco Paesani. On the accuracy of the mb-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *The Journal of chemical physics*, 145(19):194504, 2016.

[153] Tobias Morawietz, Andreas Singraber, Christoph Dellago, and Jörg Behler. How van der waals interactions determine the unique properties of water. *Proceedings of the National Academy of Sciences*, 113(30):8368–8373, 2016.

[154] Tamsin L Malkin, Benjamin J Murray, Andrey V Brukhno, Jamshed Anwar, and Christoph G Salzmann. Structure of ice crystallized from supercooled water. *Proceedings of the National Academy of Sciences*, 109(4):1041–1045, 2012.

[155] Valeria Molinero and Emily B Moore. Water modeled as an intermediate element between carbon and silicon. *The Journal of Physical Chemistry B*, 113(13):4008–4016, 2008.

[156] Arpa Hudait, Siwei Qiu, Laura Lupi, and Valeria Molinero. Free energy contributions and structural characterization of stacking disordered ices. *Physical Chemistry Chemical Physics*, 18(14):9544–9553, 2016.

[157] Thomas Koop, Beiping Luo, Athanasios Tsias, and Thomas Peter. Water activity as the determinant for homogeneous ice nucleation in aqueous solutions. *Nature*, 406(6796):611, 2000.

[158] Marc Souaille and Benoıt Roux. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer physics communications*, 135(1):40–57, 2001.

[159] JR Espinosa, E Sanz, C Valeriani, and C Vega. Homogeneous ice nucleation evaluated for several water models. *The Journal of Chemical Physics*, 141(18):18C529, 2014.

[160] Tianshu Li, Davide Donadio, Giovanna Russo, and Giulia Galli. Homogeneous ice nucleation from supercooled water. *Physical Chemistry Chemical Physics*, 13(44):19807–19813, 2011.

[161] D Quigley. Communication: Thermodynamics of stacking disorder in ice nuclei, 2014.

[162] Ulf R Pedersen, Felix Hummel, and Christoph Dellago. Computing the crystal growth rate by the interface pinning method. *The Journal of Chemical Physics*, 142(4):044104, 2015.

[163] P Taborek. Nucleation in emulsified supercooled water. *Physical Review B*, 32(9):5902, 1985.

[164] Paul J DeMott and David C Rogers. Freezing nucleation rates of dilute solution droplets measured between- 30 and- 40 c in laboratory simulations of natural clouds. *Journal of the atmospheric sciences*, 47(9):1056–1064, 1990.

[165] HR Pruppacher. A new look at homogeneous ice nucleation in supercooled water drops. *Journal of the atmospheric sciences*, 52(11):1924–1933, 1995.

[166] Nils ER Zimmermann, Bart Vorselaars, Jorge R Espinosa, David Quigley, William R Smith, Eduardo Sanz, Carlos Vega, and Baron Peters. Nacl nucleation from brine in seeded simulations: Sources of uncertainty in rate estimates. *The Journal of Chemical Physics*, 148(22):222838, 2018.

[167] Aleks Reinhardt and Jonathan PK Doye. Free energy landscapes for homogeneous nucleation of ice for a monatomic water model. *The Journal of chemical physics*, 136(5):054501, 2012.

[168] Laura Filion, Michiel Hermes, Ran Ni, and Marjolein Dijkstra. Crystal nucleation of hard spheres using molecular dynamics, umbrella sampling, and forward flux sampling: A comparison of simulation techniques. *The Journal of chemical physics*, 133(24):244115, 2010.

[169] László Gránásy, Tamás Pusztai, and Peter F James. Interfacial properties deduced from nucleation experiments: A cahn–hilliard analysis. *The Journal of chemical physics*, 117(13):6157–6168, 2002.

[170] William B Hillig. Measurement of interfacial free energy for ice/water system. *Journal of crystal growth*, 183(3):463–468, 1998.

[171] Bingqing Cheng, Jorg Behler, and Michele Ceriotti. Nuclear quantum effects in water at the triple point: Using theory as a link between experiments. *The journal of physical chemistry letters*, 7(12):2210–2215, 2016.

[172] Edgar A Engel, Bartomeu Monserrat, and Richard J Needs. Anharmonic nuclear motion and the relative stability of hexagonal and cubic ice. *Physical Review X*, 5(2):021033, 2015.

[173] T. Bartels-Rausch, V. Bergeron, J. H. E. Cartwright, R. Escribano, J. L. Finney, H. Grothe, P. J. Gutiérrez, J. Haapala, W. F. Kuhs, J. B. C. Pettersson, S. D. Price, C. I. Sainz-Díaz, D. J. Stokes, G. Strazzulla, E. S. Thomson, H. Trinks, and N. Uras-Aytemiz. Ice structures, patterns, and processes: A view across the ice-fields. *Review of Modern Physics*, 84:885–944, 2012.

[174] B. J. Murray, D. A. Knopf, and A. K. Bertram. The formation of cubic ice under conditions relevant to the Earth's atmosphere. *Nature*, 434:202–205, 2005.

[175] W. F. Kuhs, C. Sippel, A. Falenty, and T. C. Hansen. Extent and relevance of stacking disorder in “ice Ic”. *Proceedings of the National Academy of Sciences*, 109:21259–21264, 2012.

[176] M. A. Morales, J. R. Gergely, J. McMinis, J. M. McMahon, J. Kim, and D. M. Ceperley. Quantum Monte Carlo Benchmark of Exchange-Correlation Functionals for Bulk Water. *Journal of Chemical Theory and Computation*, 10:2355, 2014.

[177] J.; Galli G.; Gygi F. Zhang, C.; Wu. Structural and Vibrational Properties of Liquid Water from Van Der Waals Density Functionals. *Journal of Chemical Theory and Computation.*, 7:3054, 2011.

[178] B. Santra, A. Michaelides, and M. Scheffler. Coupled Cluster Benchmarks of Water Monomers and Dimers Extracted from Density-Functional Theory Liquid Water: The Importance of Monomer Deformations. *Journal of Chemical Physics*, 131:124509, 2009.

[179] T. E. Markland and M. Ceriotti. Nuclear quantum effects enter the mainstream. *Nature Reviews Chemistry*, 2:0109, 2018.

[180] C. P. Herrero and R. Ramírez. Configurational entropy of hydrogen-disordered ice polymorphs. *Journal of Chemical Physics*, 140:234502, 2014.

[181] Z. Raza, D. Alfè, C. G. Salzmann, J. Klimeš, A. Michaelides, and B. Slater. Proton ordering in cubic ice and hexagonal ice; a potential new ice phase – XIc. *Physical Chemistry Chemical Physics*, 13:19788–19795, 2011.

[182] J. Lekner. Energetics of hydrogen ordering in ice. *Physica B: Condensed Matter*, 252:149–159, 1998.

[183] R. Ramírez and C. P. Herrero. Quantum path integral simulation of isotope effects in the melting temperature of ice Ih. *Journal of Chemical Physics*, 133:144511, 2010.

[184] S. Habershon, T. E. Markland, and D. E. Manolopoulos. Competing quantum effects in the dynamics of a flexible water model. *Journal of Chemical Physics*, 131:024501, 2009.

[185] B. Pamuk, J. M. Soler, R. Ramírez, C. P. Herrero, P. W. Stephens, P. B. Allen, and M. V. Fernandez-Serra. Anomalous Nuclear Quantum Effects in Ice. *Physical Review Letters*, 108:193003, 2012.

[186] R. Ramírez, N. Neuerburg, M.-V. Fernández-Serra, and C. P. Herrero. Quasi-harmonic approximation of thermodynamic properties of ice Ih, II, and III. *Journal of Chemical Physics*, 137:044502, 2012.

[187] E. A. Engel, Y. Li, and R. J. Needs. First-principles momentum distributions and vibrationally corrected permittivities of hexagonal and cubic ice. *Physical Review B*, 97:054312, 2018.

[188] Y. Zhang and W. Yang. Comment on “Generalized Gradient Approximation Made Simple”. *Physical Review Letters*, 80:890, 1998.

[189] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *Journal of Chemical Physics*, 110:6158, 1999.

[190] L. Goerigk and S. Grimme. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Physical Chemistry Chemical Physics*, 13:6670, 2011.

[191] S. Grimme, J. Antony, S. Ehrlich, and S. Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements H-Pu. *Journal of Chemical Physics*, 132:154104, 2010.

[192] Lars Goerigk and Stefan Grimme. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Physical Chemistry Chemical Physics*, 13(14):6670–6688, 2011.

[193] O. Marsalek and T. E. Markland. Quantum Dynamics and Spectroscopy of Ab Initio Liquid Water: The Interplay of Nuclear and Electronic Quantum Effects. *J. Phys. Chem. Lett.*, 8:1545, 2017.

[194] Supplemental material.

[195] Markus Macher, Jiří Klimeš, Cesare Franchini, and Georg Kresse. The random phase approximation applied to ice. *The Journal of Chemical Physics*, 140(8):084502, 2014.

[196] Blazej Grabowski, Lars Ismer, Tilmann Hickel, and Jörg Neugebauer. Ab initio up to the melting point: Anharmonicity and vacancies in aluminum. *Physical Review B*, 79(13):134106, 2009.

[197] Albert Glensk, Blazej Grabowski, Tilmann Hickel, and Jörg Neugebauer. Breakdown of the arrhenius law in describing vacancy formation energies: the importance of local anharmonicity revealed by ab initio thermodynamics. *Physical Review X*, 4(1):011018, 2014.

[198] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters*, 98(14):146401, 2007.

[199] Jörg Behler. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem. Int. Ed.*, 56(42):12828, 2017.

[200] Jörg Behler. *RuNNer – A Neural Network Code for High-Dimensional Neural Network Potentials*. Universität Göttingen, 2018.

[201] G. Lippert, J. Hutter, and M. Parrinello. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theoretical Chemistry Accounts*, 103:124, 1999.

[202] DE Hare and CM Sorensen. The density of supercooled water. ii. bulk samples cooled to the homogeneous nucleation limit. *The Journal of chemical physics*, 87(8):4840–4845, 1987.

[203] Lawrie B Skinner, CJ Benmore, Joerg C Neuefeind, and John B Parise. The structure of water around the compressibility minimum. *The Journal of chemical physics*, 141(21):214507, 2014.

[204] S. Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics*, 117:1, 1995.

[205] Andreas Singraber, Jörg Behler, and Christoph Dellago. A library-based lammps implementation of high-dimensional neural network potentials. *submitted to the Journal of Chemical Theory and Computation*, 2018.

[206] PW Bridgman. The pressure-volume-temperature relations of the liquid, and the phase diagram of heavy water. *The Journal of Chemical Physics*, 3(10):597–605, 1935.

[207] K. Röttger, A. Endriss, J. Ihringer, S. Doyle, and W. F. Kuhs. Lattice constants and thermal expansion of H_2O and D_2O ice Ih between 10 and 265 K. *Acta Crystallographica B*, 50:644, 1994.

[208] M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland. Nuclear Quantum Effects in Water and Aqueous Systems: Experiment, Theory, and Current Challenges. *Chemical Reviews*, 116:7529, 2016.

[209] Masakazu Matsumoto, Takuma Yagasaki, and Hideki Tanaka. Genice: Hydrogen-disordered ice generator. *Journal of computational chemistry*, 39(1):61–64, 2018.

[210] C. Drechsel-Grau and D. Marx. Quantum Simulation of Collective Proton Tunneling in Hexagonal Ice Crystals. *Physical Review Letters*, 112:148302, 2014.

[211] Michele Ceriotti and Thomas E Markland. Efficient methods and practical guidelines for simulating isotope effects. *The Journal of chemical physics*, 138(1):014112, 2013.

[212] Bingqing Cheng, Anthony T Paxton, and Michele Ceriotti. Hydrogen diffusion and trapping in α-iron: The role of quantum and anharmonic fluctuations. *Physical review letters*, 120(22):225901, 2018.

[213] Mariana Rossi, Wei Fang, and Angelos Michaelides. Stability of complex biomolecular structures: van der waals, hydrogen bond cooperativity, and nuclear quantum effects. *The journal of physical chemistry letters*, 6(21):4233–4238, 2015.

Bingqing Cheng

Email bingqing.cheng@epfl.ch
Phone number +41 78 818 88 29
Address MXG 317, Station 12, Lausanne, CH-1015, Switzerland
Scholar page https://scholar.google.ch/citations?user=s5ZqEskAAAAJ&h

Education

Sep 2014 –	**École Polytechnique Fédérale de Lausanne** Doctor of Philosophy, Materials Science
Sep 2012 – Aug 2014	**The University of Hong Kong** Master of Philosophy, Materials Science
Sep 2008 – Jun 2012	**The University of Hong Kong & Shanghai Jiao Tong University** (Dual-)Bachelor degree, Mechanical Engineering

Employment and Research Experience

Sep 2014 –	Research assistant in the Laboratory of Computational Science and Modelling, EPFL (**Supervised by Prof. Michele Ceriotti**).
Mar 2018 – Apr 2018	Visiting scholar in the group of Prof. Jörg Behler in Georg-August-Universität Göttingen.
Apr 2018 – May 2018	Visiting scholar in the group of Prof. Christoph Dellago in the University of Vienna.
Sep 2012 – Jul 2014	Research assistant in the Department of Mechanical Engineering, the University of Hong Kong (**Supervised by Prof. A. H. W. Ngan**).

Teaching and Supervision Experience

Mar 2017 –	Co-supervised a MARVEL INSPIRE Master Project student in EPFL.
Sep 2015 – Dec 2017	Teaching Assistant for the courses "Statistical Mechanics" and "Surface Science" in EPFL.
Jun 2016	Tutor for the CECAM School "Path Integral Quantum Mechanics".
Sep 2012 – Jul 2014	Teaching assistant for the course "Introduction to Mechanical Engineering" and lab demonstrator in HKU.
Sep 2012 – Aug 2014	Co-supervised two undergraduate Final Year Project students in HKU.

Awards & Grants

1. **Early Postdoc.Mobility fellowship (Swiss National Science Foundation, 2018)**
2. The JCP poster prize (CECAM workshop on ice nucleation, 2018)
3. CSCS Production Project S787 (99,100 node hours on Piz Daint, co-PI with Michele Ceriotti). This grant already led to two publications so far.
4. International Visitor Program, ICASEC, Georg-August-Universität (EUR 500)
5. **Prize for the Best Oral Presentation (Doctoral Program in Materials Science and Engineering, EPFL, 2017).** One prize was awarded among all PhD students in the department.
6. Poster Contest Award (Marvel Review and Retreat,2017)
7. **Award for Outstanding Research Postgraduate Student (HKU, 2013-2014).** Ten awards are made each year among all the research postgraduate students in the university.
8. Mechanical Engineering Outstanding Thesis Award (HKU, 2013-2014)

9. Mechanical Engineering Outstanding Research Postgraduate Award (HKU, 2012-2013)
10. Postgraduate scholarship (HKU, 2012-2014)
11. Fong's Project Prize in Mechanical Engineering (HKU, 2012)

Computer Skills

Programming	Proficient in Fortran, C++, Python and Mathematica. I have contributed to the development of path integral molecular dynamics code i-PI.
Simulation	Proficient user of molecular dynamics package LAMMPS, enhanced sampling plugin PLUMED and i-PI.

Scientific Activities

I have independently written a grant for the allocation of computer time (99,100 node hours) at the Swiss National Supercomputing Center (CSCS).

I have refereed for journals including Physica E, Computational Materials Science, and Electronic Materials Letters.

Journal Publications

(*** Corresponding author**)

1. Bingqing Cheng*, Edgar A Engel, Jörg Behler, Christoph Dellago, Michele Ceriotti. (**2018**) *ab initio thermodynamics of liquid and solid water*. Submitted.
2. Bingqing Cheng*, Christoph Dellago, Michele Ceriotti. (**2018**) *Theoretical prediction of the homogeneous ice nucleation rate: disentangling thermodynamics and kinetics*. Physical Chemistry Chemical Physics, DOI: 10.1039/C8CP04561E.
3. Venkat Kapil, Mariana Rossi, Ondrej Marsalek, Riccardo Petraglia, Yair Litman, Thomas Spura, Bingqing Cheng, Alice Cuzzocrea, Robert H Meißner, David M Wilkins, Przemys law Juda, Sebastien P Bienvenue, Wei Fang, Jan Kessler, Igor Poltavsky, Steven Vandenbrande, Jelle Wieme, Clemence Corminboeuf, Thomas D Kuhne, David E Manolopoulos, Thomas E Markland, Jeremy O Richardson, Alexandre Tkatchenko, Gareth A Tribello, Veronique Van Speybroeck, Michele Ceriotti. (**2018**) *i-PI V2: A Universal Force Engine for Advanced Molecular Simulations.* Computer Physics Communications, DOI:10.1016/j.cpc.2018.09.020.
4. Bingqing Cheng*, Michele Ceriotti. (**2018**) *Computing the Tolman length for solid-liquid interfaces*. The Journal of Chemical Physics, 148(23): 231102.
5. Bingqing Cheng, Anthony T Paxton, Michele Ceriotti. (**2018**) *Hydrogen diffusion and trapping in a-iron: the role of quantum and anharmonic fluctuations.* Physical Review Letters, 120(22): 225901.
6. Bingqing Cheng*, Michele Ceriotti. (**2018**) *Computing the absolute Gibbs free energy in atomistic simulations: applications to defects in solids*. Physical Review B, 97(5): 054102.
7. Bingqing Cheng*, Gareth A Tribello, Michele Ceriotti. (**2017**) *The Gibbs free energy of homogeneous nucleation: from atomistic nuclei to the planar limit.* The Journal of Chemical Physics, 147(10): 104707.
8. Bingqing Cheng*, Michele Ceriotti. (**2016**) *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation*. The Journal of Chemical Physics, 146(3): 034106.
9. Bingqing Cheng, Jörg Behler, Michele Ceriotti. (**2016**) *Nuclear Quantum Effects in Water at the Triple Point: Using Theory as a Link Between Experiments.* Journal of Physical Chemistry Letters, 7(12): 2210-2215.

10. Bingqing Cheng, Gareth A Tribello, Michele Ceriotti. (**2015**) *Solid-liquid interfacial free energy out of equilibrium*. Physical Review B, 92(18): 180102.
11. Peggy S S Leung, Hing Shun Leung, Bingqing Cheng, Alfonso H W Ngan. (**2015**) *Size dependence of yield strength simulated by a dislocation-density function dynamics approach*. Modelling and Simulation in Materials Science and Engineering, 23(3): 035001.
12. Bingqing Cheng, Michele Ceriotti. (**2014**) *Direct path integral estimators for isotope fractionation ratios*. The Journal of Chemical Physics, 141(24): 244112.
13. Bingqing Cheng, Hing Shun Leung, Alfonso H W Ngan. (**2014**) *Strength of metals under vibrations - Dislocation-density-function dynamics simulations*. Philosophical Magazine, 95(16-18):1-21.
14. Hing Shun Leung, Peggy S S Leung, Bingqing Cheng, Alfonso H W Ngan. (**2014**) *A New Dislocation-density-function Dynamics Scheme for Computational Crystal Plasticity by Explicit Consideration of Dislocation Elastic Interactions*. International Journal of Plasticity, 67: 1-25.
15. Bingqing Cheng*, Alfonso H W Ngan. (**2013)** *Crystal plasticity of Cu nanocrystals during collision*. Materials Science and Engineering, 585:326-334.
16. Bingqing Cheng, Alfonso H W Ngan. (**2013**) *The sintering and densification behaviour of many copper nanoparticles: A molecular dynamics study*. Computational Materials Science, 74:1-11.
17. Bingqing Cheng*, Alfonso H W Ngan. (**2013**) *Thermally induced solid-solid structural transition of copper nanoparticles through direct geometrical conversion*. The Journal of Chemical Physics, 138(16):164314.
18. Bingqing Cheng, Alfonso H W Ngan. (**2013**) *The crystal structures of sintered copper nanoparticles: A molecular dynamics study*. International Journal of Plasticity, 47: 65-78.

(Up-to-date citation data and bibliographic indexes can be obtained at the Google scholar page: https://scholar.google.ch/citations?user=s5ZqEskAAAAJ)

(Private full texts can be found on www.researchgate.net/profile/Bingqing_Cheng)

Talks

1. Atomistic Simulation Center Seminar 2018 (Queen's University Belfast). **Invited seminar**: *ab initio thermodynamics of liquid and solid water.*
2. Theory of Condensed Matter Seminar 2018 (University of Cambridge). **Invited seminar**: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
3. Computational physics group seminar 2018 (University of Vienna). **Seminar**: *Predicting homogeneous nucleation rate from atomistic simulations.*
4. Condensed Matter Seminar 2018 (SISSA, Trieste). **Invited seminar**: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
5. Group Seminar 2018 (Georg-August-Universität Göttingen). **Seminar**: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
6. SFB ViCom Conference 2018 (Vienna). **Selected talk**: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
7. DPG Conference 2018 (Berlin). **Contributed talk**: *Computing the absolute Gibbs free energy in atomistic simulations: applications to defects in solids.*
8. MARVEL Junior Seminar 2017 (Lausanne). **Seminar**: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
9. EPFL EDMX Materials Research Day 2017 (Lausanne). **Selected talk**: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
10. DPG Conference 2017 (Dresden, Germany). **Contributed talk**: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*

Posters

1. CC2AI 2018 (Lausanne). Poster: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
2. Recent Advances in Modelling Rare Events RARE 2017 (Agra, India). Poster: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
3. Marvel Review and Retreat 2017 (Lausanne). Poster: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
4. NSF/CECAM school on Computational Materials Science: From Basics to Applications 2017 (Lausanne). Poster: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
5. CECAM School "Building links between experiments and computer simulations of crystallisation" 2017 (Lausanne). Poster: *Bridging the gap between atomistic and macroscopic models of homogeneous nucleation.*
6. VES School 2017 (Lugano). Poster: *Solid-liquid interface free energy out-of-equilibrium.*
7. CECAM School "Path Integral Quantum Dynamics" 2016 (Lausanne). Poster: *Direct path integral estimators for isotope fractionation ratios.*
8. Psi-K Conference 2015 (San Sebastian, Spain). Poster: *Solid-liquid interface free energy out-of-equilibrium.*
9. CECAM School "Computational Trends in Solvation and Transport in Liquids" 2015 (Juelich, Germany). Poster: *Direct path integral estimators for isotope fractionation ratios.*
10. PASC 2015 Conference (Zurich). Poster: *Direct path integral estimators for isotope fractionation ratios.*
11. GRC on Nano-Mechanical Interfaces 2013 (Hong Kong). Poster: *The mechanism in the sintering of Cu nanoparticles.*
12. IC4N Conference 2013 (Corfu, Greece). Poster: *Novel dislocation-free deformation mechanism in Cu nanoparticles.*

Ce document a été imprimé au Centre d'impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.