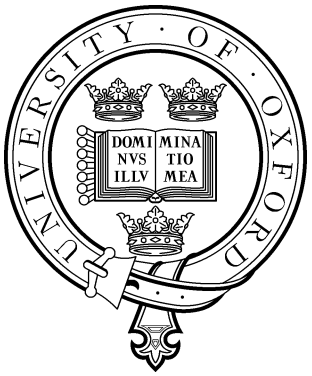


Computer Simulation of the Homogeneous Nucleation of Ice

Aleks Reinhardt
St Catherine's College
University of Oxford

Supervisor: Dr Jonathan P. K. Doye

A thesis submitted for the degree of
Doctor of Philosophy in Physical and Theoretical Chemistry

Hilary Term 2013

Abstract

Computer simulation of the homogeneous nucleation of ice
Aleks Reinhardt, St Catherine's College
Submitted for the degree of Doctor of Philosophy in Physical and Theoretical Chemistry
Hilary Term 2013

In this work, we wish to determine the free energy landscape and the nucleation rate associated with the process of homogeneous ice nucleation. To do this, we simulate the homogeneous nucleation of ice with the mW monatomic model of water and with all-atom models of water using primarily the umbrella sampling rare event method. We find that the use of the mW model of water, which has simpler dynamics compared to all-atom models of water, but is nevertheless surprisingly good at reproducing experimental data, results in very reasonable agreement with classical nucleation theory, in contrast to some previous simulations of homogeneous ice nucleation. We suggest that previous simulations did not observe the lowest free energy pathway in order parameter space because of their use of global order parameters, leading to a deviation from classical nucleation theory predictions.

Whilst monatomic water can nucleate reasonably quickly, all-atom models of water are considerably more difficult to simulate, primarily because of their slow dynamics of ice growth and the fact that standard order parameters do not work well in driving nucleation when such models are being used. In this thesis, we describe a local, rotationally invariant order parameter that is capable of growing ice homogeneously in a biassed simulation without the unnatural effects introduced by global order parameters, and without leading to non-physical chain-like growth of 'ice' clusters that results from a naïve implementation of the standard Steinhardt–Ten Wolde order parameter. We have successfully used this order parameter to force the growth of ice clusters in simulations of all-atom models of water.

However, although ice growth can be achieved, equilibrating simulations with all-atom models of water is extremely difficult. We describe several approaches to speeding up the equilibration in all-atom models of water to enable the computation of free energy profiles for homogeneous ice nucleation.

Contents

Acknowledgements

I should first like to thank my supervisor, Dr Jonathan Doye, for his invaluable guidance throughout my DPhil project. I am also very grateful to the other members and former members of our research group for their help and ideas, and especially to Drs Alexander Williamson and Flavio Romano for our innumerable useful discussions. Furthermore, Dr Ard Louis's and Dr Thomas Ouldridge's input was very helpful when we addressed the benefits and drawbacks of various simulation methods during the regular meetings of our extended group.

During the course of my DPhil project, I was involved in the preparation and publication of several papers, and I wish to thank the co-authors of these publications for their input and for the time they invested in our collaborations.

I should like to thank Dr Eva Noya and Professor Carlos Vega for their help in setting up and benchmarking our TIP4P/2005 code, particularly the Ewald summation routines, as well as supplying us with some initial configurations of cubic and hexagonal ice. I am also grateful to Dr Chantal Valeriani for her helpful suggestions at the start of the project.

Financial support from the Engineering and Physical Sciences Research Council, which funded me by means of a Doctoral Training Grant, is acknowledged with gratitude.

Finally, thanks are also due to my family for their unwavering support through the years.

Publications

This thesis is based in part on the following articles prepared during the course of the research project.

1. A Reinhardt, A J Williamson, J P K Doye, J Carrete, L M Varela and A A Louis, 'Re-entrant phase behavior for systems with competition between phase separation and self-assembly', *J Chem Phys*, **134**, 104905, 2011. (Also selected for the *Virtual Journal of Nanoscale Science and Technology*, **23**(12), 2011.)
   The following copyright notice applies: *Copyright 2011 American Institute of Physics. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the American Institute of Physics. The following article appeared in The Journal of Chemical Physics and may be found at http://dx.doi.org/10.1063/1.3557059.*
   The full article is available on arXiv:1010.4676 [cond-mat.soft].

2. A Reinhardt and J P K Doye, 'Free energy landscapes for homogeneous nucleation of ice for a monatomic water model', *J Chem Phys*, **136**, 054501, 2012.
   The following copyright notice applies: *Copyright 2012 American Institute of Physics. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the American Institute of Physics. The following article appeared in The Journal of Chemical Physics and may be found at http://dx.doi.org/10.1063/1.3677192.*
   The full article is available on arXiv:1109.6837 [cond-mat.stat-mech].

3. A Reinhardt, J P K Doye, E G Noya and C Vega, 'Local order parameters for use in driving homogeneous ice nucleation with all-atom models of water', *J Chem Phys*, **137**, 194504, 2012.
   The following copyright notice applies: *Copyright 2012 American Institute of Physics. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the American Institute of Physics. The following article appeared in The Journal of Chemical Physics and may be found at http://dx.doi.org/10.1063/1.4766362.*
   The full article is available on arXiv:1208.6033 [cond-mat.stat-mech].

4. A Reinhardt, F Romano and J P K Doye, 'Computing phase diagrams for a quasicrystal-forming patchy-particle system', *Phys Rev Lett*, accepted for publication, 2013.
   The following copyright notice applies: *Copyright 2013 by the American Physical Society.*
   The full article is available on arXiv:1302.2592 [cond-mat.stat-mech].

Glossary of abbreviations and symbols

The following abbreviations are used in this thesis.

CNT	classical nucleation theory
fus	fusion
ice I_c	cubic ice I
ice I_h	hexagonal ice I
LJ	Lennard-Jones (potential)
MC	(Metropolis) Monte Carlo
MD	molecular dynamics
mW	monatomic water (potential)
NVdE	Nada–Van der Eerden (potential)
TIP*n*P	transferable intermolecular potential with *n* points
VdW	Van der Waals

The following notation for physical quantities is used (where not defined otherwise).

A	Helmholtz energy
a	area
α	MC transition probability
β	inverse temperature, $1/k_\mathrm{B}T$
D	diffusion coefficient
Δ	isobaric-isothermal partition function
δ	Dirac delta function
δ_{ij}	Kronecker delta function
ε	relative permittivity, LJ energy
ε_0	permittivity of free space
G	Gibbs energy
γ	interfacial free energy
H	enthalpy, hamiltonian
H	Heaviside step function
h	Planck constant
I	moment of inertia
J	nucleation rate
K	equilibrium constant, kinetic energy
k	(forward) rate constant
k'	reverse rate constant
k_B	Boltzmann constant
Λ	de Broglie wavelength
m	mass
μ	chemical potential
N	number of particles
N_A	Avogadro constant
n	chemical amount, number of particles in cluster
ω	angular velocity
P	probability
p	probability, pressure, linear momentum
ϕ	packing fraction
Ψ	stationary distribution
Q	canonical partition function, order parameter
q	order parameter

R	universal gas constant
R, r	radius, position
ρ	number density, mass density
S	entropy
σ	LJ lengthscale
σ_{pw}	patch width
t	time
T	thermodynamic temperature
τ	torque
Θ	MC acceptance probability
U	internal energy, potential energy
V	volume, umbrella weight potential
v	volume per particle, velocity
Z	Zeldovich factor

1 Introduction

It is well-known that substances cooled below their thermodynamic freezing point do not necessarily freeze, especially when they are very pure. Water droplets, in particular, can be cooled to approximately $-45\,^{\circ}\mathrm{C}$ at standard pressure without freezing.[5–7] Water in such a supercooled state is metastable with respect to ice, but cannot convert into the thermodynamically stable form because of a free energy barrier associated with the nucleation process. The addition of an external nucleation seed will cause the entire volume of supercooled liquid water to freeze rapidly, with the temperature rising up to the freezing point.[8]

Homogeneous nucleation is a kinetically disfavoured process; according to classical nucleation theory, a critical cluster must spontaneously form in the supercooled liquid before crystallisation can proceed. The free energy barrier to ice nucleation without an external nucleation seed, in the framework of classical nucleation theory, arises as a result of a competition between a favourable bulk free energy difference between the phases and the unfavourable formation of an interface between the phases. The presence of a free energy barrier to nucleation makes homogeneous nucleation a rare event, as a cluster of some critical size must spontaneously form before large-scale crystallisation can occur.[9–11] This means that the average time between nucleation events is many orders of magnitude greater than the duration of the event itself.

Even though most freezing events happen heterogeneously in practice, homogeneous nucleation is interesting not only as model system, but also because it is thought to play a major rôle in processes such as the formation of ice in cirrus clouds in the upper troposphere,

governing both global climate change and short-term weather,[12–18] and may even have implications for the safety of jet aviation fuel.[19] Indeed, it is thought to be the dominant mechanism of nucleation in many atmospheric processes.[20] Homogeneous nucleation is also important in biological systems, such as in the process of the cryopreservation of biological tissues,[21;22] including sperm cells,[23] and in freeze-avoiding insects.[24;25] On the other hand, it is sometimes important for water to remain a (supercooled) liquid and not to freeze, for example in the processes of electrification and rainfall in clouds.[26–29] The interplay between supercooled droplets and crystalline ice is thought strongly to influence the release of snow, rain and hail from clouds.[26] Understanding on a molecular level when and how water freezes has recently been identified as one of the ten major open questions in ice science.[30]

Homogeneous nucleation is thus not simply a theorist's idealisation of the world, and it is worth obtaining free energy landscapes and rates and trying to elucidate the detailed microscopic mechanism of the process. However, homogeneous ice nucleation is difficult to study both experimentally and computationally. There are several reasons why this apparently straightforward problem is demanding to simulate:

(a) As a rare event, nucleation is effectively impossible to simulate in a brute force manner: rare event techniques must be employed to drive the process.

(b) Most realistic water models involve long-ranged forces and are thus computationally expensive to use.

(c) It is unclear what reaction co-ordinate is the most sensible to use; various previous studies used order parameters that can lead to non-physical influences being exerted on distant molecules, and the nucleation pathways that have been determined in such studies appear not to be the lowest free energy pathways. Furthermore, even local order parameters, when used in biassed simulations, can be difficult to apply because of the similarity of the liquid and crystalline states.

(d) The natural dynamics of supercooled water and ice are extremely slow, presumably because of the strong hydrogen bonding present in the liquid phase. This makes equilibration very difficult to achieve on a computationally tractable timescale.

In this work, we simulate the homogeneous nucleation of ice both with a monatomic model of water and with all-atom models of water, predominantly using umbrella sampling[31;32] as our rare event method of choice. Our aim is, ultimately, to determine the free energy landscape and the rate of homogeneous ice nucleation, as well as to study the phase distribution of hexagonal and cubic ice during the homogeneous nucleation process.

When simulated with a simple monatomic water (mW) model,[33] homogeneous ice nucleation can be studied using rare event methods. Although the absence of explicit hydrogen atoms in this model makes the dynamics much simpler, and hence the crystallisation tractable, the model produces remarkably good agreement with experimental thermodynamic data,[33] and so insights into the thermodynamics of ice nucleation can be obtained. However, applying the same technique to all-atom models of water is not straightforward, primarily for two reasons: the standard Steinhardt–Ten Wolde order parameter[34–37] is not capable of driving the nucleation in such models, and the dynamics of ice growth are very slow.

We have designed a local, rotationally invariant order parameter that is capable of growing ice homogeneously in a biassed simulation without the unnatural effects caused by global order parameters and without leading to non-physical chain-like growth of 'ice' clusters that results from a naïve implementation of the standard Steinhardt–Ten Wolde order parameter. We have successfully used this order parameter to force the growth of ice clusters in simulations of all-atom models of water, such as TIP4P/2005.[38]

Nevertheless, such all-atom models exhibit extremely slow dynamics, and although forcing ice growth is possible, these simulations are computationally exceptionally demanding and take a very long time to equilibrate. We have tried to circumvent this difficulty in several

ways, such as by coupling the all-atom simulation to a series of simulations with potentials progressively changing to the monatomic water case, or harnessing the increased speed of molecular dynamics simulations, which allow for collective particle motion, when determining the free energy barrier of nucleation in a hybrid Monte Carlo[39] approach.

In this thesis, we look at the basic physical interpretation of homogeneous nucleation in chapter 2, and at simulation methods used to study it in chapter 3. We then consider the properties of water and the ways it is simulated with simple potentials in chapter 4. We discuss order parameters, both in general and as applied to our systems, in chapter 5. We then present our main results in chapters 6 (on nucleating ice with a monatomic water model) and 7 (on simulating the equivalent process with all-atom models of water). We finally summarise our results and discuss the implications of our work in chapter 8.

Certain information that does not fit neatly into the main framework of the thesis is included in the appendices. In appendix A, we study the phase behaviour of a system which exhibits competition between self-assembly and phase separation using classical statistical thermodynamics. In appendix B, we study the thermodynamics of a system of patchy particles in which a quasicrystalline phase is thermodynamically stable. In appendix C, we derive the classical nucleation theory expression for the rate of nucleation. Finally, in appendix D, we look at certain aspects of simulation methods in more detail than in chapter 3.

2 Nucleation

Nucleation is the first stage of the crystallisation process: it involves the formation of a cluster of a thermodynamically more stable phase within a metastable phase. Most natural nucleation processes are heterogeneous in the sense that the stable phase grows on top of an external nucleation seed (such as a pre-existing crystal, an impurity, or the surface of the containing vessel); by contrast, homogeneous nucleation takes place entirely in the bulk metastable phase (figure 2.1). The absence of an external nucleation seed means that a cluster of the stable phase must grow spontaneously before the system can crystallise; this process involves overcoming a free energy barrier, and homogeneous nucleation is thus a rare event. This means that, for example, very pure liquids can be supercooled without transforming to the more stable solid phase.

Nucleation is typical of first-order phase transitions[40] in which fluctuations from the

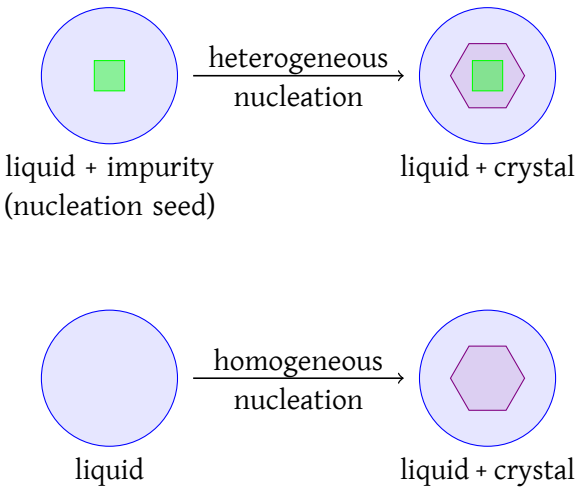


Figure 2.1: A schematic representation of heterogeneous and homogeneous nucleation.

metastable phase are disfavoured. However, a system can also be mechanically unstable; in such instances, the usual phase transformation mechanism is spinodal decomposition.[41] We briefly discuss stability criteria, binodals and spinodals in appendix A.4.3.

In this chapter, we derive the basic results of classical nucleation theory, which is the theory most often used in rationalising experimental and computational nucleation data.

2.1 Free energy barrier in classical nucleation theory

Classical nucleation theory (CNT) is the standard theoretical framework in which nucleation experiments and simulations are typically considered. Although there are significant issues with the theory (section 2.3), it nevertheless provides a convenient way of thinking about the basic physics of nucleation. Classical nucleation theory dates back to Volmer and Weber's work from the beginning of the 20th century,[42] with some modifications since.[43–46]

We define two states, A and B, as shown in figure 2.2. State A comprises only the metastable phase α, while in state B, an unstable equilibrium between the metastable phase α and a cluster of the stable[1] phase β is established.[47] To find the free energy barrier as the system is transformed from state A to state B, we find the Gibbs energy for each state at a constant temperature and a constant pressure. We can evaluate the Gibbs energy of state A

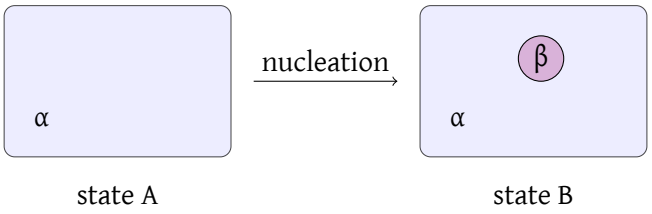


Figure 2.2: Two system states, A and B, are depicted. In state A, the entire system is made up of the metastable phase α; by contrast, in state B, a cluster of the stable phase β has formed within the metastable phase.

[1] Although we refer to phase β as the stable phase, it only needs to be more stable than phase α and need not be the most stable phase.

as[40]

$$G_{\mathrm{A}}^{\alpha} = \mu_{\mathrm{A}}^{\alpha} N, \tag{2.1}$$

and that of state B as

$$G_{\mathrm{B}}^{\alpha+\beta} = \mu_{\mathrm{B}}^{\alpha} N^{\alpha} + \mu_{\mathrm{B}}^{\beta} N^{\beta} + a\gamma, \tag{2.2}$$

where a is the surface area of the cluster of phase β.[2]

We assume that phase β is separated from phase α by a sharp boundary (*i.e.* one containing no particles) and that it is sufficiently macroscopic so that we can use classical thermodynamics; this is known as Gibbs' capillarity approximation.[7;48;49] The total number of particles is thus $N = N^{\alpha} + N^{\beta}$. We can consequently rewrite equation (2.2) as

$$G_{\mathrm{B}}^{\alpha+\beta} = \mu_{\mathrm{B}}^{\alpha} N + \left(\mu_{\mathrm{B}}^{\beta} - \mu_{\mathrm{B}}^{\alpha}\right) N^{\beta} + a\gamma. \tag{2.3}$$

We can also state that $\mu_{\mathrm{B}}^{\alpha} = \mu_{\mathrm{A}}^{\alpha} = \mu^{\alpha}$; this holds because the chemical potential of a phase is a function of the pressure and the temperature only.[47] The change in Gibbs energy on conversion from state A to state B is therefore

$$\Delta G = G_{B}^{\alpha+\beta} - G_{A}^{\alpha} = N^{\beta}\Delta\mu + a\gamma, \tag{2.4}$$

where $\Delta\mu = \mu^{\beta} - \mu^{\alpha}$. It is also usual to rewrite the number of particles in terms of the volume of the cluster using $N = \rho V$, where ρ is the number density.[3] This gives

$$\Delta G = \rho^{\beta} V^{\beta} \Delta\mu + a\gamma. \tag{2.5}$$

Since β is the more stable phase, $\Delta\mu < 0$ and thus the Gibbs energy change involved in the formation of a cluster of phase β involves a favourable bulk phase term scaling with the volume of the cluster and an unfavourable interfacial term scaling with the surface area of the cluster.

[2] We typically assume that the interfacial free energy is independent of the curvature of the interface, that it is the same as the interfacial free energy of an infinite planar interface,[7] and that it is the same for any exposed crystalline plane.

[3] This assumes that the cluster's density is the same as that of the bulk phase β; *i.e.* the cluster is incompressible.[50]

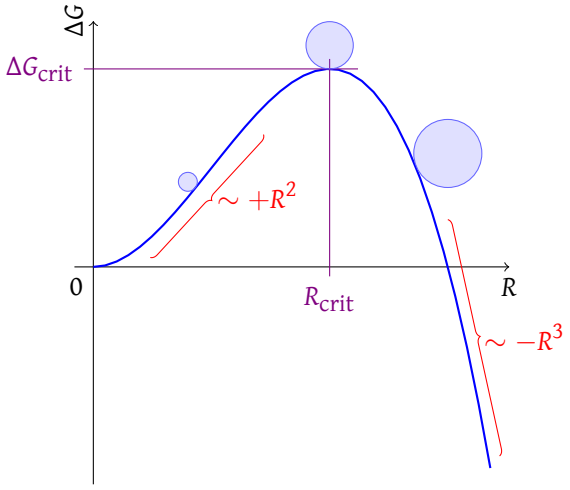


Figure 2.3: Gibbs energy change as a function of the cluster size (using equation (2.6)) associated with the formation of a cluster of a thermodynamically more stable phase within a metastable phase.

For small clusters, the surface area to volume ratio is highest, so we expect to see a free energy barrier associated with nucleation.[51] If we further assume that the cluster is spherical with radius R, which would be reasonable for macroscopic systems, since a spherical shape minimises the surface area to volume ratio, we can write

$$\Delta G = \frac{4\pi}{3} R^3 \times \rho^{\beta} \Delta\mu + 4\pi R^2 \times \gamma; \tag{2.6}$$

the form of this curve is depicted in figure 2.3. We can find the maximum in the curve by differentiation; the critical cluster size and the height of the free energy barrier are given by[50]

$$R_{\text{crit}} = -\frac{2\gamma}{\rho^{\beta}\Delta\mu} \quad \text{and} \quad \Delta G_{\text{crit}} = \frac{16\pi\gamma^3}{3\left(\rho^{\beta}\Delta\mu\right)^2}. \tag{2.7}$$

At coexistence, $\Delta\mu = 0$: this implies that the free energy barrier is infinite, and suggests that homogeneous nucleation will never be the transformation effecting the interconversion between the two equilibrium phases, even though they are in principle equally likely to occur at coexistence.[8]

It is often convenient to retain the number of particles within the largest cluster of phase β in the equations. To do this, we can look back at equation (2.4); if we continue to assume

that the cluster is spherical, we can express the surface area as

$$a = 4\pi R^2 = 4\pi (R^3)^{2/3} = 4\pi \left(\frac{3N^\beta}{4\pi\rho^\beta} \right)^{2/3}, \tag{2.8}$$

resulting in the Gibbs energy difference of[11;51]

$$\Delta G(N^\beta) = N^\beta \Delta\mu + \gamma \left[36\pi \left(\frac{N^\beta}{\rho^\beta} \right)^2 \right]^{1/3} \tag{2.9}$$

and a critical cluster size of

$$N^\beta_{\text{crit}} = -\frac{32\pi\gamma^3}{3\left(\rho^\beta\right)^2 \Delta\mu^3}. \tag{2.10}$$

2.2 Kinetics of homogeneous nucleation

The rate of nucleation depends exponentially on the height of the free energy barrier, and the rate equations obtained below are reminiscent of classical chemical kinetics. However, in classical chemical kinetics, the rate of a process depends exponentially on its activation energy via the Arrhenius equation,[40] but the activation energy itself is usually not strongly temperature-dependent; by contrast, nucleation rates depend exponentially on a free energy barrier that itself exhibits a strong dependence on the temperature of the system. To a good approximation, the temperature alone changes the nucleation free energy barrier from infinity at the binodal to zero at the spinodal point.[7] Consequently, the temperature dependence of nucleation rates is typically considerably more pronounced than that of chemical reaction rates.

As the literature is somewhat opaque when it comes to deriving the nucleation rate expression within the framework of classical nucleation theory, we provide a detailed derivation in appendix C. We show in this appendix that the steady-state CNT rate of homogeneous nucleation in a condensed matter system can be written as (equation (C.36))

$$J^\ddagger = N k_{\text{crit}} Z \exp\left[-\frac{\Delta G_{n_{\text{crit}}}}{k_B T} \right], \tag{2.11}$$

where

$$Z = \sqrt{-\frac{\Delta G''_{n_{\text{crit}}}}{2\pi k_{\text{B}} T}} \tag{2.12}$$

is the Zeldovich factor[44;52] and k_{crit} is the rate constant associated with cluster growth at the critical cluster size. This is an especially useful form of the rate equation for use in simulations.

Homogeneous nucleation is under kinetic control, and the phase that forms a cluster is not necessarily the thermodynamically stable phase under the conditions of the experiment, but typically the phase which has a lower free energy barrier associated with the formation of a cluster. This phenomenon is known as Ostwald's step rule or Ostwald's rule of stages,[53–55] since an intermediate phase is often seen to nucleate before the thermodynamically stable phase ultimately forms. Nucleation into such metastable phases has been seen in various experiments and simulations.[8] If the supercooling is very large, however, it can be the case that all phases that can potentially nucleate have similarly small nucleation free energy barriers; in such cases, the kinetic pre-factor can determine which nucleation pathway is the fastest one: the dynamically most accessible phase will nucleate, and this phase can often be the thermodynamically stable phase.[8]

2.3 Some shortcomings of classical nucleation theory

Classical nucleation theory is very simple: this is probably one of its most significant advantages, as it is easy to understand and apply. However, the agreement of nucleation rates predicted using classical nucleation theory with experiment is less than perfect.[7] Indeed, it has been argued that even when CNT predictions do match experimental results well, this may be due to a cancellation of errors;[7] however, this is likely to occur only at a particular temperature and the temperature dependence of nucleation rates predicted by classical nucleation theory is often found to be systematically wrong.[7] One example of such behaviour is the vapour-liquid

nucleation of nonane (n-C_9H_{20}), where predicted low temperature rates are a factor of 2×10^{-5} lower and high temperature rates a factor of 4×10^3 higher than experimental values.[56]

In classical nucleation theory, bulk thermodynamic properties are applied to microscopic systems, which is clearly not correct. The theory's predictions are very sensitive to the choice of value for the interfacial free energy; for example, the height of the free energy barrier (equation (2.7)) varies as γ^3. This means that the use of the interfacial free energy in the context of the capillarity approximation is especially contentious;[49] this is particularly problematic because even the critical cluster size can be relatively small and hence highly curved, and although the interfacial free energy depends on the curvature of the interfacial surface,[57] this is not accounted for in the classical theory. Furthermore, classical nucleation theory does not satisfy the law of mass action,[58] the interfacial free energy term does not disappear when the cluster considered comprises a single particle,[12] and some mechanical degrees of freedom may not be properly taken into account.[59;60] Its biggest drawback may be that it fails to describe 'non-classical' multi-stage nucleation processes.[11;51;61]

As a solution to some of these problems, various corrective factors have been introduced to the classical nucleation theory equations; moreover, several alternative theories – phenomenological, kinetic and microscopic – have been devised.[9;48;62] Nevertheless, due to its simplicity, classical nucleation theory remains the model framework in which the process of homogeneous nucleation is most often considered, and is often very successful. For example, despite its flaws, classical nucleation theory results in surprisingly good agreement with experiment in the study of nucleation of liquid water from the vapour phase.[63;64]

3 Simulation methods

In this chapter, we provide a brief introduction to the simulation methods that we use. In particular, we discuss the basic Monte Carlo simulation approach (section 3.1), followed by a discussion of umbrella sampling (section 3.2), histogram reweighting (section 3.3), determining rates from umbrella sampling simulations (section 3.4), other rare event techniques (section 3.5), hamiltonian exchange (section 3.6), molecular dynamics (section 3.7) and, finally, the hybrid Monte Carlo method (section 3.8).

The aim of this chapter is to provide a basic overview of the simulation methods used. In some cases, reasonably standard derivations that are relevant to the discussion have been placed in appendix D; however, some technical aspects, particularly in sections where few standard texts discuss the issues raised, are intentionally included in the main text.

3.1 Monte Carlo simulations

In Metropolis Monte Carlo simulations,[65] a physical system is evolved in configuration space by undergoing a series of stochastic moves (such as rotations or displacements of particles). In typical simulations of particles, we accept or reject such moves depending on the change in the potential energy of the system. As the system evolves in (Monte Carlo) time, we can calculate time averages of various thermodynamic properties; provided that the system is ergodic,[1] we can then equate these time averages to ensemble averages. However, whilst standard Monte

[1] This means that the system's free energy landscape should be such as to allow all important contributory configurations to the ensemble average to be accessed; if there are multiple free energy minima separated by large free energy barriers, special methods must be used to ensure ergodicity is achieved.

Carlo simulations entail the calculation of averaged thermodynamic observables at equilibrium, the simulation method can be adapted to find other properties of systems, such as free energies as a function of an order parameter, as discussed in section 3.2 below.

In Monte Carlo simulations, systems are not evolved using physical laws; instead, configuration space is probed stochastically by random walks around statistically important areas. It is not therefore generally possible to study dynamical properties of systems with Monte Carlo simulations,[65] although the use of moves that are not strictly physical in nature often allows for more efficient probing of configuration space to take place.[66] Nevertheless, it has been shown that Monte Carlo simulations of liquids using local moves often yield dynamic information in good agreement with molecular dynamics simulations.[67]

In the canonical thermodynamic ensemble (*i.e.* fixed N, V and T), the probability that a system is in configuration i with potential energy U_i is given by[68]

$$P_i = \frac{\exp\left[-\beta U_i\right]}{Q(N,\ V,\ T)}, \qquad \text{where} \quad Q(N,\ V,\ T) = \sum_i \exp\left[-\beta U_i\right] \tag{3.1}$$

is the canonical partition function.[2] The average of any observable $\mathcal{X}$ is then given simply by

$$\langle \mathcal{X} \rangle = \sum_i P_i \mathcal{X}_i = \frac{\sum_i \mathcal{X}_i \exp\left[-\beta U_i\right]}{Q(N,\ V,\ T)}. \tag{3.2}$$

However, since we do not know $Q(N,\ V,\ T)$ in advance of the simulation (if we did, there would be no need to perform the simulation in the first place), this equation is not very useful in practice. The Metropolis Monte Carlo algorithm provides a way of avoiding the need to evaluate $Q(N,\ V,\ T)$; a biassed walk is undertaken in configuration space, with the system spending the largest amount of time near configurations having the lowest free energy. This means that the largest contributors to the time average of observables of interest are the most dominant

[2] In principle, the total internal energy of each configuration, including the kinetic energy, should be considered. However, in Monte Carlo simulations, it is only the difference between the potential energies of configurations that contributes to the acceptance probability. At a constant temperature, the kinetic energy term will not change between configurations, and so kinetic energies are not normally considered at all in constant temperature Monte Carlo simulations.

configurations. In summary, the algorithm proceeds as follows:[66]

- Choose configuration 'new' with transition probability $\alpha(\text{old} \to \text{new})$.[3]
- Accept this configuration with acceptance probability $\Theta(\text{old} \to \text{new})$. If a move is rejected, the old configuration is restored and re-sampled.
- Add the values of observables of interest to the average sum.
- Repeat until observable averages converge.

In order to ensure that all probabilities are stationary at equilibrium, the system must satisfy what is known as *balance*,[66]

$$P_{\text{old}} \sum_i \alpha(\text{old} \to \text{new}_i)\Theta(\text{old} \to \text{new}_i) = \sum_i P_{\text{new}_i}\alpha(\text{new}_i \to \text{old})\Theta(\text{new}_i \to \text{old}), \tag{3.3}$$

which is to say that for every 'old' configuration, the number of transitions to all other configurations (new_i) must equal the number of transitions to the 'old' configuration. This is a simple statement of dynamic equilibrium, but is not especially convenient in simulations, as it would require us to keep track of many configurations at once. Instead, we can impose a more stringent condition onto the system and require it to satisfy *detailed balance*,[66]

$$P_{\text{old}}\alpha(\text{old} \to \text{new})\Theta(\text{old} \to \text{new}) = P_{\text{new}}\alpha(\text{new} \to \text{old})\Theta(\text{new} \to \text{old}), \tag{3.4}$$

where we impose a balance condition onto every pair of possible configuration transitions.[4] This is easier to use in practice, as we only require the knowledge of the two configurations under consideration, but it is worth emphasising that it is not necessary to impose detailed balance for a Monte Carlo simulation to be formally correct.[69]

In most of our simulations, the transition matrix is symmetrical (*i.e.* forward and backward transitions are equally likely, $\alpha(\text{new} \to \text{old}) = \alpha(\text{old} \to \text{new})$) and the two terms

[3] Since we choose each configuration randomly, *i.e.* without memory, this is a Markov chain.

[4] This means that any method in which it is in any given situation not possible to choose a reverse move following a move does not obey detailed balance. Consequently, the selection of which particle to move must not be systematic; it must be possible for the same particle to be moved twice in a row for detailed balance to be obeyed (but a single move need not involve only one particle).

cancel out. Using the detailed balance condition (equation (3.4)) allows us to write an explicit expression for the ratio of acceptance probabilities,

$$\frac{\Theta(\text{old} \to \text{new})}{\Theta(\text{new} \to \text{old})} = \frac{P_{\text{new}}}{P_{\text{old}}}. \tag{3.5}$$

Assuming that the Boltzmann distribution for the probability of each configuration holds, we have, therefore, an acceptance probability ratio of[66]

$$\frac{\Theta(\text{old} \to \text{new})}{\Theta(\text{new} \to \text{old})} = \exp\left[-\beta\left(U_{\text{new}} - U_{\text{old}}\right)\right]. \tag{3.6}$$

This ratio can be realised in a variety of ways; we use the original Metropolis scheme,[65] giving an acceptance probability of

$$\Theta(\text{old} \to \text{new}) = \min\left\{1, \exp\left[-\beta\left(U_{\text{new}} - U_{\text{old}}\right)\right]\right\}. \tag{3.7}$$

This means that all moves leading to a reduction in potential energy are accepted, and some moves that increase the potential energy are also accepted, with a probability that drops as the rise in potential energy increases.[5]

3.1.1 Isobaric-isothermal ensemble

In the thermodynamic limit (as $N \to \infty$), the choice of ensemble is irrelevant except for properties derived from fluctuations.[68] However, this is not necessarily the case when considering systems with a small number of particles. In nucleation studies in particular, the use of an 'open' ensemble, such as the isobaric-isothermal (NpT) ensemble or the grand canonical (μVT) ensemble, is preferable to a closed one (such as the canonical (NVT) ensemble). One reason for this is that when a (dense) crystal nucleus grows, in a constant volume simulation,

[5] To decide whether to accept a move with a probability less than unity, we compare the calculated probability to a random number generated uniformly between zero and one; if the random number is lower than the calculated probability, we accept the move. In our simulations, random numbers are generated using a method described by Marsaglia and co-workers.[70]

the density of the liquid phase will decrease[6] due to a reduction of the number of particles remaining in the liquid phase.[66] It has been shown that using the canonical ensemble as opposed to an open ensemble leads to significant errors in nucleation studies.[71] An additional advantage of using the isobaric-isothermal ensemble is that the natural potential in this system is the Gibbs energy; this allows for a simpler comparison with experimental data, which are generally also recorded at constant pressure.

The Monte Carlo procedure as presented above is followed closely in isobaric-isothermal ensemble simulations, except that the probability is given by[68]

$$P_i = \frac{\exp\left[-\beta\left(U_i + pV_i\right)\right]}{\Delta(N,\, p,\, T)}, \qquad \text{where} \quad \Delta(N,\, p,\, T) = \sum_i \exp\left[-\beta\left(U_i + pV_i\right)\right] \tag{3.8}$$

is the isobaric-isothermal partition function. Conveniently, the Gibbs energy is now given by $G = -k_\mathrm{B}T \ln \Delta(N,\, p,\, T)$, just as the Helmholtz energy is given by $A = -k_\mathrm{B}T \ln Q(N,\, V,\, T)$ in the canonical ensemble.[68]

Since the volume of the simulation box does not change when individual particles are displaced or rotated, the acceptance probability for such moves is identical to that for canonical ensemble simulations presented above. However, in addition to single particle moves, we need to allow for the volume of the simulation box to change in a constant pressure simulation.

It can be shown that trial moves in $\ln V$ yield a Metropolis Monte Carlo acceptance probability[66;72]

$$\Theta(V_\mathrm{old} \to V_\mathrm{new}) = \min\left\{1,\, \exp\left[-\beta\left(\Delta U + p\Delta V\right) + (N+1)\ln\left(\frac{V_\mathrm{new}}{V_\mathrm{old}}\right)\right]\right\}. \tag{3.9}$$

The interaction energies of the entire system must be recalculated when attempting a volume move: this makes such moves computationally expensive compared to single-particle displacement and rotation moves. In most of our simulations, we perform on average (but

[6] In simulations of ice nucleation, the solid phase is less dense than the liquid phase, but the basic argument is unchanged.

randomly) one volume move per Monte Carlo cycle.[7] Some derivations produce slightly different acceptance criteria,[73;74] but their use does not significantly change the ensemble averages for reasonably-sized systems.[75]

Finally, a word of caution about detailed balance. Although we choose new volumes by multiplying the existing volume by a scaling factor x such that $\ln V_{\text{new}} = \ln xV_{\text{old}}$,[8] we cannot simply choose x to be between $1 - X$ and $1 + X$, where $X > 0$ is the volume move size. For example, if we happened to choose the maximum volume change in the negative direction, such that $V_{\text{new}} = (1 - X)V_{\text{old}}$, then in the next step, we simply could not recover the original volume, as $(1 - X)(1 + X) = 1 - X^2 < 1$, which breaks detailed balance. We instead pick the scaling factor using $\exp((2\mathcal{R} - 1)\ln(1 + X))$, where $\mathcal{R}$ is a random number between 0 and 1, meaning that consecutive maximum volume scalings in different directions give $\frac{X+1}{X+1} = 1$. We are thus able to revert back to the original volume, as required by detailed balance.

3.2 Umbrella sampling

In a Monte Carlo simulation, the average of any observable $\mathcal{X}$ is given by equation (3.1),

$$\langle \mathcal{X} \rangle = \sum_i P_i \mathcal{X}_i = \frac{\sum_i \mathcal{X}_i \exp\left[-\beta U_i\right]}{Q(N,\ V,\ T)}. \tag{3.10}$$

However, if the probability of sampling a certain region of phase space is low, the statistical significance of the results is compromised. This is sometimes termed non-ergodicity. A method which can help alleviate this problem is the umbrella sampling technique.[31] The basic idea is to introduce some weight function w in order to increase the probability of sampling in the low probability (high free energy) regions by forcing the system to stay in some specified region of

[7] A Monte Carlo cycle corresponds to N Monte Carlo steps, where N is the number of particles in the system.

[8] In other words, we are not attempting volume sampling by 'adding' a volume to the existing volume, but are instead 'adding' to $\ln V$ in the sense that $\ln(xV) = \ln V + \ln x$. This is accounted for in the acceptance criterion.[66;72]

phase space. Then, we can express the time average of an observable $\mathcal{X}$ as[76]

$$\langle \mathcal{X} \rangle = \frac{\sum_i \mathcal{X}_i \exp\left[-\beta U_i\right]}{\sum_i \exp\left[-\beta U_i\right]} = \frac{\sum_i \mathcal{X}_i \frac{w_i}{w_i} \exp\left[-\beta U_i\right]}{\sum_i \frac{w_i}{w_i} \exp\left[-\beta U_i\right]} = \frac{\sum_i \frac{\mathcal{X}_i}{w_i} w_i \exp\left[-\beta U_i\right]}{\sum_i \frac{1}{w_i} w_i \exp\left[-\beta U_i\right]}, \tag{3.11}$$

and so we have[31]

$$\langle \mathcal{X} \rangle = \frac{\langle \mathcal{X}/w \rangle_w}{\langle 1/w \rangle_w}, \tag{3.12}$$

where the average is taken over a biassed configuration (*i.e.* one where the Boltzmann exponent is multiplied by the weight function w).

Provided that we choose a suitable weight function w, we can constrain the simulation to sample a particular region of interest (called a *window*). This is useful because it allows us to speed up equilibration in each window.[47;66;76;77]

We can express the relevant free energy (in the canonical ensemble, this is the Helmholtz energy) as a function of the order parameter Q[9] through the stationary distribution using[68]

$$A(Q) = -k_\mathrm{B} T \ln \Psi(Q), \tag{3.13}$$

relative to some (unimportant) zero. The stationary distribution can be expressed as the average probability of finding the simulation at a particular Q, and can thus be expressed with a Dirac δ-function as[77]

$$\Psi(Q) = \left\langle \delta(Q - Q(\boldsymbol{r}^N)) \right\rangle, \tag{3.14}$$

where Q is the order parameter of interest and $Q(\boldsymbol{r}^N)$ is the 'actual' order parameter in a simulation (in other words, the above function is simply a normalised counter). Let us choose $w = \exp\left[-\beta V(Q)\right]$, where $V(Q)$ is some function of the order parameter. Following on from equation (3.12),

$$\left\langle \delta(Q - Q(\boldsymbol{r}^N)) \right\rangle = \frac{\left\langle \delta(Q - Q(\boldsymbol{r}^N))/w \right\rangle_w}{\langle 1/w \rangle_w}, \tag{3.15}$$

[9] To avoid confusion between the canonical partition function and the order parameter, we denote the partition function with the variables held constant explicitly stated in this chapter, *i.e.* $Q(N, V, T)$.

giving in expanded form

$$\left\langle \delta(Q - Q(\boldsymbol{r}^N)) \right\rangle = \frac{\left\langle \exp\left[\beta V(Q(\boldsymbol{r}^N))\right] \delta(Q - Q(\boldsymbol{r}^N)) \right\rangle_w}{\left\langle \exp\left[\beta V(Q(\boldsymbol{r}^N))\right] \right\rangle_w} \tag{3.16}$$

$$= \exp\left[\beta V(Q)\right] \times \frac{\left\langle \delta(Q - Q(\boldsymbol{r}^N)) \right\rangle_w}{\left\langle \exp\left[\beta V(Q(\boldsymbol{r}^N))\right] \right\rangle_w}, \tag{3.17}$$

where the last equality follows from the fact that V is a function of the (selected) parameter Q and, multiplied by the Dirac delta function, will result in either the functional value at Q or zero. Note that the subscript w means that we are measuring these averages in a biassed ensemble.

We need not, in fact, calculate the average in the denominator, as we are only interested in the relative free energies. Taking the logarithm of equation (3.17) gives us

$$\ln \Psi(Q) = \beta V(Q) + \ln \left\langle \delta(Q - Q(\boldsymbol{r}^N)) \right\rangle_w - \ln \left\langle \exp\left[\beta V(Q(\boldsymbol{r}^N))\right] \right\rangle_w. \tag{3.18}$$

Since the final term in the above equation is not dependent on Q, but is instead an average over all $Q(\boldsymbol{r}^N)$, it is simply a constant that can be disregarded when calculating free energy differences.

To evaluate the remaining average, we will use a Metropolis Monte Carlo scheme. We recall that in Metropolis Monte Carlo, we add to an average after every step the value of the quantity we are searching for. For example, in the k-th step, the average of the quantity $\mathcal{X}$ will be

$$\mathcal{X}_k = \frac{1}{k} \sum_{i=1}^{k} \mathcal{X}(\boldsymbol{r}_i^N), \tag{3.19}$$

where each step is accepted or rejected with a Boltzmann probability. In this non-Boltzmann sampling, therefore, we can use the same approach, except insofar that we use a non-Boltzmann acceptance criterion analogous to equation (3.7),

$$\Theta(\text{old} \to \text{new}) = \min\left\{\exp\left[-\beta\left(U_{\text{new}} - U_{\text{old}} + \Delta V\right)\right]\right\}, \tag{3.20}$$

where ΔV is the change in the biassing potential. This means that, in practice, we will simply

need to count the number of steps that a simulation spends at each value of the order parameter, $s(Q)$, and then calculate

$$A(Q) = -V(Q) - k_B T \ln s(Q) + \text{constant}. \tag{3.21}$$

We remark here that this procedure works equally well in the isobaric-isothermal (NpT) ensemble; the only thing that changes is the Metropolis Monte Carlo acceptance criterion to allow for volume moves to take place and the relevant free energy is the Gibbs energy. Furthermore, having two (or more) order parameters allows us to extend the method into two (or more) dimensions with no conceptual changes.

When we use windows to sample only in a limited range of order parameter space, the free energies thus calculated must be matched up. Although relatively complex approaches exist for multi-dimensional umbrella sampling, such as the weighted histogram analysis method,[78–80] there are two considerably simpler approaches that can be used in one-dimensional umbrella sampling. Generally, in nucleation studies, we wish to find the Gibbs energy associated with nucleation relative to the supercooled liquid; therefore, the Gibbs energy of the liquid phase, as determined by Q, can be shifted to zero. Once we have constructed a free energy profile in each window, we can attempt to match one of the overlapping points (for each pair of adjacent windows) and see if the rest of the overlapped region then also matches up. If it does not do so reasonably, this is a good indication that the system is not yet well equilibrated. A less arbitrary method of matching the Gibbs energies of adjacent windows is to calculate the (discrete) gradients, $\frac{dG(Q)}{dQ}$, fit a polynomial to the gradients across all relevant windows, and then integrate this polynomial to obtain $G(Q)$.[47;77] However, if the system is well equilibrated, the two methods should give essentially the same answer; if the system is not equilibrated sufficiently, then the polynomial fit approach is not likely to produce an accurate answer.

In our simulations, we typically use adaptive umbrella sampling biassing weights[32]

rather than the more traditional quadratic bias potentials.[47] This means that we adaptively change the weights associated with each order parameter value in the window after every simulation iteration to take into account the previous weights and the number of steps spent at that order parameter value in the previous iteration. In principle, the weight $V(Q)$ associated with each value of the order parameter is expected to tend to the negative of the free energy at that order parameter. However, we do limit the maximum change per update in order to minimise hysteresis effects; that is, we update weights using

$$V_{\text{new}}(Q) = V_{\text{old}}(Q) + xk_{\text{B}}T \ln s(Q) \tag{3.22}$$

and we find x such that the maximum change, $xk_{\text{B}}T \ln(s_{\text{max}})$, corresponds to a few $k_{\text{B}}T$. We remove this constraint and set $x = 1$ once the free energies no longer fluctuate significantly between iterations.

To test whether equilibrium has been attained after an umbrella sampling iteration, we need to ensure that there is, within reason, equal sampling across order parameter values in each window, and that the order parameter values fluctuate quickly. A further check to ensure that we have sufficiently equilibrated each window is to ascertain that the potential enthalpy as a function of the order parameter is consistent in the overlapping regions.

3.2.1 Extracting other quantities from umbrella sampling simulations

As mentioned, we typically wish to find the potential enthalpy ($H = U + pV$, where U is the potential energy) as a function of the order parameter. If we record a histogram of the enthalpy as a function of the order parameter, we can then use the formula given above in equation (3.12),

$$\langle H(Q) \rangle = \frac{\langle H(Q)/w \rangle_w}{\langle 1/w \rangle_w} = \frac{\left\langle H(Q) e^{\beta V(Q)} \right\rangle_w}{\left\langle e^{\beta V(Q)} \right\rangle_w}, \tag{3.23}$$

and we note that the average value of each of these, as measured in the umbrella sampling simulation, is just the sum over all steps of the function in question divided by the total number

of steps. As we are interested only in a given value of Q, however, we only take the steps at which this condition is fulfilled into account. This gives

$$\langle H(Q)\rangle \approx \frac{\sum_{h=\text{each } H} \mathcal{C}(h,\, Q)\mathrm{e}^{\beta V(Q)}h}{\sum_{h=\text{each } H} \mathcal{C}(h,\, Q)\mathrm{e}^{\beta V(Q)}} = \frac{\sum_{h=\text{each } H} \mathcal{C}(h,\, Q)h}{\sum_{h=\text{each } H} \mathcal{C}(h,\, Q)}, \tag{3.24}$$

where $\mathcal{C}$ is a counter function, and both the numerator and the denominator can be obtained from the histograms. These values are *absolute*, and cannot be shifted by an arbitrary constant as was done for the Gibbs energy above. In this case, the imposed umbrella weights have cancelled out: we are in practice just calculating the average H as gathered from the simulations, without regard to the weights. This is because we are interested in the enthalpy at a given Q, and not how we got to that value of Q; ultimately, $H(Q)$ is just the average enthalpy as a function of the order parameter as evaluated from the simulation, irrespective of whether the simulation was biassed or not. Nevertheless, to ensure that good statistics are obtained, the order parameter Q must be visited frequently, and this can typically only be achieved in biassed simulations.

We may sometimes be interested in other quantities that do depend on the weights. For example, we might wish to find the probability distribution of an alternative order parameter that was not used in the umbrella sampling procedure itself. Up to a constant factor, the probability of finding an order parameter z is

$$P(z) = \left\langle \delta(z - z(\boldsymbol{r}^N)) \right\rangle. \tag{3.25}$$

In order to find this, we again use the definitions above to give

$$\left\langle \delta(z - z(\boldsymbol{r}^N)) \right\rangle = \frac{\left\langle \delta(z - z(\boldsymbol{r}^N))/w \right\rangle_w}{\langle 1/w \rangle_w} \tag{3.26}$$

$$= \frac{\left\langle \delta(z - z(\boldsymbol{r}^N)) \exp(\beta V(Q(\boldsymbol{r}^N))) \right\rangle_w}{\langle \exp(\beta V(Q(\boldsymbol{r}^N))) \rangle_w} \tag{3.27}$$

$$\approx \frac{\sum_{q=\text{each } Q} \mathcal{C}(z,\, q)\mathrm{e}^{\beta V(q)} \cdot 1}{\sum_{h=\text{each } z} \sum_{q=\text{each } Q} \mathcal{C}(h,\, q)\mathrm{e}^{\beta V(q)}}. \tag{3.28}$$

Since we will generally be interested in the free energy as a function of any new order parameter, we normally will not need to evaluate the denominator, as it will be encompassed in the constant term that we ignore upon taking logarithms.

3.3 Histogram reweighting

We may sometimes wish to gain an estimate of an observable under conditions similar to, but slightly different from, those in which we ran our simulations. We can do this by means of histogram reweighting.[81–83]

In the isobaric-isothermal ensemble, the average of some observable $\mathcal{X}$ is given by

$$\langle \mathcal{X} \rangle = \sum_{i=1}^{\mathcal{N}} P_i \mathcal{X}_i = \frac{\sum_{i=1}^{\mathcal{N}} \mathcal{X}_i \exp\left[-\beta \left(U_i + pV_i\right)\right]}{\Delta(N,\, p,\, T)}, \tag{3.29}$$

where $\mathcal{N}$ is the total number of configurations. If we assume that the probability of a certain value of $\mathcal{X}_i$ at a given potential enthalpy ($H_i = U_i + pV_i$) is correct as obtained from a Monte Carlo simulation at a constant inverse temperature β, then the relevant probabilities at a different inverse temperature, β', are going to be given by

$$P_i' = \frac{\exp\left[-\beta' H_i'\right]}{\Delta(N,\, p,\, T')}. \tag{3.30}$$

However, a configuration's potential enthalpy depends only on the configuration itself and not on the temperature, so $H_i' = H_i$, and we can write the ratio between the known probability P_i and the probability at the new temperature as

$$\frac{P_i'}{P_i} = \frac{\Delta(N,\, p,\, T)}{\Delta(N,\, p,\, T')} \times \exp\left[-(\beta' - \beta) H_i\right]. \tag{3.31}$$

Therefore

$$\langle \mathcal{X} \rangle_{\beta'} = \sum_{i=1}^{\mathcal{N}} P_i' \mathcal{X}_i \tag{3.32}$$

$$= \sum_{i=1}^{\mathcal{N}} \frac{\Delta(N,\, p,\, T)}{\Delta(N,\, p,\, T')} \times \exp\left[-(\beta' - \beta)H_i\right] (P_i \mathcal{X}_i) \tag{3.33}$$

$$= \frac{\Delta(N,\, p,\, T)}{\Delta(N,\, p,\, T')} \sum_{i=1}^{\mathcal{N}} \exp\left[-(\beta' - \beta)H_i\right] (P_i \mathcal{X}_i) \tag{3.34}$$

$$= \frac{\Delta(N,\, p,\, T)}{\Delta(N,\, p,\, T')} \left\langle \exp\left[-(\beta' - \beta)H\right] \mathcal{X} \right\rangle_\beta , \tag{3.35}$$

where in the last step, we use the same notation as in the umbrella sampling section above, with the subscripted β showing that the average is evaluated at this temperature. To find the ratio of the partition functions, let us suppose for the time being that we are interested in the observable unity (*i.e.* $\mathcal{X}_i = 1$ for every i at every temperature). Then

$$\frac{\Delta(N,\, p,\, T')}{\Delta(N,\, p,\, T)} = \left\langle \exp\left[-(\beta' - \beta)H\right] \right\rangle_\beta , \tag{3.36}$$

which suggests that[82;83]

$$\langle \mathcal{X} \rangle_{\beta'} = \frac{\left\langle \exp\left[-(\beta' - \beta)H\right] \mathcal{X} \right\rangle_\beta}{\left\langle \exp\left[-(\beta' - \beta)H\right] \right\rangle_\beta}, \tag{3.37}$$

again mirroring the relevant expression in the umbrella sampling section above. In other words, we have now established that, just like with umbrella sampling simulations, we should be able to extract the unweighted data (*i.e.* the mean value of the observable at the specified temperature T') from a simulation run at a different temperature. To be able to use this, we need to ensure that the temperature difference is not too large; this ensures that the configurations relevant to the new temperature T' are sufficiently sampled at the simulation temperature T.[83]

We measure averages in a Monte Carlo simulation by simply counting the number of occurrences of particular observable values,

$$\langle \mathcal{X} \rangle = \sum_{i=1}^{\mathcal{N}} P_i \mathcal{X}_i \approx \frac{1}{n} \sum_{i=1}^{n} \mathcal{X}_i, \tag{3.38}$$

where n is the number of steps. This means that we can recast equation (3.37) as

$$\langle \mathcal{X} \rangle_{\beta'} = \frac{\left\langle \exp\left[-(\beta' - \beta)H\right] \mathcal{X} \right\rangle_{\beta}}{\left\langle \exp\left[-(\beta' - \beta)H\right] \right\rangle_{\beta}} \approx \frac{\sum_{i=1}^{n} \exp\left[-(\beta' - \beta)H_i\right] \mathcal{X}_i}{\sum_{i=1}^{n} \exp\left[-(\beta' - \beta)H_i\right]}. \tag{3.39}$$

In practice, we must assemble a histogram of the number of occurrences of enthalpy and observable-of-interest pairs. We generally have to bin continuous variables to discretise them; this can introduce some error.[84] Using this histogram approach, the sum is not performed over every step, but instead over every pair of H and $\mathcal{X}$ that occurred in the simulation, multiplied by the 'degeneracy' of the configuration (*i.e.* the number of steps the configuration was visited). If $\mathcal{C}(H, \mathcal{X})$ is the counter of occurrences of the enthalpy H and the observable $\mathcal{X}$ at the same time, then

$$\langle \mathcal{X} \rangle_{\beta'} \approx \frac{\sum_{h=\text{each } H} \sum_{x=\text{each } \mathcal{X}} \mathcal{C}(h, x) \exp\left[-(\beta' - \beta)h\right] x}{\sum_{h=\text{each } H} \sum_{x=\text{each } \mathcal{X}} \mathcal{C}(h, x) \exp\left[-(\beta' - \beta)h\right]}. \tag{3.40}$$

We use this method to estimate the height of the free energy barrier of nucleation at temperatures other than the temperature at which we performed the simulation. We can treat each value of the Gibbs energy as a function of the order parameter to be an 'observable' and apply the above result. We use the stationary distribution as the basis for constructing the Gibbs energy profile,

$$G(Q) = -V(Q) - k_{\mathrm{B}} T \ln\left[\left\langle \delta(Q - Q(\boldsymbol{r}^N)) \right\rangle_w\right] + \text{constant}, \tag{3.41}$$

where $V(Q)$ is the umbrella biassing potential. We are therefore interested in $\left\langle \delta(Q - Q(\boldsymbol{r}^N)) \right\rangle_w$ at a different temperature. The histograms we must therefore keep track of in simulations are those of $[H, Q]$ pairs, and we can calculate

$$\left\langle \delta(Q - Q(\boldsymbol{r}^N)) \right\rangle_{w,\beta'} \propto \sum_{h=\text{each } H} \mathcal{C}(h, Q) \exp\left[-(\beta' - \beta)(h + V(Q))\right] \tag{3.42}$$

from an umbrella sampling simulation.[10] This allows us to construct the entire Gibbs energy

[10] The denominator is not important, as it is constant for any given temperature, and so would be encompassed in

profile; as we are generally using *biassed* simulations ($V(Q) \neq 0$), we need to subtract the relevant umbrella sampling weights in the same manner as we do in the calculation at the simulation temperature.

3.4 Nucleation rates

Nucleation barriers calculated in umbrella sampling simulations can be used as a starting point in the calculation of the rate of nucleation in a Bennett–Chandler-type approach.[85] We assume that the crystal nucleation rate J is related to the free energy barrier ΔG through an Arrhenius-type equation[50;52;86–88]

$$J = \kappa P(n_{\text{crit}}), \tag{3.43}$$

where n_{crit} is the size of the critical cluster, $P(n_{\text{crit}})$ denotes the probability that a spontaneous fluctuation from the supercooled liquid will result in the formation of a critical nucleus, and κ is the kinetic pre-factor. The probability of being at the top of the free energy barrier is given by $P(n_{\text{crit}}) = \exp(-\beta \Delta G_{n_{\text{crit}}})$. We show in section 2.2 and appendix C that the kinetic pre-factor can be written as[52]

$$\kappa = Z \rho_{\text{liq}} f_{n_{\text{crit}}}, \tag{3.44}$$

where Z is the Zeldovich factor, $f_{n_{\text{crit}}}$ is the rate at which particles are attached to the critical cluster, and ρ_{liq} is the number density of the supercooled liquid.[11] The Zeldovich factor, Z, is given by

$$Z = \sqrt{-\frac{\Delta G''_{n_{\text{crit}}}}{2\pi k_{\text{B}} T}}, \tag{3.45}$$

where $\Delta G''_{n_{\text{crit}}}$ refers to the second derivative of ΔG evaluated at the critical nucleus size. The attachment rate $f_{n_{\text{crit}}}$ can be expressed as the diffusion of the cluster size at the top of the

the 'constant' term of $G(Q)$. Notice the addition of the $V(Q)$ term – the probability of a configuration depends also on the umbrella weight in non-Boltzmann simulations.[31]

[11] The rate equations given in section 2.2 and appendix C are particle-based; in constant pressure simulations and experimental practice, however, concentration-based rates, as given here, are more natural.

barrier[50;87;88] by a modified Einstein relation,[66]

$$f_{n_{\text{crit}}} = \frac{\left\langle \{\Delta n(t)\}^2 \right\rangle}{2t}, \tag{3.46}$$

and the variance of the cluster size is

$$\left\langle \{\Delta n(t)\}^2 \right\rangle = \left\langle (n(t) - n_{\text{crit}})^2 \right\rangle . \tag{3.47}$$

We measure this variance in molecular dynamics simulations (section 3.7) at the top of the barrier,[12] where all simulations are taken into account, not just those that remain close to the critical size.[88] As the attachment and detachment rates are, by construction, equal at the top of the barrier, we can measure the average of all changes to the cluster size. The value of the attachment rate is different at short times compared to the long-time behaviour; at short times, we observe local motion of particles in or near the cluster, whilst long-time behaviour is caused by the physical rearrangement of the liquid phase required to increase (or decrease) the cluster size more permanently. As a result, it is important to measure the variance in cluster size over a longer time period.

Although this is considerably simpler than the full Bennett–Chandler scheme,[66;85;89] it has been shown to agree very well with other methods of calculating rates in crystal nucleation, such as forward flux sampling and direct measurement from molecular dynamics.[88]

3.5 Other ice nucleation rare event methods

Quigley and Rodger studied homogeneous ice nucleation using the molecular dynamics-based metadynamics technique.[90;91] In a metadynamics simulation, a (normally gaussian) potential is added to the free energy each time a point is visited in phase space (as quantified by some order parameter). Since low free energy regions are naturally visited more often, more and

[12] To do this, we must first generate many uncorrelated configurations at the top of the barrier; these can be taken from the umbrella sampling simulations.

more potential energy is added to those points and this eventually levels out the free energy surface, thereby allowing the entire phase space to be sampled.[92]

We initially studied[93] ice nucleation using forward flux sampling,[50;94–97] and this technique has also been used in the study of ice nucleation of a monatomic water model by Li and co-workers.[98] In this method, a series of interfaces in order parameter space is created and the system is driven across these by starting large numbers of simulations from each previous interface towards the one we are interested in. This allows us to calculate conditional probabilities of crossing each interface and thus allows for an easy calculation of a nucleation rate. By calculating the 'stationary distribution' in order parameter space, we can also compute the free energy barrier of the process. As we make reference to the forward flux sampling technique in other chapters, we describe the technique in more detail in appendix D.5.

Other approaches, such as Wang–Landau sampling[99] or transition interface sampling,[100] can also be used; however, the basic approach is similar to umbrella sampling and forward flux sampling, respectively, and so we do not describe them in detail here.

3.6 Hamiltonian replica exchange

Entirely unphysical moves can be introduced into Monte Carlo simulations; these can then allow us to sample phase space much more quickly than more physical moves would allow for.[101] One such possible move is a hamiltonian replica exchange.[102–108]

Hamiltonian exchange is a variant of the parallel tempering technique, where ergodicity over a free energy barrier is achieved by running a simulation in identical, non-interacting copies ('replicas') of a system in parallel at several different temperatures, including temperatures at which the barrier is easily crossed by the system. By swapping configurations amongst these parallel runs (or equivalently, by swapping temperatures amongst configurations), the low-temperature system of interest can achieve ergodicity and can cross barriers that it

otherwise could not in a reasonable simulation timescale.[107]

In hamiltonian exchange,[105] we generalise this idea so that instead of running simulations at various temperatures (which is not necessarily conducive to speeding up nucleation simulations, as higher temperatures involve higher free energy barriers), we change the hamiltonian of the system. This allows us to couple our system to one in which the process of interest occurs more readily.[66;101;102;109] In principle, we wish to increase the real time rate at which the system explores phase space; despite the increase in the number of simulations whose results are not of any physical interest (we are only interested in one hamiltonian), it is sometimes the case that the gain in sampling efficiency is greater than the time wasted on non-physical simulations.[66]

Let us copy the initial configuration amongst M replicas, each with a different hamiltonian. Since there are no interactions between the replicas, we can define an overall *extended ensemble* containing all M replicas, and its partition function will be the product of the individual replica partition functions[66] (which can be canonical or isobaric-isothermal partition functions, or indeed partition functions of another ensemble[66;110]). This means that the overall weight for a given extended configuration is simply the product of all the replicas' Boltzmann factors. If we exchange two randomly chosen (but adjacent) replicas, we need to recalculate the overall weight of the extended configuration.[13] Just like in section 3.1, in order to ensure that we obtain an equilibrium distribution for the extended ensemble (and hence in the individual system ensembles), we impose the detailed balance condition onto this exchange move, and assume that the Boltzmann distribution holds.

[13] In classical parallel tempering, where the only change in the hamiltonian is the temperature at which the replicas run, this 'recalculation' is especially simple, as the potential energies are known, and therefore such temperature swap moves are in fact computationally very inexpensive.[66]

The total probability of an extended ensemble state A is

$$p_{\mathrm{A}} = \frac{\prod_{i=1}^{M} \exp\left[-\beta_i H_i\right]}{Q_{\mathrm{extended}}}, \tag{3.48}$$

where β_i is the reduced temperature of replica i, H_i is the relevant energy of replica i (depending on the replica's hamiltonian and the positions and orientations of the particles), and Q_{extended} is the partition function of the extended ensemble. For a hamiltonian exchange move from state A to state B, we must obey detailed balance (see section 3.1),

$$p_{\mathrm{A}} \Theta(\mathrm{A} \to \mathrm{B}) = p_{\mathrm{B}} \Theta(\mathrm{B} \to \mathrm{A}). \tag{3.49}$$

Assuming that the Boltzmann distribution holds, we can write, for the system represented in figure 3.1, that

$$\frac{\Theta(\mathrm{A} \to \mathrm{B})}{\Theta(\mathrm{B} \to \mathrm{A})} = \frac{p_{\mathrm{B}}}{p_{\mathrm{A}}} = \frac{\exp\left[-\beta_1 H_1(\boldsymbol{r}_1) - \beta_2 H_2(\boldsymbol{r}_3) - \beta_3 H_3(\boldsymbol{r}_2) - \beta_4 H_4(\boldsymbol{r}_4)\right]}{\exp\left[-\beta_1 H_1(\boldsymbol{r}_1) - \beta_2 H_2(\boldsymbol{r}_2) - \beta_3 H_3(\boldsymbol{r}_3) - \beta_4 H_4(\boldsymbol{r}_4)\right]}, \tag{3.50}$$

which conveniently simplifies into

$$\frac{\Theta(\mathrm{A} \to \mathrm{B})}{\Theta(\mathrm{B} \to \mathrm{A})} = \exp\left[\beta_2 H_2(\boldsymbol{r}_2) + \beta_3 H_3(\boldsymbol{r}_3) - \beta_2 H_2(\boldsymbol{r}_3) - \beta_3 H_3(\boldsymbol{r}_2)\right]. \tag{3.51}$$

Thus, in general, for a swap between the configuration associated with $\boldsymbol{r}_i$, initially governed

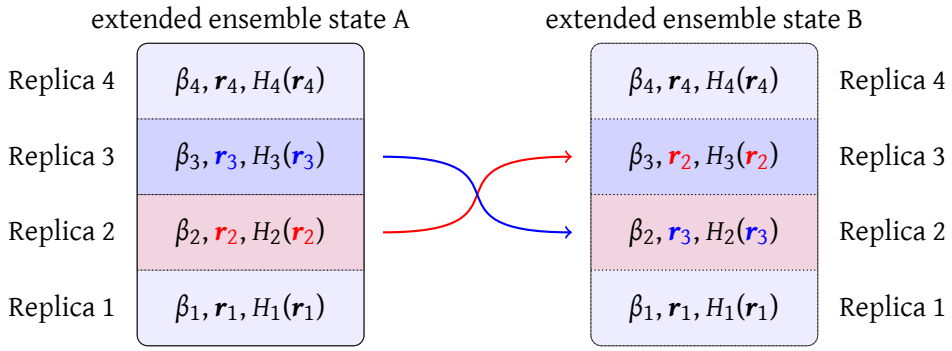


Figure 3.1: A schematic showing the exchange of configurations, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$, between two replicas, 2 and 3. The configuration vector $\boldsymbol{r}_i$ should be understood to refer to both the positions and the orientations of all the particles.

by the hamiltonian H_i, and $\boldsymbol{r}_j$, initially governed by the hamiltonian H_j, we can write a Metropolis-style acceptance probability akin to equation (3.7),[101;102]

$$\Theta(\text{old} \to \text{new}) = \min\{1, \exp[-\mathcal{H}]\}, \tag{3.52}$$

where

$$\mathcal{H} = \beta_i H_i(\boldsymbol{r}_j) + \beta_j H_j(\boldsymbol{r}_i) - \beta_i H_i(\boldsymbol{r}_i) - \beta_j H_j(\boldsymbol{r}_j). \tag{3.53}$$

The hamiltonians will include any umbrella sampling terms; furthermore, in simulations where we use hamiltonian exchange in an isobaric-isothermal ensemble, we may in certain cases need to add the standard volume change term to this acceptance criterion.[110;111] In particular, the acceptance criterion depends on $\mathcal{H}$ as[14]

$$\begin{aligned}\mathcal{H} = {} & \beta_i U_i(\boldsymbol{r}_j) + \beta_j U_j(\boldsymbol{r}_i) - \beta_i U_i(\boldsymbol{r}_i) - \beta_j U_j(\boldsymbol{r}_j) \\ & + \left\{\beta_i p_i V_j + \beta_j p_j V_i - \beta_i p_i V_i - \beta_j p_j V_j\right\},\end{aligned} \tag{3.54}$$

where the terms in curly brackets reflect the change in volume. On the assumption that the simulations are performed at the same inverse temperature ($\beta = \beta_i = \beta_j$) and the same pressure ($p = p_i = p_j$), which is the case in our simulations, but need not be the case generally, then

$$\begin{aligned}\mathcal{H}/\beta &= U_i(\boldsymbol{r}_j) + U_j(\boldsymbol{r}_i) - U_i(\boldsymbol{r}_i) - U_j(\boldsymbol{r}_j) + p\left\{V_j + V_i - V_i - V_j\right\} \\ &= U_i(\boldsymbol{r}_j) + U_j(\boldsymbol{r}_i) - U_i(\boldsymbol{r}_i) - U_j(\boldsymbol{r}_j).\end{aligned} \tag{3.55}$$

Conveniently, if the pressures and temperatures of all the replicas are the same, we no longer have to worry about any volume terms in the acceptance criterion.

3.7 Molecular dynamics simulations

Molecular dynamics (MD) simulations are an alternative to the Monte Carlo approach described above. Instead of sampling stochastically in configuration space, molecular dynamics simulations

[14] The logarithmic terms in the pressure acceptance term as seen in subsection 3.1.1 cancel out in equation (3.54) provided that the number of particles in the two replicas is identical.

aim to evolve a physical system dynamically in time by numerically integrating Newton's equations of motion.[66] This has the important advantage over Monte Carlo simulations that we can reasonably claim to be looking at the 'natural' dynamics of a system. However, molecular dynamics simulations necessitate the calculation of forces on particles and are thus typically computationally more expensive than the corresponding Monte Carlo simulations.[66] We describe some common MD integration approaches in appendix D.3.

In Monte Carlo simulations, the correct thermodynamic ensemble is enforced by the Metropolis acceptance criterion. In pure molecular dynamics, it can be shown[66] that for an isolated system, the total energy of the system is conserved, and so the basic thermodynamic ensemble sampled by such simulations is the microcanonical (*NVE*) rather than the canonical (*NVT*) ensemble of basic Monte Carlo. This is often not convenient when comparing results to experiments conducted at constant temperature and pressure, and various barostats and thermostats[112–116] have been developed to couple the system of interest to, say, a heat bath.[15]

3.8 Hybrid Monte Carlo

It has been shown that molecular dynamics (MD) simulations of ice-water interfaces appear to offer a significant reduction in the computer time required compared to Monte Carlo simulations.[117;118] The choice of MD simulations over MC simulations is therefore appealing; however, as we are interested in the free energy landscape of ice nucleation, we wish to continue to use umbrella sampling and thus Monte Carlo simulations. As a compromise, we couple Monte Carlo with MD simulations in a hybrid Monte Carlo approach.[39;119–122] We run an overarching Monte Carlo simulation, but instead of using ordinary random particle moves, we use global dynamic moves by means of short MD simulations. Provided that

[15] Although most of our molecular dynamics simulations are part of a hybrid Monte Carlo approach (section 3.8), in which the thermodynamic ensemble is maintained by an overarching Monte Carlo simulation, this is not the case in some Bennett–Chandler-type simulations (section 3.4) we run to find the nucleation rate, where we couple the system to a barostat and a thermostat as described for each system we study.

the MD integrator is time-reversible and symplectic and that the choice of momenta from the Maxwell–Boltzmann distribution is accounted for in the Metropolis acceptance criterion, detailed balance is obeyed[39;119–122] irrespective of the fact that the MD hamiltonian does not incorporate an umbrella sampling term.

3.8.1 Detailed balance

It is not immediately obvious whether the hybrid Monte Carlo technique obeys detailed balance (section 3.1). Indeed, especially if we use an umbrella potential in the overarching Monte Carlo scheme but not in the trial molecular dynamics runs, we can imagine that a pre-critical cluster in a brute force MD simulation is very likely to collapse rapidly. It may seem that this implies that reducing the cluster size is more favourable than increasing it, and this may appear to violate detailed balance. However, detailed balance is obeyed even if different hamiltonians are used for MD and MC steps, provided that the molecular dynamics algorithms used are area-preserving and time-reversible.[39;119–122] Indeed, the requirement for time reversibility allows us to understand intuitively why we are obeying detailed balance: if in the step following an accepted trial move we choose velocities that are exactly the negatives of the ones generated by the previous MD run, we can return to the original configuration.[119] A cluster that predominantly shrinks will simply lead to more rejected moves, but this does not break detailed balance. We show that hybrid Monte Carlo simulations obey detailed balance more formally in appendix D.2.

From chapter 3.1, we have the detailed balance condition (equation (3.6)),

$$P_{\text{old}}\alpha(\text{old} \to \text{new})\Theta(\text{old} \to \text{new}) = P_{\text{new}}\alpha(\text{new} \to \text{old})\Theta(\text{new} \to \text{old}), \tag{3.56}$$

where P_{state} refers to the probability of being in a given configuration, α is the probability of attempting the move from one configuration to another and Θ is the acceptance probability of moving from one configuration to another. In section 3.1, we assumed that the transition

probability α was symmetric and it thus cancelled out in the acceptance criterion. This condition is not fulfilled in hybrid Monte Carlo simulations, and so we need to account for these transition probabilities explicitly. The acceptance criterion for the MD move step is thus given by[39;123;124]

$$\Theta(\text{old} \rightarrow \text{new}) = \min\left\{1,\, \exp\left[-\beta\Delta U - \beta p \Delta V\right] \frac{\alpha(\text{new} \rightarrow \text{old})}{\alpha(\text{old} \rightarrow \text{new})}\right\}, \tag{3.57}$$

where ΔU is the change in the potential energy and ΔV the change in volume. Although an MD trajectory is deterministic and so the probability of reaching a given final configuration from an initial configuration is unity, we need to take into account the probability of generating the velocities associated with each particle; the particles' kinetic energies are chosen from a Maxwell–Boltzmann distribution (or, equivalently, the velocities in each dimension are chosen from a normal distribution with mean zero and a standard deviation scaled to produce the correct temperature), and so the transition probability α is given by

$$\alpha(\text{old} \rightarrow \text{new}) = \exp\left[-\beta K^{\text{old}}\right], \tag{3.58}$$

where K is the kinetic energy. The overall acceptance criterion is therefore

$$\Theta(\text{old} \rightarrow \text{new}) = \min\left\{1,\, \exp\left[-\beta(\Delta U + \Delta K + p\Delta V)\right]\right\}. \tag{3.59}$$

The extension to umbrella sampling simulations is straightforward.

However, even if detailed balance is in principle obeyed and the correct ensemble is being sampled, there is the issue of move generation that may not be capable of sampling all the configurations we wish it to sample. For example, if we are interested in a window in which the cluster of interest almost invariably melts when simulated in an unconstrained MD simulation, then, even if we reject essentially all the moves, we may never produce trial moves that increase the size of the cluster. In this case, even though in the limit of an infinite simulation we may sample correctly, we face the usual problem of ergodicity; we may simply not be able to produce moves which would probe the entire phase space we wish to sample. This means that the MD

cycle in the MC framework should not have too different a hamiltonian (for example, we cannot choose a molecular dynamics temperature that is much higher than the one imposed by the MC acceptance criterion, as the vast majority of moves would end up being rejected) and should be capable of producing configurations throughout the umbrella sampling window of interest. In theory, this is no different from problems facing 'ordinary' Monte Carlo moves; however, the problem appears at first glance to be more likely to occur with MD moves, not least because the system can change much more in a short MD simulation than following a random single particle move.[16] Nevertheless, provided that we can sample the entire (umbrella sampling) window, the sampling will indeed correspond to the desired ensemble.

3.8.1.1 Thermostats and barostats

Although we enforce the temperature and pressure to correspond to the correct thermodynamic ensemble by means of an overarching Monte Carlo scheme, it may sometimes be beneficial to use a thermostat or barostat within the short MD simulations as well: for example, Faller and de Pablo have suggested that using molecular dynamics to ensure the equilibration of the box volume is faster than the simple volume scaling method introduced in section 3.1.1.[123;126]

However, many simple thermostats and barostats do not fulfil the time-reversibility requirement for hybrid Monte Carlo to maintain detailed balance. For example, in the commonly employed Berendsen weak coupling barostat,[113] we calculate the virial pressure following an MD time step and then scale the centre of mass co-ordinates and the box dimensions towards the target pressure p_{target} with a scaling factor s such that[113]

$$s^3 = 1 + \frac{\Delta t}{\tau_p}\left(p(t) - p_{\text{target}}\right), \tag{3.60}$$

where τ_p is some characteristic coupling time. To resolve the issue of this volume rescaling not

[16] Although hybrid Monte Carlo is typically used with timesteps larger than those used in pure MD simulations,[122] we find that in order to ensure that the system is well-sampled (*i.e.* that the move acceptance rate is reasonably high), we must utilise both a short MD sequence duration and a short MD timestep in our simulations. Short timesteps are in general required when dealing with ice phases.[125]

being time-reversible, we can choose randomly whether τ_p is positive or negative prior to every MD sequence.[123] To help speed up convergence, Faller and de Pablo suggest that the molecular dynamics target pressure should be chosen randomly from within a range near p_{target}; the Monte Carlo acceptance criterion will still ensure the correct constant pressure behaviour of the overall ensemble.[123] Indeed, this pressure range is required to help maintain reversibility, as we cannot in general predict the pressure of the previous step from the current pressure: a uniform choice of the target pressure in a reasonably large range around p_{target} helps to enable a reasonable scaling factor for the reverse move to be obtained.[126] However, we do note that this is not a guarantee that detailed balance will be obeyed and thus such an approach should be used with caution.

We find that in our simulations, using a time-reversible Berendsen barostat does not appear to help in speeding up volume equilibration, and so we randomly choose between 'ordinary' Monte Carlo volume moves and MD moves in our hybrid Monte Carlo simulations to ensure that the isobaric-isothermal ensemble is being sampled.

3.8.2 Assigning linear and angular velocities

A common way of generating velocities at the desired temperature in molecular dynamics simulations is to choose velocity vector components from a normal distribution with zero mean and unit standard deviation[17] for every particle.[66] We then calculate the kinetic energy,

$$K = \sum_i \frac{|\boldsymbol{p}_i|^2}{2m_i}. \tag{3.61}$$

The instantaneous temperature arising from this kinetic energy is given by equipartition[68] as $k_{\text{B}}T = 2K/f$, where f is the number of degrees of freedom. If we wish to impose an external temperature, we can scale each velocity vector to give $\boldsymbol{v}_i^{\text{scaled}} = \boldsymbol{v}_i\sqrt{T_{\text{ext}}/T}$. Calculating the new

[17] We generate normally distributed random numbers using the Box–Muller approach.[127]

instantaneous temperature gives

$$k_{\mathrm{B}}T_{\mathrm{scaled}} = \frac{2K_{\mathrm{scaled}}}{f} = \frac{2}{f}\sum_i \frac{m_i}{2}\left|\boldsymbol{v}_i\right|^2 \frac{T_{\mathrm{ext}}}{T} = \frac{T_{\mathrm{ext}}}{f}\sum_i m_i\left|\boldsymbol{v}_i\right|^2 \frac{k_{\mathrm{B}}f}{2\sum_i \frac{m_i}{2}\left|\boldsymbol{v}_i\right|^2} = k_{\mathrm{B}}T_{\mathrm{ext}}, \quad (3.62)$$

as required.[66] We thus have a Maxwell–Boltzmann distribution of velocities at exactly the imposed temperature, T_{ext}.

However, in hybrid Monte Carlo, the velocities *must not be scaled* to an externally determined 'correct' temperature in this way. Doing so would mean that we would begin each MD simulation at a particular temperature, but, unless we were to run the MD simulation in an isokinetic ensemble, it is unlikely that the final temperature will always correspond exactly to the imposed temperature. This would clearly violate detailed balance. Instead, we select each velocity from a normal distribution with zero mean, but with the correct standard deviation ($\sigma = \sqrt{k_{\mathrm{B}}T_{\mathrm{ext}}/m}$) as relevant to the Maxwell–Boltzmann distribution at the chosen temperature.

A common implementation of velocity selection in many MD codes is atom-based, whereby each atom is assigned a random velocity corresponding to the Maxwell–Boltzmann distribution. However, for rigid molecules, including most simple water models, this is clearly not wholly appropriate, as a constraints algorithm[18] will immediately dampen out the velocities corresponding to vibrations and the effective temperature will therefore be considerably lower than expected. Instead, angular velocities should be generated and, if necessary, the atomic linear velocities can simply be determined by evaluating the appropriate cross product.

The rotational contribution to the kinetic energy is given by

$$K_{\mathrm{rot}} = \frac{1}{2}\left[I_{xx}\omega_x^2 + I_{yy}\omega_y^2 + I_{zz}\omega_z^2\right], \quad (3.63)$$

[18] Bond distances and angles of rigid molecules must remain fixed after an MD move. This can be achieved either by using true rigid body dynamics, where not only forces but also torques are calculated,[128–130] or by the addition of 'internal forces' that keep bond distances and angles at their predetermined values after a move using a Lagrange multiplier approach.[131] In our simulations of rigid water molecules, we use a quaternion-based rigid body dynamics approach as detailed in appendix D.4.

where ω_j are the components of the angular velocity vector in the principal (molecular) frame of reference and I_{jj} are the principal moments of inertia. We can compare this expression to the linear case,

$$K_{\text{trans}} = \frac{1}{2}\left[mv_x^2 + mv_y^2 + mv_z^2\right], \tag{3.64}$$

and can thus determine that, by analogy, the standard deviation should be $\sigma = \sqrt{k_B T/I_{jj}}$ for each angular velocity component ω_j. Thus, the principal frame angular velocity vector components can be chosen to be of the form $\omega_j = \mathcal{G}\sqrt{k_B T/I_{jj}}$, where $\mathcal{G}$ is a random number drawn from a normal distribution with mean zero and a standard deviation of unity, and is scaled by the desired standard deviation ($\sqrt{k_B T/I_{jj}}$) to correspond to the Maxwell–Boltzmann distribution for the angular velocity.

3.9 Periodic boundaries, tail corrections and long-range interactions

In our simulations, we typically use periodic boundary conditions (figure 3.2) so that no artificial interfaces affect the simulation results; this allows us to simulate 'bulk' systems with finite numbers of particles. In such simulations, we usually choose a cutoff radius r_c for interparticle interactions of less than half the box length to ensure only the nearest image of any particle is considered.[66]

A consequence of truncating particle interactions at a spherical cutoff is that, if the potential is not identically nought at distances greater than r_c, we have introduced an error. However, we can correct for this by using what is known as a tail correction,[66]

$$U^{\text{tail}} = \frac{N\rho}{2}\int_{r_c}^{\infty} U(r)g(r)4\pi r^2\,\mathrm{d}r. \tag{3.65}$$

Generally, r_c will be chosen such that $g(r)$, the radial distribution function, is unity at all $r \geq r_c$. The halving ensures we do not double count every contribution.[66] This tail correction depends

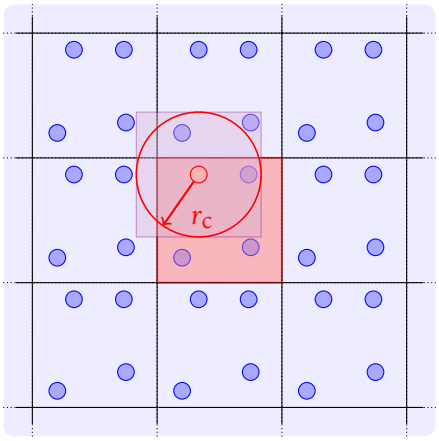


Figure 3.2: In simulations using periodic boundary conditions, we need to ensure that no particle interacts with itself, and a particle interacts with (at most) the nearest image of every other particle. The maximum possible cutoff radius satisfying this is labelled r_c and corresponds to a half of the simulation box size. The simulation box is shaded in red.

on the density, and so it changes with volume moves, but stays constant in canonical ensemble simulations.

The integral in equation (3.65) will only converge for $U(r) \propto r^{-n}$ if $n > 3$. An example of such a short-ranged potential is the Lennard-Jones potential, which scales as r^{-6} and r^{-12}, and the above integral thus converges. However, for potentials $U(r) \propto r^{-n}$ where $n \leq 3$, such as coulombic interactions, the tail correction expression diverges, and we need to consider interactions with all periodic images explicitly. The traditional approach of doing this is to use Ewald summation.[132;133] We describe the basic idea behind the approach below; an accessible derivation is given in the book of Frenkel and Smit.[66]

In an electrically neutral system in a cubic box of size L with positive and negative charges to give a total of N particles, the coulombic contribution to the overall system energy is

$$U^{\text{coul}} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{\boldsymbol{n}}{}' \frac{q_i q_j}{4\pi\varepsilon_0} \frac{1}{|\boldsymbol{r}_{ij} + \boldsymbol{n}L|}, \tag{3.66}$$

where the variable $\boldsymbol{n}$ locates the periodic box and the prime indicates that we do not take into account the particle i itself (*i.e.* the sum is over all particles j and periodic images $\boldsymbol{n}$, except for

$j = i$ if $\boldsymbol{n} = \mathbf{0}$).

The sum in equation (3.66) does not converge absolutely; however, it is possible to rewrite this using a decomposition into three components that we can evaluate. We consider every point charge to be surrounded by a diffuse shell of the opposite charge – typically chosen to follow a normal distribution –, and so at distances far from the point charge, all the charge is screened and there is no coulombic interaction with that point charge. The real space sum is readily evaluated in this way. However, to recover the original problem, we must also remove the screening charges introduced (figure 3.3). Since in a periodic box, the sum of all screening charges represents a smoothly varying periodic function, we can evaluate it in Fourier space instead. Additionally, we need a self-interaction correction term to take into account the fact that the Fourier sum ought not to have included the interaction of each particle with itself. The Ewald sum of coulombic energy is given by[66]

$$\begin{aligned} 4\pi\varepsilon_0 U^{\text{coul}} = {} & \frac{2\pi}{V}\sum_{\boldsymbol{k}\neq 0}\frac{|\rho(\boldsymbol{k})|^2}{|\boldsymbol{k}|^2}\exp\left[-\frac{|\boldsymbol{k}|^2}{4\alpha}\right] && \text{Fourier sum} \\ & - \sqrt{\frac{\alpha}{\pi}}\sum_{i=1}^{N} q_i^2 && \text{self-interaction correction} \\ & + \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}\frac{q_i q_j}{2r_{ij}}\left(1-\operatorname{erf}\left(\sqrt{\alpha}r\right)\right) && \text{real space sum,} \end{aligned} \tag{3.67}$$

where

$$\rho(\boldsymbol{k}) = \sum_{i=1}^{N} q_i \exp(\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}_i) = \sum_{i=1}^{N} q_i\left[\cos(\boldsymbol{k}\cdot\boldsymbol{r}_i) + \mathrm{i}\sin(\boldsymbol{k}\cdot\boldsymbol{r}_i)\right], \tag{3.68}$$

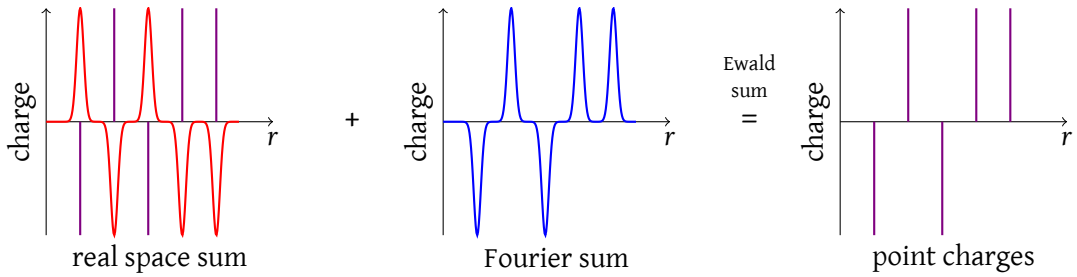


Figure 3.3: A pictorial representation of Ewald summation.[66;134]

$\boldsymbol{k} = 2\pi\boldsymbol{q}/L$, and $\boldsymbol{q}$ are the reciprocal vectors. When using molecular dynamics simulations, these expressions are differentiated when evaluating the force acting on every charge.[135]

4 Water and ice

It should come as no surprise that water is one of the most studied substances in the physical sciences:[136] it is, after all, intrinsically linked to life itself and is fundamental not only to our existence, but also to a huge number of natural processes we encounter every day. It is the only common substance which manifests itself in all three basic states of matter – vapour, liquid and solid – under natural conditions.[29;136] However, despite having been studied since the dawn of civilisation, water continues to surprise us with its myriad of unusual properties.[136]

A water molecule comprises two hydrogen atoms and one oxygen atom with an H-O-H angle of 104.52° (figure 4.1),[137] just short of the tetrahedral angle of $\arccos\left(-1/3\right) = 109.47°$. Its bent shape, which can be rationalised by a simple Walsh diagram using molecular orbital theory,[138] and the different electronegativities of oxygen and hydrogen give it a polar character with a comparatively high dipole moment of 6.2×10^{-30} C m (or 1.85 D).[137] Oxygen's high electronegativity allows water to engage in strong hydrogen bonding with a bond enthalpy of nearly $10\,k_\mathrm{B}T$ at room temperature.[139]

Hydrogen bonding is a major contributor to water's anomalous behaviour when

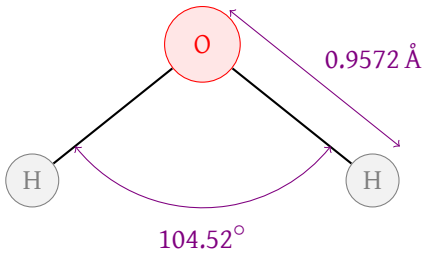


Figure 4.1: The geometry of an unbound water molecule. The angle and bond length were deduced from the rotational components of the rovibrational spectrum.[137]

compared to similar inorganic compounds. For example, water has higher melting and boiling points than H_2S despite the latter's greater molecular mass;[137] water has a negative change in volume upon melting (and hence a negative solid-liquid coexistence line gradient in the p-T phase diagram);[137] it exhibits an increase in diffusivity on isothermal compression;[140] and in its liquid state, there is a density maximum at 4 °C.[137] Such anomalous behaviour is not, however, a unique feature of water; for example, germanium and silicon metals crystallise into tetrahedron-based solids with a density lower than that of the melt,[137;141] while In_2Te_3 exhibits a density maximum above its freezing point.[137] Many aspects of this anomalous behaviour are not caused by hydrogen bonding directly, but by the tetrahedrality that, in water, arises as a consequence of the hydrogen bonding.[141]

There are many ways in which water molecules can form a hydrogen-bonded network; this results in a complex solid-state phase diagram for water (figure 4.2). The amount of hydrogen bonding taking place is significant in the liquid phase as well. We note that the enthalpy of fusion of ice at atmospheric pressure is 6.01 kJ mol^{-1},[137] whilst the enthalpy of vaporisation at the triple point is 51.059 kJ mol^{-1};[144] in the latter process, hydrogen bonds

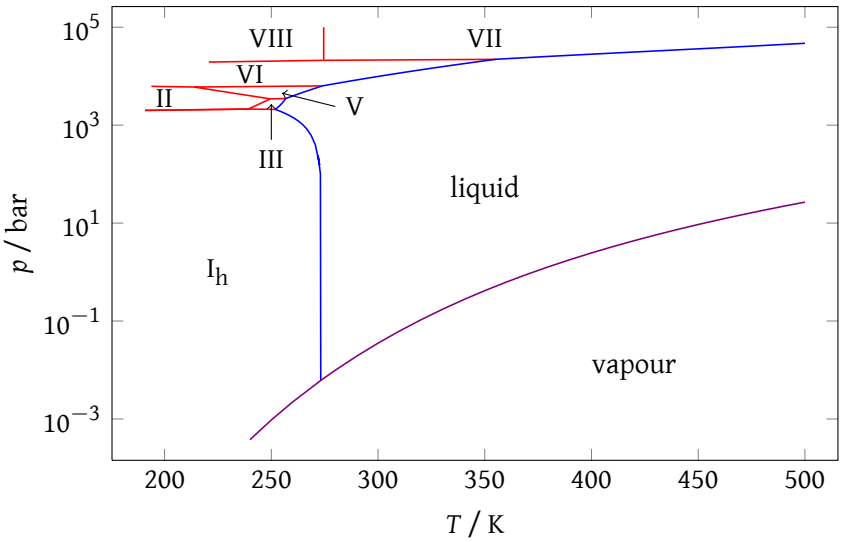


Figure 4.2: A region in the experimental phase diagram of water;[142;143] water is somewhat unusual in having a large number of distinct solid phases that are thermodynamically stable under different conditions.

clearly must be broken, but the low enthalpy of fusion suggests that the majority of hydrogen bonds survive in the liquid phase, although perhaps on a transient timescale.[29]

In this chapter, we look at some properties of supercooled water (section 4.1) and of cubic and hexagonal ice (section 4.2). We describe some of the effects of electric and magnetic fields on ice nucleation (section 4.3) and summarise the experimental approaches to studying homogeneous ice nucleation (section 4.4). Finally, we look at how water can be treated in computer simulations (section 4.5), in particular at water models (subsection 4.5.1) and previous simulations of ice nucleation (subsection 4.5.2).

4.1 Supercooled water

As discussed in section 2.1, there is a free energy barrier associated with nucleation. This means that very pure water can be cooled considerably below its freezing point: at atmospheric pressure, small water droplets can be cooled down to approximately $-45\,^{\circ}\mathrm{C}$ without freezing,[5] and at pressures of about 2000 bar, water can be supercooled to temperatures in the region of $-70\,^{\circ}\mathrm{C}$.[5] The temperature at which liquid water spontaneously freezes under certain conditions is known as the homogeneous nucleation temperature.[29] The structure of liquid water between the homogeneous nucleation temperature and the freezing temperature is not qualitatively different from the structure of thermodynamically stable water; however, the hydrogen bonding becomes more pronounced as the temperature is decreased.[145]

Below about $-137\,^{\circ}\mathrm{C}$, melting amorphous (glassy) ice also yields a liquid, the so-called 'ultraviscous water'.[29] It is unclear whether this liquid is in some way different from ordinary supercooled water, and the presence of a liquid-liquid critical point is a contentious issue.[29;146–149] Above this temperature, but below the homogeneous nucleation temperature, liquid water is difficult to obtain for an appreciable amount of time,[1] and the region is sometimes

[1] Under unusual conditions, such as when tiny liquid clusters are cooled in supersonic flow,[150] water can remain

known as a 'no-man's land'.[29] However, this difficulty is in principle only a practical limit, since the absence of homogeneous nucleation is a kinetic phenomenon, not a thermodynamic one: in practice, it is simply unfeasible to perform measurements in the very short time between cooling and freezing.[29]

4.2 Water in the solid state

4.2.1 Structure of ice I

As depicted in the phase diagram of water (figure 4.2), water can adopt a variety of thermodynamically stable solid structures: they differ in the positions of the oxygen atoms, the lattice, whether ice networks inter-penetrate, and whether the protons are ordered or disordered. In addition, water also forms several amorphous solid phases (glasses) at low temperatures. A recent review of the complex phase behaviour of ice was written by Salzmann and co-workers;[151] a review of the current state of the art in ice science more generally is also available.[152]

In this work, however, we are concerned predominantly with the behaviour of water under atmospheric conditions, where only the ice I phase is relevant. However, ice I exists in two modifications; these are known as hexagonal ice (I_h), the thermodynamically stable phase, and cubic ice (I_c), which is metastable with respect to hexagonal ice. Hexagonal ice adopts the lonsdaleite structure and cubic ice the diamond structure in the oxygen atom arrangement.[137] In other words, cubic ice is a cubic close packing (with an ABCABC. . . sequence of layers, *i.e.* a face-centred cubic structure) of oxygens with additional oxygens in one half of the tetrahedral holes; hexagonal ice is its hexagonal analogue (with an ABABAB. . . sequence of layers). An equivalent alternative view is that the ABABAB. . . sequence of hexagonal ice is of 'double' layers

in the liquid state even at these temperatures. However, such 'nanoparticles' of substances often have different thermodynamic behaviour from the bulk, as we discuss in section 4.4.

of two inter-penetrating hexagonal close packing structures, or cubic close packing structures in the case of cubic ice.[154] In terms of the simple ionic structures, cubic ice has the sphalerite (zinc blende) structure and hexagonal ice the wurtzite structure (figure 4.3), with both zinc and sulphur replaced by oxygen. In a polyhedron representation, the (inter-penetrating) OO_4 tetrahedra share all four corners with neighbouring tetrahedra.[153] The hydrogens are then placed such that they satisfy the Bernal–Fowler ice rules (see below); this can be achieved in a variety of ways, and the proton disorder that results is an important feature of many phases of ice, leading to a residual entropy even at absolute zero.[155] Representative structures of hexagonal and cubic ice are depicted in figure 4.4. At a more local level, hexagonal ice comprises layers of six-rings of oxygens in the chair form, and in the plane perpendicular to this, layers of

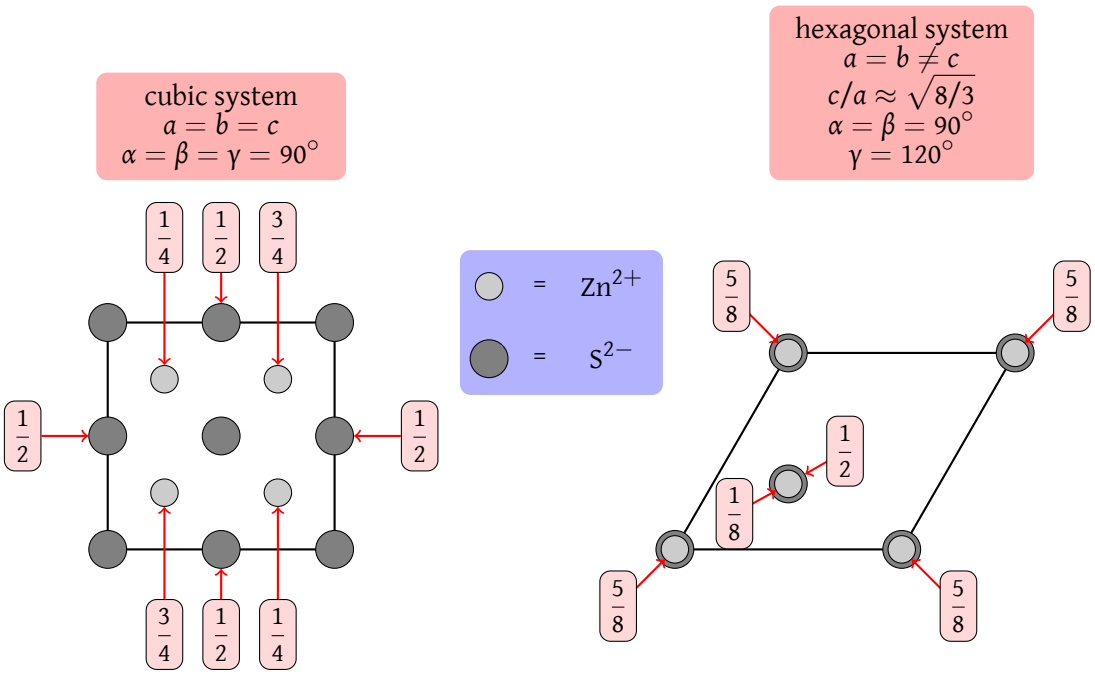


Figure 4.3: The sphalerite (left) and wurtzite (right) structures of zinc sulphide in plan view along the c axis.[153] Heights above the $z = 0$ plane are expressed as fractions of c. If both Zn and S are replaced by oxygen, these two structures represent cubic and hexagonal ice, respectively. In hexagonal ice, the c/a ratio is 1.62806, similar to the ideal tetrahedral value of 1.63299, and is almost independent of temperature.[137] The lattice parameters are $a = 4.519$ Å and $c = 7.357$ Å at 253 K for hexagonal, and $a = 6.358$ Å at 78 K for cubic ice.[137]

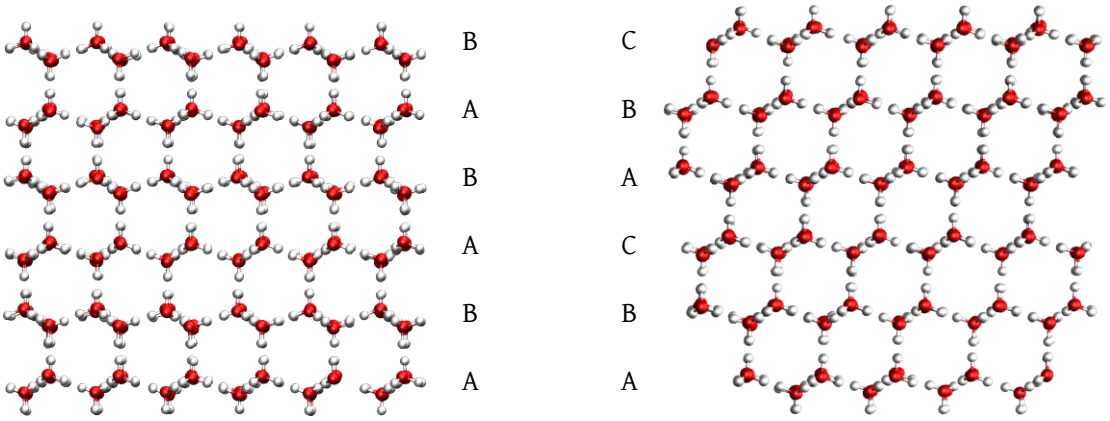


Figure 4.4: A side-by-side comparison of hexagonal (left) and cubic (right) close packing in the respective forms of ice I. Oxygens are depicted in red and hydrogens in white. While oxygens are regularly ordered, there is noticeable proton disorder.

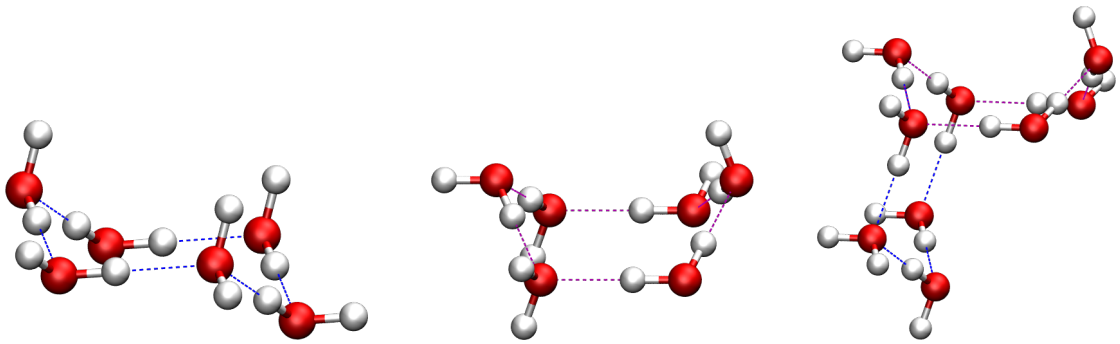


Figure 4.5: Representative extracts of atoms from a structure of hexagonal ice. Both chair (left) and boat (centre) arrangements of oxygens can be extracted from hexagonal ice, and these can be seen to be the components of larger structures (right). On the other hand, there are no boat arrangements of oxygens in cubic ice. Oxygens are depicted in red and hydrogens in white.

oxygens in the boat form (figure 4.5). By contrast, cubic ice has no oxygens arranged in the boat form.

The basal (0001) planes of hexagonal ice and the (111) planes of cubic ice are identical in the local environment:[154;156] in other words, following a hexagonal stacking of layers, ABABAB, the next layer can easily be displaced and we can form ABABABC and so forth with little additional strain introduced. Such an occurrence is termed a stacking fault.[156;157] The microscopic mechanism of stacking fault formation has been investigated by Pirzadeh and

Kusalik using computer simulation.[158]

A hydrogen involved in a hydrogen bond does not lie at the midpoint of the two oxygens it is bonded to: the intramolecular O-H covalent bond is much shorter than the intermolecular H-O hydrogen bond. In such a hydrogen bond, the molecule to which hydrogen is bonded covalently is termed the hydrogen bond donor, and the other molecule participating in the hydrogen bonding is termed the hydrogen bond acceptor.[137] A water molecule engaged in full hydrogen bonding with four neighbours is thus a hydrogen bond donor in two of the hydrogen bonds and a hydrogen bond acceptor in the remaining two. Although the bond angle in water is less than the tetrahedral angle, the arrangement of water molecules in 'perfect' ice I is in fact nearly ideally tetrahedral; the hydrogen bonds are not completely straight.[137;152]

4.2.2 Ice rules

In ideal ice I,

- each molecule's oxygen lies at the centre of a tetrahedron and its hydrogens point towards two corners of the tetrahedron,
- two hydrogens from adjacent oxygens are hydrogen-bonded to this central oxygen,
- there is only one hydrogen between any two adjacent oxygen atoms, and
- the probability of any configuration satisfying the above criteria is equal.

These conditions are known as the Bernal–Fowler ice rules.[155;159] There are six orientations (figure 4.6) for a given water molecule that satisfy the hydrogen bonding requirements; this leads to a residual entropy in ice associated with the multiple configurations permissible by the ice rules of $S_m(0) \approx R \ln\left(3/2\right) \approx 3.4\,\mathrm{J\,K^{-1}\,mol^{-1}}$.[40;155] In fact, below about 72 K, a proton ordered form of ice I_h, ice XI, is stable relative to hexagonal ice,[137] although ice I is typically not spontaneously converted into ice XI except under special conditions, since the transformation into ice XI requires defects which violate the ice rules and are energetically disfavoured.[137] Algorithms

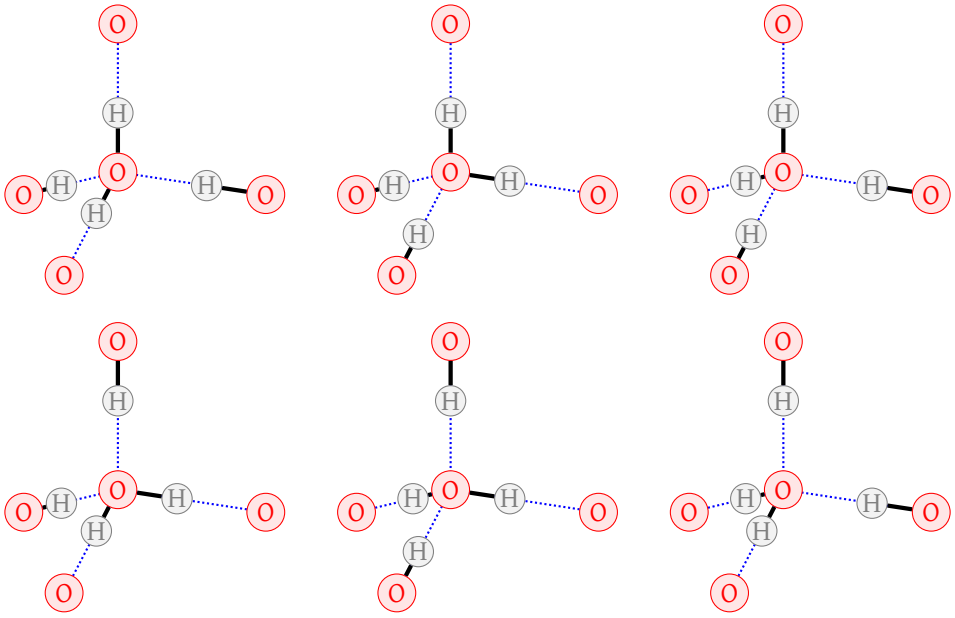


Figure 4.6: The six possible orientations of hydrogens around a water molecule with respect to its four neighbours satisfying the Bernal–Fowler ice rules. Oxygens are shown in red and hydrogens in grey; covalent bonds are shown with a thick black line.

for constructing proton disordered ice I configurations in simulations are available.[160;161]

In real ice, the ice rules are not completely satisfied. At temperatures above absolute zero, entropically-driven Schottky and Frenkel defects naturally occur,[153] and similar considerations lead to orientational Bjerrum defects (figure 4.7), where there are two hydrogens (D defects) or no hydrogens (L defects) along an O-O connection.[137] Ionic defects, where a pair of H_3O^+ and OH^- ions is formed in the lattice, and extrinsic defects due to impurities can also occur in real ice.[137;152]

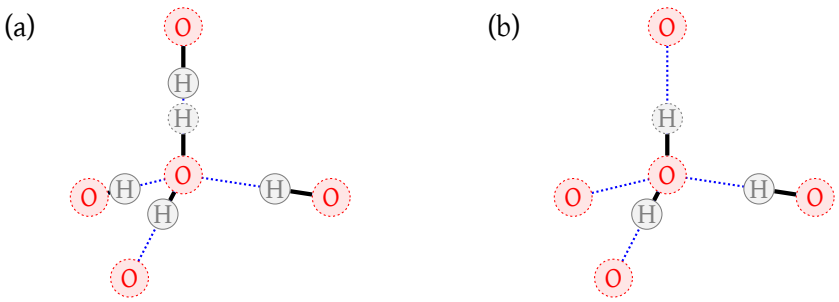


Figure 4.7: Bjerrum defects in ice: (a) depicts a D (*doppelt*) defect and (b) depicts an L (*leer*) defect.

4.2.3 Hexagonal and cubic ice

At atmospheric pressure below the freezing point, the thermodynamic form of water is hexagonal ice, but, as stated above, an important metastable phase in these conditions is cubic ice. Cubic ice appears to form as the predominant (if metastable) phase between 130 K and 150 K;[29] it can be formed for example by warming amorphous[29] or high-pressure phases (ices II to IX) in liquid nitrogen, by water vapour expansion at appropriate temperatures[162–164] or by rapid quenching of aqueous aerosol droplets on a cryoplate.[165] It is gradually converted into hexagonal ice between about 160 K and 240 K.[154;166]

Due to the ease of stacking faults developing in hexagonal ice (see above), it has been suggested that pure cubic ice may in fact not exist at all[167] or may be much rarer than previously thought.[168] Especially ice grown in mesopores is frequently identified as being cubic; however, the additional peaks observed in X-ray or electron diffraction patterns in some situations may simply reflect pronounced stacking defects in hexagonal ice.[167] Indeed, diffraction patterns can be deceiving: Moore and Molinero have studied the theoretical diffraction patterns of the ice formed in molecular dynamics simulations of the mW model at 180 K.[169] They find that cubic and hexagonal ice form in layers in an approximate (but non-uniform) ratio of 2 : 1. Despite this, the calculated diffraction pattern is almost the same as that of cubic ice, with the characteristic hexagonal peaks being absent.[169] The experimental results in which cubic ice was determined to form using diffraction methods must therefore be interpreted with some caution. Kuhs and co-workers have studied the structural characteristics of experimental ice and have established that even if cubic ice may involve many stacking faults, such stacking faults are not random and their proportion increases with temperature.[170] Nevertheless, cubic ice appears to have been observed unambiguously in other experiments without using X-ray diffraction patters to aid in the recognition process, such as using atomically resolved images of single crystals in the solid phase using electron microscopy[171] or scanning tunnelling microscopy.[172] Both

forms of ice, as well as 'mixed' forms, have been observed in nucleation experiments in various conditions.[168;173–176]

Cubic ice is thought to play an important rôle in certain atmospheric processes.[175;177–180] Cubic ice is always the metastable form; however, the experimentally measured enthalpy change on the conversion of cubic into hexagonal ice is only between 13 and 50 J mol^{-1}.[137;181] Curiously, relative humidities[2] greater than 100 per cent have been observed in aircraft measurements in the atmosphere even in apparently cloud-free regions;[182] furthermore, supersaturations over 30 % persist within ice clouds.[178] Cubic ice is the less stable form of ice and has an approximately 10 % higher vapour pressure than hexagonal ice.[179] It has therefore been suggested that cubic ice may form in preference to hexagonal ice during homogeneous nucleation in the upper atmosphere;[175] this has been corroborated by several nucleation experiments mimicking atmospheric conditions (see section 4.4). However, although cubic ice forms as the predominant phase below 190 K,[175] which is the temperature within polar stratospheric clouds and clouds in the tropical tropopause region,[175] as the crystal nucleus grows, some heat associated with crystallisation is generated; because heat dissipation in the atmosphere is slow, larger nuclei are almost invariably annealed into hexagonal ice or converted to hexagonal ice by nucleation on the surface of cubic ice nuclei.[175] Nevertheless, cubic ice has been invoked in explanations of certain atmospheric phenomena, such as rationalising what is known as the Scheiner halo around the Sun, which may occur due to the refraction of light passing through the octahedral cubic ice crystals in the atmosphere,[183;184] and the growth of certain snow flakes.[156;185]

Murray and Bertram suggest that the initial homogeneous nucleation of ice generally occurs in the form of cubic ice; larger droplets crystallise into hexagonal ice that may grow on

[2]The relative humidity is defined as the vapour pressure of liquid water divided by the equilibrium vapour pressure over hexagonal ice. Relative humidities over 100 % thus imply the formation of cubic ice or ice incorporating impurities such as HNO_3 that increase its vapour pressure.[175]

the surface of a cubic nucleation seed, in agreement with Ostwald's step rule.[176] This behaviour is consistent with Takahashi's model of broken hydrogen bonds,[156] which suggests that cubic ice has a lower interfacial free energy in contact with liquid water than any of the I_h surfaces. This idea has been used to show that small droplets (below about 15 nm in radius) will favour cubic ice.[186;187]

Finally, although cubic ice is always the metastable form experimentally, this is not necessarily the case in computer simulations. In the TIP4P model of water,[188] cubic ice may, depending on the precise conditions of the simulation, be more stable than hexagonal ice,[189] although the energy difference may be less than the measurement uncertainty.[190] Interestingly, in grand canonical Monte Carlo simulations of TIP4P/2005 water, ice layers that grew heterogeneously on a surface were not typically hexagonal ice even if the underlying surface was crystallographically commensurate with hexagonal ice.[191]

4.3 Electric and magnetic fields

Whether electric fields play a rôle in determining the rate of ice nucleation is a matter of some contention. Initial simulations indicated that the rate of nucleation is significantly increased in a strong (5×10^9 V m^{-1}) static homogeneous electric field,[192;193] and electric fields have also been used to rationalise the greater rate of nucleation seen in simulations near surfaces, where an orientational ordering of molecules at the surface produces a net electric field,[194;195] and in heterogeneous nucleation.[196] The motivation for electric field induced nucleation comes from the observation that freezing can experimentally be catalysed by local electric fields when water is in contact with chiral amino acid crystals, even when their surfaces cannot catalyse nucleation heterogeneously.[197] Although more recent experimental measurements suggest that low- and medium-strength electric fields (up to 1.6×10^5 V m^{-1}) do not affect the rate of homogeneous nucleation,[198] it is difficult to realise the strong fields studied in simulations

in macroscopic experiments,[196;198] so these measurements do not necessarily invalidate the simulation results. Some heterogeneous freezing experiments indicate an increase in freezing propensity when surface electric fields of the order of 10^7 V m^{-1} are involved; this is sometimes known as the electrofreezing effect.[199–201] Due to the contrasting experimental evidence, it would be very interesting to elucidate exactly how much of an effect electric fields really have in simulations. The simulations of Svishchev and Kusalik[192;193] used extremely small simulation boxes in which crystallisation is much more tractable due to the considerable finite size effects; similarly, subsurface nucleation can also be rationalised by the greater diffusion the molecules enjoy near the surface,[194] and so the simulation results can be explained without necessarily involving electric fields. Furthermore, the phase diagram of classical models of water is slightly perturbed in electric fields,[202;203] and dipole scaling must be used to obtain good agreement with experiment.[203] However, Yan and Patey have recently simulated the nucleation of ice using all-atom models of water on a surface with a rapidly decaying electric field (of the order of 10^9 V m^{-1} in strength) and have observed that such strong microscopic electric fields act as a nucleating force;[204] they suggest that certain heterogeneous nucleants such as kaolinite induce nucleation through electric fields rather than because they provide a good structural match to the ice lattice.[196;204] Interestingly, although water has no permanent magnetic moment, magnetic fields of 0.5 T have also been reported to change the rate of nucleation significantly;[205] another recent simulation study has looked at the combined effect of electric and magnetic fields and has observed that the field strengths required to facilitate ice growth depend strongly on the crystal plane of ice that is exposed.[206] Oscillating electric fields, by contrast, can inhibit nucleation, perhaps by disturbing the water molecules as they begin to form an ice nucleus.[207;208] There appear to be many questions that remain unanswered in this area, and further experimental and theoretical investigation into the effects of both electric and magnetic fields on nucleation is undoubtedly warranted.[208]

4.4 Ice nucleation experiments

As nucleation is a rare event, it is difficult to study not only computationally, but also experimentally; indeed, Oxtoby points out that many experiments that have been performed seem not to be independently reproducible.[7] Homogeneous nucleation is particularly sensitive to a large number of factors; provided that experiments are well-designed, it can thus act as a 'very sensitive probe of microscopic events'.[12] For example, in the heteromolecular homogeneous nucleation of H_2SO_4 and H_2O, the addition of 5 parts per trillion of NH_3 increases the nucleation rate by 16 orders of magnitude.[209]

In most experimental studies, homogeneous ice nucleation is studied in small water droplets. The primary reason for this is that water must be very pure to ensure that homogeneous nucleation is being observed, and separating the total volume into small droplets means that most of these droplets will be free of impurities: those that are not will freeze early on (typically by at least an order of magnitude of separation in time) and can be assumed to have undergone heterogeneous nucleation.[12;176;210] The experimental rate of nucleation depends on the volume of the droplets, as for larger droplets, the statistical probability of a spontaneous nucleation event taking place within the volume of the droplet is correspondingly higher. However, extremely small water droplets can exhibit different thermodynamic behaviour from the bulk due to the enhanced rôle of surfaces;[211] this is sometimes known as the Gibbs–Thomson effect.[212] Considerable effort has been expended studying aqueous solutions and water in confinement in order to study processes at temperatures where nucleation intervenes. However, such results must be interpreted with great care, as the behaviour of confined or non-pure water can differ considerably from bulk supercooled water.[213] The minimum droplet size that still exhibits 'bulk' nucleation behaviour is unclear. There have also been some suggestions in the literature that 'homogeneous' ice nucleation as studied by many experimental groups is not necessarily always truly homogeneous, as it is extremely difficult to remove any potential

heterogeneous nucleant and indeed to ascertain that such nucleants have been removed.[61;214]

Many experiments aimed to mimic clouds, such as by studying nucleation in aerosols or droplets suspended in air by electrodynamic traps. For example, Hagen and co-workers studied nucleation in an expansion cloud chamber, in which moist gas is first expanded and then cooled. Water droplets were nucleated from the gas phase, and these were then observed, using photography, to freeze.[26;215] The volume of the liquid droplets was estimated from previous vapour-liquid nucleation[215] and droplet growth results.[26] Wood and co-workers studied relatively monodisperse droplets produced by an inkjet printer cartridge that were allowed to fall under gravity in a tube that was warm at the top and cold at the bottom; the height at which freezing occurs (tracked by backscattered light depolarisation) was then mapped to a temperature.[216] Earle and co-workers studied cryogenic laminar aerosol flow in a vertical tube with infrared spectroscopy used to monitor freezing.[217;218] Finally, Baumgärtel, Leisner and co-workers studied supercooled levitated microdroplets of water (and heavy water) in electrodynamic traps, determining the size and phase of the droplet by studying the depolarisation of Mie scattering from a He-Ne laser.[174;219–224] To levitate droplets, small amounts of charges are introduced on the surface of the droplets, but these are said not to affect the rate of nucleation.[224]

Often, to minimise the effect of external surfaces, water droplets were studied in emulsions. For example, Wood and Walton studied water droplets in an oil emulsion which they cooled in a copper cell with liquid nitrogen via a copper bar, and they then regulated the temperature using an additional heater.[210] Similarly, Butorin and Skripov studied nucleation in an oil emulsion.[225] They repeatedly melted and refroze droplets to improve statistical confidence,[225] which has been shown to be very important in nucleation experiments.[226] Both Michelmore and Franks[227] and Taborek[228] studied nucleation in oil emulsions with sorbitan tristearate acting as the emulsifier; nucleation rates were determined using calorimetry in both

cases. Taborek was able to supercool the water droplets before they froze to a larger extent than many other studies; it was later suggested[27;221;222] that the greater supercooling achieved was due to the surfactant used to create the oil-water emulsion 'inhibiting' the nucleation: this is a curious claim, however, given that the homogeneous nucleation temperature is generally interpreted to be the lowest possible one.[229]

Some experiments also studied water placed on a surface;[230–232] this is often chosen to be a hydrophobic surface, such as a polyethylene foil, in order to minimise the amount of contact.[230] Edd and co-workers studied monodisperse droplets (whose diameter varied by less than 2 %) trapped in a microfluidic device created using soft lithography; the array of droplet traps of the device was wetted by a fluorinated oil, and the device was then cooled on a cryostage.[231] One difficulty in this set-up is that the time lag between the cooling of the cold stage and the temperature reaching the desired value within the water droplets must be taken into account; the authors suggest that this, while corrected for to a degree, may have led to a systematic error in the results.[231] Another microfluidic device to study homogeneous ice nucleation has recently been devised by Stan and co-workers.[233] They suspended monodisperse water droplets in a moving stream of liquid fluorocarbon (perfluoromethyldecalin with a water-insoluble surfactant) cooled in-flow and measured the temperature with linear arrays of thermometers on both the top and the bottom of the flow channel; they found nucleation rates in general agreement with electrodynamic trap experiments.[233] In many of these studies, the point at which droplets froze was recorded by a microscope coupled to a high-speed camera.[231–233]

At lower temperatures, smaller droplets were generated in supersonic de Laval nozzles.[150;173;234] In these experiments, nucleation was tracked by electron diffraction[150;173] and by FTIR coupled with small-angle X-ray scattering.[234] Manka and co-workers established that nucleation in their set-up always proceeds via the liquid phase; there is never any direct

vapour-ice transition observed.

Good historical surveys of the earliest experiments of ice nucleation have been provided by McDonald,[235] Mossop[236] and Mason,[237] and some more recent analyses of the available experimental data have been written by Pruppacher[27] and by Jeffery and Austin.[5]

It is instructive to examine the experimental predictions for nucleation rates and the interfacial free energies. We summarise some recent results in table 4.1, giving the relevant temperature range and water droplet size, as well as, where available, predictions for the nucleation rate and interfacial free energy. The rate estimates in table 4.1 are provided at a temperature of 237 K, which was chosen because it is within the temperature range of the majority of the experimental conditions. However, even where empirical relationships for the rate are provided (table 4.2), it is important to realise that such equations are only applicable in the region of the experiment. For example, many of the early experiments would predict extremely high nucleation rates if extrapolated to, say, 200 K, but Huang and Bartell's experimental rates are considerably lower (approximately 10^{30} m^{-3} s^{-1} at 200 K). As Oxtoby remarked, nucleation is 'one of the few areas of science in which agreement of predicted and measured rates to within several orders of magnitude is considered a major success.'[7] Indeed, the wide spread of the rates summarised in table 4.1 suggests that agreement within several orders of magnitude is a considerable success in experimental work itself! It has been suggested that the current state of the art experimental techniques should be considered to have an uncertainty in the reported rate of two orders of magnitude;[238] Murray and co-workers suggested that the principal reason for this great variation in reported rates is the uncertainty in temperature measurements,[232] and this has recently been corroborated by the experiments of Riechers and co-workers.[239]

The spread of experimental nucleation rates is, however, greater even than two orders of magnitude. There are several possible reasons why different experimental studies have led

Study	T/K	Droplet diameter/m	Rate/$m^{-3}\,s^{-1}$	γ/mJ m^{-2}
Thomas[245] 1952	232.7	assumed 5×10^{-5}	—	30.8
Wood[210] 1970	235.9 to 237.6	2×10^{-6} to 3×10^{-5}	3.6×10^{12}	24.2
Butorin[225] 1972	235.2 to 239.1	2×10^{-5} to 5×10^{-4}	$^{*}8.75 \times 10^{10}$	28.7
Hagen[26;215] 1980, 1981	228 to 233	2×10^{-9} to 4×10^{-7}	4.26×10^{16}	—
Michelmore[227] 1982	233 to 236.5	1×10^{-6} to 1.5×10^{-5}	1.8×10^{11}	—
Taborek[228] 1985	235 to 240	3×10^{-6} to 3×10^{-4}	1.9×10^{10}	28.3
Stoyanova[230] 1994	249 to 251	$\sim 1.2 \times 10^{-3}$	3×10^{10}	11.2
Bartell[150;173] 1994, 1995	200	$\sim 6 \times 10^{-9}$	—	(I_c) 21.6
Pruppacher[27] 1995	229 to 244	—	$^{*}10^{13}$	—
Baumgärtel, Leisner[174;219–224] 1999–2012	~237	3×10^{-5} to 1×10^{-4}	6.3×10^{11}	29.2
Wood[216] 2002	235 to 237	$\sim 7 \times 10^{-5}$	1.1×10^{13}	—
Edd[231] 2009	235 to 238	5×10^{-5}	5.2×10^{8}	33.4
Murray[232] 2010	234.9 to 236.7	1×10^{-5} to 4×10^{-5}	1.9×10^{12}	(I_c) 20.8
Earle[217;218] 2010, 2011	234.8 to 236.2	2×10^{-6} to 6×10^{-6}	8.6×10^{11}	—
Manka[234] 2012	202 to 215	6×10^{-9} to 1.2×10^{-8}	—	15.6
Riechers[239] 2013	236.5 to 237.9	5.3×10^{-5} to 9.6×10^{-5}	2.3×10^{11}	—

Table 4.1: Summary of experimental data. The reported rates are at 237 K, calculated either using the reported fits to the experimental data or, where denoted by an asterisk (*), by our fits to the experimental data reported in the relevant study. These fits are given in table 4.2. The interfacial free energies were generally estimated using classical nucleation theory fits and relate to the temperature of the experiments.

Study	Rate
Wood[210] 1970	$J/\mathrm{m}^{-3}\,\mathrm{s}^{-1} = 10^{52.5}\exp\left(-1.6\times10^{12}\,\mathrm{K}^5/(T^3(T_{\mathrm{fus}}-T)^2)\right)$
Butorin[225] 1972	$^*\log_{10}(J/\mathrm{m}^{-3}\,\mathrm{s}^{-1}) = 322.96 - 1.3165T/\mathrm{K}$
Hagen[26;215] 1980, 1981	$J/\mathrm{m}^{-3}\,\mathrm{s}^{-1} = 5.92\times10^{38}T/\mathrm{K}\times\exp\left(1.260\times10^5/(T/\mathrm{K}) - 588.1\right)$
Michelmore[227] 1982	$J/\mathrm{m}^{-3}\,\mathrm{s}^{-1} = 4\times10^{49}\times\exp\left(-1.010T_{\mathrm{fus}}^5/(T^3(T_{\mathrm{fus}}-T)^2)\right)$
Taborek[228] 1985	$\log_{10}(J/\mathrm{cm}^{-3}\,\mathrm{s}^{-1}) = 365.7 - 1.525T/\mathrm{K}$
Stoyanova[230] 1994	$\ln(J/\mathrm{cm}^{-3}\,\mathrm{s}^{-1}) = 15.2 - (8.5\times10^{10}\,\mathrm{K}^5/(T^3(T_{\mathrm{fus}}-T)^2))$
Pruppacher[27] 1995	$^*\log_{10}(J/\mathrm{cm}^{-3}\,\mathrm{s}^{-1}) = -0.0066(T/\mathrm{K})^3 + 4.6(T/\mathrm{K})^2 - 1083T/\mathrm{K} + 84597$
Baumgärtel, Leisner[174;219–224] 1999–2012	$\log_{10}(J/\mathrm{cm}^{-3}\,\mathrm{s}^{-1}) = -1.46(T-T_{\mathrm{fus}})/\mathrm{K} - 46.98$
Wood[216] 2002	$\log_{10}(J/\mathrm{m}^{-3}\,\mathrm{s}^{-1}) = 6 - 1.8154(T-T_{\mathrm{fus}})/\mathrm{K} - 58.589$
Edd[231] 2009	$\log_{10}(J/\mathrm{pL}^{-1}\,\mathrm{s}^{-1}) = -1.912(T-T_{\mathrm{fus}})/\mathrm{K} - 75.4$
Murray[232] 2010	$\ln(J/\mathrm{cm}^{-3}\,\mathrm{s}^{-1}) = -2.92T/\mathrm{K} + 706.5$
Earle[217;218] 2010, 2011	$J/\mathrm{m}^{-3}\,\mathrm{s}^{-1} = 6.98\times10^{38}T/\mathrm{K}\times\exp\left(-832.108 + 181228/(T/\mathrm{K})\right)$
Riechers[239] 2013	$\ln(J/\mathrm{cm}^{-3}\,\mathrm{s}^{-1}) = -3.549(T/\mathrm{K} - 235) + 19.44$

Table 4.2: Summary of rate equations used to obtain the rate estimate in table 4.1, as reported in the relevant study or, where denoted by an asterisk (*), by our fits to the experimental data reported.

to such hugely different nucleation rates. While even small differences in water purity, for example, can lead to vastly different results, an alternative rationalisation of this behaviour is that nucleation happens at the surface of the liquid droplets rather than in the bulk in certain experimental studies.[240] It has been suggested that for tetrahedral liquids in general, surface crystallisation is favoured,[241] and in particular, the nucleation of ice in clouds occurs at or near the surface of liquid droplets.[240] Although Sigurbjörnsson and Signorell suggest that current experimental techniques cannot reliably distinguish between the two modes of nucleation,[238] surface nucleation has been shown to be preferred in simulations of ice nucleation in small systems with the Nada–Van der Eerden 6-site water potential.[194;195;242] By contrast, careful monitoring in experiments has not shown surface freezing in ice nucleation in electrodynamic traps.[220;221] The size of the water droplets that undergo freezing may play an important rôle in this regard,[221] with smaller droplets favouring surface-induced nucleation; Duft and Leisner suggest that surface-based nucleation cannot be dominant in drops larger than 4×10^{-6} m in diameter,[220] and this is corroborated experimentally by Earle and co-workers,[217;218] who suggest that surface-based nucleation dominates for droplets whose diameters are less than $\sim 1 \times 10^{-5}$ m.[218] It has also been shown that acoustic fields, such as in experiments studying the rate of nucleation in an ultrasonic acoustic trap, appear to favour surface nucleation.[243;244] However, while many of the more recent experimental studies are careful to distinguish between the two modes of nucleation, most of the older literature simply assumed that nucleation happens in the bulk, and so particular care should be taken when interpreting such data.

It is worth bearing in mind that although the studies performed in laboratory conditions were generally undertaken at atmospheric pressure, several experimental studies mimicking nucleation in clouds used expansion chambers, and it is especially unclear what the pressure in such 'atmospheric' experiments was. Furthermore, small droplets of water can have a higher internal (Laplace) pressure than their surroundings.[40] The differences in the experimental

nucleation rates, however, are so great that a relatively small pressure change would not likely be sufficient to rationalise such discrepancies.

Homogeneous ice nucleation rates have also been measured in real clouds,[246;247] although such data cannot easily be interpreted as being related to the nucleation of pure water, as there are many other potential nucleants in the atmosphere. The effect of electrolytes, often called soluble cloud condensation nucleants in atmospheric science,[247] is an interesting problem in its own right. It is important to bear in mind that these external species do not necessarily imply a heterogeneous nucleation pathway, as they are not macroscopic crystals obviating the need for a critical nucleus to grow spontaneously, but can simply change the homogeneous nucleation pathway. Homogeneous ice nucleation as a function of water activity has been studied,[14;28;248;249] and it has been suggested that the water activity, defined as the ratio of the vapour pressure of the solution to the vapour pressure of pure liquid water, affects nucleation in a similar way to changing the pressure.[28] However, pure water nucleation is nevertheless of importance in atmospheric science, in particular in cumulus clouds: even though water droplets may initially form from the vapour phase on soluble condensation nucleants, the droplets reach a size so large before ice nucleation occurs that the resulting solution is so dilute as to be effectively pure water.[250]

The experimental rates of ice nucleation tend to be several orders of magnitude higher than those obtained using classical nucleation theory,[174] although the comparison is somewhat dubious without an independent measure of the interfacial free energy. It has been suggested that a speed-up can perhaps be attributed to the fact that, at odds with the CNT premise, it is unlikely that the liquid phase molecules are uncorrelated: the liquid is a network of hydrogen bonded water molecules, and small rearrangements in the liquid phase may rapidly become ice-like.[222] However, this highly ordered hydrogen-bonded liquid structure, especially at lower temperatures, slows diffusion considerably; so much so, in fact, that it has been suggested that

the Gibbs energy of diffusion activation corresponding to the transfer of water molecules across the ice-water interface may be rate-determining.[6;27;222] However, other experimental work suggests the temperature dependence of the nucleation rate primarily reflects the behaviour of the free energy barrier of nucleation rather than diffusion activation.[174]

The ice-liquid interfacial free energy, also shown in table 4.1, likewise varies considerably amongst the experimental studies, from as low as 11.2 mJ m^{-2} to as high as 33.4 mJ m^{-2}.[230;231] In the majority of these studies, the interfacial free energy was determined by fitting experimental rates to classical nucleation theory, and so the large variation in the resulting values for γ is not surprising. However, values for the interfacial free energy thus obtained should be taken with a pinch of salt. Murray and Bertram have reanalysed several studies of homogeneous ice nucleation on the assumption that ice I_c nucleates in preference to ice I_h;[232] this can lead to an appreciable difference in the estimated value: for example, their own data give rise to an interfacial free energy of 26.8 mJ m^{-2} on the assumption of hexagonal ice nucleation as opposed to 20.8 mJ m^{-2} that they reported for cubic ice.[232] In some nucleation studies, the variation of the interfacial free energy as a function of the temperature was studied; for example, Wood and Walton found[210] that their agreement with classical nucleation theory is very good if the interfacial free energy is taken to decrease linearly (with the constant of proportionality being $\sim$0.211 mJ m^{-2} K^{-1}). Similarly, Huang and Bartell assumed[150;173] that γ varies as $T^{0.3}$, while Murray and Bertram's reanalysis of the available data[232] suggests that $\gamma \propto T^{0.97}$.

4.5 Simulations of ice crystallisation

Experiments are at the present time not capable of enabling us to gain any microscopic insight into the process of homogeneous ice nucleation. Simulations, on the other hand, are able to provide us with at least an approximate view of the molecular-level mechanism of the process.

In simulations, we are typically interested in finding nucleation rates, which can be compared to experiment, and evaluating free energy barriers, as well as more generally in the structural and dynamic properties of the growing ice nucleus.

In this section, we first briefly discuss water models: having a good water model that can reproduce experimental properties is crucial if we are looking not just at the basic underlying physical behaviour, but actually trying to understand the behaviour of the actual experimental system. We then briefly look at what has been achieved in simulations of ice nucleation thus far.

4.5.1 Simulating water

Many computational approaches have been developed to study water and aqueous systems, from *ab initio* simulations of individual molecules to the use of implicit solvents. A simple empirical model that reproduces a large number of experimental results at a reasonable computational cost is the TIP4P/2005 model,[38] which seems to be the best of the 'simple' all-atom models available at present, and is one that works well across several phases[251–253] and at large supercoolings.[254;255] This is the model we use in most of our 'all-atom' simulations. In the following subsubsection (4.5.1.1), we describe this model; in subsubsection 4.5.1.2, we then discuss some situations where the model may not be sufficient. Finally, we describe the simpler monatomic water potential we have also used in subsubsection 4.5.1.3.

4.5.1.1 The TIP4P/2005 model

In the TIP4P/2005 model, water is treated as a rigid molecule comprising four sites: an oxygen Lennard-Jones site, two positive point charges representing the hydrogens, and an additional negative charge slightly displaced from the oxygen position. The geometry of a TIP4P/2005 molecule is depicted in figure 4.8: the O-H bond length and the H-O-H bond angle are the same as in the experimentally determined geometry (shown in figure 4.1). A possible set of centre-of-mass principal frame co-ordinates corresponding to the specification is given in

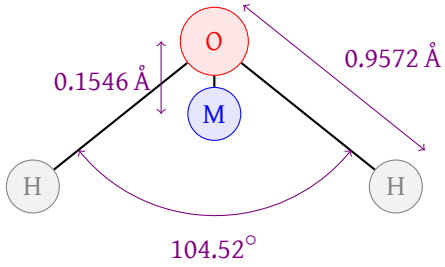


Figure 4.8: A TIP4P/2005 water molecule.[38] The oxygen (O) is a Lennard-Jones centre; the hydrogens (H) are positive point charges; M is a negative point charge. For clarity, the O-M distance depicted is exaggerated relative to the O-H and H-H distances.

table 4.3. Many of the model's properties, and how such properties compare with similar models, have been determined.[251;252;256–259]

The Lennard-Jones potential is given by

$$U^{\mathrm{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{4.1}$$

where r is the distance between two oxygen atoms, σ is the minimum distance at which the potential is zero, and ε is the depth of the energy well. The values of σ and ε for the TIP4P/2005 model are given in table 4.4; these were parameterised such that the temperature at which liquid water exhibits a maximum in density was one of the properties used in the fitting process,[38] and this seems to be an especially relevant property that makes this model of water better than others at predicting other experimental properties.[251] The Lennard-Jones potential is truncated at some cutoff radius r_c and a tail correction is calculated using equation (3.65).

In the majority of our simulations, we use a cutoff radius of $r_c = 8.5$ Å. To take into account the long-range coulombic interactions, we use Ewald summation[3] (section 3.9) with

site	x / Å	y / Å	z / Å
O	0	0	0.06557
H_A	0.7570	0	−0.5203
H_B	−0.7570	0	−0.5203
M	0	0	−0.08903

Table 4.3: The TIP4P/2005 geometry in a centre-of-mass principal frame of reference.

[3] In certain hybrid Monte Carlo simulations, we use a smooth particle mesh Ewald[260] code borrowed from the

ε / k_BK	σ / Å	q_H / e	q_M / e
93.20	3.1589	0.5564	−1.1128

Table 4.4: TIP4P/2005 model parameters:[38] ε and σ refer to the Lennard-Jones potential parameters; q_H and q_M are the charges on the H and M sites, respectively.

$\alpha = (1.1/\sigma)^2 \approx \left(0.35\,\text{Å}^{-1}\right)^2$. The self-interaction term in equation (3.67) can be evaluated per molecule as $U^{\text{self}} = -\sqrt{\frac{\alpha}{\pi}}\left(2{q_H}^2 + {q_M}^2\right)$. We also remove the intramolecular coulombic interactions (M-H_A, M-H_B and H_A-H_B),

$$U^{\text{intra}} = 2q_Hq_M\frac{(1-\text{erf}(\alpha r_{HM}))-1}{r_{HM}} + {q_H}^2\frac{(1-\text{erf}(\alpha r_{HH}))-1}{r_{HH}}, \tag{4.2}$$

giving an overall correction term that we apply to every molecule in the system of $(U^{\text{intra}} + U^{\text{self}})/k_B = -691.038\,\text{K}$.

4.5.1.2 Other all-atom models of water

There are several models of water of a similar complexity in common use. These are predominantly based on a Lennard-Jones potential determining the molecular size located at the position of the oxygen atom, and point charges on the hydrogens and elsewhere accounting for the anisotropy in the interactions. The TIP*n*P family of models preserves the experimental H-O-H bond angle; when $n = 3$,[188] the negative charge is also placed on the oxygen site. When $n = 4$,[188] the negative charge is displaced slightly with respect to the oxygen position towards the two hydrogens, while in the $n = 5$ model,[263;264] two negative charges are arranged as 'lone pairs' analogous to the classical VSEPR picture of a water molecule. The six-site NVdE potential[262;265] combines both the lone pairs and the TIP4P-like displaced negative charge site; the H-O-H bond angle is 108° in this potential (figure 4.9). Finally, the very popular SPC/E (simple point charge) model of water[266] is analogous to TIP3P, except that the H-O-H angle is tetrahedral rather than corresponding to the experimental bond angle.

Generally, these models are parameterised so that they reproduce exactly a certain set

MOSCITO MD package.[261]

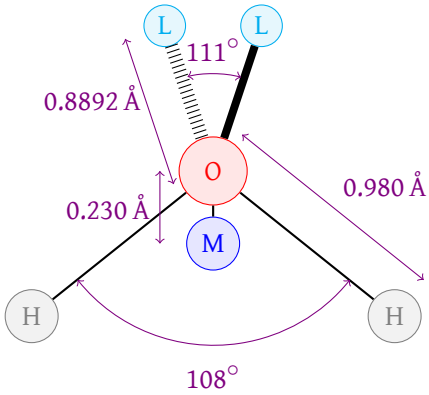


Figure 4.9: A water molecule as represented by the NVdE model.[262] The oxygen (O) and hydrogens (H) are Lennard-Jones centres; the hydrogens (H) are also positive point charges; M and L sites are negative point charges. For clarity, the O-M distance depicted is exaggerated relative to the O-H and H-H distances.

of experimental results;[251] their quality can then be determined by computing how well they reproduce other properties that were not used in the parameterisation.[251] Such an analysis shows that TIP4P-type models outperform the simpler ones,[253] and TIP4P/2005 is likely to be the best of such simple models of water.[251;252] The TIP4P-family of models includes TIP4P, TIP4P/2005 and TIP4P/Ice; these all share the same basic geometry of the atomic sites, but their LJ and electrostatic parameters and the position of the M site were fitted to reproduce different experimental properties, and TIP4P/2005 is the best in this family as a general model of water.[251] Given that we are studying ice nucleation, we might be tempted to use the TIP4P/Ice model, which was specifically parameterised to reproduce the behaviour of the ice phases well;[267] however, whilst this model works well for the various ice phases, it does not reproduce liquid properties as well as TIP4P/2005,[38;251] and so the use of TIP4P/2005 as a general model performing consistently well across the condensed phases is preferable.

It is possible to incorporate features such as flexibility, polarisability and nuclear quantum effects in an empirical potential, but this comes at a high computational cost; Vega and Abascal estimate that such an empirical model would be 1050 times more costly to simulate than a simple rigid model.[252] At a greater cost still, it may also be possible to use a hybrid

quantum mechanics/molecular mechanics approach to simulate a region of water in quantum detail using density functional theory.[268]

A recent review[269] of some water models suggests that many aspects of the models are subtle and may not be seen in standard comparisons to experimental data; interpreting simulation results as being indicative of real water, especially concerning hydrogen bonding, may be dangerous, as the choice of model may affect certain aspects of bonding. In particular, although one may argue that water is a polarisable and flexible molecule and any model of it must account for this,[252] more complex models do not necessarily result in a better description of the system; for example, the polarisable SWM4-NDP model[270] exhibits a tendency towards co-planar hydrogen bond acceptance in the liquid[269] and a destabilisation of the ice phase,[125] whilst the inclusion of flexibility can lead to a degraded phase diagram,[271] and generally does not lead to a substantial improvement in performance.[272;273] Using explicit nuclear quantum effects may often, but not always, result in only a small change in the system's properties due to a cancellation of intra- and intermolecular quantum effects.[271;274;275]

In experiments on supercooled heavy water (D_2O), it was observed that the same degree of supercooling results in faster nucleation than in light water, but the gradient in an Arrhenius plot is unchanged (*i.e.* there is no kinetic isotope effect in the classical sense).[222;228] This suggests that the formation of the critical nucleus does not involve the breaking or forming of bonds involving protons; it is water molecules rearranging rather than reacting.[222] This is important for our simulations of nucleation, as it lends support to the use of classical models of water at reasonable nucleation temperatures. We must of course balance the computational tractability of a calculation with a physically reasonable description. Although the TIP4P/2005 model works well even at very large supercoolings,[255] below about 150 K, classical models such as TIP4P/2005 and TIP4P/Ice begin to break down.[254;276;277] We therefore do not attempt to simulate ice nucleation at such severe supercoolings in order to avoid needing to simulate

quantum behaviour. While the TIP4PQ/2005 model,[276] which is an analogue of TIP4P/2005 reparameterised to be used in path integral simulations, produces the best phase diagram calculated so far,[277] the TIP4P/2005 model continues to be one of the best, and certainly the best of the simple models, in capturing the overall behaviour of water,[269] but as in any simulation, care must naturally be exercised when interpreting simulation results for a water model to real, experimental water.

Finally, we might be interested in simulating ice nucleation with additional molecules or ions present, such as various dissolved solutes present in atmospheric conditions.[17] It is not clear *a priori* that additional species can simply be inserted into water simulated by a given potential. Most water potentials were parameterised to reproduce the properties of pure water, and introducing species other than water can lead to behaviour that was not foreseen by the model designers. For example, even if there is a point charge of a certain magnitude on the H-sites of common water models, the addition of a purely ionic species into the mixture will not necessarily lead to ordinary ionic behaviour; the ionic site in the water model incorporates in non-obvious ways the experimental properties of water. Indeed, there is no core repulsion on the hydrogen site in the TIP4P/2005 model, so a purely ionic term without an interaction with the oxygen's Lennard-Jones site would have no repulsion and would lead to a divergent potential energy. Such considerations must also be borne in mind in simulations of biological molecules; a 'better' water model will not necessarily give rise to more physical behaviour if the remaining species are not reparameterised accordingly.[278] If such a reparameterisation is undertaken, then a TIP4P-type model with explicit three-body terms in the potential has been proposed as being especially suitable for such heterogeneous environments.[279]

4.5.1.3 The mW model

The monatomic water (mW) potential[33] is a coarse-grained model of water in which there are no explicit hydrogens: the tetrahedrality is achieved by three-body terms in the potential. It is

essentially a modification of the three-body Stillinger–Weber potential[280] for silicon with a greater weight given to tetrahedrality. It comprises both two-body and three-body terms,

$$U = \sum_i \sum_{j>1} U_2(r_{ij}) + \sum_i \sum_{j\neq i} \sum_{k>j} U_3(r_{ij}, r_{ik}, \theta_{jik}), \tag{4.3}$$

where

$$U_2(r_{ij}) = A\varepsilon \left(B \left[\frac{\sigma}{r_{ij}}\right]^p - \left[\frac{\sigma}{r_{ij}}\right]^q \right) \exp\left(\frac{\sigma}{r_{ij} - a\sigma}\right) \quad \text{and} \tag{4.4}$$

$$U_3(r_{ij}, r_{ik}, \theta_{jik}) = \lambda\varepsilon \left(\cos\theta_{jik} + \frac{1}{3}\right)^2 \exp\left(\frac{\gamma\sigma}{r_{ij} - a\sigma}\right) \exp\left(\frac{\gamma\sigma}{r_{ik} - a\sigma}\right). \tag{4.5}$$

The potential is identical in form to the Stillinger–Weber potential for silicon,[280] but the parameterisation is slightly different. In particular, $A = 7.049556277$, $B = 0.6022245584$, $p = 4$, $q = 0$, $a = 1.8$ and $\gamma = 1.2$ are identical to the parameters for silicon, whilst $\lambda = 23.15$, $\varepsilon/k_B = 3114.42238$ K and $\sigma = 2.3925$ Å are different;[33;281] this reflects the greater tetrahedrality of water and ensures that the energy scale is correct. We cut off the interactions at $r_c = a\sigma$.[4]

Despite being monatomic, the model provides a surprisingly good structural and thermodynamic representation of water;[33] however, unlike all-atom water models, it crystallises relatively readily.[282] The absence of explicit hydrogens appears to change dramatically the rate of nucleation. This could be, for example, because of a change in the free energy barrier to nucleation or because the dynamics of the process are much quicker. Indeed, it is the unnaturally fast diffusion in the model which is most at odds with experiment.[282] A possible rationalisation of this behaviour may be that the complication of hydrogen bond flipping is no longer an issue. For example, if one molecule were to flip in TIP4P/2005 water, to continue obeying the ice rules, a whole series of water molecules would need to flip, which would

[4] A word of caution: although in principle there is a cutoff at $r_c = a\sigma = 4.3065$ Å, at which point the exponential terms become discontinuous, in practice, we have to use a slightly smaller cutoff of 4.30264 Å. The exponentials evaluate, in Fortran's double precision, to zero at this new cutoff, whilst higher values of r cause an underflow. The same considerations apply to the exponentials in the three-body term; the cutoff for three-body terms is 4.3018 Å due to the additional γ factor in the exponential. Spurious results can be obtained if this is not noticed and taken into account.

clearly be an unlikely event, although it occurs naturally in water.[283;284] When hydrogens are removed, this complication no longer exists. We make use of the model's enhanced crystallisability to compute the free energy profile for nucleation in the mW model and hence to understand better the thermodynamics of ice nucleation.

The model does not contain any long-range forces and so does not require the use of the computationally-demanding Ewald summation. The structure arises through three-body terms rather than through the competition between electrostatic and Lennard-Jones contributions as in the TIP4P/2005 model. However, three-body terms make the implementation of the mW model in a Monte Carlo simulation somewhat more challenging than using two-body potentials would be. It is prohibitively expensive in terms of computer memory to track all three body interactions in a separate array (as, at least in the most naïve implementation, it would grow as N^3 with the number of particles and would be unsustainable in terms of memory requirements). Instead, we implement neighbour lists with heap sorting (appendix D.6) to keep track of past and present neighbours and recalculate some three-body energies as required.

4.5.2 Simulations of ice nucleation

Ice crystallisation has been studied in a significant number of simulations using all-atom models of water;[90;117;147;157;158;168;191–196;204;206;262;285–299] however, despite the obvious importance of the process of ice nucleation (chapter 1), there have not been many successful attempts at simulating it so far. Crystallisation has been observed in an individual molecular dynamics trajectory at a fixed density lower than the liquid state density,[286] and this remains the only successful brute-force simulation of ice nucleation with an all-atom model. Nucleation under 'special' conditions has been observed, such as in simulations of very small numbers of particles in a strong electric field[192;193] or near a surface;[194;195;285] however, in these studies, the ice clusters that grew quickly spanned the simulation box.

If we use the experimental results summarised in table 4.1 to assume a fast nucleation rate of $10^{13}\ \mathrm{m}^{-3}\ \mathrm{s}^{-1}$, a single nucleation event within a volume of a reasonably-sized simulation box of $(50\ \text{Å})^3$ would take of the order of 10^4 years *of simulation time* in a brute force simulation; this translates to about 10^{25} s of computer time, eight orders of magnitude longer than the age of the universe. Clearly, brute force simulations are unlikely to be successful, at least at supercoolings similar to the ones studied in the majority of experiments. An alternative approach is to use rare event methods to compute the free energy landscapes for nucleation. Such free energy landscapes have recently been calculated in several studies of homogeneous nucleation from bulk supercooled liquid water. Radhakrishnan and Trout used two-dimensional umbrella sampling,[287;288] Quigley and Rodger used a metadynamics approach,[90] and Brukhno and co-workers used replica exchange umbrella sampling.[168;291] All these simulations used the TIP4P water model and presented results at a temperature of 180 K and at standard pressure,[5] which represent an approximately 22 % supercooling. The ice form generated in these simulations has varied from pure cubic ice,[90] to pure hexagonal ice,[287] and to mixtures with predominantly hexagonal ice.[291] We describe these simulations in considerable detail in section 5.2 in the context of the order parameters used.

[5] The actual pressures used are 1 atm[90] and 1 bar,[287;291] but they are for all intents and purposes equivalent within the simulation error.

5 Order parameters

In order to monitor the process of nucleation, we require a quantitative measure that can distinguish how far along the process has advanced. The quantity describing this is usually known as a reaction co-ordinate or an order parameter.[1] In this chapter, we first review briefly the various approaches to the classification of particles that have been used in nucleation studies (section 5.1), including a detailed analysis of the reasons why we believe global order parameters are unsuitable for driving nucleation (subsection 5.1.2). We then examine the results of the previous simulations of homogeneous ice nucleation in the light of their choice of order parameter (section 5.2), before presenting the order parameter we used in the majority of our simulations and the reasons why we believe it to be more appropriate than the previously used order parameters (section 5.3).

5.1 Order parameters in nucleation studies

5.1.1 Spectral analysis on the sphere

In the nucleation literature, classification of particles as being solid-like or liquid-like is often based on the Steinhardt classification parameter[34;35]

$$q_l(i) = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^{+l} |q_{lm}(i)|^2 \right]^{1/2}, \tag{5.1}$$

[1] We make a distinction between an order parameter and a classification parameter: an order parameter is what we use as a measure of crystallinity of the system (be it local or global in nature), whilst a classification parameter serves to classify a molecule or particle individually. In much of the literature, what we call 'classification parameters' are termed 'order parameters', blurring the distinction between the two concepts.

where

$$q_{lm}(i) = \frac{1}{N_{\text{neighs}}(i)} \sum_{j=1}^{N_{\text{neighs}}(i)} Y_{lm}\left(\theta_{ij}, \varphi_{ij}\right), \tag{5.2}$$

$Y_{lm}\left(\theta_{ij}, \varphi_{ij}\right)$ are the spherical harmonics, θ and φ are the polar angles measured in an arbitrary laboratory frame of reference and $N_{\text{neighs}}(i)$ is the number of neighbours of particle i.[2] It is important to note that all $q_l(i)$ are rotationally invariant regardless of the choice of l. There is no radial component in this scheme; one can be introduced if necessary. However, a limited radial dependence arises through our definition of neighbours. The neighbours of a particle can be defined in various ways; from energy (or bonding) considerations and a fixed-distance spherical cutoff to Voronoi polyhedra[300] and the novel SANN algorithm.[301] The choice of neighbourhood definition can significantly affect simulations;[302] for order parameters of the type described below, Van Meel and co-workers suggest that simple fixed-distance spherical cutoffs are the most suitable choice.[301]

The spherical harmonics form a complete orthonormal system, so any real function on the unit sphere can be expressed using these functions in a Laplace series.[303] We can consider this to be a sort of spherical harmonic analogue of a Fourier transform of neighbour density on the unit sphere, such that

$$\rho\left(\theta_{ij}, \varphi_{ij}\right) = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} A_{lm} Y_{lm}^{\star}\left(\theta_{ij}, \varphi_{ij}\right), \tag{5.3}$$

and, since ρ is a delta function at certain values of θ and φ, the 'Fourier' coefficient $A_{lm} \propto Y_{lm}\left(\theta_{ij}, \varphi_{ij}\right)$,[34] and so the quantities expressed in equation (5.2) are essentially neighbour-averaged expansion coefficients.[34] This spherical harmonic transform technique is sometimes known as spectral analysis on the sphere; it is widely used in a variety of fields.[304–308]

[2] Although we use the index variable j in all equations of this type to represent neighbour identities, it should be understood to represent particle identities associated with the neighbour number j. It would perhaps be formally more correct to refer to the j-th element of a neighbour list, but this would reduce the readability of many equations without any substantial informative value.

Although we can obtain an excellent description of the environment about a particle by expanding the neighbour density in a Laplace series, in nucleation studies we are often hoping to compare this environment to a symmetrical crystalline system. In this case, for computational efficiency, we can simply pick the value of l that best corresponds to the symmetry of the crystalline system and compute the expansion coefficients for that l only.[34] It is both convenient and computationally less expensive to replace the complex spherical harmonics with their real analogues (see appendix D.7);[309] we continue to denote complex conjugates in the following for generality, but they can be dropped if real spherical harmonics are used.

5.1.2 Global order parameters

In many studies, the local Steinhardt classification parameters are averaged across the system to give global Steinhardt order parameters; these are typically expressed as the magnitude of the vector sum of the local classification parameters averaged over all N particles in the system, namely[34]

$$Q_l = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^{+l} \left| \frac{1}{N_{\text{tot}}} \sum_{i=1}^{N} N_{\text{neighs}}(i) q_{lm}(i) \right|^2 \right]^{1/2}, \tag{5.4}$$

where $N_{\text{tot}} = \sum_{i=1}^{N} N_{\text{neighs}}(i)$. Although local classification parameters q_l are a good measure of the local order about a particle, they are generally non-zero both in the solid phase and in the liquid phase, as the liquid is often reasonably well-ordered, especially when considering only its first neighbour shell. However, the vectors add incoherently in the liquid phase and the global order parameter Q_l averages out to zero for large systems, whilst it does not do so in the solid phase.[35] As a result, the increase in the magnitude of Q_l can be used to track how solid-like a system is. Global order parameters of this type have been used in two previous studies of homogeneous ice nucleation.[90;287] However, global order parameters are not ideal for use in nucleation studies,[35] particularly in studies where the system is not only tracked, but driven to increase the value of a global order parameter.

First of all, in nucleation studies, we often wish to perform calculations in such a way as to enable us to compare the results to classical nucleation theory. To do this, we need to know the size of the largest crystalline cluster, usually by knowing how many particles there are in the cluster. However, with a global order parameter, we have no knowledge of what the size of the largest crystalline cluster is;[3] indeed, the physical meaning of any particular value of a global order parameter is not only system-size dependent, but physically hard to interpret, and a free energy landscape calculated as a function of a global order parameter likewise does not have a clear physical interpretation in terms of nucleation.

Secondly, the interfacial free energy is in competition with the more favourable entropy arising from a larger number of smaller clusters. It can be shown[35] that this entropy can play a significant rôle in small systems that can be simulated on computers: in the early stages of nucleation, many small nuclei are always more favourable than one single nucleus comprising the same number of particles. There is, however, a crossover to the expected behaviour once a certain nucleus size has been passed. This means that, when comparing results to classical nucleation theory, a 'global' measure of crystallinity – which effectively induces an entropic break-up of small clusters – is inappropriate.[35]

Finally, we suggest that the pathways produced when the nucleation process is driven by global order parameters may be inconsistent with the natural nucleation pathways (*i.e.* those pathways that would occur for an unbiassed system given sufficient time), particularly so in the case of ice. Since there is no distinction between the liquid and the solid states of particles when global order parameters are used, driving the system to increase its global order parameter can potentially induce an orientational coherence even in the liquid state. This suggests that a particle is influenced not only by its neighbours, but potentially by a crystalline cluster that is very far removed from it, even though such a long-range interaction has no basis in

[3] As an extreme example, we can imagine two clusters orientated such that their local vectors are exactly the opposites of each other; such a system might have a low global order parameter even if it is in fact well-ordered.

reality. Furthermore, when rare event techniques are applied to a system, it is relatively easy to compensate for arbitrarily large free energy barriers; there is a danger, therefore, that if the natural, lowest free energy pathway is dynamically slow (having accounted for the free energy barrier associated with the process itself), such a pathway may not be observed and a higher free energy pathway could be found instead provided that it is dynamically faster and its higher free energy has been negated by a rare event method. We also suggest that, for ice nucleation, the use of global order parameters in driving nucleation may lead to precisely such high free energy pathways; we substantiate this hypothesis in chapters 6 and 7.

5.1.3 Local order parameters

Even though the local Steinhardt classification parameters defined above cannot distinguish between solid and liquid particles on their own, there are a few approaches that allow us to do this without sacrificing their local nature. It is often convenient to calculate the dot products of the individual local classification parameters expressed in vector form, $\boldsymbol{q}_l(i)$, whose $(2l+1)$ components are the Steinhardt parameters $q_{lm}(i)$ for $m \in [-l, l] \cap \mathbb{Z}$, with the equivalent vectors of a particle's neighbours.[35] We calculate the rotationally invariant function $d_l(i,j) = \hat{\boldsymbol{q}}_l(i) \cdot \hat{\boldsymbol{q}}_l^{\star}(j)$, where i and j are neighbours; this dot product value ranges between -1 and $+1$. By plotting the distribution of d_l values for the liquid and the crystalline phases, a critical threshold d_c can be determined as the first point where the probability of being in the solid phase is non-zero.[35] The number of crystalline connections is then defined as[35]

$$n_{\text{connections}}(i) = \sum_{j=1}^{N_{\text{neighs}}(i)} \mathrm{H}(d_l(i,j) - d_c), \tag{5.5}$$

where H is the Heaviside step function. The number of connections should be higher in the solid phase than in the liquid phase, and a criterion involving a threshold minimum number of connections to distinguish between the two phases is often a good classification

parameter.[35;310]

Another procedure involves calculating the neighbour-averaged contribution,[311;312]

$$\langle q_{lm}(i)\rangle = \frac{1}{N_{\text{neighs}}(i)+1}\sum_{j=0}^{N_{\text{neighs}}(i)} \hat{q}_{lm}(j), \tag{5.6}$$

where j runs over all the neighbours of particle i, and includes the particle itself (when $j = 0$).[4] The average local bond classification parameter is then given by

$$\langle q_l(i)\rangle = \left[\frac{4\pi}{2l+1}\sum_{m=-l}^{+l} |\langle q_{lm}(i)\rangle|^2\right]^{1/2}, \tag{5.7}$$

and particle i is classified as being solid-like or liquid-like depending on whether the value of $\langle q_l(i)\rangle$ is higher or lower than a given threshold. In both approaches mentioned, the second neighbour shell is effectively taken into account through the use of local Steinhardt vectors of the first neighbour shell, either by averaging or by taking dot products.

Any two particles belong to the same crystalline cluster if they are both classified as being solid-like and are located within a certain fixed distance of one another (that is, they are neighbours). Once all the particles have been classified, the size of the largest such cluster is normally calculated in nucleation studies; this then acts as the overall (local) order parameter.

To obtain the size of the largest cluster, we assume that each solid particle starts in its own 'cluster' and each cluster is numbered and its size recorded. We then loop through the particles using the following algorithm:

- Determine if the current particle is solid.
- Look through all its neighbours.
- If any of the neighbours is also solid and the particles are not yet labelled as being in the same cluster, then merge the clusters. We do this by relabelling the cluster number of all the particles which have the higher number into the lower cluster number, and also

[4] Note that unlike in references 311 and 312, we use components of the normalised vector $\hat{\boldsymbol{q}}_l$, rather than divide by the number of neighbours. This leads to a very slight improvement in the order parameter's capacity to distinguish between liquid water and ice.

updating the size of the lower number cluster to the sum of both, and setting the higher number cluster size to zero.

– Once the loop is complete, all molecules will have a cluster number assigned to them, and the largest cluster can readily be identified.

5.1.4 Other order parameters

Other types of order parameter have been suggested in the literature. A potentially useful classification parameter in ice nucleation might be the Chau–Hardwick tetrahedrality parameter[313;314]

$$q_{\mathrm{CH}}(i) = \frac{3}{32} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\hat{\boldsymbol{r}}_{ij} \cdot \hat{\boldsymbol{r}}_{ik} + 1/3\right)^2, \tag{5.8}$$

where $\hat{\boldsymbol{r}}_{ij}$ are unit vectors from particle i to its j-th neighbour ($j \in \{1, 2, 3, 4\}$). If all the angles are perfectly tetrahedral, each dot product will give a value of $\cos\left(\arccos\left(-1/3\right)\right) = -1/3$, which gives an ideal value of the Chau–Hardwick parameter of zero. However, in ice nucleation, this classification parameter is of only marginal use on a local scale, as supercooled liquid water is very much tetrahedral itself. It is possible to define a global analogue of this classification parameter (sometimes denoted ζ) by averaging over the q_{CH} values of the individual particles;[90] however, even though this quantity can distinguish between bulk ice and liquid water, it is not ideal for driving nucleation, as it is a global order parameter.

It is very useful for an order parameter used in driving nucleation to be local in nature; however, that an order parameter is local is not sufficient for it to be a valid metric used for driving the nucleation process. For example, in their ice nucleation simulations,[168;291] Brukhno and co-workers used a maximum director projection approach, whereby they find which vector in a pre-determined canonical orientation (such a vector is then called a 'director') best matches each neighbour interparticle vector by calculating which interparticle vector-director pair results in the maximum scalar projection, taking care to include either a cubic or a hexagonal

arrangement of neighbours, but not a mix. They then calculate the mean of the maximum scalar projections of the interparticle vectors with this optimal set of directors and calculate the number of ice-like connections by calculating how many pairs of neighbours produce a mean maximum projection greater than some threshold value. Although these order parameters do permit driving the growth of ice in a fixed orientation with respect to the simulation box, the rotational bias inherent in the procedure induces a non-local and non-physical orientational coherence in the growing ice cluster. Furthermore, because this order parameter induces crystallisation with a specific orientation with respect to the simulation box, this bias is likely to lead to a significant error in the free energy barrier, because it excludes many other possible pathways (which could have a lower free energy). We suggest that, as a result, this order parameter is not suitable to study ice nucleation.

It is also possible to envisage a classification parameter based only on geometric considerations, such as the distances and angles between particles, the dihedral angles between molecules, the number of hydrogen bond donors and acceptors, and deviations from linearity of hydrogen bonds;[93] however, such classification parameters are not transferable, are likely to deteriorate in quality when the temperature or another parameter is changed, and, most importantly, failed to grow ice in a coherent long-range manner in our simulations.[93] The dot product classification parameters that we use instead include information about the second shell of neighbours of every particle, and this appears to be necessary in order to be able to grow ice with sufficient long-range correlations.

More complex order parameters used to track larger molecule crystallisation have also been proposed,[315] whilst an overview of many simpler order parameters used in vapour-liquid nucleation studies was produced by Senger and co-workers.[316]

5.2 Analysis of order parameters used in previous simulations

We briefly introduced the previous simulations of homogeneous ice nucleation from the bulk supercooled liquid water in subsection 4.5.2, and we have presented the order parameters they used in this chapter. Having introduced and discussed the order parameters, we are now in a position to examine more closely the results and conclusions of the previous simulations of homogeneous ice nucleation.

In their simulations of homogeneous ice nucleation, Radhakrishnan and Trout used two-dimensional umbrella sampling,[287;288] Quigley and Rodger used a metadynamics approach,[90] and Brukhno and co-workers used replica exchange umbrella sampling.[168;291] Both Radhakrishnan and Trout and Quigley and Rodger used global order parameters of the type described in subsection 5.1.2, whilst Brukhno and co-workers used a maximum director approach described in subsection 5.1.4. We have already discussed the reasons why global and rotationally-biassed order parameters are not entirely suitable to the study of homogeneous nucleation.

The free energy barriers computed in these studies at a $\sim$20 % supercooling are intriguing. Brukhno and co-workers find a free energy barrier of approximately 130 $k_{\mathrm{B}}T$, although this value should be largely discounted because of their use of a non-rotationally-invariant order parameter, and, as they themselves note, because their system was not fully equilibrated.[291] The barriers reported by Radhakrishnan and Trout (63 $k_{\mathrm{B}}T$) and Quigley and Rodger (79 $k_{\mathrm{B}}T$)[5] are more interesting. If the equilibrium free energy landscapes have been obtained, these values should be lower bounds to the free energy barrier one would get for the 'perfect' order parameter because of configurations at the top of the barrier that are not truly representative of the intermediates between the two states. As one would generally expect an order parameter

[5] We note here that the system size in Quigley and Rodger's simulation is in the range where a cylindrical cluster that spans the box has the lowest surface to volume ratio for ice clusters at the top of the barrier.[90] However, this can only act to lower the free energy barrier compared to a spherical critical cluster.

that follows the largest crystalline cluster within the system to provide a better representation of the intermediates, one would expect the free energy barrier obtained with such an approach to be larger than that obtained using global orientational order parameters. Classical nucleation theory can provide an estimate of the former; the CNT expression for the Gibbs energy barrier of a spherical cluster comprising N particles is given by equation (2.9),

$$\Delta G(N) = -N\Delta_{\mathrm{fus}}\mu + \gamma\left[36\pi\left(\frac{N}{\rho}\right)^2\right]^{1/3}, \tag{5.9}$$

where ρ is the number density of the crystalline phase, $\Delta_{\mathrm{fus}}\mu$ is the change in chemical potential on fusion and γ is the interfacial free energy between the two phases. An estimation of the classical nucleation theory result for TIP4P water at 180 K gives a free energy barrier of $\Delta G \approx 35\, k_{\mathrm{B}}T$ and a critical nucleus size of 106. To obtain these, we use the simple approximation[12] that $\Delta_{\mathrm{fus}}\mu(T) \approx \Delta_{\mathrm{fus}}H_{\mathrm{m}}((T_{\mathrm{fus}} - T)/T_{\mathrm{fus}})$, and we insert the appropriate values for the TIP4P model.[6] We note that one generally expects the interfacial free energy to decrease with decreasing temperature, as has been inferred from experiment,[210;320] and so the use of γ from coexistence likely leads to an overestimation of this barrier.

One is tempted to ask why there is such a big difference in the barrier, and in the opposite direction to what is expected. The first option is that classical nucleation theory's shortcomings are responsible for a severe underestimation of the free energy barrier; however, although it is well-known that there are deficiencies in the framework of CNT (section 2.3), such a big difference nevertheless seems surprising. Another option might be that the free energy landscapes computed in simulations are not fully at equilibrium; to find the free energy barrier of nucleation, not only the end points, but the entire path space must be equilibrated. This may be especially difficult in ice growth simulations, which are very slow even from a planar

[6] The enthalpy of fusion for the TIP4P model is $\Delta_{\mathrm{fus}}H = 1.05\,\mathrm{kcal\,mol^{-1}}$;[317] we assume the density of ice is only slightly increased at 180 K at $\rho(\mathrm{I_h}) = 0.945\,\mathrm{g\,mL^{-1}}$ from the melting point density of $\rho(\mathrm{I_h}) = 0.940\,\mathrm{g\,mL^{-1}}$;[317] the melting temperature is $T_{\mathrm{fus}} = 232\,\mathrm{K}$;[317] and we use the measured interfacial free energy of $\gamma = 24\,\mathrm{mJ\,m^{-2}}$,[318] which was calculated for TIP4P water with a sharp cutoff, rather than with Ewald summation (the non-use of Ewald summation has been shown not to affect the results significantly[319]).

ice-liquid interface where there is no free energy barrier to nucleation.[117;321] Such path space equilibration may have been a particular problem in the previous simulations, as the nature of the order parameters may bias the system to locate more easily pathways that are high in free energy, but where the dynamics along the reaction co-ordinate (once the free energy barrier has been negated) are more computationally tractable to observe. For example, the coherence induced by the use of global order parameters may make it easier for water molecules to join the ice crystal if they can 'feel' the orientation of the crystal, even if this is not the natural mode of growth.

We have thus hypothesised that the use of global order parameters may have led to the sampling of non-physical nucleation pathways in the previous studies, and we explore this hypothesis further in chapters 6 and 7.

5.3 Order parameters for homogeneous ice nucleation

The choice of an order parameter to drive ice nucleation is not trivial: as discussed above, it is preferable that it be local; it must be forgiving enough to be able to induce the growth of a small ice cluster; and it must be strict enough to ensure that the structure grown is actually ice-like and has, ultimately, long-range order. In our simulations, we use a variation of the dot product approach described above that is commonly used in studies of tetrahedral liquids.[36;37;322] We choose to use $l = 3$, since the $l = 3$ spherical harmonics are the ones best describing tetrahedrality. A plot of the distribution of $d_3(i, j)$, shown in figure 5.1, might tempt us to classify as ice any molecule which has, say, a value of $d_3(i, j) \leq -0.82$. We could then define

$$n_{\text{connections}}(i) = \sum_{j=1}^{N_{\text{neighs}}(i)} \mathrm{H}(-0.82 - d_3(i, j)), \tag{5.10}$$

and every molecule with more than 3 connections could be classified as ice. Fully-formed crystals and liquid water are very well differentiated in this way; however, the principal difference

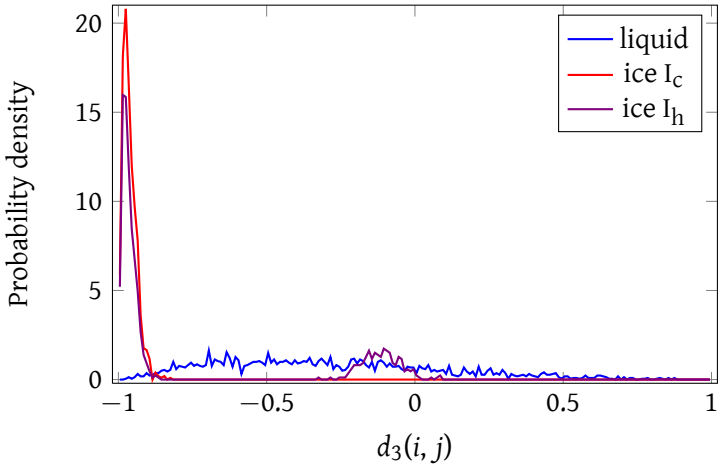


Figure 5.1: A typical probability density distribution for all pairs of $d_3(i, j) = \hat{\boldsymbol{q}}_3(i) \cdot \hat{\boldsymbol{q}}_3^{\star}(j)$, where the centres of mass of molecules i and j are within 3.5 Å of each other. The three states depicted were equilibrated at 200 K (using the TIP4P/2005 water model) and the ice structures are not, therefore, 'perfect'. This figure is analogous to those in references 36 and 37.

between cubic ice (with only staggered bonds) and hexagonal ice (with one eclipsed bond, which is reflected in a $d_3(i, j)$ value near -0.1) is not accounted for: using a classification criterion as defined by equation (5.10) would consequently unfairly bias us towards cubic ice, as we would accept imperfectly formed cubic ice (with three suitable connections) as ice-like, but imperfectly formed hexagonal ice with three suitable connections would not be accepted as ice-like if two of these connections were in the $d_3(i, j) \leq -0.82$ region and one were in the vicinity of -0.1.[37] We thus need to account for the eclipsed bond, and we define a classification parameter as

$$n_{\text{connections}}(i) = \sum_{j=1}^{N_{\text{neighs}}(i)} \Gamma(d_3(i, j)), \tag{5.11}$$

where

$$\Gamma(x) = \begin{cases} 1 & \text{if } [(x < -0.825) \vee (-0.23 < x < 0.01)], \\ 0 & \text{otherwise.} \end{cases} \tag{5.12}$$

These limiting values were chosen to encompass $d_3(i, j)$ regions (figure 5.1) where the probability density function for either ice phase has a value greater than 0.1. We classify a molecule as ice-like if $n_{\text{connections}} \geq 3$ and as liquid-like otherwise. This gives perfect identification in both

equilibrated cubic and equilibrated hexagonal ice. Although it appears that liquid water, having a considerable proportion of molecules in the d_3 region where ice I_h has its second peak, might be mistakenly classified as ice, it is in fact almost nowhere that it has a sufficient number of connections with this value of d_3, and we find a misidentification rate in supercooled liquid water of approximately 0.8 %, suggesting that this is an excellent classification parameter. The order parameter we use to track the progress of nucleation is the size of the largest cluster of molecules classified as ice, where two molecules belong to the same crystalline cluster if they are both ice-like and their centres of mass are within 3.5 Å of each other.

We can distinguish between hexagonal or cubic ice using this approach. Provided that the above criterion is fulfilled, we can be slightly more generous in awarding 'hexagonal' status, in case the hexagonal bond is the one that is mis-formed. We classify an ice molecule as being hexagonal if any of its $d_3(i, j)$ connections is between -0.5 and 0.25. This is very generous compared to the classification scheme in equation (5.12), but a cubic bond will be nowhere near this range. Provided that the molecule is ice-like, which we test with the more rigorous test above, the precise cutoff values we choose do not especially matter, provided that a bond is, roughly speaking, eclipsed.

Unfortunately, whilst this classification procedure, regardless of the precise details of the parameterisation of the limiting values, works well for the mW potential presented in chapter 6, it does not do so in simulations of all-atom models of water. When used in the form presented above in umbrella sampling[31] or forward flux sampling simulations,[95] natural fluctuations in the system often result in molecules satisfying the order parameter even if they are not really ice-like. When forced to grow with a biassing potential, 'chains' form more easily than real ice grows, even though such chains actually represent an abuse of the order parameter, and 'ice' structures as depicted in figure 5.2 are commonplace. The real issue is not just that such chains form, but that when they do form, the system is not subsequently able

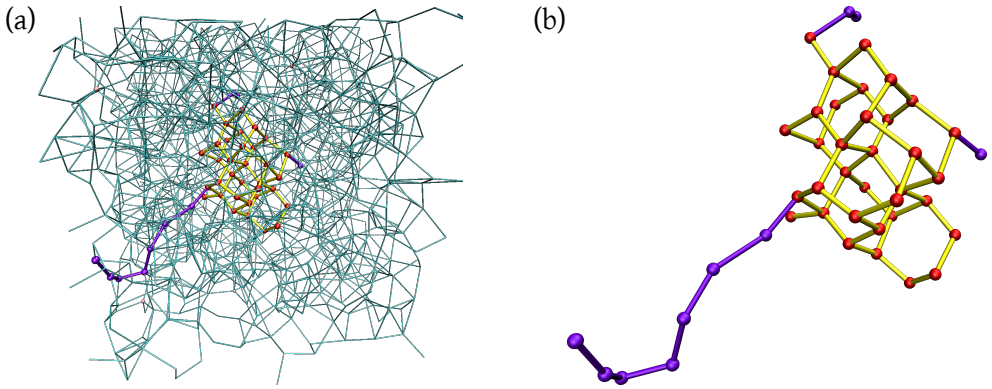


Figure 5.2: An example of non-ice-like chain growth in TIP4P/2005 umbrella sampling simulations when using the order parameter without chain removal as described in the text. The whole simulation box is depicted in (a), while only the ice-like molecules are shown in (b). The system has 1000 molecules at 240 K, starting from a 24-molecule cluster. Molecules classified as being part of the largest crystalline cluster are shown in red and violet; there are 45 molecules in this cluster. Molecules whose centres of mass are within 3.5 Å are connected with lines. Molecules shown in violet would be removed from the largest crystalline cluster on application of the chain removal algorithm described in the text.

to transform to the correct (compact) ice structure, and ice growth is arrested. For example, in forward flux sampling simulations, the probability of reaching the next interface along the reaction co-ordinate rapidly approaches zero once the system exhibits predominantly chains, and in umbrella sampling simulations, the system is frustrated so much that it fails to grow further even when using extremely large biassing potentials. This suggests that chain growth of this type is not a natural feature of ice nucleation.

Such chain growth can be observed with all variants of local order parameters we have tried. In order to alleviate this problem, we (a) classify any molecule with more than four neighbours (within 3.5 Å) as being liquid, and (b) explicitly exclude molecules belonging to chains from the largest crystalline cluster. We achieve the latter by removing any molecules with only one neighbour belonging to the largest cluster from the largest cluster, except if that single neighbour is connected to three further molecules in the largest cluster. This allows 'chains' comprising a single molecule to form and thus allows ice to grow. We iterate the

procedure until no further molecule is removed.

Removing chains could be problematic in the initial stages of nucleation: it is impossible for an ice structure smaller than a single chair (or boat) not to be formed of chains, and small rings may form instead if forced. However, due to the similarity of liquid water and ice, it is possible to wait for a boat or chair cluster to form spontaneously in a simulation, and limit umbrella sampling to systems with clusters larger than ∼10 molecules. Doing so does not appear to be a significant limitation when the critical cluster is expected to comprise over 100 molecules.

Although neighbour-averaged classification parameters are excellent at distinguishing between the phases of systems without much structure in the liquid phase,[311;323] it is less clear whether the same applies to well-structured liquids like water. Using $l = 4$ and $l = 6$ (as depicted in figure 5.3) gives better separation between the phases than does using $l = 3$. Liquid water and hexagonal ice are less well-separated in the $\langle q_4 \rangle$-$\langle q_6 \rangle$ plane than in the Lennard-Jones case; nonetheless, a choice of $\langle q_6(i) \rangle > 0.7$ as an ice-liquid boundary would appear to be reasonable. Calculating the size of the largest cluster with this method results

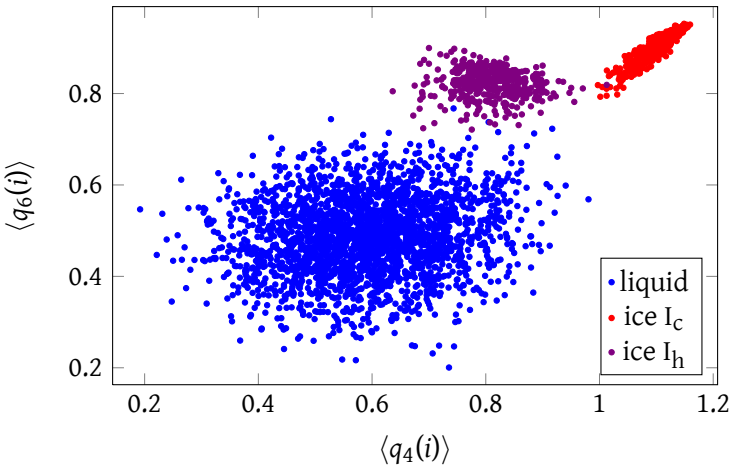


Figure 5.3: Neighbour-averaged order parameters for systems of ices I_h and I_c and liquid water. All systems were equilibrated at 200 K using the TIP4P/2005 water model, and they contain different numbers of molecules. The neighbour cutoff distance was 3.5 Å.

on average in only slightly smaller clusters compared to those resulting from the dot product approach, and the formation of chains is as problematic as when using the latter. We show in chapter 6 that for the mW monatomic model of water,[33] a dot product approach leads to a free energy profile that is almost entirely consistent with classical nucleation theory. By contrast, including 'surface' molecules in the largest ice cluster, as attempted in other studies with the same model of water,[98;324] appears to reduce the agreement with classical nucleation theory (see chapter 6). Lechner and co-workers have recently used a combination of dot product vectors and their neighbour-averaged classification parameters to study the rôle of the surface and the bulk terms when comparing simulation data to classical nucleation theory for a soft-core colloid model.[323;325] Using neighbour-averaged classification parameters reduces the size of the critical cluster when compared to one calculated using a dot product approach in their work.[323;325] The clusters we have analysed with both approaches show only a small difference in cluster size; on average, the neighbour-averaged clusters are slightly smaller, but for individual configurations, the converse can also hold. Nonetheless, given that the number of molecules classified as being part of the cluster by the dot product approach is itself a rather conservative estimate of the cluster size (chapter 6), and since it would be computationally extremely expensive to perform two-dimensional umbrella sampling with an additional order parameter, we restrain ourselves, for the time being, to using the dot product approach only. However, the neighbour-averaged approach is an attractive alternative, and ensuring that clusters are of essentially the same size with both approaches is a useful confirmation that the exact details of the order parameter do not appear to change the outcome significantly.

6 Homogeneous ice nucleation with a monatomic water model

Here, we present simulations which we hope present a step towards understanding the homogeneous nucleation of ice. We wish to look at a model of water that does not have the full complexity of TIP4P-type models, namely the monatomic model of water (mW) proposed by Moore and Molinero[33] (subsubsection 4.5.1.3), which has already been used to study both the supercooled liquid[148;326] and ice crystallisation (in the bulk[98;282;324] and in confinement[36;327–331]). Furthermore, some nucleation rates and critical cluster sizes have very recently been reported from molecular dynamics simulations near the liquid transformation temperature of 202 K[324] and in the range 220 K to 240 K.[98]

In this chapter, we first summarise the methods used in our simulations of nucleation using the mW model (section 6.1). We then present our results: the free energy landscape (subsection 6.2.1), the pathway of nucleation (subsection 6.2.2) and the nucleation rate (subsection 6.2.3). Finally, we suggest some conclusions arising from these simulations (section 6.3).

6.1 Simulation details

To simulate the system, we use the standard Metropolis Monte Carlo approach as discussed in section 3.1. In addition, we use the umbrella sampling method with windowing (section 3.2) in the isobaric-isothermal ensemble to calculate the free energy profile. In the umbrella sampling

simulations, we iteratively update the weights depending on the number of steps spent at each value of the order parameter in the previous iteration.

In simulations, care must be taken to ensure that the ice clusters do not begin to span the box. Such configurations are a non-physical consequence of the periodic boundary conditions and should be excluded from our sampling of the nucleation free energy profile, but can arise due to fluctuations to a more cylindrical shape. To reduce the probability of spanning clusters arising, we simulate the nucleation process in a cubic box, as if we allow box dimensions to vary independently, random fluctuations in the liquid phase often produce long cuboid boxes in which the formation of spanning clusters is more likely and which can thus crystallise readily, albeit in a non-physical manner. We should note that the spanning of clusters only happens very occasionally in our simulations because we use a sufficiently large number of particles that for the size of the crystalline clusters we consider, their surface to volume ratio is minimised by a spherical rather than a cylindrical shape.

The order parameter we use is the size of the largest ice cluster in the system. It is similar to the order parameter we presented for TIP4P/2005 nucleation in section 5.3, but not identical to it. We use Steinhardt-style local classification parameters[34;35] similar to those used previously in studies of tetrahedral liquids;[36;37] namely, we use the $(2l+1)$-dimensional real vector order parameter $\boldsymbol{q}_l(i)$ with components $m \in [-l,\, l] \cap \mathbb{Z}$ given by

$$q_{lm}(i) = \frac{1}{N_{\text{neighs}}(i)} \sum_{j=1}^{N_{\text{neighs}}(i)} S_{lm}\left(\theta_{ij},\, \varphi_{ij}\right), \tag{6.1}$$

where S_{lm} are the real spherical harmonics (appendix D.7), $N_{\text{neighs}}(i)$ is the number of neighbours of particle i, θ and φ are the polar angles, and we use $l = 3$. The sum in j is over all neighbours within 3.6 Å. We then calculate $d_l(i,j) = \hat{\boldsymbol{q}}_l(i) \cdot \hat{\boldsymbol{q}}_l(j)$. The resulting quantity ranges between -1 and $+1$, with cubic ice exhibiting a peak at -1 and hexagonal ice exhibiting a large peak at -1 (corresponding to staggered bonds) and a small peak near -0.1 (corresponding to eclipsed

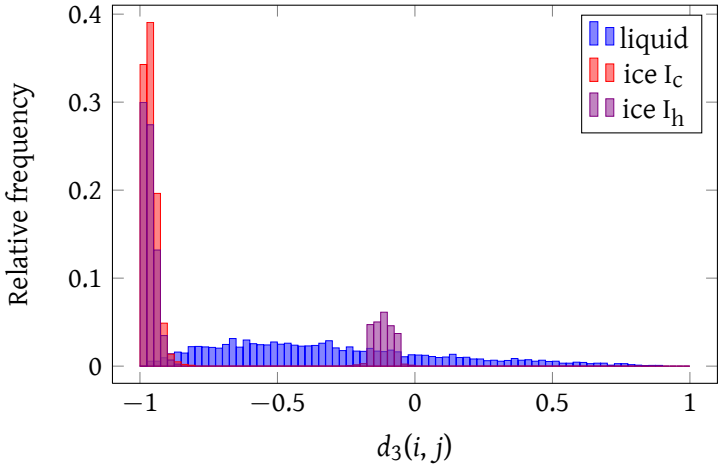


Figure 6.1: A typical probability distribution for all pairs of $d_3(i, j) = \hat{\boldsymbol{q}}_3(i) \cdot \hat{\boldsymbol{q}}_3(j)$, where particles i and j are within 3.6 Å of each other. The three states depicted were equilibrated using the mW potential at 220 K and the ice structures are not, therefore, 'perfect'. This figure is analogous to those in references 36 and 37, and to figure 5.1 corresponding to TIP4P/2005 water.

bonds), as depicted in figure 6.1. As in section 5.3, we must ensure that we do not bias simulations towards cubic ice, and our overall classification parameter is thus defined as

$$n_{\mathrm{con}}(i) = \sum_{j=1}^{N_{\mathrm{neighs}}(i)} \Gamma(d_3(i, j)), \tag{6.2}$$

where

$$\Gamma(x) = \begin{cases} 1 & \text{if } [(x < -0.82) \vee (-0.145 < x < -0.065)], \\ 0 & \text{otherwise.} \end{cases} \tag{6.3}$$

The order parameter we use in umbrella sampling simulations to track nucleation is the size of the largest cluster of particles classified as ice, where two particles belong to the same crystalline cluster if they are both ice-like and within 3.6 Å of each other. In summary, the differences between this order parameter and the one presented in section 5.3 are that (a) neighbours are defined to be within 3.6 Å rather than 3.5 Å, (b) the limiting values in equation (6.3) are slightly different from those in equation (5.12), and (c) we do not specifically remove chains or require that an ice particle should have no more than four neighbours. Although the order parameter described in section 5.3 is likely to be a better order parameter, the scheme presented here is

sufficient for simulating nucleation in the mW model of water, in which crystallisation is more facile. To calculate nucleation rates, we perform MD simulations using the LAMMPS molecular dynamics code[332] in a Bennett–Chandler-type approach,[10;52;86;88] as described in section 3.4.

6.2 Results

6.2.1 Free energy landscape

In our Monte Carlo simulations, we simulated 1400 mW particles at 1 bar pressure and 220 K, which corresponds to a 20 % supercooling. The umbrella sampling used partially overlapping windows of various sizes until results were consistent and the criteria for equilibration given in section 3.2 were deemed to have been fulfilled. On average, this involved approximately 5×10^{10} Monte Carlo steps for each window. The simulations involved three distinct scenarios: in one case, crystal clusters were grown directly from the supercooled liquid; in the other two cases, we started simulations in the initial few umbrella sampling windows with seed clusters of suitably equilibrated hexagonal and cubic ice, respectively. These were then allowed to shrink and grow, and clusters that grew into overlapping window regions were taken as initial starting points for the umbrella sampling simulations in later windows. We must emphasise that the cubic and hexagonal ice simulations were not constrained in any way to form a particular phase of ice; we merely introduced small clusters taken from a particular crystal structure into the liquid, and the growth and shrinkage proceeded unconstrained from those starting configurations. The free energy profiles corresponding to crystal nucleus growth from the supercooled liquid and from equilibrated hexagonal clusters are shown in figure 6.2; for clarity, we have not plotted the results for nucleation starting from cubic clusters, as the results are essentially identical to the hexagonal ice case.

For the simulations in which we used hexagonal and cubic seed clusters, the five seed clusters chosen for each system comprised approximately 35 particles. In the first umbrella

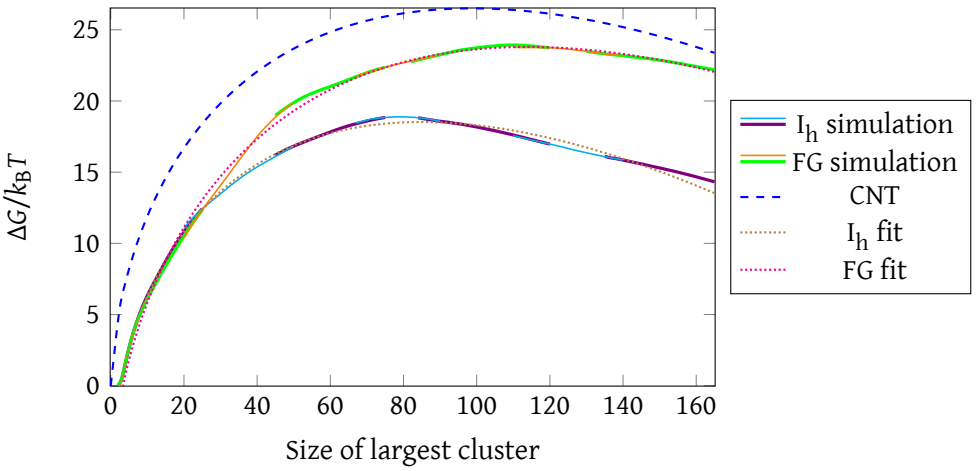


Figure 6.2: The free energy profile of mW nucleation as a function of the size of the largest crystalline cluster in the system. Simulation results from different windows are depicted in alternating colours to show their overlap. The dashed line corresponds to the classical nucleation theory prediction; dotted lines depict fits to the simulation data. The free energy profiles for ice nucleation seeded with hexagonal crystal clusters (I_h) and ice nucleation directly from the supercooled liquid (FG) are shown.

sampling window, we allowed the cluster to melt completely and then to regrow within the window. The point $\Delta G = 0$ corresponds to the supercooled liquid state for all simulations. Matching the free energy difference between the first and the second window is somewhat problematic, as the equilibrated state in the first window is, given that the clusters are allowed to melt completely, the same as in the simulations grown from the supercooled liquid directly. However, we suggest the free energy difference between the simulations with different starting points arises at a later stage, when the clusters cease to be interconvertible. We justify this assertion by performing umbrella sampling simulations on an intermediate window between the first and the second one, in which the cluster is not allowed to melt completely; we observe no difference between the free energies reported in figure 6.2 and in this overlapping window, which suggests that, although the resulting ice nuclei are somewhat different between the equilibrated states in the overlapping regions of the first two umbrella sampling windows in seeded simulations, they have the same free energy.

Some snapshots of the nucleation process are shown in figure 6.3. Interestingly, the

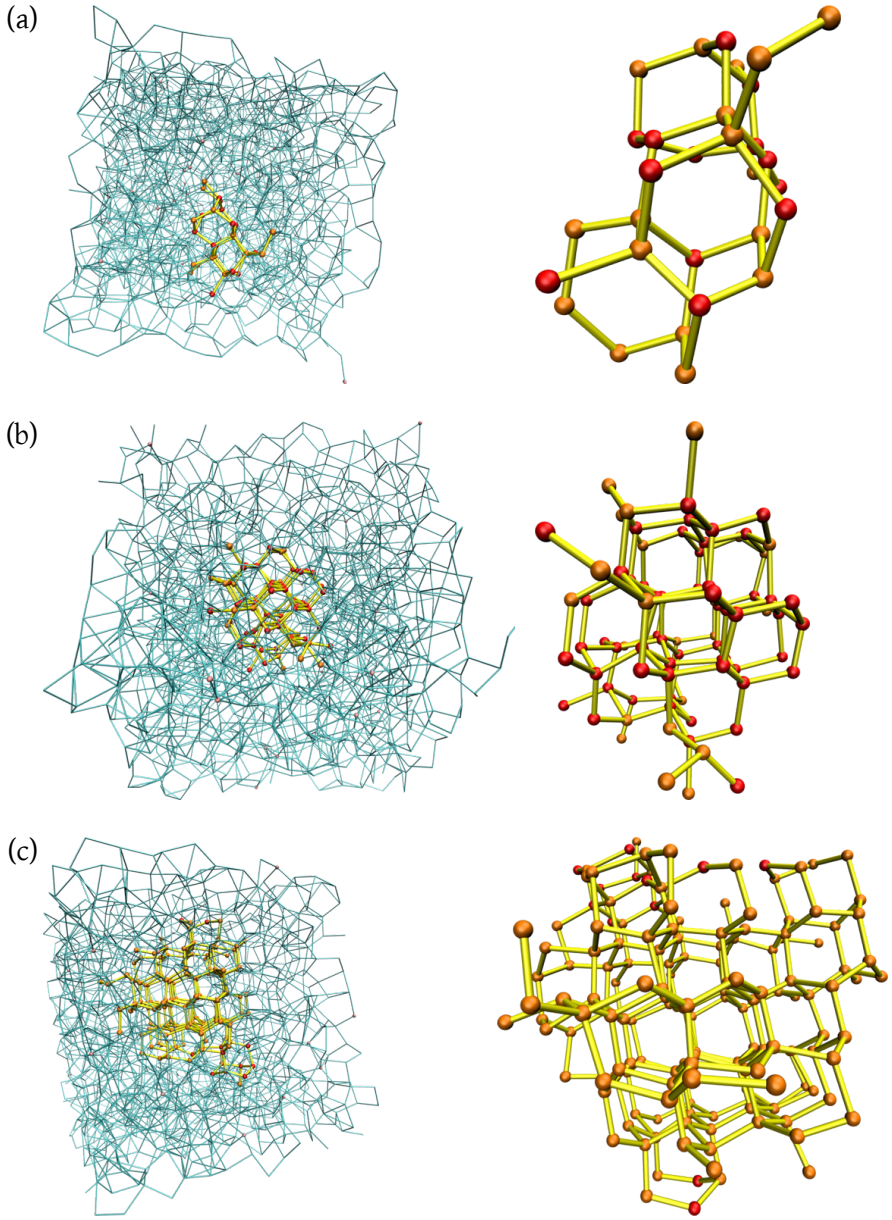


Figure 6.3: Representative nucleation snapshots from umbrella sampling simulations. Particles within 3.6 Å are connected with lines; the lines are yellow within the largest ice cluster and cyan elsewhere. Spheres represent particles classified as ice: red spheres correspond to cubic ice, orange spheres correspond to hexagonal ice and pink spheres (in the left column only) correspond to ice particles not within the largest crystalline cluster. In (a), a 30-particle cluster as nucleated from the supercooled liquid is shown. In (b), an 83-particle cluster as grown in a simulation initially seeded with an equilibrated cubic ice cluster is shown. In (c), a 165-particle cluster as grown in a simulation initially seeded with an equilibrated hexagonal ice cluster is shown. In each case, the left and right pictures depict the same cluster from different perspectives; one within the liquid framework and one showing solely the largest crystalline cluster.

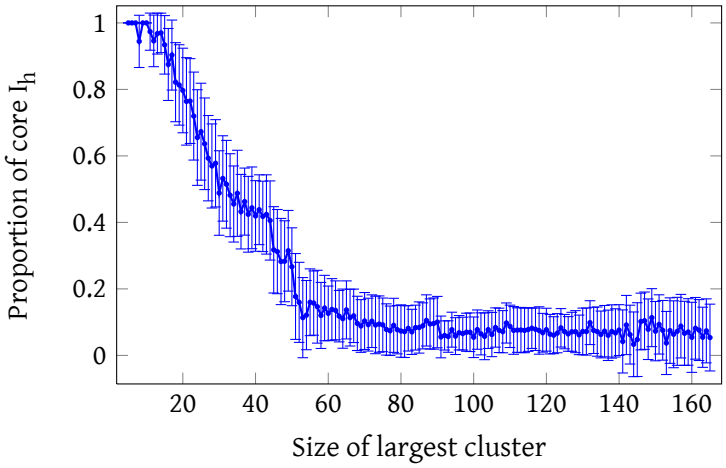


Figure 6.4: The proportion of core ice particles classified as hexagonal for the set of simulations in which the crystalline cluster was grown directly from the supercooled liquid. Error bars show the standard deviation for the population of configurations at each ice cluster size.

type of ice that grows during this nucleation process depends on the starting point. If we begin a simulation from the supercooled liquid, the ice clusters that form are initially mixed phases of hexagonal and cubic ice. As these 'mixed' clusters grow, however, they become predominantly cubic (except for surface particles, for which a classification as cubic or hexagonal is less well defined[1]); the proportion of core[2] ice particles classified as being hexagonal drops from almost unity to approximately 10 % by the time the crystal nucleus comprises 50 particles (figure 6.4). Starting from pre-formed cubic clusters in the lower umbrella sampling windows also results in primarily cubic ice growth. By contrast, starting the simulation with a pre-formed hexagonal cluster close in size to, but slightly smaller than, the critical size, results in essentially pure hexagonal ice growth. This suggests that in this model, ice prefers to grow in the phase of the underlying crystal nucleus. We have not been able to find a significant difference in the nucleation free energy profiles corresponding to cubic and hexagonal ice. However, our

[1] For example, if an eclipsed bond is formed with a particle classified as being liquid, it is unclear whether this should qualify a particle as being hexagonal.

[2] Core ice particles are ice particles in the largest crystalline cluster whose four neighbours (particles within 3.6 Å) all also belong to the cluster; such particles are classified as being hexagonal if they have one eclipsed and three staggered connections to neighbours.

simulations show that the free energy barrier to nucleation of freely grown (mainly cubic) ice is $\sim 24\, k_B T$ compared to $\sim 18\, k_B T$ for pre-formed (hexagonal or cubic) ice, and the critical size is ~ 114 compared to ~ 85 (see figure 6.2). The primary difference arises in the early stages of nucleation, where the ice grown directly from the supercooled liquid relaxes from mixed clusters to a much purer form of cubic ice. Nevertheless, the 'cubic' ice grown from a supercooled liquid has more hexagonal defects than the corresponding cubic ice grown from a small perfectly cubic crystal nucleus.

In a very recent study[98] of homogeneous ice nucleation in the mW model, Li and co-workers studied the kinetics of nucleation using forward flux sampling at a range of temperatures, including at 220 K. They found a critical crystal nucleus size of 265 at this temperature, which seems at first glance to be a considerably larger cluster than the sizes we obtain. However, the order parameter used in their study was different from the one we use; the principal difference is that all the neighbours of their classified ice particles were counted as being ice-like in their work (called 'surface' particles), whilst we treat such particles as liquid-like. We calculated the sizes of our critical clusters using their order parameters (utilising $l = 6$ spherical harmonics, a real-space neighbour cutoff of 3.2 Å and a dot product cutoff of 0.5). When surface particles are not included, the cluster size is approximately equal to the values we obtained; however, including surface particles gives critical sizes of approximately 150 and 200 (corresponding to sizes of 85 and 114 without surface particles, respectively). Although this is still less than their reported 265, the disagreement is not nearly as significant as it might initially appear. We can rationalise the discrepancy by noting that our system is able to relax locally due to the equilibration afforded it by umbrella sampling, whilst their system is driven forward aggressively by the forward flux sampling algorithm.

Similarly, Moore and Molinero have recently performed crystallisation simulations near the liquid transformation temperature;[324] they estimate that the critical nucleus at 208 K

comprises 120 particles, whereas our simulations suggest, following histogram reweighting (section 3.3), that the critical nucleus at that temperature contains approximately 50 particles for the simulations grown directly from the supercooled liquid. As with the study by Li and co-workers, the discrepancy arises primarily from Moore and Molinero's inclusion of what they call 'intermediate ice' in their largest cluster.

The question of whether 'surface' particles are ice-like or liquid-like is not easy to answer, and whether restrictive or generous order parameters are more realistic is unclear; this is indeed a rather general problem affecting all nucleation studies.[323] It is certainly the case in our experience that particles near the surface of 'core' ice can be very neatly arranged and thus reminiscent of the ice phase; but equally, there are many examples of particles adjacent to the ice cluster that have a structure totally incompatible with ice (such as having more than five neighbours). Nevertheless, although the choice of order parameter will affect the appearance of the free energy landscape (and the sampling efficiency), it should in principle have little effect on the barrier height and no effect on the rate of nucleation, provided that it is still a good parameter to describe the nucleation process.

Moore and Molinero studied the crystallisation of supercooled mW model water at 180 K, where they observed ice formed from the supercooled liquid to consist of intercalated layers of cubic and hexagonal ice in the approximate ratio 2 : 1.[169;282] We have not seen such behaviour in our simulations; an obvious possible reason for this difference is that their simulations were run at a lower temperature: indeed, they were run below the liquid transformation temperature of 202 K,[324] where the structure of the liquid phase changes. However, Li and co-workers observed the ice grown in their simulations to be initially hexagonal, but transforming to a mixture of cubic and hexagonal ice in the approximate ratio 1 : 1 in later stages of the nucleation at temperatures including 220 K.[98] Umbrella sampling allows clusters both to grow and to shrink and thus to equilibrate locally, whereas their observations may be a result of kinetics

dominating the growth pathway. In this regard, it is noteworthy that we have occasionally observed a hexagonal defect on cubic ice leading to further hexagonal ice growth and vice versa, and faster growth encourages more defects. Curiously, brute-force simulations by Russo and Tanaka suggest that water simulated using the mW potential at 206 K crystallises predominantly in a cubic fashion.[333]

It is also interesting to note that the system can become trapped in different regions of path space. This is the case not only for 'pure' cubic and hexagonal ice clusters, for which we would expect it to be difficult to interconvert once growth has started, but also between the ice clusters grown directly from the supercooled liquid, which were predominantly cubic, and the lower free energy pure cubic ice clusters. Such non-equilibrium effects are not new in studies of nucleation; for example, analogous behaviour has been observed in the crystal nucleation of a binary suspension of oppositely charged colloids.[334]

From classical nucleation theory, we can estimate the free energy of nucleation using equation (2.9). We can estimate $\Delta_{\mathrm{fus}}\mu$ by performing Gibbs–Helmholtz integration from the coexistence point ($T_1 = 274.6\,\mathrm{K}$[33]). To do this, we can evaluate

$$\frac{G_{\mathrm{m}}(T_2, p)}{T_2} = \text{constant} - \int_{T_1}^{T_2} \frac{\langle U\rangle_{\mathrm{m}} + p\,\langle V\rangle_{\mathrm{m}}}{T^2}\,\mathrm{d}T, \tag{6.4}$$

where $\langle U\rangle_{\mathrm{m}}$ is the mean molar potential energy and $\langle V\rangle_{\mathrm{m}}$ is the mean molar volume measured in a simulation, averaged over a reasonably large number of Monte Carlo cycles. We perform such simulations over a range of temperatures for both the liquid state and the hexagonal ice state and then find best-fit equations for the mean potential energy and the mean volume, and integrate them appropriately.[3] Using this approach in a series of short simulations, we find that the chemical potential difference (figure 6.5) between the two phases is $\Delta_{\mathrm{fus}}\mu(220\,\mathrm{K})/k_{\mathrm{B}} \approx 118\,\mathrm{K}$ per particle (consistent with the value obtained by Jacobson and co-workers[335]), in surprisingly good agreement, considering the degree of supercooling, with the value obtained from the simple

[3] For a longer discussion, see subsection 7.1.1.

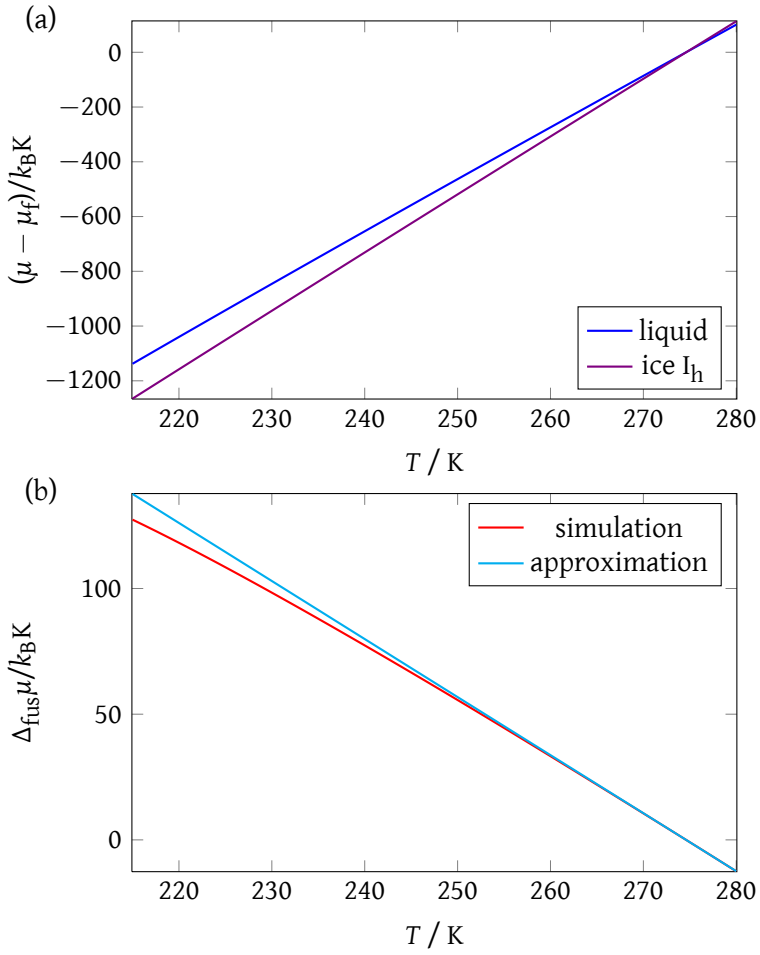


Figure 6.5: In (a), the approximate chemical potential difference of the liquid and the ice I_h phases from the chemical potential at the melting point (μ_f) for the mW model is shown as a function of temperature. In (b), the difference between the two curves depicted in (a) is shown alongside the simple approximation that $N\Delta_{fus}\mu \approx \Delta_{fus}H\left(1 - T/T_{fus}\right)$. This subfigure is similar to figure 8 of reference 335.

approximation that $\Delta_{fus}\mu(T) \approx \Delta_{fus}H_m(1 - T/T_{fus})$,[12] where we use $\Delta_{fus}H = 1.26\,\text{kcal mol}^{-1}$ and $T_{fus} = 274.6\,\text{K}$,[33] giving $\Delta_{fus}\mu(220\,\text{K})/k_B \approx 126\,\text{K}$ per particle. We are not aware of any calculated values of the interfacial free energy term for the mW model and so we use that determined for the TIP4P model, $24\,\text{mJ m}^{-2}$.[318] The classical nucleation theory free energy profile is shown for comparison in figure 6.2; it is qualitatively similar to the computed profiles, although the barrier is somewhat higher. One aspect to bear in mind is that the CNT profile passes through the origin, whereas for our order parameter, the size of the largest crystalline cluster has an average value of two in the supercooled water. Interestingly, a line of best fit

corresponding to $\Delta G(N) = -(\Delta_{\text{fus}}\mu)N + t_2 N^{2/3} + t_3 N^{1/3} + t_4$, where t_i are parameters determined by regression ($\Delta G(N)/k_\text{B}T = -0.54N + 3.45N^{2/3} + 0.84N^{1/3} - 6.13$), suggests the interfacial free energy attributable to the lowest free energy pathway is $\gamma \approx 22.9\,\text{mJ}\,\text{m}^{-2}$ (also shown in figure 6.2).[4] A similar analysis for the system grown directly from the supercooled liquid, where we fit all four parameters, including $\Delta_{\text{fus}}\mu$, suggests that $\Delta_{\text{fus}}\mu(220\,\text{K})/k_\text{B} \approx 120\,\text{K}$ and $\gamma \approx 26.2\,\text{mJ}\,\text{m}^{-2}$; in other words, the chemical potential does not change significantly, but the interfacial free energy becomes somewhat less favourable for 'mixed' crystal growth. Given the physically reasonable values obtained from these fits, we can conclude that classical nucleation theory appears to apply to a reasonable degree to the nucleation of mW ice; an accurate numerical determination of the interfacial free energy for the mW model would help in assessing this observation.

It is also interesting to note that the enthalpy associated with the growth of the crystal nucleus is a monotonically downhill function of the nucleus size (figure 6.6); this result implies that the barrier to nucleation is primarily entropic in nature. This suggests that the

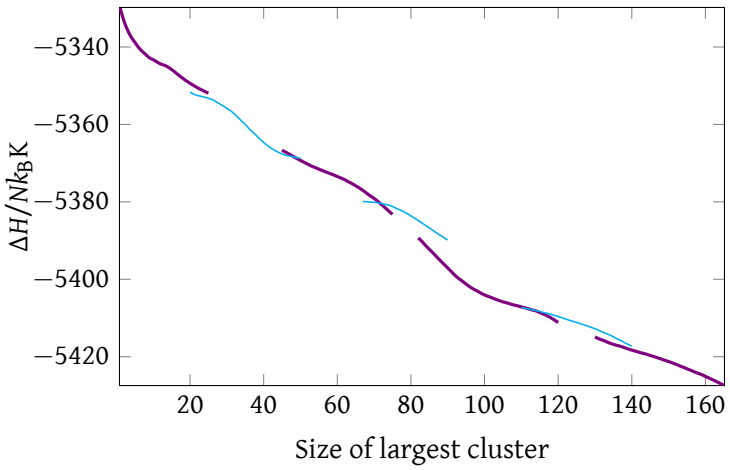


Figure 6.6: The enthalpy per particle as a function of the size of the largest crystalline cluster for the system seeded with a hexagonal ice cluster. Simulation results from different windows are depicted in alternating colours to show their overlap.

[4] We assume that t_2 corresponds directly to $\gamma(36\pi/\rho^2)^{1/3}$; the remaining two fitted terms serve primarily to shift the curve appropriately, as the simulation data curve goes through zero at a cluster size of 2; this is a consequence of the classification parameter not being well-suited to extremely small clusters.

formation of an ice-liquid surface is enthalpically favourable (or at least neutral), but entropically unfavourable. Since $\left(\frac{\partial \gamma}{\partial T}\right)_p = -S_{\text{interface}}$, where $S_{\text{interface}}$ is the entropy change per unit area upon the formation of an interface, this finding is consistent with experiment, where the interfacial free energy was found to decrease with temperature[150;173;210;232] (section 4.4).

Li and co-workers estimated the interfacial free energy from a fit of the nucleation rate to the CNT result as a function of the temperature, assuming that γ is temperature-independent and that the critical nucleus is spherical. They reported a value of $\gamma = 31.01\,\text{mJ}\,\text{m}^{-2}$,[98] which is higher than what we have determined. In order to compare our results with theirs, we have calculated the order parameter used by Li and co-workers for a large number (∼60 000) of independent configurations taken along the nucleation pathway for the simulations started from the supercooled liquid; we weighted each value with the associated free energy from figure 6.2. This produced an approximate free energy profile corresponding to nucleation measured with their order parameter (figure 6.7). Fitting a CNT-like expression, $\Delta G(N)/k_\text{B}T = t_1 N + t_2 N^{2/3} + t_3 N^{1/3} + t_4$, to this free energy profile allows us to calculate the

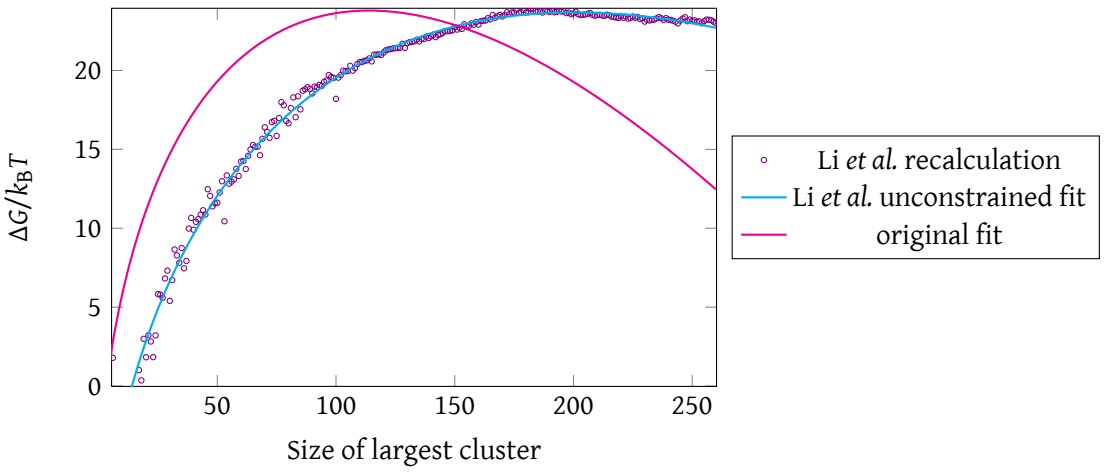


Figure 6.7: The free energy profile of mW nucleation as a function of the size of the largest crystalline cluster in the system as recalculated when using the order parameter proposed by Li and co-workers,[98] and the original fit for comparison. The height of the free energy barrier is the same for both approaches, but the size of the largest crystalline cluster is rather different.

interfacial free energy associated with the process. If we fix t_1 to the bulk chemical potential change ($t_1 = -\Delta_{\mathrm{fus}}\mu(220\,\mathrm{K})/k_{\mathrm{B}}T$), we find that $\gamma \approx 34.9\,\mathrm{mJ\,m^{-2}}$, in reasonable agreement with the result of Li and co-workers. By contrast, if we fit all four t_i parameters and interpret them using classical nucleation theory (shown in figure 6.7), we find that $\Delta_{\mathrm{fus}}\mu(220\,\mathrm{K})/k_{\mathrm{B}} \approx 81.52\,\mathrm{K}$ and $\gamma \approx 20.8\,\mathrm{mJ\,m^{-2}}$.

We suggest that this fitting process may provide a measure of how reasonable the order parameter used to track the nucleation process is; an unconstrained fit should reproduce $\Delta_{\mathrm{fus}}\mu$ reasonably when fitting data to CNT. This is the case when using our order parameter, but not when using the order parameter of Li and co-workers; their order parameter is too generous and the result is a chemical potential difference that does not appear to favour ice as much as it should, and a consequently lower interfacial free energy, since much of the ice cluster is liquid-like. If the assumption is that $\Delta_{\mathrm{fus}}\mu$ can be approximated by its bulk value in this case, then the interfacial free energy will necessarily be overestimated in compensation. We suggest, therefore, that counting 'surface' particles as being part of the ice cluster may result in an order parameter that is not well suited to being utilised in a fit to classical nucleation theory.

Since the mW potential is a reparameterisation of the Stillinger–Weber potential for silicon, it is worthwhile also briefly to compare the free energy profile to that calculated for the nucleation of Stillinger–Weber silicon, which Beaucage and Mousseau simulated in the canonical ensemble at a 25 % supercooling and for which they determined the critical nucleus size to be approximately 175.[336] Although the conditions and the order parameter used to track the nucleation process differ from our simulations, we note that the increased weight given to tetrahedrality in the mW potential appears to result in a smaller critical nucleus size. This is not unexpected in the light of classical nucleation theory, as the greater similarity of the liquid to the solid presumably decreases the interfacial free energy.

From a visual inspection of configurations, the crystalline clusters seemed to be fairly

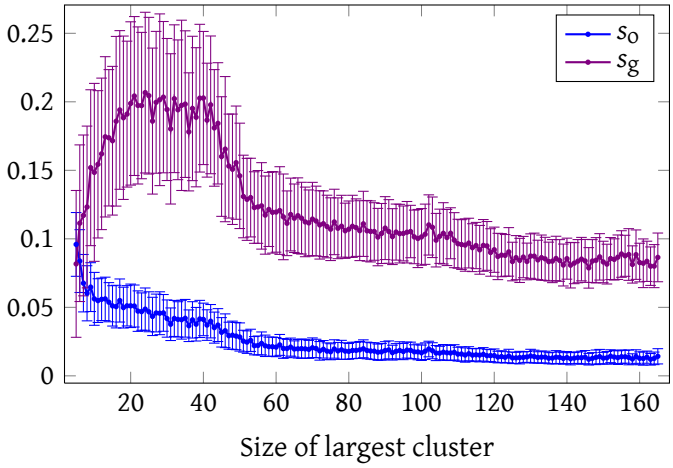


Figure 6.8: A plot of the two sphericity parameters against the size of the largest crystalline cluster calculated for snapshots of the system along the local order parameter used to drive nucleation in our system. Error bars show the standard deviation for the population of configurations at each cluster size. The diagram shows results for nucleation from a supercooled liquid; the other two systems behave analogously.

spherical. To quantify this observation, we have computed two measures of the sphericity of the clusters. The first is based upon the moment of inertia tensor, which has elements

$$I_{ij} = \sum_{k=1}^{N} m_k (r_k^2 \delta_{ij} - r_{ki} r_{kj}), \tag{6.5}$$

where r_{ki} is the i-th component of the vector between the cluster's centre of mass and particle k, r_k is the overall magnitude of this vector and δ_{ij} is the Kronecker delta. i and j refer to cartesian co-ordinates in an arbitrary laboratory frame of reference. The eigenvalues of this tensor give the principal components of the moment of inertia; we use these to quantify the orientational sphericity using the asphericity parameter[337]

$$s_o = \frac{(I_{xx} - I_{yy})^2 + (I_{xx} - I_{zz})^2 + (I_{yy} - I_{zz})^2}{2(I_{xx} + I_{yy} + I_{zz})^2}, \tag{6.6}$$

which ranges between zero for a perfectly spherical cluster and unity for extremely elongated ones. Any cluster with $s_o \lessapprox 0.1$ appears very much spherical by visual inspection. As the clusters grow in our simulations, their sphericity increases, as can be seen in figure 6.8, although this is to some extent a result of the discrete nature of clusters being felt more at smaller sizes. We also

characterise sphericity in terms of the radius of gyration, namely using[326]

$$s_{\mathrm{g}} = \frac{R_{\mathrm{g}}/\text{Å}}{1.5c^{1/3}} - 1, \text{ where } R_{\mathrm{g}}^2 = \frac{1}{c^2}\sum_{i}^{c}\sum_{j>i}^{c}\left|\boldsymbol{r}_i - \boldsymbol{r}_j\right|^2, \tag{6.7}$$

c is the size of the cluster and the $1.5c^{1/3}$ factor reflects the approximate radius of gyration of a perfect sphere of ice at the simulation temperature. This parameter tends to zero for perfect spheres. This measure again confirms that the crystalline clusters are reasonably spherical.

We remark here that while these parameters measure spherical symmetry, they do not principally measure compactness: a spherical cluster with small strings growing out of it evenly on all sides will appear to be more spherically symmetric than a compact oval-shaped cluster, even though we might expect the oval cluster to be more favourable (that is, having a lower surface area to bulk volume ratio). Nevertheless, we have not encountered any especially pathological clusters in our visual inspection of configurations.

6.2.2 Nucleation pathway

In order to compare further the nucleation behaviour we observe to that of Quigley and Rodger,[90] we have evaluated the global order parameters that they used along the nucleation pathways computed in this work. We used the Steinhardt-style Q_6 and Chau–Hardwick-style[313;314] tetrahedrality parameters as defined by Quigley and Rodger,[90] including their smoothing function, to ensure comparability of results. We thus have[5]

$$Q_l = \left(\frac{4\pi}{2l+1}\sum_{m=-l}^{+l}\left|\frac{1}{4N}\sum_{i=1}^{N}\sum_{j\neq i}^{N} f(r_{ij}) S_{lm}\left(\theta_{ij}, \varphi_{ij}\right)\right|^2\right)^{1/2}, \tag{6.8}$$

where N is the number of particles, $f(r_{ij})$ is the smoothing function, and we again use real spherical harmonics (the results with complex spherical harmonics are, of course, identical),

[5] This global order parameter is essentially identical to the global order parameter described by equation (5.4), except for the assumption that every particle has exactly four neighbours (*i.e.* N_{tot} in equation (5.4) is just $4N$ here) and the addition of a smoothing function.

and[6]

$$\zeta = \frac{1}{4N} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{\substack{k>j \\ k \neq i}}^{N} f(r_{ij}) f(r_{ik}) \left(\hat{\boldsymbol{r}}_{ij} \cdot \hat{\boldsymbol{r}}_{ik} + 1/3 \right)^2 , \tag{6.9}$$

where $\hat{\boldsymbol{r}}_{ij}$ is the unit vector from particle i to particle j. The smoothing function is defined as

$$f(r) = \begin{cases} 1 & \text{if } r \leq 3.1\,\text{Å}, \\ \frac{1}{2}\left(\cos \frac{(r/\text{Å} - 3.1)\pi}{0.4} + 1 \right) & \text{if } 3.1\,\text{Å} < r \leq 3.5\,\text{Å}, \\ 0 & \text{otherwise.} \end{cases} \tag{6.10}$$

The bulk ζ parameter for supercooled mW liquid water at 220 K tends to $\sim$0.21. For equilibrated cubic and hexagonal ice at 220 K, ζ approaches 0.026, while Q_6 approaches 0.44 for hexagonal and 0.51 for cubic ice.

It is also important to note that as these global order parameters measure the average order *of the entire system*, an ice cluster of a given size will produce different values of Q_6 and ζ depending on the overall size of the system.[338] Therefore, as we wish to compare our results to figure 3 of reference 90, where the number of particles in the simulation was 576, we have calculated the global order parameters in such a way that we have only taken into account the nearest 576 particles from the centre of mass of the crystal nucleus as identified by our local order parameter. The mean values of Q_6 and ζ along our pathway for the nucleation from a cubic ice seed are depicted in figure 6.9. At the top of our free energy barrier, we find $Q_6 \approx 0.10$ and $\zeta \approx 0.16$. By contrast, Quigley and Rodger report that the saddle point of their two-dimensional free energy landscape occurs at $Q_6 \approx 0.22$ and $\zeta \approx 0.19$.[90] Furthermore, in our simulations, the order parameters ζ and Q_6 vary roughly linearly as the ice cluster grows, as would be expected for the growth of a crystal nucleus into a relatively unperturbed liquid. By contrast, Quigley and Rodger initially observe an increase in Q_6 with little change in ζ. The nucleation pathway we

[6] We remark here that Quigley and Rodger suggest that ζ tends to unity for perfect tetrahedral networks;[91] however, since the tetrahedral angle is $\arccos(-1/3)$, the sum in equation (6.9) will clearly tend to zero when the network is tetrahedral. We therefore assume that Quigley and Rodger actually reported their results as a function of $1 - \zeta$.

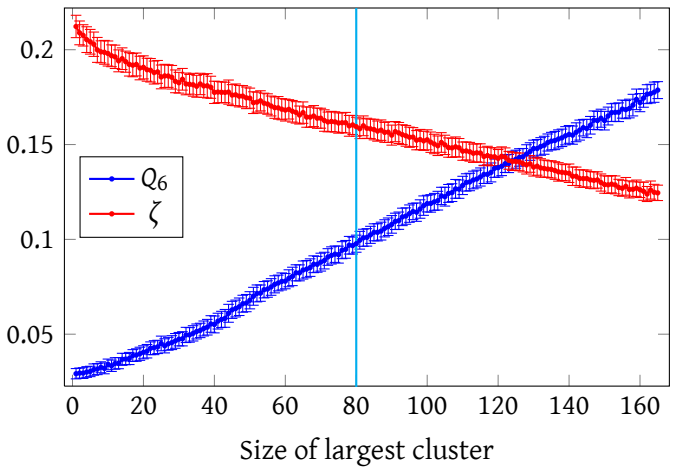


Figure 6.9: The global order parameters Q_6 and ζ calculated along the local order parameter used to drive nucleation in our system, for simulations initially seeded with a cubic ice nucleus. Error bars show the standard deviation for the population of configurations at each cluster size. The cyan solid line indicates the approximate location of the nucleation barrier as calculated by local order parameters. The results depicted here refer to the 576 particles nearest the centre of mass of the ice nucleus for comparison with reference 90, and not to the full 1400-particle system.

observe therefore appears to differ significantly from the one reported by Quigley and Rodger. The sigmoidal form of their Q_6-ζ free energy landscape suggests their pathway involves an initial increase in the Q_6 orientational order of the whole system, not just of a crystal nucleus. This 'pre-ordering' then appears to make crystallisation more tractable on computational timescales, even though this is not the natural pathway of nucleation.

Although it is difficult to claim that our order parameter reflects the 'natural' nucleation pathway, we can attempt to track the progress of nucleation as it happens spontaneously, without umbrella sampling, in order to be able to say whether the pathway we have identified is a reasonable one. An obvious difficulty with this approach is that the very reason why we used a rare event technique is that the process is too slow to observe in a brute force simulation. As a compromise, we can attempt to decrease the temperature of the simulation slightly such that the nucleation can be observed in a brute force simulation; we assume that the mechanism is not overly different at lower temperatures. An issue that can arise here is that at lower temperatures

that are closer to the liquid transformation temperature,[324] density fluctuations result in an unstable environment and any results obtained may no longer be representative of nucleation at the temperature for which we have computed a free energy profile. Nevertheless, between about 202 K and 220 K, we expect that the free energy of nucleation will decrease sufficiently so that we may be able to study brute force nucleation without observing any undesired behaviour close to the liquid transformation point. We estimate the free energy barriers using a histogram reweighting technique (section 3.3); an illustration of these estimates for several temperatures is given in figure 6.10. It is clear that the further we depart from the temperature of the simulation, the less representative the configurations we have actually sampled are of the (estimated) simulation at the temperature of interest, and hence the barriers to nucleation are considerably less smooth. However, the figure also reveals some small systematic deviations, which suggests that while the windows were sufficiently equilibrated in the free energy to give a smooth transition at 220 K, the enthalpy equilibration may not have been perfect (as can indeed readily be seen in figure 6.6); nevertheless, this is a small effect and, as we are not proposing to use the data at different temperatures for anything other than illustrative purposes, it is not a significant problem.

We suggest that the energy barrier at 210 K may be sufficiently low to allow us to observe spontaneous nucleation events. In 50 separate simulations of 1400 mW particles at 210 K and 1 bar, each comprising 8.5×10^8 Monte Carlo steps following equilibration, we observed a single nucleation event.[7] In this simulation, a crystal nucleus grew relatively quickly after an initial period of little growth. We tracked the three relevant order parameters (the size of the largest crystalline cluster as defined earlier, as well as Quigley and Rodger's global Q_6 and ζ

[7] It is also the case for ice growth simulations using the mW potential that molecular dynamics simulations are faster in computer time than the corresponding Monte Carlo simulations (section 7.1). Moore and Molinero suggest a typical waiting time of 200 ns for spontaneous nucleation at 208 K with this model, and this agrees well with our MD simulations at 210 K. The behaviour in terms of the three relevant order parameters of these MD simulations was very similar to the Monte Carlo trajectory discussed in the main text.

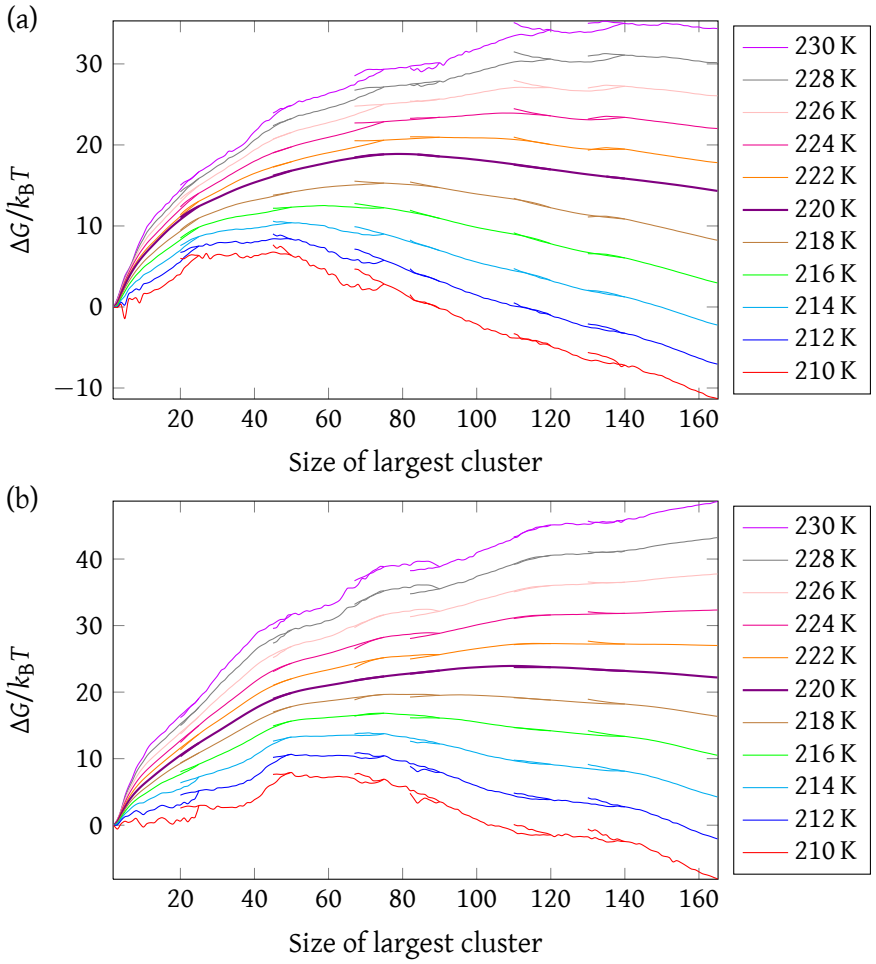


Figure 6.10: Histogram-reweighted free energy barriers to nucleation for (a) ice nucleation seeded with hexagonal crystal clusters and (b) ice nucleation directly from the supercooled liquid.

parameters) as functions of the step number; the data for the successful trajectory are shown in figure 6.11. Although there is considerable statistical fluctuation and we ought not to read too much into this diagram, as it tracks a single trajectory, we can nevertheless see a nucleation event approximately half way into the simulation, and the change in all three order parameters is approximately linear. Notice that, although there was only one large nucleus in the system, the nucleation is not sudden even once the critical nucleus size is reached; this reflects the slow kinetics of ice growth. We have also evaluated averages of Q_6 and ζ as functions of the size of the largest crystalline cluster; these are shown in figure 6.12 and correspond very well indeed to the data obtained in umbrella sampling simulations (figure 6.9). We thus believe that this

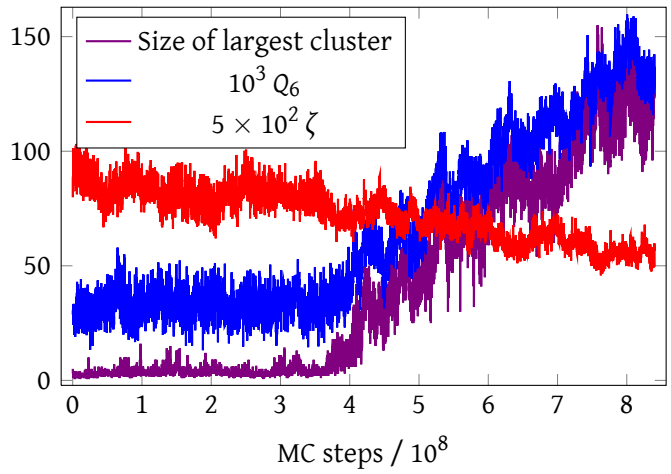


Figure 6.11: The size of the largest crystalline cluster and of the global order parameters Q_6 and ζ (evaluated for the nearest 576 particles from the centre of mass of the crystalline nucleus) along a single trajectory of a brute force simulation of 1400 particles at 210 K.

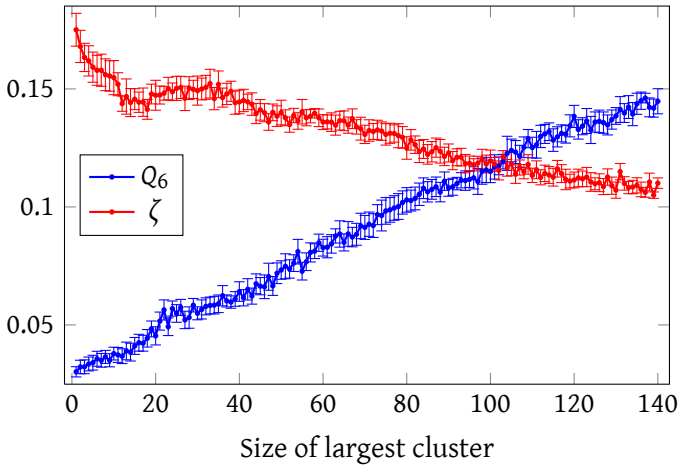


Figure 6.12: The global order parameters Q_6 and ζ calculated as functions of the local order parameter used in our umbrella sampling simulations, tracked in a brute force simulation of 1400 particles at 210 K. Error bars show the standard deviation for the population of configurations at each cluster size. The results depicted here refer to the 576 particles nearest the centre of mass of the ice nucleus for comparison with reference 90, and not to the full 1400-particle system.

brute-force behaviour is a good indication that applying umbrella sampling does not change the natural nucleation pathway.

6.2.3 Nucleation rate

Configurations from the top of the barrier obtained in Monte Carlo umbrella sampling simulations were taken as starting points in isobaric-isothermal MD simulations (with a Nosé–Hoover barostat and thermostat with relaxation times 2 ps and 1 ps, respectively, and a time step of 5 fs). Velocity components were assigned from a normal distribution with mean zero and scaled to correspond to the simulation temperature of 220 K.

For the system nucleating from a seed hexagonal crystal, we ran 40 simulations from distinct starting configurations at the top of the barrier. Of these, 17 exhibited nucleus growth and the remainder nucleus shrinkage. This roughly 1 : 1 ratio of growing and shrinking suggests that we are approximately at the top of the barrier. We calculate the mean attachment rate as a function of the time across the simulations (figure 6.13). Taking the average of the points at $t > 1 \times 10^{-10}$ s gives a long-time attachment rate of $f_{n_{\text{crit}}} = 1.175 \times 10^{13}\,\text{s}^{-1}$. The Zeldovich factor is $Z^2 = \left|\Delta G''(n_{\text{crit}})\right| / 2\pi k_\text{B} T = 0.00216/2\pi = 0.000344$, where the

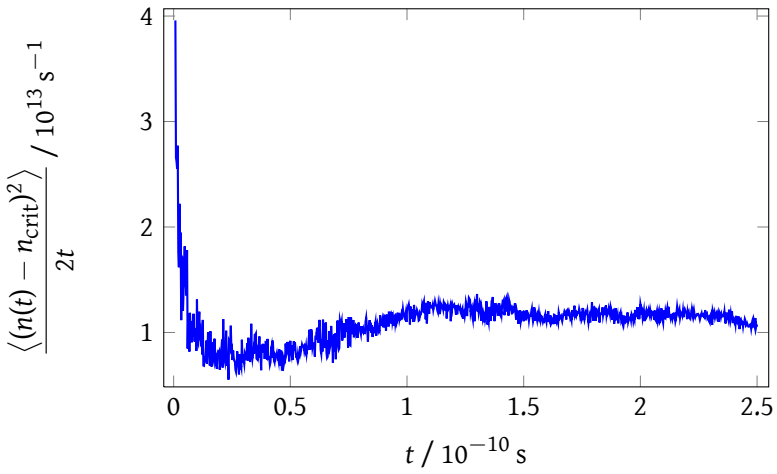


Figure 6.13: The mean local-time attachment rate as a function of time, averaged over all 40 MD simulations of 2520 mW particles starting from an 85-particle crystal nucleus.

second derivative of the Gibbs energy was evaluated from the CNT fit given above, and the number density of the liquid is $\rho_{liq} = 3.3 \times 10^{28}\,m^{-3}$. This gives a kinetic pre-factor of $\kappa = Z\rho_{liq}f_{n_{crit}} = 7.17 \times 10^{39}\,m^{-3}\,s^{-1}$. The probability of reaching the top of the free energy barrier is $P(n_{crit}) = \exp\left[-\Delta G(n_{crit})/k_B T\right] = 9.0 \times 10^{-9}$. The rate of mW ice nucleation at 220 K is therefore estimated to be $J = 6.5 \times 10^{31}\,m^{-3}\,s^{-1}$.

We performed analogous simulations starting from the critical cluster in the system which was grown directly from the supercooled liquid and has a critical nucleus size of 114. Of the 40 MD trajectories started from this point, 20 exhibited shrinking and 20 exhibited growth of the largest crystal nucleus. The attachment rate was found to be $1.07 \times 10^{13}\,s^{-1}$, the second derivative of ΔG evaluated at the top of the barrier was $0.00159\,k_B T$ and the probability of reaching the top of the barrier was $P(n_{crit}) = 4.6 \times 10^{-11}$, giving a rate of nucleation of $J = 2.6 \times 10^{29}\,m^{-3}\,s^{-1}$.

Using forward flux sampling[94;95;97] with molecular dynamics,[339;340] Li and co-workers have determined[98] nucleation rates for the homogeneous nucleation of ice with the mW model ranging between $2.148 \times 10^{25}\,m^{-3}\,s^{-1}$ at 220 K and $1.672 \times 10^{-7}\,m^{-3}\,s^{-1}$ at 240 K, 5–10 orders of magnitude below the experimental values at the temperatures where these have been reported (see section 4.4). Our results are somewhat higher than those reported by Li and co-workers. To an extent, we can rationalise this difference by noting that their use of forward flux sampling ensures a purely kinetically driven process; by contrast, our systems are locally equilibrated. This is not usually a problem with algorithms such as forward flux sampling because local equilibration is generally considerably faster than progression to a higher value of the order parameter serving as a reaction co-ordinate; however, in systems such as water, local equilibration can be very slow. The slow dynamics can cause the forward flux sampling algorithm to result in poor sampling.[341] This serves to rationalise the difference in the calculated rate between the system we have grown directly from the supercooled liquid and

the one simulated by Li and co-workers.

A second issue is that there is no straightforward way for a simulation method to choose the pathway of fastest growth if this difference arises only for somewhat larger crystal nuclei; indeed, the free energy profiles we have calculated look the same for small clusters. In forward flux sampling, the order parameter space is staged; however, if all the pathways in the initial stages are equally fast, then the pathways chosen in forward flux sampling will be the ones that are the dominant ones in numbers, not the ones that may, ultimately, prove to be faster. In real experimental systems, there is no staging which can force the system up a higher free energy path, as a fluctuation will generally be rare but fast; real-life nucleation will thus always follow the lowest free energy pathway unless the different pathways are only very slightly different in free energy. In forward flux sampling, once an ultimately slower pathway is chosen, the simulation is extremely unlikely to be able to backtrack and choose a faster pathway. We suggest that this is why Li and co-workers did not see the nucleation of a pure phase of ice, even though our results suggest that the rate of nucleation of pure hexagonal ice is two orders of magnitude higher than that of mixed ice. A greedy simulation algorithm such as forward flux sampling, despite sampling natural dynamics, is even more likely to proceed down the 'wrong' pathway than simulation methods which allow for local equilibration, such as umbrella sampling. Nevertheless, umbrella sampling is by no means immune to this problem: indeed, the hexagonal crystal that we have grown was only possible because we effectively imposed that nucleation pathway by seeding the system.

6.3 Conclusions

We have simulated the homogeneous nucleation of ice as a function of the size of the largest crystalline cluster using the mW potential. The free energy barriers we have calculated agree reasonably well with classical nucleation theory, and the interfacial free energies extracted

from fits to the computed free energy profiles appear to be sensible. We have observed that the crystalline clusters grown directly from the supercooled liquid tend to be cubic, although they have a higher free energy than those clusters grown from ideal crystalline seeds.

A comparison with the free energy landscapes computed in the previous studies of homogeneous ice nucleation[90;287;291] at a similar undercooling (∼20 %), albeit with the TIP4P model, reveals that the free energy barrier calculated in this work is considerably lower. We suggest that previous simulations were not at equilibrium in path space; that is, the lowest free energy paths between the liquid and the crystal were not located. The high free energy barriers previously reported may be a result of non-local[90;287] and non-rotationally-invariant[291] order parameters used to track the progress of nucleation. We therefore hypothesise more generally that care should be exercised when using global orientational order parameters, because there may be a tendency for the pathways that are easiest to locate with rare event techniques not actually to be those with the lowest free energy, particularly for systems such as water, where crystal growth is slow.

The simulations we have performed allow us to learn some lessons for future simulations of all-atom models of water. If the mW potential is a reasonable model for the energetics of water, as is strongly suggested by the work of Molinero and co-workers,[33] the difficulty in simulating ice nucleation in all-atom models such as TIP4P may be more due to the slow kinetics of the process than an insurmountable free energy barrier. We suggest that, to test this hypothesis, a study of the homogeneous nucleation of ice using an all-atom water potential and driven by a local order parameter would be a worthwhile, if difficult, endeavour. This is the topic of the following chapter.

7 Homogeneous ice nucleation with all-atom water models

The crystallisation of ice has been studied in a large number of simulations (subsection 4.5.2); nevertheless, whereas several simulations of homogeneous nucleation using the mW coarse-grained water model[33] have been reasonably successful,[98;169;282] including ours (chapter 6), all-atom simulations have been less so. The mW potential is a good representation of the structure and the thermodynamics of water; by contrast, it has unrealistically fast dynamics and no representation of hydrogens, and so the use of an all-atom model would offer considerable further insight into the process. However, Matsumoto and co-workers' single MD trajectory of TIP4P water nucleating into ice remains the only successful brute-force simulation of homogeneous ice nucleation with an all-atom model.[286] Other simulations have used small system sizes or looked at conditions that are not representative of homogeneous nucleation from bulk liquid water. While brute-force simulations of a rare event are unlikely to be successful (subsection 4.5.2), the use of rare event methods can potentially allow us to compute free energy landscapes for nucleation. We have already discussed the simulations of Radhakrishnan and Trout,[287;288] who used umbrella sampling, and of Quigley and Rodger,[90] who used metadynamics, and why the use of global order parameters in driving homogeneous ice nucleation can lead to non-physical nucleation pathways (section 5.2).

The principal difficulty in simulating homogeneous ice nucleation using all-atom models of water seems to be that the natural dynamics associated with ice crystallisation are extremely slow. For example, even the growth by a few layers of a planar ice-water interface below the

freezing point, which is a purely downhill process in free energy, can take weeks of computer time with a TIP4P-type model.[117] The fastest rate of growth of such an interface is just below the freezing point,[292;293;296] but this is where the free energy barrier to nucleation is very large, which makes homogeneous nucleation a difficult process to simulate.

It is a surprising state of affairs that modern simulation methods have so far not been able to capture the fundamental physical behaviour of the homogeneous nucleation process of ice. Some of the outstanding problems are how to simulate the homogeneous nucleation of ice using a local measure of order and the determination of a free energy landscape using such an order parameter. In this chapter, we attempt to address both these aspects: we first apply the order parameter presented in chapter 5.3 to overbiassed homogeneous ice nucleation simulations at a low supercooling, at which the dynamics of nucleation are reasonably fast, in order to demonstrate that the order parameter we have developed is capable of both tracking and driving the homogeneous nucleation process (section 7.1). We then investigate several approaches to achieving equilibrium sampling of ice nucleation at lower temperatures, including hamiltonian exchange (section 7.2) and the use of histogram reweighting from simulations run at high temperature (low supercooling), at which we demonstrate that it is feasible to equilibrate the ice nucleation process (section 7.3).

7.1 Driven ice nucleation from supercooled TIP4P/2005 water

All-atom models of water have particularly slow dynamics, especially for crystal growth, making any possible improvement in computational speed worth considering. It has been shown that ice crystal growth occurs more rapidly (in computer time) in molecular dynamics (MD) simulations than in corresponding Monte Carlo ones,[117;118] which may suggest that the collective motion possible in MD simulations helps to speed up the dynamics of cluster reorganisation, and thus aids us in driving crystallisation. The choice of MD simulations over MC simulations is

therefore appealing; however, as we are ultimately interested in the free energy landscape of ice nucleation, we wish to continue to use umbrella sampling and thus Monte Carlo simulations.

While the majority of simulations presented in this section used the standard Metropolis Monte Carlo (MC) approach (section 3.1) in the isobaric-isothermal ensemble coupled with umbrella sampling (section 3.2) and the order parameters presented in section 5.3, we have recently begun to couple Monte Carlo with MD simulations in a hybrid Monte Carlo approach (section 3.8), where short MD simulations replace single particle rotational and translational Monte Carlo moves. Provided that the MD integrator is time reversible and symplectic and that the choice of momenta from the Maxwell–Boltzmann distribution is accounted for in the Metropolis acceptance criterion, detailed balance is obeyed[39;119–122] irrespective of the fact that the MD hamiltonian does not incorporate an umbrella sampling term (section 3.8). We implement the symplectic and time reversible quaternion-based algorithm of Miller III and co-workers[129;130] to simulate rigid body rotations (appendix D.4).[1] We are able to drive nucleation using both methods; however, simulations are considerably faster in real time when the hybrid Monte Carlo approach is used,[2] which confirms the importance of collective motion for nucleation.

7.1.1 Driving force for TIP4P/2005 nucleation

Classical nucleation theory predicts (equation (2.7)) that the height of the free energy barrier to nucleation depends on the interfacial free energy between the phases in question (to the third

[1] We also initially tried to perform MD simulations of rigid bodies by first performing a non-constrained move and then applying a constraints algorithm; however, we encountered significant difficulties doing so. While reasonably long MD simulations using an analytic constraints algorithm[131] result in the correct behaviour, short MD simulations of only a few steps within a hybrid Monte Carlo scheme invariably led to a (catastrophic) decrease in the potential energy of the system.

[2] In our hybrid Monte Carlo simulations, we chose the MD timestep and the number of MD steps per Monte Carlo step to result in an approximately 40 % acceptance ratio. For example, at 240 K, we used an MD timestep of 0.75 fs and 13 MD steps per MC step. A constant pressure is achieved by attempting an MC volume move on average once every four MC steps. As a point of comparison, nucleation from a cubic ice seed in a standard Monte Carlo simulation took approximately six times longer than the growth from a hexagonal seed to similar final sizes in a hybrid Monte Carlo simulation.

power) and the chemical potential difference between the phases (to the second power). Since as the temperature decreases, the chemical potential favours the crystalline phase more and more, and the interfacial free energy also decreases,[210;320] we might expect that homogeneous nucleation will be increasingly more favourable the further the system is supercooled. However, while this holds to an extent, we also need to consider the effect of the system's kinetics on the speed at which the process of nucleation takes place (notwithstanding the free energy barrier to nucleation itself). Indeed, at very low temperatures, the 'liquid' phase can be so dynamically arrested that we might describe it as a glass.[29]

It is difficult to quantify the relevant dynamic behaviour, as self-diffusion (the coefficient of which has been determined for the TIP4P/2005 model of water even at very large supercoolings[255;257]) alone is not likely to be the only dynamical process slowing down the nucleation. It is also relatively difficult to measure the interfacial free energy, especially at a significant supercooling; however, some interfacial free energies have been determined in simulations for several models of water with various levels of agreement amongst each other.[318;319;342;343]

What we can, however, do reasonably straightforwardly, is to estimate the difference in chemical potential using a simple approach. For a single-component system, the chemical potential is equal to the molar Gibbs energy. The Gibbs–Helmholtz equation is[40] $\left(\frac{\partial(G/T)}{\partial T}\right)_p = -\frac{H}{T^2}$, and so, at constant pressure,

$$\frac{G(T_2, p)}{T_2} = \frac{G(T_1, p)}{T_1} - \int_{T_1}^{T_2} \frac{H(T)}{T^2}\,\mathrm{d}T. \tag{7.1}$$

From a series of isobaric-isothermal simulations, we can therefore calculate the Gibbs energy relative to some standard Gibbs energy. Coexistence lines have been determined for several ice phases by thermodynamic integration from various 'ideal' states and an excellent review for the determination of such data has recently been published by Vega and co-workers.[190] Since the

melting point of TIP4P/2005 water has been well characterised,[252] and we know that at that point, the chemical potentials of the two species are identical, this can act as the starting point for our thermodynamic integration (*i.e.* the point at which $\Delta G = 0$).

In an isobaric-isothermal simulation, we can evaluate

$$\frac{G_\mathrm{m}(T_2, p)}{T_2} = \text{constant} - \int_{T_1}^{T_2} \frac{\langle U\rangle_\mathrm{m}(T) + p\,\langle V\rangle_\mathrm{m}(T)}{T^2}\,\mathrm{d}T, \tag{7.2}$$

where $\langle U\rangle_\mathrm{m}$ is the mean molar potential energy[3] and $\langle V\rangle_\mathrm{m}$ is the mean molar volume measured in a simulation, averaged over some number of MC cycles. We can perform such simulations at several temperatures for both a liquid state and an ice state, and then find best-fit equations for the potential energy and the volume and integrate them appropriately to give the chemical potential difference, as we have done in subsection 6.2.1 for the mW model. However, an equation of state has been reported by Noya and co-workers for ice $\mathrm{I_h}$ using the TIP4P/2005 model of water.[254] The number density of ice $\mathrm{I_h}$ at 1 bar pressure is given by[254]

$$\begin{aligned}\rho_{\mathrm{I_h}}(T)/\text{Å}^3 = {} & \left(-3.65556\times10^{-10}(T/\mathrm{K})^2 - 1.59648\times10^{-7}(T/\mathrm{K}) + 0.999849\right)^{0.0453293} \\ & \times \left(9.35249\times10^{-12}(T/\mathrm{K})^3 - 8.81404\times10^{-9}(T/\mathrm{K})^2\right. \\ & \left. - 2.58957\times10^{-6}(T/\mathrm{K}) + 0.0318622\right),\end{aligned} \tag{7.3}$$

and its potential energy is given by[4] $U(T)/k_\mathrm{B} = -7643.01\,\mathrm{K} + 3.6938T$. These fits are consistent with our simulation results within statistical error. For the liquid, analogous measurements have been undertaken,[255;344;345] and we have calculated a polynomial fit[5] to the results reported by Pi and co-workers,[255] giving a number density expression of

$$\begin{aligned}\rho_\mathrm{liq}(T)/\text{Å}^3 = {} & 1.09634\times10^{-12}(T/\mathrm{K})^5 - 1.22799\times10^{-9}(T/\mathrm{K})^4 + 5.38691\times10^{-7}(T/\mathrm{K})^3 \\ & - 0.000115505(T/\mathrm{K})^2 + 0.0121095(T/\mathrm{K}) - 0.465873\end{aligned} \tag{7.4}$$

[3] In principle, to find G, we would need to include a kinetic energy term as well. However, as we are only interested in the *difference* between ice and water, and the kinetic energy from equipartition is the same in both states, this term would ultimately disappear.

[4] The potential energy is fitted by us to the potential energies at 1 bar pressure reported by Noya and co-workers in the supporting information to reference 254.

[5] The fits are very reasonable for our estimation, but do not follow the data points perfectly; they should not be used for anything other than estimates.

and a potential energy expression of

$$U(T)/k_\mathrm{B}\mathrm{K} = 1.32996 \times 10^{-7}(T/\mathrm{K})^5 - 0.000149451(T/\mathrm{K})^4 + 0.065877(T/\mathrm{K})^3 - 14.2004(T/\mathrm{K})^2 + 1500.94(T/\mathrm{K}) - 69273.7. \tag{7.5}$$

Using these expressions and the melting point[6] of TIP4P/2005 ice of 250.5 K,[117] we can calculate the relevant chemical potential difference as a function of the temperature: this is depicted in figure 7.1(a).

The form of the curves shown in figure 7.1(a) may at first glance appear unusual;

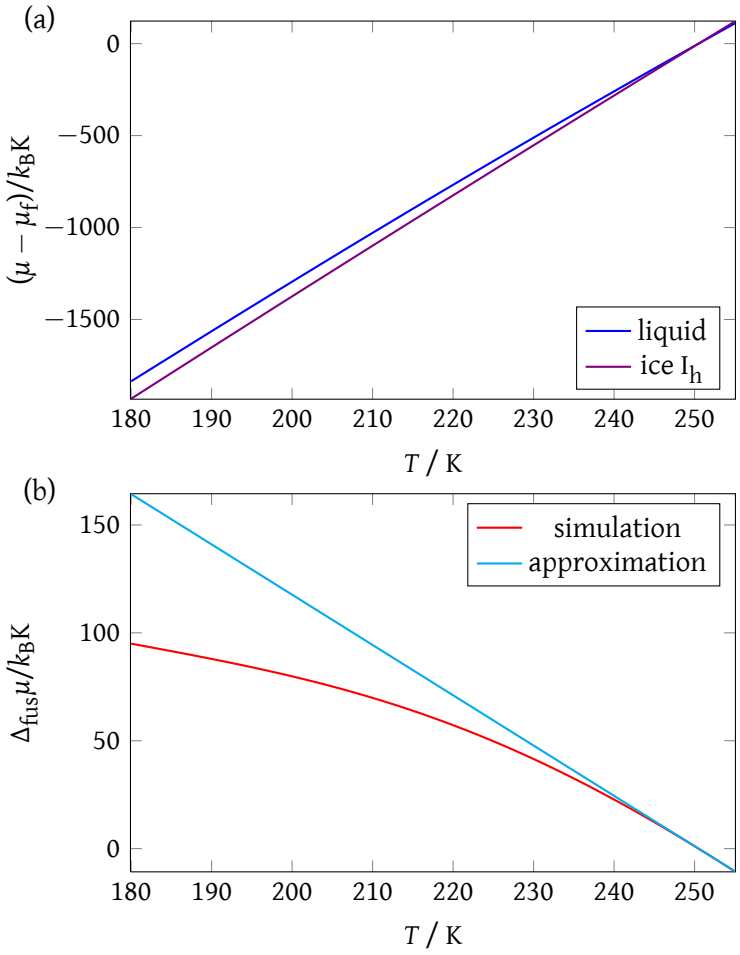


Figure 7.1: In (a), the chemical potential difference of the liquid and the ice $\mathrm{I_h}$ phases from the chemical potential at the melting point (μ_f) for the TIP4P/2005 model is shown as a function of temperature. In (b), the difference between the two curves depicted in (a) is shown alongside the simple approximation that $N\Delta_\mathrm{fus}\mu \approx \Delta_\mathrm{fus}H\left(1 - T/T_\mathrm{fus}\right)$.

[6] The melting point is reported as 250.5 K in reference 117, but has been revised upwards to 252 K in newer work.[252] For our estimate here, this small difference is not very significant.

from the fundamental equation for the Gibbs energy, we know that $\left(\frac{\partial G}{\partial T}\right)_p = -S$, and since the entropy is always non-negative, G should be a monotonically decreasing function.[271;346] However, the curves shown in figure 7.1 are not the absolute chemical potentials; the majority of the temperature dependence arises from the absolute value of the Gibbs energy at coexistence in equation (7.2), which we have arbitrarily set to zero. However, that temperature dependence is the same for both phases, and since it is only the *difference* in the chemical potential that we are interested in (figure 6.5(b)), this need not be taken into account. The free energy calculations performed by Gao and co-workers on the analogous TIP4P model show that, when calculated absolutely, the expected behaviour is reproduced under the conditions of interest.[346]

The change in chemical potential is given by $N\Delta_{\text{fus}}\mu(T) = \Delta_{\text{fus}}H - T\Delta_{\text{fus}}S$. On the assumption that $\Delta_{\text{fus}}H$ and $\Delta_{\text{fus}}S$ are temperature independent, and using the identity $\Delta_{\text{fus}}S = \Delta_{\text{fus}}H/T_{\text{fus}}$, where T_{fus} is the equilibrium freezing temperature, we can estimate[12] the change in chemical potential as

$$N\Delta\mu_{\text{fus}} \approx \Delta_{\text{fus}}H\left(1 - \frac{T}{T_{\text{fus}}}\right). \tag{7.6}$$

For the TIP4P/2005 model, $\Delta_{\text{fus}}H/Nk_{\text{B}} = 584\,\text{K}$ per molecule[252] and $T_{\text{fus}} = 250.5\,\text{K}$.[117] Curiously, compared to the case of the mW potential (figure 6.5(b)), the chemical potential of fusion as a function of the temperature determined above (and depicted in figure 7.1(b)) deviates more rapidly from this simple approximation. Unfortunately, the chemical potential difference, which is the driving force for nucleation, changes very gradually with the supercooling, making the study of the homogeneous nucleation of ice especially challenging. Indeed, even at a 20 % supercooling at 200 K, the chemical potential driving force is just $\sim 0.4\,k_{\text{B}}T$ per particle, which is a weak driving force for nucleation.[37]

7.1.2 Nucleation pathways

We have run overbiassed umbrella sampling simulations in a variety of systems with TIP4P/2005 water. Typical simulations involved between 1900 and 2500 water molecules and three distinct scenarios were considered: growth from a seed hexagonal ice cluster, growth from a seed cubic ice cluster and growth directly from the supercooled liquid water.

The approach we used was to bias the umbrella sampling weights to favour larger clusters. We typically chose these weights to correspond roughly to the negative of the free energy predicted by classical nucleation theory for a given cluster size: the precise potentials we used were changed somewhat as the simulations progressed.[32] However, these weights must not be changed too rapidly, since if the umbrella weights bias the growth to be too quick, then the system can begin to grow defective crystal nuclei that cannot repair themselves by shrinking and regrowing: the weights must be sufficiently small to allow clusters to grow and shrink throughout each umbrella sampling window.

Several snapshots of ice growth in such simulations are depicted in figure 7.2. Provided that umbrella sampling does not attempt to drive the nucleation too quickly, the resulting ice clusters appear to be reasonable: for example, they do not span the simulation box and are compact. This suggests that the order parameter presented above is a suitable order parameter both to track and to drive the process of homogeneous ice nucleation.

It is intriguing to note that it appears that in what we believe to be overbiassed driving of nucleation, cubic ice seeds seem to grow in a cubic fashion, and conversely, hexagonal ice seeds result in the growth of hexagonal ice. Growth directly from the supercooled liquid is somewhat more tricky: because the order parameter we use cannot track clusters smaller than 6 molecules, and cannot readily track the growth of very small clusters of ice, it was necessary to wait for small clusters to form spontaneously. These were then biassed to grow further, although with a very gentle set of umbrella weights, and only when the growth became spontaneous with

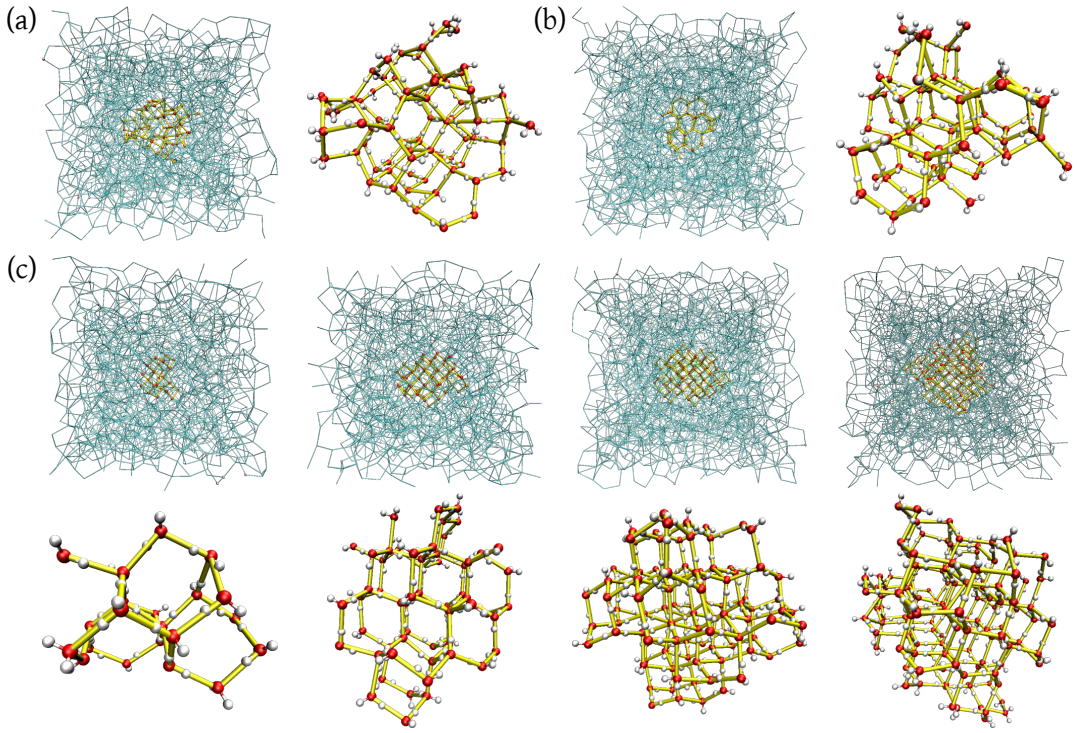


Figure 7.2: Representative nucleation snapshots from umbrella sampling simulations of TIP4P/2005 water. In each case, two pictures depict the same cluster from different perspectives; one within the liquid framework (in cyan) and one showing solely the largest crystalline cluster. In the former, spheres represent centres of mass of molecules classified as ice: red spheres correspond to cubic ice, orange spheres correspond to hexagonal ice and pink spheres correspond to ice molecules not within the largest crystalline cluster. Pictures representing solely the largest cluster depict both the oxygen (red) and the hydrogens (white) of each molecule. In (a), an 82-molecule ice cluster grown from the supercooled liquid at 240 K is shown; in (b), a 73-molecule ice cluster grown from a small cluster of I_h ice at 240 K is shown; and in (c), a series of ice clusters of increasing size (comprising 23, 60, 107 and 145 molecules from left to right) grown from a small cluster of I_c ice at 200 K is depicted. There are 1900 molecules in the system in (a) and the first two configurations of (c), and 2500 molecules in (b) and the last two configurations of (c). Simulations of nucleation from a hexagonal seed (shown in (b)) were undertaken using the hybrid Monte Carlo approach, and the rest by a standard Metropolis Monte Carlo approach. $p = 1$ bar.

that set of weights did we progress to higher umbrella sampling windows. Analogously to what we observed in mW simulations (subsection 6.2.1), ice grown directly from the supercooled liquid contains both cubic and hexagonal ice patterns, with cubic ones dominating, but less so than in the corresponding mW clusters: for example, 60-molecule ice clusters in this work were classified to have approximately 70 % core cubic ice, whilst the mW analogues were about 90 % cubic. Whether this is a result of a true difference between the TIP4P/2005 and mW models of water or simply a consequence of overbiassed non-equilibrated driving in this work is unclear and warrants further investigation. The ice clusters observed in this work are roughly spherical (figure 7.3), and this sphericity follows the same trends as for the mW model nucleation reported in the previous chapter.

Matsumoto and co-workers looked at some properties of the clusters they observed along the nucleation pathway of their spontaneous MD nucleation trajectory.[286] Their clusters comprise molecules that are connected by a network of long-lived hydrogen bonds, and thus include the majority of the cluster surface. Nevertheless, non-compact clusters were observed

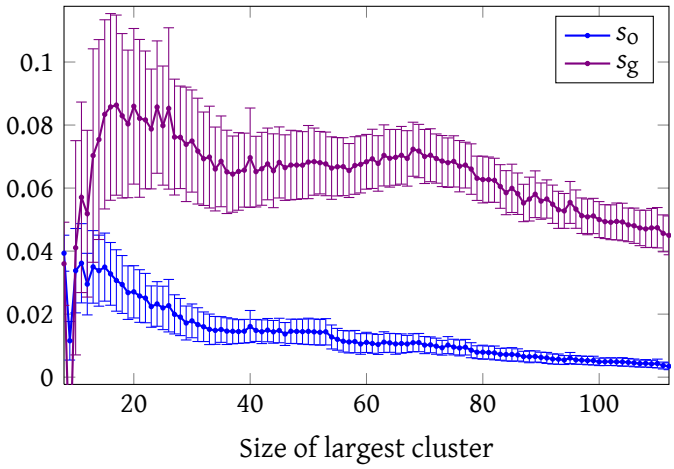


Figure 7.3: A plot of the two sphericity parameters introduced in subsection 6.2.1 against the size of the largest crystalline cluster calculated for snapshots of the system along the local order parameter used to drive nucleation in our system. Error bars show the standard deviation for the population of configurations at each cluster size. The diagram shows results for nucleation from a cubic seed at 200 K.

not to lead to successful nucleation pathways, whereas the clusters associated with nucleation were much more compact and exhibited only few chains (figure 4 of reference 286). We can compare this behaviour to what we observe in driven nucleation simulations. Although the order parameter we use to drive nucleation does not count molecules in chains as belonging to the largest crystalline cluster, it does not, in principle, suppress such growth, and so if chain growth were a natural feature associated with nucleation, we might expect chains to grow nevertheless. In order to investigate whether chains are such a feature of the nucleation process or, as we suggested previously, an artefact of the order parameter coupled with the slow dynamics of ice growth, we have calculated the numbers of molecules that are classified as belonging to chains along the nucleation pathway as driven by the order parameter presented above. In fact, chains are usually present on the surface of the growing ice nuclei; however, the absolute numbers of molecules involved are small: for example, in the set of simulations started from a hexagonal seed cluster, a cluster of 50 molecules has on average only 3.2 molecules belonging to chains, and such chains grow reasonably uniformly on all sides of the crystalline cluster surface. Although certain configurations do exist with longer chains, chain growth does not seem to play a crucial rôle in ice nucleation as driven by our order parameter, which is consistent with the behaviour observed in spontaneous nucleation by Matsumoto and co-workers.

Although we made no serious attempt to equilibrate the nucleation simulations, it may nevertheless be possible to gain physical insight into the nucleation pathway from our simulations. In order to compare the pathway observed with that reported in our simulations using the mW potential (subsection 6.2.2) and with Quigley and Rodger's simulation of TIP4P nucleation,[90] we calculate the Steinhardt-style Q_6 and Chau–Hardwick-style[313;314] tetrahedrality parameters as defined by Quigley and Rodger,[90] including their smoothing function, as defined in subsection 6.2.2 (equations (6.8)–(6.10)). We take into account the nearest 576 molecules from the centre of mass of the largest crystalline cluster as determined by the

local order parameter defined above in order to ensure that these results can be compared to the previous work. The resulting diagram for the simulation of nucleation from a cubic ice seed nucleus is depicted in figure 7.4. Although the curves are rather noisy, as can be expected from a set of simulations that have not been equilibrated, and there are clear minor variations in slope corresponding to different umbrella sampling windows, we can nevertheless observe that the two order parameters plotted change roughly linearly as the cluster size increases. This linearity, which reflects the growth of an ice nucleus into a largely unperturbed liquid, is consistent with the nucleation pathway we reported for the mW model of water in the previous chapter, although the actual values of the global order parameters suggest that the system studied here is less well ordered than its mW analogue. This is perhaps not surprising considering that the mW systems we studied previously were very well equilibrated. Importantly, the pathway is rather different from that observed in Q_6-ζ space by Quigley and Rodger,[90] as their free energy landscape involves an initial increase in the Q_6 orientational order before the tetrahedrality parameter ζ begins to change, which implies that the entire system, rather than just a crystalline

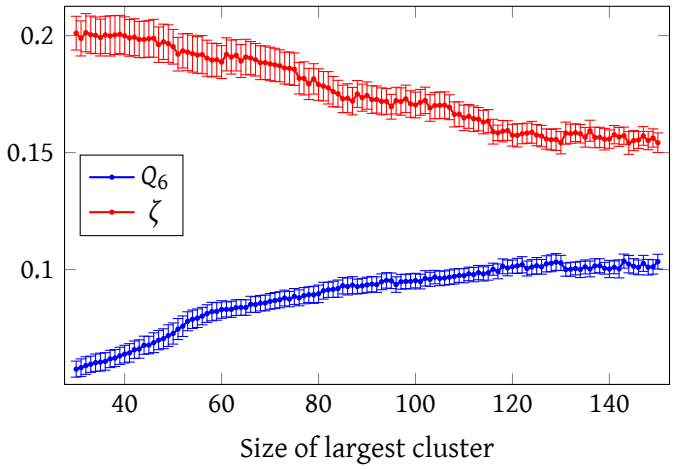


Figure 7.4: The global order parameters Q_6 and ζ calculated as a function of the size of the largest crystalline cluster, the order parameter used to drive nucleation in this work, for the system seeded with a cubic ice nucleus. Error bars show the standard deviation for the population of configurations at each cluster size. The results depicted here refer to the 576 particles nearest the centre of mass of the ice nucleus. $T = 200\,\mathrm{K}$, $p = 1\,\mathrm{bar}$.

nucleus, becomes more ordered prior to the nucleation event. Since the free energies associated with the nucleation pathway in Quigley and Rodger's study are considerably higher than the ones implied by our (not completely equilibrated) results, this adds further weight to our contention that global order parameters may locate pathways that are not fully consistent with the natural nucleation pathways. It must be emphasised that local order parameters do not necessarily result in more natural reaction co-ordinates; however, if a pathway can be found that has a lower free energy barrier associated with it, then such a pathway will be favoured over one that has a significantly higher free energy barrier.

7.1.3 Discussion

We have introduced an order parameter that is capable of tracking and driving the homogeneous nucleation of ice with the TIP4P/2005 water model. The order parameter is local in nature and thus does not exhibit the anomalous behaviour associated with global order parameters. We believe it to be the first rotationally invariant order parameter of this kind that is capable of driving homogeneous ice nucleation in simulations of an all-atom model of water. One of the major difficulties in the development of such an order parameter is that the time that is required to confirm whether an order parameter is fit for purpose is very significant, given that the dynamics at reasonable supercoolings are so slow. In particular, it is important that the umbrella sampling weights not be increased too quickly even if it appears that no growth is forthcoming: one must exercise a considerable degree of patience when performing such nucleation simulations.

The order parameters we have introduced represent the first step in obtaining free energy landscapes and nucleation rates for the homogeneous nucleation of ice from simulations. However, there remain considerable challenges ahead. The basic problem is that the dynamics of ice nucleation for all-atom models are excruciatingly slow,[117;321] making equilibration

very difficult. This problem is illustrated by figure 7.5, which shows that at a reasonable 20 % supercooling (200 K) in the TIP4P/2005 model, a crystalline cluster of approximately 220 molecules in a system of 2500 molecules neither shrinks nor grows. At this temperature, classical nucleation theory would predict that a 220-molecule crystalline cluster is post-critical and so we might expect to see it grow. However, using a cluster that classical nucleation theory would predict to be pre-critical in size does not alter the system's static behaviour. Although only the first 1 ns of this pure MD simulation is shown in figure 7.5, the results are no different when simulated up to 70 ns, taking approximately a month of CPU time for each trajectory. Furthermore, to allow us to equilibrate a system, we would require not one, but a large number of freezing and melting events to be able to be simulated in such a period of computer time.

One control parameter at our disposal is, of course, the temperature. The fastest rate of growth of an ice-water interface is generally some 10 K below the freezing point,[292;293;296;297] where the increased freezing driving force of cooler systems optimally balances their slower dynamics. The faster dynamics are illustrated by the reasonably rapid melting of the crystalline

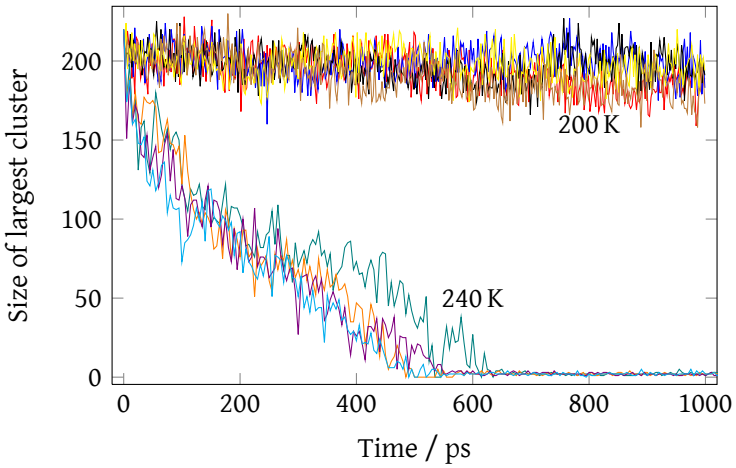


Figure 7.5: MD simulations of melting. The starting point is a crystalline cluster comprising approximately 220 molecules embedded in supercooled liquid water. The curves exhibiting melting were simulated at 240 K, whilst the remaining ones were simulated at 200 K. These simulations entailed 2500 TIP4P/2005 water molecules. Note that the melting point of TIP4P/2005 ice is 252 K.[252] $p = 1$ bar.

cluster at 240 K (∼5 % supercooling) as depicted in figure 7.5. One consideration that must be borne in mind when choosing a suitable temperature at which to perform simulations is that the majority of experimental rates have been reported at temperatures corresponding to supercoolings of between 27 % and 10 % (section 4.4); these also encompass the atmospherically relevant conditions. While higher temperature simulations may be easier to run in some ways due to the expedited dynamics, raising the temperature is not without its problems. For example, while we can negate the higher nucleation free energy barrier with umbrella sampling, it is not just the barrier height, but also the critical cluster size that increases with increasing temperature. Increasing the temperature would thus require us to simulate considerably larger systems than are computationally affordable in order to avoid spurious finite size effects: for illustration, at 200 K, the critical cluster for TIP4P/2005 water is predicted by classical nucleation theory to encompass 10^2 molecules, while at 240 K, this rises to 10^4 molecules.[7] We could attempt to extrapolate the free energy barrier to lower temperatures using histogram reweighting[81] based on the results of small cluster simulations at higher temperatures. However, the calculation of nucleation rates requires the simulation of critical clusters, and so must be performed at sufficiently low temperatures so that the critical cluster is small enough to be feasible to simulate.

How does one, then, successfully simulate the homogeneous nucleation process and obtain a free energy landscape and nucleation rate? The simplest strategy is simply to wait for a very long time: however, given how computationally challenging the process is, this may involve an inordinate amount of computer time. A second approach is to use more efficient simulation algorithms; indeed, as discussed above, the use of the hybrid Monte Carlo approach

[7] We use the enthalpy of melting, density and melting temperature results summarised by Vega and Abascal[252] for the TIP4P/2005 model and the interfacial free energy for the basal plane of TIP4P ice reported by Davidchack and co-workers[319] to estimate the critical cluster sizes. We note that using the TIP4P value for the interfacial free energy is only a first approximation and may not be very well suited to the TIP4P/2005 model of water. For example, the vapour-liquid surface tension is appreciably larger for TIP4P/2005 than for the TIP4P model;[252] determining whether the same applies to the ice-liquid interfacial free energy would be an interesting study in its own right.

gives a significant advantage over standard Monte Carlo simulations in terms of simulation speed. Other tricks of the trade that might be advantageous include the use of hamiltonian exchange[105] to couple the system of interest to one that is dynamically faster (as we attempt to do in section 7.2), or the use of reaction fields in place of the computationally expensive Ewald summation.[347]

Finally, the water potential we use is another parameter of the system that is under our control. It has been suggested that there are few differences in the dynamics of ice melting of most common all-atom water models,[348] and so a possible solution to the equilibration problem may be to use a model of water that is not necessarily the best at describing most experimental properties, but one whose dynamics are computationally faster; examples might include the TIP5P(-E)[263;264] model[8] and the Nada–Van der Eerden model.[262;265] Both the NVdE and the TIP5P-E models exhibit a more tetrahedral structure than TIP4P/2005 water, both in the liquid[349] and in the ice phases,[125] and this is likely to stabilise the ice phase relative to the liquid.[125] This additional tetrahedrality likely arises as a consequence of the presence of explicit 'lone pairs' in these models.[349] Even though such potentials may be computationally more demanding than TIP4P-analogues on a per-step basis, the dynamics of ice growth might nevertheless be faster.[157;265]

In summary, the order parameter we presented in chapter 5 is capable not only of tracking, but also of driving the homogeneous nucleation of ice with the TIP4P/2005 water model. The order parameter is local in nature and thus does not exhibit the anomalous behaviour associated with global order parameters. However, although we have demonstrated that this order parameter is capable of forcing the growth of ice using an all-atom model of water, we have also shown that the equilibration of TIP4P/2005 ice growth at a 20 % supercooling is an extremely

[8] Unfortunately, the density of TIP5P ice is too high, and the stable phase at atmospheric pressure is ice II (a rhombohedral proton-ordered relatively high-density phase that forms naturally at high pressures from ice I_h [137]) rather than ice I_h,[252] making it potentially less suitable in the study of the homogeneous nucleation of ice at this pressure.

slow process. The development of a seemingly rigorous order parameter may help us to advance our understanding of ice nucleation; in particular, we have further corroborated our hypothesis that the difficulty in simulating ice nucleation using all-atom models such as TIP4P/2005 is more a result of the slow dynamics of the process rather than of an overwhelmingly large free energy barrier. This work thus represents a stepping stone towards the successful determination of a free energy landscape and nucleation rate for homogeneous ice nucleation for all-atom water models. We look at some approaches to speeding up the dynamics in the remainder of this chapter.

7.2 Hamiltonian replica exchange

We have successfully simulated the nucleation of ice using the mW model (chapter 6); we might thus imagine that it could be possible to couple the two systems in order to harness the faster dynamics of mW model nucleation in a hamiltonian exchange scheme (section 3.6) in the hope that the additional computational power required would be outweighed by a significant increase in the sampling of different regions of phase space.

In our implementation of hamiltonian exchange, we run Monte Carlo simulations for each replica serially. If we use the replica-dependent potential energy

$$U_i(\boldsymbol{r}_i) = (1-\lambda_i)U^{\mathrm{TIP4P/2005}}(\boldsymbol{r}_i) + \lambda_i U^{\mathrm{mW}}(\boldsymbol{r}_i) + U_i^{\mathrm{US}}(\boldsymbol{r}_i), \qquad \text{where } \lambda_i \in \mathbb{R} \cap [0,\,1]\,, \tag{7.7}$$

such that $\lambda_0 = 0$ and $\lambda_M = 1$, where $M+1$ is the total number of replicas, then using the expression for $\mathcal{H}$ given in equation (3.55) gives

$$\begin{aligned}\mathcal{H}/\beta = (\lambda_j - \lambda_i)\left[U^{\mathrm{mW}}(\boldsymbol{r}_i) - U^{\mathrm{TIP4P/2005}}(\boldsymbol{r}_i) - U^{\mathrm{mW}}(\boldsymbol{r}_j) + U^{\mathrm{TIP4P/2005}}(\boldsymbol{r}_j)\right] \\ + U_i^{\mathrm{US}}(\boldsymbol{r}_j) + U_j^{\mathrm{US}}(\boldsymbol{r}_i) - U_i^{\mathrm{US}}(\boldsymbol{r}_i) - U_j^{\mathrm{US}}(\boldsymbol{r}_j).\end{aligned} \tag{7.8}$$

Although this may appear complicated, it is in fact not very expensive to calculate, as the TIP4P/2005 and mW potential energies are already known from the simulation (since both of

them are required to obtain the overall energy for each replica[9]) and, even though the umbrella sampling weights may appear to be complicated at first glance, the order parameter, which is computationally the most demanding part of the umbrella sampling aspect of the simulation, does not need to be recalculated, as it depends solely on the positions and is therefore known for each simulation. It is worth emphasising, however, that the umbrella sampling weights do not cancel out, as the different replicas will have different weights associated with particular choices of the order parameter.[10] Whenever a step is chosen to involve a swap move, we choose the replica numbers to exchange by randomly choosing replica i within the permissible range ($i \in [0, M] \cap \mathbb{Z}$), then randomly select j to be within some integer number x of i (in our simulations, $x = 1$), but such that j is still in the same range as i (in other words, for $i = 0$, we can only choose $j = 1$); this is accounted for in the transition matrix to ensure that detailed balance is obeyed.

One issue that arises when coupling the mW potential with an all-atom model such as TIP4P/2005 is whether the mW site should be mapped onto the TIP4P/2005 oxygen site or the TIP4P/2005 centre-of-mass site. It is not clear which would be a better representation; however, because of the large mass of the oxygen atom relative to the hydrogen atoms, the centre of mass of the water molecule is fairly close to the oxygen site, and this issue might therefore not be especially worrying. However, there is another, much more significant issue with coupling the TIP4P/2005 model with the mW model: in the mW representation, water molecules have no explicit hydrogens and hence no explicit orientations: water's tetrahedral structure arises from the three-body terms. It is not immediately clear if and how we might

[9] A practical question that arises here is whether we need to calculate the mW energy contribution even in the case where $\lambda = 0$, or the TIP4P/2005 contribution when $\lambda = 1$, where these do not contribute at all to the overall potential energy. This depends on how often we wish to attempt swap moves. If we do so relatively frequently, then it is beneficial to keep both energies calculated, as only a relatively small proportion of the overall calculations needs to be performed when we move a single particle, whereas a complete recalculation of one of the potentials would be required in a swap move otherwise. Updating the energy gradually may therefore be better than recalculating, depending on how many steps there are on average between swap attempts for that simulation.

[10] In practice, this involves a simple lookup for the correct umbrella sampling weight potential to use in each case.

be able to generate orientations for the mW potential such that a swap move would respect detailed balance (section 3.1). For example, imagine that we are in a two state extended ensemble with full TIP4P/2005 and full mW potentials. One option of generating moves from the mW state to the TIP4P/2005 state that are likely to be accepted is to find the optimum orientations for TIP4P/2005 when the move is proposed. However, if we accept a move from TIP4P/2005 to mW with a hydrogen bond system that is not optimal, and in the next step attempt a reverse move with an optimal hydrogen bond network, this clearly violates detailed balance. A completely random choice of orientations would also not obey detailed balance, as the TIP4P/2005→mW move would certainly not have a completely random orientation of hydrogen bonds; furthermore, such a scheme would result in the rejection of the majority of trial swap moves. However, whilst these issues can be a problem when moving between a potential with no explicit orientational dependence to one that does have it (that is, moving from $\lambda = 1$ to $\lambda < 1$), there is no need, in principle, to proceed all the way to the 'pure' mW state in a hamiltonian exchange scheme when trying to speed up a TIP4P/2005 simulation, provided that an intermediate state exhibits sufficiently fast dynamics.

In figure 7.6, the TIP4P/2005 and mW potential energies are plotted for a liquid state of 300 particles at 1 bar pressure and 200 K. The waters have random orientations at the mW end of the combined hamiltonian, and so the TIP4P/2005 energy corresponding to the pure mW hamiltonian is very high. By contrast, the mW energy does not change considerably when a pure TIP4P/2005 configuration is subjected to the pure mW hamiltonian, as the centres of mass correspond to relatively similar (but not identical) low energy states. It may perhaps seem strange why TIP4P/2005 has such a low potential energy compared to the mW model; the primary reason for this is just that there are more interactions in the larger cutoff radius used with the TIP4P/2005 model. In principle, because we are changing hamiltonians, it is not the difference in the absolute energies that is the most important (we calculate the energies in

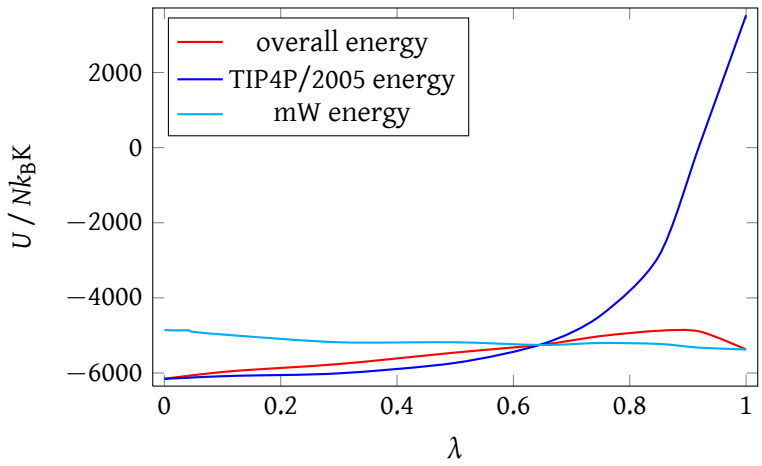


Figure 7.6: The averaged potential energies per particle of the TIP4P/2005 and mW potentials in a system of 300 liquid molecules supercooled to 200 K as functions of λ (with $\lambda = 0$ corresponding to a pure TIP4P/2005 hamiltonian and $\lambda = 1$ to a pure mW hamiltonian), each point having been equilibrated, within reason, using the relevant (mixed) potential. $p = 1$ bar.

both systems using both hamiltonians, so the majority of the difference is in any case cancelled out), but more so the difference in the energy of the equilibrium configurations, as a low mW energy configuration may well lead to a high TIP4P/2005 energy and vice versa. If we start with an equilibrated TIP4P/2005 supercooled liquid,[11] we therefore expect a relatively high acceptance rate for swap moves initially, but once the remaining replicas have had the chance to equilibrate to their own low-energy state consistent with the (mixed) hamiltonian, then the swap acceptance rate drops very considerably.

A schematic diagram of potential energy against λ is shown in figure 7.7. In a swap move between $\lambda = 0.2$ and $\lambda = 0.4$, the mW energy is increased and the TIP4P/2005 energy is decreased compared to the equilibrated state; however, at the same time, the relative importance of mW increases and that of TIP4P/2005 decreases, making this more disfavoured than might appear at first glance. A swap to lower values of λ would involve a similar process,

[11] The original formulation of the replica exchange scheme suggests that we 'replicate' one configuration in all simulations – hence the name of the method –, but of course if we wish to obey the Boltzmann distribution in the extended ensemble, we must equilibrate the systems before any sampling occurs, and so in practice, we start from saved configurations corresponding to (locally) equilibrated states for each of the replicas.

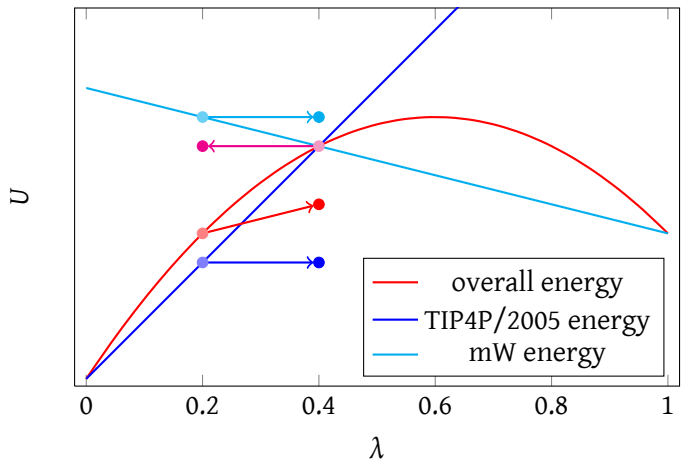


Figure 7.7: A schematic diagram showing the variation of potential energies of the TIP4P/2005 (blue) and mW (cyan) potentials as functions of λ (with $\lambda = 0$ corresponding to a pure TIP4P/2005 hamiltonian and $\lambda = 1$ to a pure mW hamiltonian). Arrows show what happens when equilibrium configurations are swapped between the two potentials. The energies corresponding to the mW and TIP4P/2005 potentials following a swap move of course do not change, but the weights associated with them do. The red arrow shows the new total energy for one of the replicas (*i.e.* the sum of the old TIP4P/2005 and mW energies, each appropriately weighted by the new value of λ) and the magenta arrow the total energy of the other replica involved in the swap move.

where the TIP4P/2005 energy is higher compared to the equilibrated state, and the TIP4P/2005 contribution to the total energy is also higher, and conversely for the mW contribution (although in the case illustrated, for simplicity, since the energies at $\lambda = 0.4$ happen to be the same in this hypothetical example, there is no overall change in energy for the second replica). The overall energy change for the swap is therefore positive and swap moves will be unlikely to be accepted. This is not a feature unique to this system: it is generally the case that if we equilibrate the system to one hamiltonian, the energy of the pure hamiltonian we are coupling with will be higher in this non-optimal configuration than it would be for its own equilibrated system. Therefore, this is a generic feature of hamiltonian exchange.

However, whilst we have performed several trial hamiltonian exchange simulations, it is unfortunately the case that an enormous number of replicas must be simulated in order

to get reasonable exchange acceptance rates.[12] Although we have parallelised the simulation code using OpenMP,[350;351] the resulting code was still so slow that not even the $\lambda = 0.5$ system showed any signs of equilibrating within a reasonable time frame. Part of the reason why the acceptance rate is likely to be rather small is that the use of mW and TIP4P/2005 models, notwithstanding the difference in temperature scales (we could easily simulate the mW potential at a higher temperature than the TIP4P/2005 equivalent, for example), results in somewhat different structures even of liquid water. These subtle but real differences in structure are hinted at by the correlation functions depicted in figure 7.8. These functions do not vary monotonically as λ is varied, which seems to imply an underlying structural difference in the liquids, and is likely to be the cause of the strange behaviour where the angular distribution function is roughly zero for $\theta \lessapprox \pi/6$ for both the mW and the TIP4P/2005 models, but not for certain mixed potentials. Moreover, the shape of the angular distribution function near $\theta = \pi/2$ is fairly different for the two potentials. Another interesting feature to notice is the kink in the radial distribution function of water simulated with the mW potential just short of 4 Å; indeed, even though the radial distribution functions for mW and TIP4P/2005 water have both been favourably compared to experiment,[33;38] the comparisons have been to *different* experimental data,[352;353] and the two structures are different enough not to be fully compatible.

Although using the mW potential in a hamiltonian exchange scheme did not prove to be fruitful, we also investigated using other potential hamiltonians in such a scheme. Instead of coupling to a different model of water altogether, it might be possible to modify the TIP4P/2005 model in a way that would encourage the growth of ice. To do this, it is instructive to investigate the rôle of the Lennard-Jones site in the model, as the large Lennard-Jones contribution, which is spherically symmetrical, is likely to be responsible for the metastability of the liquid water

[12] For example, at 200 K, we used 21 replicas between $\lambda = 0$ and $\lambda = 0.5$ with uneven λ gaps to optimise the acceptance rate in a system of 300 liquid molecules, and the best swap acceptance rates we recovered were of the order of 10 %; however, the replicas were not equilibrated very well within their total simulation time, and so the swap acceptance rates would be expected to decrease further.

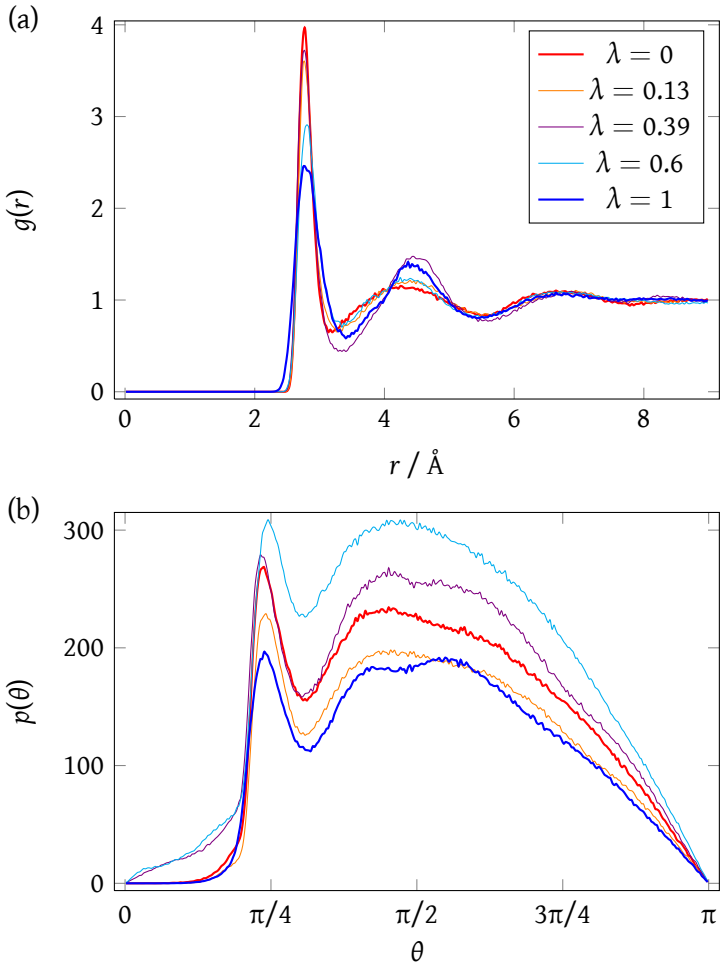


Figure 7.8: In (a), the centre-of-mass radial distribution function is shown for a liquid state at 200 K comprising 300 molecules for several values of λ. In (b), the unnormalised angular distribution function is shown for the same system, calculated for all pairs of neighbours within 4.3 Å of every molecule. Note that this means the number of particles included within the cutoff is different for each λ.

compared to the more ordered ice. In liquid TIP4P/2005 water, the first oxygen neighbour is located, on average, at roughly 2.8 Å from the oxygen in the water molecule under consideration, the precise value depending somewhat on the temperature. The Lennard-Jones parameter σ, at which the potential is zero, is $\sigma = 3.16$ Å, meaning that there is a relatively large repulsive region (and the first neighbouring oxygen is in fact within this repulsive region), and a large attractive region for $r > \sigma$. The minimum in the Lennard-Jones part of the potential is at $r = 2^{1/6}\sigma = 3.56$ Å.

The Lennard-Jones energy contribution is positive (*i.e.* it is on average repulsive) in water systems.[255] However, this contribution is less disfavourable for liquid water, even though liquid water is denser than ice and we might thus expect there to be more disfavouring at close distances. This is because there are also more neighbours in the attractive region, and these can compensate for the additional repulsive energy. Thus, the Lennard-Jones contribution disfavours the formation of ice, as shown by Lynden-Bell and Debenedetti,[354] who established that the reduction in the Lennard-Jones contribution in relation to the electrostatic contribution in the SPC/E model of water leads to the evolution of local order from spherically symmetric to tetrahedral, and the transition involves a regime of minimum order.[354]

In the hope that we might favour the ice phases, we therefore considered the behaviour of the system as we reduceed the Lennard-Jones attractive region and allowed electrostatic forces to be relatively stronger. To do this, we truncated and shifted the Lennard-Jones potential in the water model at some $r = r_c$ that was less than 8.5 Å, meaning that there were no attractions in some region that had previously been attractive. We also rescaled σ to retain the point at which the interaction becomes repulsive, and rescaled ε to retain the depth of the well. Unfortunately, while it may be that tetrahedral order is favoured by reducing the Lennard-Jones contribution to the potential, the dynamics of the ice nucleation process continue to be far too slow using such models, which implies that using a hamiltonian exchange scheme with a more tetrahedral version of the TIP4P/2005 model would not be especially advantageous.

In summary, whilst hamiltonian replica exchange may be beneficial to use in some circumstances, we have not been able to find a situation in which the extra simulation effort required would offset the gain in equilibration speed for the system of interest.

7.3 High temperature simulations

As discussed in subsection 7.1.3, in particular with reference to figure 7.5, simulations of ice and water at significant supercoolings with all-atom models of water are very slow indeed. One approach which could give us an estimate of low temperature free energies, while still being computationally tractable to simulate, would be to simulate the homogeneous nucleation process using the TIP4P/2005 model of water at temperatures at which the dynamics are reasonably fast and at which equilibration can therefore be achieved, and then use histogram reweighting (section 3.3) to estimate the free energy barrier to nucleation at the supercooling of interest.

We have run simulations analogous to those presented in subsection 7.1.2 using a hybrid Monte Carlo approach (section 3.8) with adaptive umbrella sampling (section 3.2). These simulations started with a seed hexagonal and a seed cubic ice crystal at 240 K and 1 bar with the TIP4P/2005 model of water, which represents an approximately 5 % supercooling. The starting umbrella weights corresponded to the negatives of the free energy barrier to nucleation estimated from classical nucleation theory, as discussed in section 7.1.3. A free energy profile for nucleation from such simulations is presented in figure 7.9.

To test whether equilibrium has been attained after an umbrella sampling iteration, we need to ensure that there is approximately equal sampling across order parameter values and that the order parameter values fluctuate quickly. We can also calculate the enthalpy as a function of the order parameter and ensure that it behaves sensibly (*i.e.* it does not suddenly jump by a large amount anywhere; see for example figure 7.11(b)). These equilibration criteria are fulfilled by the simulations whose results are presented in figure 7.9, and so we have obtained equilibrium results for ice nucleation at low supercoolings.

Although we use windowing with umbrella sampling, as described in section 3.2, in order to speed up the convergence, once we have configurations throughout the region of

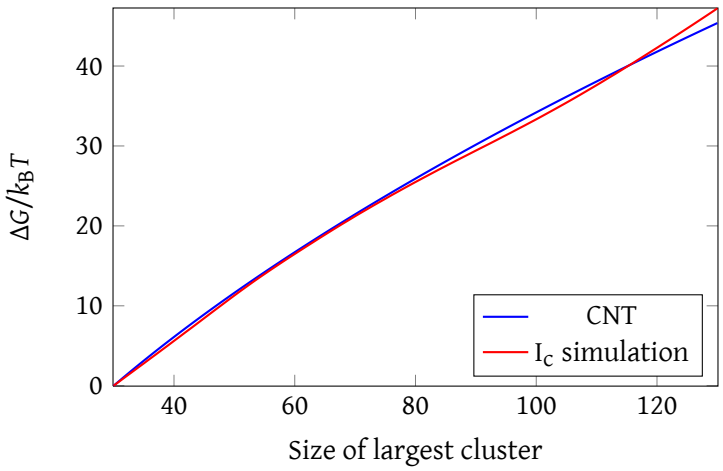


Figure 7.9: The free energy profile of TIP4P/2005 ice nucleation as a function of the size of the largest crystalline cluster in the system for simulations seeded with a cubic ice seed from hybrid Monte Carlo simulations relative to the free energy at a crystalline cluster size of 30. The hexagonal nucleation seed simulation results are analogous. The classical nucleation theory estimate is also shown. $T = 240\,\mathrm{K}$, $p = 1\,\mathrm{bar}$, 2500 molecules.

interest, we no longer window the umbrella sampling simulations. The reason for this is that the order parameter we use can fluctuate to a fairly large extent (see *e.g.* figure 7.5), and so the individually-calculated values are not necessarily truly representative of the 'mean' cluster size associated with that configuration at that temperature. This means that window edges are prone to errors (indeed, this is what appears to be happening at the extreme values of the crystalline cluster size in figure 7.9[13]). Whilst windowing is very important in order to speed up convergence,[76] if we start from an equilibrium, or a near equilibrium, set of configurations, not using windowing in such simulations is not problematic.

The simulation results shown in figure 7.9 appear to agree very well with the classical nucleation theory prediction; indeed, the agreement is almost suspiciously good. If we fit the data for clusters comprising 40 to 110 molecules to a CNT-like expression for the Gibbs

[13] The equivalent cubic seed simulations that were run in a differently-sized overall window (30–110 as opposed to 30–130), for example, show a deviation from CNT similar to the one observed here at large cluster sizes, but this deviation occurs at correspondingly smaller cluster sizes. This suggests that the deviation is likely to be a windowing effect rather than a true feature of the free energy landscape.

energy as a function of the size of the largest crystalline cluster,[14] as we did when studying mW nucleation in subsection 6.2.1, we find that $\Delta_{\mathrm{fus}}\mu(240\,\mathrm{K})/k_{\mathrm{B}} \approx 29.7\,\mathrm{K}$ and $\gamma \approx 23.5\,\mathrm{mJ\,m^{-2}}$; these values compare favourably to the $\Delta_{\mathrm{fus}}\mu(240\,\mathrm{K})/k_{\mathrm{B}} \approx 27.7\,\mathrm{K}$ estimated using equation (7.6) and $\gamma \approx 24.5\,\mathrm{mJ\,m^{-2}}$ calculated for the basal plane of TIP4P ice.[319] Given that we started the simulations with umbrella weights corresponding to classical nucleation theory once windowing had been removed, we could envisage a situation where the clusters essentially remain at their original size because of the slow dynamics of the ice cluster growth/shrinkage process. While this is certainly always a conceivable issue in simulations of water, it does not appear to play a significant rôle in these hybrid Monte Carlo simulations at 240 K. We attempted equivalent simulations with starting weights corresponding to classical nucleation theory weights at 235 K and at 245 K (with the simulation temperature remaining at 240 K); clusters were observed to shrink over time in the former and to grow over time in the latter set of simulations. This suggests that the classical nucleation theory estimate at the temperature of simulation does in fact yield a reasonable approximation to the free energy barrier associated with the simulated process. However, the critical cluster size at this temperature comprises of the order of 1.5×10^4 molecules (subsection 7.1.3), and so good agreement in this very early stage of the nucleation process at this temperature is no guarantee that the critical cluster size or the height of the free energy barrier to nucleation are also well estimated by CNT.

Although these results suggest that using classical nucleation theory may give a (surprisingly) good estimate for the free energy profile at 240 K, we cannot use this information in practice to obtain nucleation rates, as we cannot simulate the critical cluster size. In order to obtain free energies at lower temperatures, we can consider using histogram reweighting, as in the mW case detailed in subsection 6.2.2.

[14] We fitted to the data for clusters comprising 40 to 110 molecules only in order to avoid windowing effects. We shifted the data such that the Gibbs energy at a cluster size of 40 matched that predicted by classical nucleation theory.

High temperature simulations have considerably faster dynamics than low temperature simulations. To some extent, this is a consequence of the somewhat different structure of the liquid phase at various temperatures, which can speed up the orientational dynamics in the less tetrahedral high temperature simulations.[355] Although the free energy barrier to molecular reorientation can vary by up to 4 kcal mol^{-1} depending on the microscopic structure of the liquid,[355] it is unlikely to exceed the nucleation free energy barrier: however, the variation of the structure with temperature may imply that the real low temperature configurations are captured less well in high temperature simulations; any histogram reweighting is just an approximation to the actual behaviour at low temperatures.

In order to use histogram reweighting, we must be able to sample the configurations that are important at low temperature in our high-temperature simulations. One way of estimating how well we are sampling is to consider the potential enthalpy distribution as a function of the order parameter. To do this, we need to collect potential enthalpy-order parameter pair counts (section 3.3). We show a heat map of such a histogram in figure 7.10(a); we notice that the potential enthalpy is roughly uniformly sampled, except that there is an unusual region of high potential enthalpy for clusters comprising between about 50 and about 85 molecules which is visited a small but not insignificant number of times. If slices are taken at given values of the largest crystalline nucleus size, we see that the enthalpies sampled follow roughly a gaussian distribution (figure 7.10(b)), except for the anomalous region of high potential enthalpy, where there is a distinct non-gaussian shoulder.

Applying the histogram reweighting method to the free energy barrier to nucleation depicted in figure 7.9, we obtain results as shown in figure 7.11(a). The curves are rather noisy, but there are several features worth noticing. First, going down in temperature yields a reasonable-looking curve at 235 K, but the curve at 230 K is considerably more noisy. The curve at 243 K is very unusual indeed; however, we can notice that the region where the curve behaves

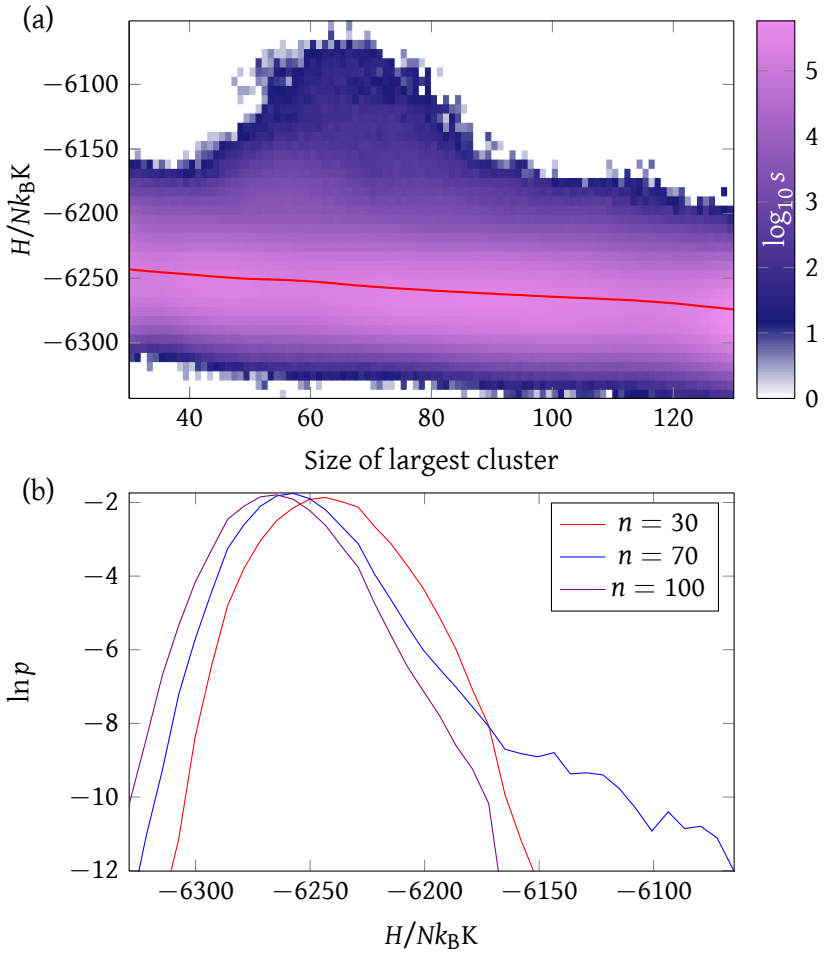


Figure 7.10: (a) Potential enthalpy-order parameter histogram for a simulation of ice nucleation from a cubic seed. s is the number of occurrences of a potential enthalpy-order parameter pair. $T = 240\,\mathrm{K}$, $p = 1\,\mathrm{bar}$, 2500 molecules. The mean value of the enthalpy as a function of the cluster size is shown by the red curve. In (b), slices are taken from the heat map in (a) and the normalised probability p is plotted as a function of the enthalpy for three values of the largest cluster size.

unexpectedly corresponds to the region in order parameter space where a reasonably large set of high potential enthalpy configurations were sampled (figure 7.10(a)). Whilst the non-gaussian tail in the enthalpy probability (figure 7.10(b)) at first glance appears to be favourable when estimating properties at higher temperatures, where the 'natural' enthalpy is in fact higher, it is important that the enthalpies are not just sampled, but are sampled with the correct weight for histogram reweighting to work correctly; the unexpectedly large probability in the non-gaussian shoulder in the enthalpy distribution means that on reweighting, the entropy

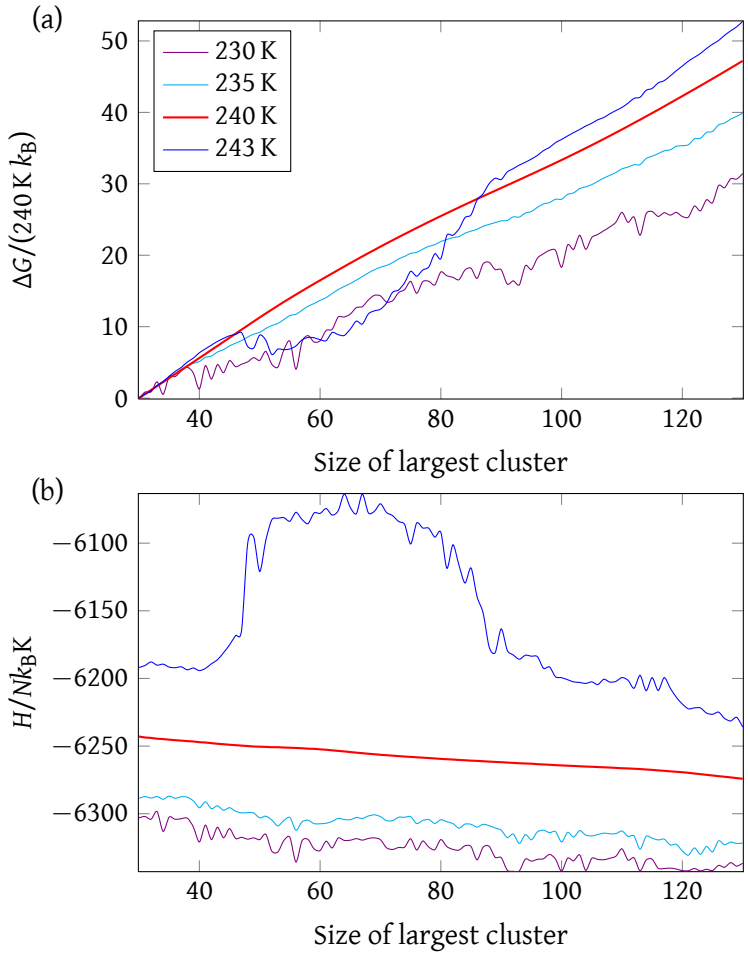


Figure 7.11: Histogram-reweighted (a) free energy barriers to nucleation and (b) mean potential enthalpy for ice nucleation seeded with cubic crystal clusters as functions of the size of the largest crystalline cluster. The temperature of simulation was 240 K. $p = 1$ bar, 2500 molecules.

of these states is overestimated, leading to a decrease in ΔG in this region as the temperature increases. Nevertheless, as we are not especially interested in high temperature estimates, and there are no non-gaussian kinks in the distributions at the lower end of the enthalpy scale, this result is not too concerning.

It is also instructive to apply histogram reweighting to the enthalpy itself; this is shown in figure 7.11(b). We can see that the curves are mainly at the very edges of the well-visited region in figure 7.10(a): the enthalpies are simply not especially well representative of the configurations at temperatures other than the temperature of the simulation. Indeed, the same

holds for the density. For example, typical liquid density fluctuations are up to $\pm 0.01\,\mathrm{g\,mL^{-1}}$ at coexistence. This is comparable to the liquid phase density change on cooling the supercooled liquid from 240 K to 230 K, where the equation of state given above (equation (7.4)) predicts a density change of $0.012\,\mathrm{g\,mL^{-1}}$. Consequently, it may be very difficult to get reliable histogram reweighted estimates of the free energy more than a few degrees away from the simulation temperature.

One way of attempting to circumvent this difficulty is to apply a bias to the enthalpy of the simulated system so that high-temperature simulations sample configurations with enthalpies more representative of low temperatures. We have done this by applying an additional umbrella potential that depends on the enthalpy as a function of the size of the largest crystalline cluster. Whilst progress has been slow, such simulations have allowed us to obtain smoother estimates of the enthalpy and the free energy as a function of cluster size at temperatures down to 230 K; nevertheless, the enthalpy will need to be decreased by a very significant further amount in order to obtain equilibrium estimates at temperatures corresponding to critical cluster sizes that we can then simulate directly in a simulation.

Although the order parameter we presented in chapter 5 appears to be a reasonable order parameter to use when driving the process of homogeneous ice nucleation, the slow dynamics of ice growth at low temperatures continue to pose a very significant obstacle to the determination of free energy landscapes and nucleation rates. The use of advanced simulation methods does offer new insights into the process; nevertheless, the successful calculation of nucleation rates continues to be a challenge.

8 Summary and conclusion

The homogeneous nucleation of ice from supercooled liquid water is an important process, but one which is difficult to study both in experiment and computationally. Brute-force simulations are unlikely to be able to be used successfully in an investigation of the process; this is because homogeneous nucleation is a rare event, and the rate of nucleation is sufficiently small so that a typically-sized simulation box simulated for an accessible period of time will not result in any nucleation events.

In order to study the process, rare event techniques must be employed. In our Monte Carlo simulations, we used predominantly umbrella sampling, which allows us to bias the system of interest into visiting phase space that it would not normally visit, and to extract thermodynamic data of interest from such biassed non-Boltzmann simulations. However, to be able to apply umbrella sampling to a rare event, we require a means of quantifying the extent to which the process of interest has occurred: such a quantity is known as an order parameter. In this work, we developed a local order parameter which enables us to distinguish liquid water from ice, and which is capable not only of tracking, but also of driving homogeneous ice nucleation using both the monatomic and all-atom models of water. Importantly, the use of this order parameter does not result in the growth of 'chains' that can otherwise arrest the growth of ice clusters.

We have demonstrated that, when water is simulated with a simple monatomic potential, whose dynamics are fast, we can successfully extract a free energy profile and calculate nucleation rates. The free energy we obtained was reasonably consistent with classical

nucleation theory, which is at odds with several earlier studies of homogeneous ice nucleation. We rationalised this disparity by considering the differences in the order parameters used. In contrast to the previous studies, the order parameter we developed is local and rotationally invariant, and we have shown that its use in driving homogeneous ice nucleation is likely to result in free energy pathways considerably closer to the natural nucleation pathways than those reported previously. In particular, using global order parameters appears to result in a nucleation pathway that is not the lowest free energy pathway, but simply one that, once its higher free energy has been compensated for by means of a rare event simulation method, is a dynamically accessible pathway.

We have also shown that classical nucleation theory is a reasonable estimate for the free energy barrier to ice nucleation in simulations of TIP4P/2005 water. Simulations using umbrella sampling weights just slightly larger than the negatives of the free energy predicted by classical nucleation theory allow us to grow small ice clusters to considerably larger sizes. Such clusters behave analogously to those studied in the monatomic water model simulations, even when they are not technically at equilibrium. Our equilibrated simulations of ice nucleation at small supercoolings show remarkable agreement between the simulation estimate of the free energy barrier to nucleation and that predicted by classical nucleation theory at the same temperature.

We have shown that the principal difficulty in simulating the process of homogeneous ice nucleation using all-atom models of water is the slowness of ice growth dynamics, irrespective of the free energy barrier to nucleation itself. An important finding is that the use of hybrid Monte Carlo, which uses short molecular dynamics simulations in place of single-particle Monte Carlo moves, results in a marked decrease in the time required to simulate the growth of ice, which further corroborates the suggestion in the literature that collective motion is an important feature of the process. Simulation methods such as hamiltonian exchange and histogram reweighting can be used to attempt to facilitate the equilibration process; however,

we have shown that regardless of any special methods used, the process of ice growth from the liquid is particularly slow, and more work is needed to determine a free energy profile and nucleation rates at temperatures at which it is computationally tractable to simulate the critical cluster size.

Our work opens up several interesting questions which it may be useful to address in future work. For example, while classical nucleation theory estimates are very reasonable in predicting suitable free energy barriers to nucleation for TIP4P/2005 model simulations, our estimates used a TIP4P model value for the ice-liquid interfacial free energy. It would be interesting to know how this value compares to its TIP4P/2005 analogue, especially if calculated as a function of the supercooling temperature.

A future avenue of research could use the approach we have outlined in this work to study the homogeneous nucleation of ice from various atmospherically relevant dilute solutions. Charged particles have been shown to influence the vapour-liquid nucleation process significantly,[356] and it may be of particular interest to atmospheric scientists to study the liquid-solid analogue. A simulation approach similar to ours could also be used to study heterogeneous nucleation, such as nucleation in the presence of hard flat surfaces, shaped surfaces (such as square or conical troughs), or surfaces patterned with charge, or even to simulate bacterium-induced nucleation. In the light of the conflicting experimental evidence discussed previously, it may also be interesting to simulate more carefully the effects of electric fields on the nucleation rate.

In conclusion, we have developed a local order parameter that is capable of driving ice nucleation, and we have demonstrated its applicability to simulating homogeneous ice nucleation. The study of homogeneous ice nucleation is a challenging problem, and much remains to be discovered. We hope, however, that our results and the approach we have presented here will continue to be useful in simulating homogeneous ice nucleation, as well

as find use in similar applications. We trust that we have shed some light on the process of homogeneous ice nucleation in this work.

References

[1] A Reinhardt, A J Williamson, J P K Doye, J Carrete, L M Varela and A A Louis, 'Re-entrant phase behavior for systems with competition between phase separation and self-assembly', *J Chem Phys*, **134**, 104905, 2011.

[2] A Reinhardt and J P K Doye, 'Free energy landscapes for homogeneous nucleation of ice for a monatomic water model', *J Chem Phys*, **136**, 054501, 2012.

[3] A Reinhardt, J P K Doye, E G Noya and C Vega, 'Local order parameters for use in driving homogeneous ice nucleation with all-atom models of water', *J Chem Phys*, **137**, 194504, 2012.

[4] A Reinhardt, F Romano and J P K Doye, 'Computing phase diagrams for a quasicrystal-forming patchy-particle system', *Phys Rev Lett*, accepted for publication, 2013, arXiv:1302.2592.

[5] C A Jeffery and P H Austin, 'Homogeneous nucleation of supercooled water: Results from a new equation of state', *J Geophys Res-Atmos*, **102**, 25269–25279, 1997.

[6] H R Pruppacher and J D Klett, *Microphysics of clouds and precipitation*. Springer, Dordrecht, 2nd edition, 2010.

[7] D W Oxtoby, 'Nucleation of first-order phase transitions', *Acc Chem Res*, **31**, 91–97, 1998.

[8] P R ten Wolde and D Frenkel, 'Homogeneous nucleation and the Ostwald step rule', *Phys Chem Chem Phys*, **1**, 2191–2196, 1999.

[9] A Laaksonen, V Talanquer and D W Oxtoby, 'Nucleation: Measurements, theory, and atmospheric applications', *Annu Rev Phys Chem*, **46**, 489–524, 1995.

[10] S Auer and D Frenkel, 'Numerical simulation of crystal nucleation in colloids', *Adv Polym Sci*, **173**, 149–208, 2005.

[11] J Anwar and D Zahn, 'Uncovering molecular processes in crystal nucleation and growth by using molecular simulation', *Angew Chem, Int Ed*, **50**, 1996–2013, 2011.

[12] D W Oxtoby, 'Homogeneous nucleation: Theory and experiment', *J Phys: Condens Mat*, **4**, 7627–7650, 1992.

[13] M B Baker, 'Cloud microphysics and climate', *Science*, **276**, 1072–1078, 1997.

[14] S Benz, K Megahed, O Möhler, H Saathoff, R Wagner and U Schurath, '*T*-dependent rate measurements of homogeneous ice nucleation in cloud droplets using a large atmospheric simulation chamber', *J Photoch Photobio A*, **176**, 208–217, 2005.

[15] D A Hegg and M B Baker, 'Nucleation in the atmosphere', *Rep Prog Phys*, **72**, 056801, 2009.

[16] P Spichtinger and D J Cziczo, 'Impact of heterogeneous ice nuclei on homogeneous freezing events in cirrus clouds', *J Geophys Res*, **115**, D14208, 2010.

[17] B J Murray, D O'Sullivan, J D Atkinson and M E Webb, 'Ice nucleation by particles immersed in supercooled cloud droplets', *Chem Soc Rev*, **41**, 6519–6554, 2012.

[18] V I Khvorostyanov and J A Curry, 'Parameterization of homogeneous ice nucleation for cloud and climate models based on classical nucleation theory', *Atmos Chem Phys*, **12**, 9275–9302, 2012.

[19] B J Murray, S L Broadley and G J Morris, 'Supercooling of water droplets in jet aviation fuel', *Fuel*, **90**, 433–435, 2011.

[20] K Sassen and S Benson, 'Ice nucleation in cirrus clouds: A model study of the homogeneous and heterogeneous modes', *Geophys Res Lett*, **27**, 521–524, 2000.

[21] M Toner, E G Cravalho and M Karel, 'Thermodynamics and kinetics of intracellular ice formation during freezing of biological cells', *J Appl Phys*, **67**, 1582–1593, 1990.

[22] J Karlsson, E Cravalho, I B Rinkes, R Tompkins, M Yarmush and M Toner, 'Nucleation and growth of ice crystals inside cultured hepatocytes during freezing in the presence of dimethyl sulfoxide', *Biophys J*, **65**, 2524–2536, 1993.

[23] G John Morris, E Acton, B J Murray and F Fonseca, 'Freezing injury: The special case of the sperm cell', *Cryobiology*, **64**, 71–80, 2012.

[24] K E Zachariassen and E Kristiansen, 'Ice nucleation and antinucleation in nature', *Cryobiology*, **41**, 257–279, 2000.

[25] K E Zachariassen, E Kristiansen, S A Pedersen and H T Hammel, 'Ice nucleation in solutions and freeze-avoiding insects—homogeneous or heterogeneous?', *Cryobiology*, **48**, 309–321, 2004.

[26] D E Hagen, R J Anderson and J L Kassner, 'Homogeneous condensation—freezing nucleation rate measurements for small water droplets in an expansion cloud chamber', *J Atmos Sci*, **38**, 1236–1243, 1981.

[27] H R Pruppacher, 'A new look at homogeneous ice nucleation in supercooled water drops', *J Atmos Sci*, **52**, 1924–1933, 1995.

[28] T Koop, B Luo, A Tsias and T Peter, 'Water activity as the determinant for homogeneous ice nucleation in aqueous solutions', *Nature*, **406**, 611–614, 2000.

[29] P G Debenedetti, 'Supercooled and glassy water', *J Phys: Condens Mat*, **15**, R1669–R1726, 2003.

[30] T Bartels-Rausch, 'Chemistry: Ten things we need to know about ice and snow', *Nature*, **494**, 27–29, 2013.

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

[32] M Mezei, 'Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias', *J Comput Phys*, **68**, 237–248, 1987.

[33] V Molinero and E B Moore, 'Water modeled as an intermediate element between carbon and silicon', *J Phys Chem B*, **113**, 4008–4016, 2009.

[34] P J Steinhardt, D R Nelson and M Ronchetti, 'Bond-orientational order in liquids and glasses', *Phys Rev B*, **28**, 784–805, 1983.

[35] P-R ten Wolde, M J Ruiz-Montero and D Frenkel, 'Simulation of homogeneous crystal nucleation close to coexistence', *Faraday Discuss*, **104**, 93–110, 1996.

[36] E B Moore, E de la Llave, K Welke, D A Scherlis and V Molinero, 'Freezing, melting and structure of ice in a hydrophilic nanopore', *Phys Chem Chem Phys*, **12**, 4124–4134, 2010.

[37] F Romano, E Sanz and F Sciortino, 'Crystallization of tetrahedral patchy particles *in silico*', *J Chem Phys*, **134**, 174502, 2011.

[38] J L F Abascal and C Vega, 'A general purpose model for the condensed phases of water: TIP4P/2005', *J Chem Phys*, **123**, 234505, 2005.

[39] S Duane, A D Kennedy, B J Pendleton and D Roweth, 'Hybrid Monte Carlo', *Phys Lett B*, **195**, 216–222, 1987.

[40] P W Atkins and J de Paula, *Atkins' Physical Chemistry*. Oxford University Press, Oxford, 7th edition, 2001.

[41] P G Debenedetti, *Metastable liquids: Concepts and principles*. Princeton University Press, Princeton, 1996.

[42] M Volmer and A Weber, 'Keimbildung in übersättigten Gebilden', *Z Phys Chem*, **119**, 277–301, 1926.

[43] R Becker and W Döring, 'Kinetische Behandlung der Keimbildung in übersättigten Dämpfen', *Ann Phys Berlin*, **416**, 719–752, 1935.

[44] J B Zeldovich, 'On the theory of new phase formation; cavitation', *Acta Physicochim URSS*, **18**, 1–22, 1943.

[45] D Turnbull and J C Fisher, 'Rate of nucleation in condensed systems', *J Chem Phys*, **17**, 71–73, 1949.

[46] J Frenkel, *Kinetic theory of liquids*. Dover, New York, 1955.

[47] S Auer and D Frenkel, 'Numerical prediction of absolute crystallization rates in hard-sphere colloids', *J Chem Phys*, **120**, 3015–3029, 2004.

[48] K F Kelton, 'Crystal nucleation in liquids and glasses', *Solid State Phys*, **45**, 75–177, 1991.

[49] E Ruckenstein and Y S Djikaev, 'Recent developments in the kinetic theory of nucleation', *Adv Colloid Interfac*, **118**, 51–72, 2005.

[50] C Valeriani, *Numerical studies of nucleation pathways of ordered and disordered phases*. Thesis, Universiteit van Amsterdam / AMOLF, Amsterdam, 2007.

[51] J Anwar and D Zahn, 'Atomistisches Verständnis der Keimbildung und des Kristallwachstums durch molekulare Simulationen', *Angew Chem, Ger Edit*, **123**, 2042–2061, 2011.

[52] S Auer and D Frenkel, 'Prediction of absolute crystal-nucleation rate in hard-sphere colloids', *Nature*, **409**, 1020–1023, 2001.

[53] W Ostwald, 'Studien über die Bildung und Umwandlung fester Körper', *Z Phys Chem*, **22**, 289–330, 1897.

[54] I N Stranski and D Totomanow, 'Keimbildungsgeschwindigkeit und Ostwaldsche Stufenregel', *Z Phys Chem*, **163**, 399–408, 1933.

[55] J Nývlt, 'The Ostwald rule of stages', *Cryst Res Technol*, **30**, 443–449, 1995.

[56] C-H Hung, M J Krasnopoler and J L Katz, 'Condensation of a supersaturated vapor. VIII. The homogeneous nucleation of n-nonane', *J Chem Phys*, **90**, 1856–1865, 1989.

[57] R C Tolman, 'The effect of droplet size on surface tension', *J Chem Phys*, **17**, 333–337, 1949.

[58] W G Courtney, 'Remarks on homogeneous nucleation', *J Chem Phys*, **35**, 2249–2250, 1961.

[59] J Lothe and G M Pound, 'Reconsiderations of nucleation theory', *J Chem Phys*, **36**, 2080–2085, 1962.

[60] V Talanquer and D W Oxtoby, 'Dynamical density functional theory of gas–liquid nucleation', *J Chem Phys*, **100**, 5190–5200, 1994.

[61] R P Sear, 'The non-classical nucleation of crystals: Microscopic mechanisms and applications to molecular crystals, ice and calcium carbonate', *Int Mater Rev*, **57**, 328–356, 2012.

[62] M Miller, *Rotation and nucleation of atomic clusters*. Thesis, University of Cambridge, Cambridge, 1996.

[63] Y Viisanen, R Strey and H Reiss, 'Homogeneous nucleation rates for water', *J Chem Phys*, **99**, 4680–4692, 1993.

[64] I J Ford, 'Statistical mechanics of nucleation: A review', *J Mech Eng Sci*, **218**, 883–899, 2004.

[65] N Metropolis, A W Rosenbluth, M N Rosenbluth, A H Teller and E Teller, 'Equation of state calculations by fast computing machines', *J Chem Phys*, **21**, 1087–1092, 1953.

[66] D Frenkel and B Smit, *Understanding molecular simulation: From algorithms to applications*. Elsevier Academic Press, San Diego, London, 2nd edition, 2002.

[67] H E A Huitema and J P van der Eerden, 'Can Monte Carlo simulation describe dynamics? A test on Lennard-Jones systems', *J Chem Phys*, **110**, 3267–3274, 1999.

[68] D A McQuarrie, *Statistical mechanics*. University Science Books, Sausalito, 2000.

[69] V I Manousiouthakis and M W Deem, 'Strict detailed balance is unnecessary in Monte Carlo simulation', *J Chem Phys*, **110**, 2753–2756, 1999.

[70] G Marsaglia, A Zaman and W W Tsang, 'Toward a universal random number generator', *Stat Probabil Lett*, **9**, 35–39, 1990.

[71] J Wedekind, D Reguera and R Strey, 'Finite-size effects in simulations of nucleation', *J Chem Phys*, **125**, 214505, 2006.

[72] R Eppenga and D Frenkel, 'Monte Carlo study of the isotropic and nematic phases of infinitely thin hard platelets', *Mol Phys*, **52**, 1303–1334, 1984.

[73] G J M Koper and H Reiss, 'Length scale for the constant pressure ensemble: Application to small systems and relation to Einstein fluctuation theory', *J Phys Chem*, **100**, 422–432, 1996.

[74] D S Corti, 'Monte Carlo simulations in the isothermal-isobaric ensemble: The requirement of a 'shell' molecule and simulations of small systems', *Mol Phys*, **100**, 1887–1904, 2002.

[75] N G Almarza, 'A cluster algorithm for Monte Carlo simulation at constant pressure', *J Chem Phys*, **130**, 184106, 2009.

[76] D Chandler, *Introduction to modern statistical mechanics*. Oxford University Press, New York, 1987.

[77] J S van Duijneveldt and D Frenkel, 'Computer simulation study of free energy barriers in crystal nucleation', *J Chem Phys*, **96**, 4655–4668, 1992.

[78] S Kumar, J M Rosenberg, D Bouzida, R H Swendsen and P A Kollman, 'The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method', *J Comput Chem*, **13**, 1011–1021, 1992.

[79] S Kumar, J M Rosenberg, D Bouzida, R H Swendsen and P A Kollman, 'Multidimensional free-energy calculations using the weighted histogram analysis method', *J Comput Chem*, **16**, 1339–1350, 1995.

[80] J D Chodera, W C Swope, J W Pitera, C Seok and K A Dill, 'Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations', *J Chem Theory Comput*, **3**, 26–41, 2007.

[81] A M Ferrenberg and R H Swendsen, 'New Monte Carlo technique for studying phase transitions', *Phys Rev Lett*, **61**, 2635–2638, 1988.

[82] P B Conrad and J J de Pablo, 'Comparison of histogram reweighting techniques for a flexible water model', *Fluid Phase Equilibr*, **150-151**, 51–61, 1998.

[83] M Plischke and B Bergersen, *Equilibrium statistical physics*. World Scientific, London, 3rd edition, 2006.

[84] A M Ferrenberg, D P Landau and R H Swendsen, 'Statistical errors in histogram reweighting', *Phys Rev E*, **51**, 5092–5100, 1995.

[85] D Chandler, 'Statistical mechanics of isomerization dynamics in liquids and the transition state approximation', *J Chem Phys*, **68**, 2959–2970, 1978.

[86] S Auer and D Frenkel, 'Crystallization of weakly charged colloidal spheres: A numerical study', *J Phys: Condens Mat*, **14**, 7667–7680, 2002.

[87] S A Auer, *Quantitative prediction of crystal nucleation rates for spherical colloids: A computational study*. Thesis, Universiteit van Amsterdam / AMOLF, Amsterdam, 2002.

[88] L Filion, M Hermes, R Ni and M Dijkstra, 'Crystal nucleation of hard spheres using molecular dynamics, umbrella sampling, and forward flux sampling: A comparison of simulation techniques', *J Chem Phys*, **133**, 244115, 2010.

[89] M J Ruiz-Montero, D Frenkel and J J Brey, 'Efficient schemes to compute diffusive barrier crossing rates', *Mol Phys*, **90**, 925–942, 1997.

[90] D Quigley and P M Rodger, 'Metadynamics simulations of ice nucleation and growth', *J Chem Phys*, **128**, 154518, 2008.

[91] D Quigley and P M Rodger, 'A metadynamics-based approach to sampling crystallisation events', *Mol Simulat*, **35**, 613–623, 2009.

[92] A Laio and F L Gervasio, 'Metadynamics: A method to simulate rare events and reconstruct the free en-

ergy in biophysics, chemistry and material science', *Rep Prog Phys*, **71**, 126601, 2008.

[93] A Reinhardt, *Computer simulation of homogeneous ice nucleation*. Master's thesis, University of Oxford, 2009.

[94] R J Allen, D Frenkel and P R ten Wolde, 'Forward flux sampling-type schemes for simulating rare events: Efficiency analysis', *J Chem Phys*, **124**, 194111, 2006.

[95] R J Allen, D Frenkel and P R ten Wolde, 'Simulating rare events in equilibrium or nonequilibrium stochastic systems', *J Chem Phys*, **124**, 024102, 2006.

[96] C Valeriani, R J Allen, M J Morelli, D Frenkel and P R ten Wolde, 'Computing stationary distributions in equilibrium and nonequilibrium systems with forward flux sampling', *J Chem Phys*, **127**, 114109, 2007.

[97] R J Allen, P B Warren and P R ten Wolde, 'Sampling rare switching events in biochemical networks', *Phys Rev Lett*, **94**, 018104, 2005.

[98] T Li, D Donadio, G Russo and G Galli, 'Homogeneous ice nucleation from supercooled water', *Phys Chem Chem Phys*, **13**, 19807–19813, 2011.

[99] F Wang and D P Landau, 'Efficient, multiple-range random walk algorithm to calculate the density of states', *Phys Rev Lett*, **86**, 2050–2053, 2001.

[100] T S van Erp and P G Bolhuis, 'Elaborating transition interface sampling methods', *J Comput Phys*, **205**, 157–181, 2005.

[101] A Bunker and B Dünweg, 'Parallel excluded volume tempering for polymer melts', *Phys Rev E*, **63**, 016701, 2000.

[102] Y Sugita, A Kitao and Y Okamoto, 'Multidimensional replica-exchange method for free-energy calculations', *J Chem Phys*, **113**, 6042–6051, 2000.

[103] A Mitsutake, Y Sugita and Y Okamoto, 'Generalized-ensemble algorithms for molecular simulations of biopolymers', *Biopolymers (Pept Sci)*, **60**, 96–123, 2001.

[104] Y Iba, 'Extended ensemble Monte Carlo', *Int J Mod Phys C*, **12**, 623–656, 2001, arXiv:cond-mat/0012323v2.

[105] H Fukunishi, O Watanabe and S Takada, 'On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction', *J Chem Phys*, **116**, 9058–9067, 2002.

[106] A Mitsutake, Y Sugita and Y Okamoto, 'Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test', *J Chem Phys*, **118**, 6664–6675, 2003.

[107] D J Earl and M W Deem, 'Parallel tempering: Theory, applications, and new perspectives', *Phys Chem Chem Phys*, **7**, 3910–3916, 2005.

[108] J Vreede, M G Wolf, S W de Leeuw and P G Bolhuis, 'Reordering hydrogen bonds using Hamiltonian replica exchange enhances sampling of conformational changes in biomolecular systems', *J Phys Chem B*, **113**, 6484–6494, 2009.

[109] A W Wilber, J P K Doye and A A Louis, 'Self-assembly of monodisperse clusters: Dependence on target geometry', *J Chem Phys*, **131**, 175101, 2009.

[110] Q Yan and J J de Pablo, 'Hyper-parallel tempering Monte Carlo: Application to the Lennard-Jones fluid and the restricted primitive model', *J Chem Phys*, **111**, 9509–9516, 1999.

[111] T Okabe, M Kawata, Y Okamoto and M Mikami, 'Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble', *Chem Phys Lett*, **335**, 435–439, 2001.

[112] H C Andersen, 'Molecular dynamics simulations at constant pressure and/or temperature', *J Chem Phys*, **72**, 2384–2393, 1980.

[113] H J C Berendsen, J P M Postma, W F van Gunsteren, A DiNola and J R Haak, 'Molecular dynamics with coupling to an external bath', *J Chem Phys*, **81**, 3684–3690, 1984.

[114] S Nosé, 'A unified formulation of the constant temperature molecular dynamics methods', *J Chem Phys*, **81**, 511–519, 1984.

[115] W G Hoover, 'Canonical dynamics: Equilibrium phase-space distributions', *Phys Rev A*, **31**, 1695–1697, 1985.

[116] G J Martyna, M E Tuckerman, D J Tobias and M L Klein, 'Explicit reversible integrators for extended systems dynamics', *Mol Phys*, **87**, 1117–1157, 1996.

[117] R García Fernández, J L F Abascal and C Vega, 'The melting point of ice I_h for common water models calculated from direct coexistence of the solid-liquid interface', *J Chem Phys*, **124**, 144506, 2006.

[118] M M Conde, C Vega and A Patrykiejew, 'The thickness of a liquid layer on the free surface of ice as obtained from computer simulation', *J Chem Phys*, **129**, 014702, 2008.

[119] D W Heermann, P Nielaba and M Rovere, 'Hybrid molecular dynamics', *Comput Phys Commun*, **60**, 311–318, 1990.

[120] B Mehlig, D W Heermann and B M Forrest, 'Hybrid Monte Carlo method for condensed-matter systems', *Phys Rev B*, **45**, 679–685, 1992.

[121] A Brass, B J Pendleton, Y Chen and B Robson, 'Hybrid Monte Carlo simulations theory and initial comparison with molecular dynamics', *Biopolymers*, **33**, 1307–1315, 1993.

[122] M Tuckerman, *Statistical mechanics: Theory and molecular simulation*. Oxford University Press, Oxford, 2010.

[123] R Faller and J J de Pablo, 'Constant pressure hybrid molecular dynamics–Monte Carlo simulations', *J Chem Phys*, **116**, 55–59, 2002.

[124] C Desgranges and J Delhommelle, 'Phase equilibria of molecular fluids via hybrid Monte Carlo Wang–Landau simulations: Applications to benzene and *n*-alkanes', *J Chem Phys*, **130**, 244109, 2009.

[125] I Gladich and M Roeselová, 'Comparison of selected polarizable and nonpolarizable water models in molecular dynamics simulations of ice I_h', *Phys Chem Chem Phys*, **14**, 11371–11385, 2012.

[126] R Faller and J J de Pablo, 'Erratum: "Constant pressure hybrid molecular dynamics–Monte Carlo simulations"', *J Chem Phys*, **119**, 7605–7605, 2003.

[127] W H Press, S A Teukolsky, W T Vetterling and B P Flannery, *Numerical recipes in Fortran 90: The art of parallel scientific computing*. Cambridge University Press, Cambridge, 2nd edition, 1996.

[128] M P Allen and G Germano, 'Expressions for forces and torques in molecular simulations using rigid bodies', *Mol Phys*, **104**, 3225–3235, 2006.

[129] T F Miller III, M Eleftheriou, P Pattnaik, A Ndirango, D Newns and G J Martyna, 'Symplectic quaternion scheme for biophysical molecular dynamics', *J Chem Phys*, **116**, 8649–8659, 2002.

[130] H Kamberaj, R J Low and M P Neal, 'Time reversible and symplectic integrators for molecular dynamics simulations of rigid molecules', *J Chem Phys*, **122**, 224114, 2005.

[131] S Miyamoto and P A Kollman, 'SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models', *J Comput Chem*, **13**, 952–962, 1992.

[132] P P Ewald, 'Die Berechnung optischer und elektrostatischer Gitterpotentiale', *Ann Phys Leipzig*, **369**, 253–287, 1921.

[133] S W de Leeuw, J W Perram and E R Smith, 'Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants', *Proc R Soc Lond A*, **373**, 27–56, 1980.

[134] A Y Toukmaji and J A Board, 'Ewald summation techniques in perspective: A survey', *Comput Phys Commun*, **95**, 73–92, 1996.

[135] Z A Rycerz and P W M Jacobs, 'Ewald summation in the molecular dynamics simulation of large ionic systems', *Mol Sim*, **8**, 197–213, 1992.

[136] P Ball, *H_2O: A biography of water*. Phoenix, London, 2000.

[137] V F Petrenko and R W Whitworth, *Physics of ice*. Oxford University Press, Oxford, 1999.

[138] J E Huheey, E A Keiter and R L Keiter, *Inorganic chemistry: Principles of structure and reactivity*. Harper Collins, New York, 4th edition, 1993.

[139] M F Chaplin, 'Water's hydrogen bond strength', in R M Lynden-Bell, S Conway Morris, J D Barrow, J L Finney and C L Harper Jr, editors, *Water and life: The unique properties of H_2O*, 69–86. CRC Press, Boca Raton, 2010. arXiv:0706.1355.

[140] K R Harris and L A Woolf, 'Pressure and temperature dependence of the self diffusion coefficient of water and oxygen-18 water', *J Chem Soc, Faraday Trans 1*, **76**, 377–385, 1980.

[141] W Hujo, B Shadrack Jabes, V Rana, C Chakravarty and V Molinero, 'The rise and fall of anomalies in tetrahedral liquids', *J Stat Phys*, **145**, 293–312, 2011.

[142] B Hardy, 'ITS-90 formulations for vapor pressure, frostpoint temperature, dewpoint temperature and enhancement factors in the range −100 to 100 °C', in *Proceedings of the Third International Symposium on Humidity and Moisture*, 214–221. London, 1998.

[143] J Aragones, M Conde, E Noya and C Vega, 'The phase diagram of water at high pressures as obtained by computer simulations of the TIP4P/2005 model: The appearance of a plastic crystal phase', *Phys Chem Chem Phys*, **11**, 543–555, 2009.

[144] D M Murphy and T Koop, 'Review of the vapour pressures of ice and supercooled water for atmospheric applications', *Q J Roy Meteor Soc*, **131**, 1539–1565, 2005.

[145] J C Dore, 'Hydrogen-bond networks in supercooled liquid water and amorphous/vitreous ices', *J Mol Struct*, **237**, 221–232, 1990.

[146] P H Poole, F Sciortino, U Essmann and H E Stanley, 'Phase behaviour of metastable water', *Nature*, **360**, 324–328, 1992.

[147] M Yamada, S Mossa, H E Stanley and F Sciortino, 'Interplay between time-temperature transformation and the liquid-liquid phase transition in water', *Phys Rev Lett*, **88**, 195701, 2002.

[148] D T Limmer and D Chandler, 'The putative liquid-liquid transition is a liquid-solid transition in atomistic models of water', *J Chem Phys*, **135**, 134503, 2011.

[149] V Holten, C E Bertrand, M A Anisimov and J V Sengers, 'Thermodynamics of supercooled water', *J Chem Phys*, **136**, 094507, 2012.

[150] J Huang and L S Bartell, 'Kinetics of homogeneous nucleation in the freezing of large water clusters', *J Phys Chem*, **99**, 3924–3931, 1995.

[151] C G Salzmann, P G Radaelli, B Slater and J L Finney, 'The polymorphism of ice: Five unresolved questions', *Phys Chem Chem Phys*, **13**, 18468–18480, 2011.

[152] 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 icefields', *Rev Mod Phys*, **84**, 885–944, 2012.

[153] A R West, *Basic solid state chemistry*. Wiley, Chichester, 2nd edition, 1999.

[154] G P Johari, 'On the coexistence of cubic and hexagonal ice between 160 and 240 K', *Philos Mag B*, **78**, 375–383, 1998.

[155] L Pauling, 'The structure and entropy of ice and of other crystals with some randomness of atomic arrangement', *J Am Chem Soc*, **57**, 2680–2684, 1935.

[156] T Takahashi, 'On the role of cubic structure in ice nucleation', *J Cryst Growth*, **59**, 441–449, 1982.

[157] M A Carignano, 'Formation of stacking faults during ice growth on hexagonal and cubic substrates', *J Phys Chem C*, **111**, 501–504, 2007.

[158] P Pirzadeh and P G Kusalik, 'On understanding stacking fault formation in ice', *J Am Chem Soc*, **133**, 704–707, 2011.

[159] J D Bernal and R H Fowler, 'A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions', *J Chem Phys*, **1**, 515–548, 1933.

[160] J A Hayward and J R Reimers, 'Unit cells for the simulation of hexagonal ice', *J Chem Phys*, **106**, 1518–1529, 1997.

[161] V Buch, P Sandler and J Sadlej, 'Simulations of H_2O solid, liquid, and clusters, with an emphasis on ferroelectric ordering transition in hexagonal ice', *J Phys Chem B*, **102**, 8641–8653, 1998.

[162] G D Stein and J A Armstrong, 'Structure of water and carbon dioxide clusters formed via homogeneous nucleation in nozzle beams', *J Chem Phys*, **58**, 1999–2003, 1973.

[163] G Torchet, P Schwartz, J Farges, M F de Feraudy and B Raoult, 'Structure of solid water clusters formed in a free jet expansion', *J Chem Phys*, **79**, 6196–6202, 1983.

[164] J P Devlin, C Joyce and V Buch, 'Infrared spectra and structures of large water clusters', *J Phys Chem A*, **104**, 1974–1977, 2000.

[165] E Mayer and A Hallbrucker, 'Cubic ice from liquid water', *Nature*, **325**, 601–602, 1987.

[166] T C Hansen, M M Koza, P Lindner and W F Kuhs, 'Formation and annealing of cubic ice: II. Kinetic study', *J Phys: Condens Mat*, **20**, 285105, 2008.

[167] K Morishige and H Uematsu, 'The proper structure of cubic ice confined in mesopores', *J Chem Phys*, **122**, 044711, 2005.

[168] T L Malkin, B J Murray, A V Brukhno, J Anwar and C G Salzmann, 'Structure of ice crystallized from supercooled water', *Proc Natl Acad Sci USA*, **109**, 1041–1045, 2012.

[169] E B Moore and V Molinero, 'Is it cubic? Ice crystallization from deeply supercooled water', *Phys Chem Chem Phys*, **13**, 20008–20016, 2011.

[170] W F Kuhs, C Sippel, A Falenty and T C Hansen, 'Extent and relevance of stacking disorder in "ice I_c"', *Proc Natl Acad Sci USA*, **109**, 21259–21264, 2012.

[171] K Kobayashi, M Koshino and K Suenaga, 'Atomically resolved images of I_h ice single crystals in the solid phase', *Phys Rev Lett*, **106**, 206101, 2011.

[172] K Thürmer and N C Bartelt, 'Growth of multilayer ice films and the formation of cubic ice imaged with STM', *Phys Rev B*, 195425, 2008.

[173] L S Bartell and J Huang, 'Supercooling of water below the anomalous range near 226 K', *J Phys Chem*, **98**, 7455–7457, 1994.

[174] B Krämer, O Hübner, H Vortisch, L Wöste, T Leisner, M Schwell, E Rühl and H Baumgärtel, 'Homogeneous nucleation rates of supercooled water measured in single levitated microdroplets', *J Chem Phys*, **111**, 6521–6527, 1999.

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

[176] B J Murray and A K Bertram, 'Formation and stability of cubic ice in water droplets.', *Phys Chem Chem Phys*, **8**, 186–192, 2006.

[177] D M Murphy, 'Dehydration in cold clouds is enhanced by a transition from cubic to hexagonal ice', *Geophys Res Lett*, **30**, 2230, 2003.

[178] T Peter, C Marcolli, P Spichtinger, T Corti, M B Baker and T Koop, 'Atmosphere: When dry air is too humid', *Science*, **314**, 1399–1402, 2006.

[179] J E Shilling, M A Tolbert, O B Toon, E J Jensen, B J Murray and A K Bertram, 'Measurements of the vapor pressure of cubic ice and their implications for atmospheric ice clouds', *Geophys Res Lett*, **33**, L17801, 2006.

[180] A J Heymsfield, G Thompson, H Morrison, A Bansemer, R M Rasmussen, P Minnis, Z Wang and D Zhang, 'Formation and spread of aircraft-induced holes in clouds', *Science*, **333**, 77–81, 2011.

[181] Y Paul Handa, D D Klug and E Whalley, 'Difference in energy between cubic and hexagonal ice', *J Chem Phys*, **84**, 7009–7010, 1986.

[182] E J Jensen, J B Smith, L Pfister, J V Pittman, E M Weinstock, D S Sayres, R L Herman, R F Troy, K Rosenlof, T L Thompson, A M Fridlind, P K Hudson, D J Cziczo, A J Heymsfield, C Schmitt and J C Wilson, 'Ice supersaturations exceeding 100 % at the cold tropical tropopause: Implications for cirrus formation and dehydration', *Atmos Chem Phys*, **5**, 851–862, 2005.

[183] E Whalley, 'Scheiner's halo: Evidence for ice I_c in the atmosphere', *Science*, **211**, 389–390, 1981.

[184] E Whalley, 'Cubic ice in nature', *J Phys Chem*, **87**, 4174–4179, 1983.

[185] T Kobayashi, Y Furukawa, T Takahashi and H Uyeda, 'Cubic structure models at the junctions in polycrystalline snow crystals', *J Cryst Growth*, **35**, 262–268, 1976.

[186] G P Johari, 'Water's size-dependent freezing to cubic ice', *J Chem Phys*, **122**, 194504, 2005.

[187] W X Zhang, C He, J S Lian and Q Jiang, 'Selected crystallization of water as a function of size', *Chem Phys Lett*, **421**, 251–255, 2006.

[188] W L Jorgensen, J Chandrasekhar, J D Madura, R W Impey and M L Klein, 'Comparison of simple potential functions for simulating liquid water', *J Chem Phys*, **79**, 926–935, 1983.

[189] C J Fennell and J D Gezelter, 'Computational free energy studies of a new ice polymorph which exhibits greater stability than ice I_h', *J Chem Theory Comput*, **1**, 662–667, 2005.

[190] C Vega, E Sanz, J L F Abascal and E G Noya, 'Determination of phase diagrams via computer simulation: Methodology and applications to water, electrolytes and proteins', *J Phys: Condens Mat*, **20**, 153101, 2008.

[191] S J Cox, S M Kathmann, J A Purton, M J Gillan and A Michaelides, 'Non-hexagonal ice at hexagonal surfaces: The role of lattice mismatch', *Phys Chem Chem Phys*, **14**, 7944–7949, 2012.

[192] I M Svishchev and P G Kusalik, 'Crystallization of liquid water in a molecular dynamics simulation', *Phys Rev Lett*, **73**, 975–978, 1994.

[193] I M Svishchev and P G Kusalik, 'Electrofreezing of liquid water: A microscopic perspective', *J Am Chem Soc*, **118**, 649–654, 1996.

[194] L Vrbka and P Jungwirth, 'Homogeneous freezing of water starts in the subsurface', *J Phys Chem B*, **110**, 18126–18129, 2006.

[195] E Pluhařová, L Vrbka and P Jungwirth, 'Effect of surface pollution on homogeneous ice nucleation: A molecular dynamics study', *J Phys Chem C*, **114**, 7831–7838, 2010.

[196] J Y Yan and G N Patey, 'Heterogeneous ice nucleation induced by electric fields', *J Phys Chem Lett*, **2**, 2555–2559, 2011.

[197] M Gavish, J L Wang, M Eisenstein, M Lahav and L Leiserowitz, 'The role of crystal polarity in alpha-amino acid crystals for induced nucleation of ice', *Science*, **256**, 815–818, 1992.

[198] C A Stan, S K Y Tang, K J M Bishop and G M Whitesides, 'Externally applied electric fields up to 1.6×10^5 V/m do not affect the homogeneous nucleation of ice in supercooled water', *J Phys Chem B*, **115**, 1089–1097, 2011.

[199] H R Pruppacher, 'Electrofreezing of supercooled water', *Pure Appl Geophys*, **104**, 623–634, 1973.

[200] I Braslavsky and S G Lipson, 'Electrofreezing effect and nucleation of ice crystals in free growth experiments', *Appl Phys Lett*, **72**, 264–266, 1998.

[201] D Ehre, E Lavert, M Lahav and I Lubomirsky, 'Water freezes differently on positively and negatively charged surfaces of pyroelectric materials', *Science*, **327**, 672–675, 2010.

[202] J L Aragones, L G MacDowell and C Vega, 'Dielectric constant of ices and water: A lesson about water interactions', *J Phys Chem A*, **115**, 5745–5758, 2011.

[203] J L Aragones, L G MacDowell, J I Siepmann and C Vega, 'Phase diagram of water under an applied electric field', *Phys Rev Lett*, **107**, 155702, 2011.

[204] J Y Yan and G N Patey, 'Molecular dynamics simulations of ice nucleation by electric fields', *J Phys Chem A*, **116**, 7057–7064, 2012.

[205] V Aleksandrov, A Barannikov and N Dobritsa, 'Effect of magnetic field on the supercooling of water drops', *Inorg Mater*, **36**, 895–898, 2000.

[206] H Hu, H Hou and B Wang, 'Molecular dynamics simulations of ice growth from supercooled water when both electric and magnetic fields are applied', *J Phys Chem C*, **116**, 19773–19780, 2012.

[207] T H Jackson, A Ungan, J K Critser and D Gao, 'Novel microwave technology for cryopreservation of biomaterials by suppression of apparent ice formation', *Cryobiology*, **34**, 363–372, 1997.

[208] B Wowk, 'Electric and magnetic fields in cryopreservation', *Cryobiology*, **64**, 301–303, 2012.

[209] M Kulmala, P Korhonen, I Napari, A Karlsson, H Berresheim and C O'Dowd, 'Aerosol formation during parforce: Ternary nucleation of H_2SO_4, NH_3, and H_2O', *J Geophys Res*, **107**, 8111, 2002.

[210] G R Wood and A G Walton, 'Homogeneous nucleation kinetics of ice from water', *J Appl Phys*, **41**, 3027–3036, 1970.

[211] K K Nanda, 'Size-dependent melting of nanoparticles: Hundred years of thermodynamic model', *Pramana-J Phys*, **72**, 617–628, 2009.

[212] M Perez, 'Gibbs–Thomson effects in phase transformations', *Scripta Mater*, **52**, 709–712, 2005.

[213] F Aliotta, M Pochylski, R C Ponterio, F Saija, G Salvato and C S Vasi, 'Structure of bulk water from raman measurements of supercooled pure liquid and licl solutions', *Phys Rev B*, **86**, 134301, 2012.

[214] P Wilson, 'Supercooling of water', in P Wilson, editor, *Supercooling*. InTech, Rijeka, 2012.

[215] R J Anderson, R C Miller, J L Kassner and D E Hagen, 'A Study of Homogeneous Condensation-Freezing Nucleation of Small Water Droplets in an Expansion Cloud Chamber', *J Atmos Sci*, **37**, 2508–2520, 1980.

[216] S E Wood, M B Baker and B D Swanson, 'Instrument for studies of homogeneous and heterogeneous

ice nucleation in free-falling supercooled water droplets', *Rev Sci Instrum*, **73**, 3988–3996, 2002.

[217] M E Earle, T Kuhn, A F Khalizov and J J Sloan, 'Volume nucleation rates for homogeneous freezing in supercooled water microdroplets: Results from a combined experimental and modelling approach', *Atmos Chem Phys*, **10**, 7945–7961, 2010.

[218] T Kuhn, M E Earle, A F Khalizov and J J Sloan, 'Size dependence of volume and surface nucleation rates for homogeneous freezing of supercooled water droplets', *Atmos Chem Phys*, **11**, 2853–2861, 2011.

[219] P Stöckel, H Vortisch, T Leisner and H Baumgärtel, 'Homogeneous nucleation of supercooled liquid water in levitated microdroplets', *J Mol Liq*, **96-97**, 153–175, 2002.

[220] D Duft and T Leisner, 'Laboratory evidence for volume-dominated nucleation of ice in supercooled water microdroplets', *Atmos Chem Phys*, **4**, 1997–2000, 2004.

[221] P Stöckel, I M Weidinger, H Baumgärtel and T Leisner, 'Rates of homogeneous ice nucleation in levitated H_2O and D_2O droplets', *J Phys Chem A*, **109**, 2540–2546, 2005.

[222] P Kabath, P Stöckel, A Lindinger and H Baumgärtel, 'The nucleation of ice in supercooled D_2O and H_2O', *J Mol Liq*, **125**, 204–211, 2006.

[223] H Baumgärtel and H W Zimmermann, 'The homogeneous nucleation in supercooled water. An examination using statistics and irreversible thermodynamics', *J Mol Liq*, **164**, 178–186, 2011.

[224] D Rzesanke, J Nadolny, D Duft, R Muller, A Kiselev and T Leisner, 'On the role of surface charges for homogeneous freezing of supercooled water microdroplets', *Phys Chem Chem Phys*, **14**, 9359–9363, 2012.

[225] G T Butorin and V P Skripov, 'Crystallization of supercooled water', *Sov Phys Crystallogr*, **17**, 322–326, 1972.

[226] P W Wilson and A D J Haymet, 'The spread of nucleation temperatures of a sample of supercooled liquid is independent of the average nucleation temperature', *J Phys Chem B*, **116**, 13472–13475, 2012.

[227] R W Michelmore and F Franks, 'Nucleation rates of ice in undercooled water and aqueous solutions of polyethylene glycol', *Cryobiology*, **19**, 163–171, 1982.

[228] P Taborek, 'Nucleation in emulsified supercooled water', *Phys Rev B*, **32**, 5902–5906, 1985.

[229] S Balibar and F Caupin, 'Nucleation of crystals from their liquid phase', *C R Phys*, **7**, 988–999, 2006.

[230] V Stoyanova, D Kashchiev and T Kupenova, 'Freezing of water droplets seeded with atmospheric aerosols and ice nucleation activity of the aerosols', *J Aerosol Sci*, **25**, 867–877, 1994.

[231] J F Edd, K J Humphry, D Irimia, D A Weitz and M Toner, 'Nucleation and solidification in static arrays of monodisperse drops', *Lab Chip*, **9**, 1859–1865, 2009.

[232] B J Murray, S L Broadley, T W Wilson, S J Bull, R H Wills, H K Christenson and E J Murray, 'Kinetics of the homogeneous freezing of water', *Phys Chem Chem Phys*, **12**, 10380–10387, 2010.

[233] C A Stan, G F Schneider, S S Shevkoplyas, M Hashimoto, M Ibanescu, B J Wiley and G M Whitesides, 'A microfluidic apparatus for the study of ice nucleation in supercooled water drops', *Lab Chip*, **9**, 2293–2305, 2009.

[234] A Manka, H Pathak, S Tanimura, J Wölk, R Strey and B E Wyslouzil, 'Freezing water in no-man's land', *Phys Chem Chem Phys*, **14**, 4505–4516, 2012.

[235] J E McDonald, 'Homogeneous nucleation of supercooled water drops', *J Meteorol*, **10**, 416–433, 1953.

[236] S C Mossop, 'The freezing of supercooled water', *P Phys Soc Lond B*, **68**, 193, 1955.

[237] B Mason, 'The supercooling and nucleation of water', *Adv Phys*, **7**, 221–234, 1958.

[238] Ó F Sigurbjörnsson and R Signorell, 'Volume versus surface nucleation in freezing aerosols', *Phys Rev E*, **77**, 051601, 2008.

[239] B Riechers, F Wittbracht, A Hutten and T Koop, 'Homogeneous ice nucleation rate of water droplets produced in a microfluidic device and the role of temperature uncertainty', *Phys Chem Chem Phys*, Accepted Manuscript, 2013.

[240] A Tabazadeh, Y S Djikaev and H Reiss, 'Surface crystallization of supercooled water in clouds', *Proc Natl Acad Sci USA*, **99**, 15873–15878, 2002.

[241] T Li, D Donadio, L M Ghiringhelli and G Galli, 'Surface-induced crystallization in supercooled tetrahedral liquids', *Nat Mater*, **8**, 726–730, 2009.

[242] L Vrbka and P Jungwirth, 'Molecular simulations of water freezing: Brine rejection and homogeneous nucleation', in W F Kuhs, editor, *Physics and chemistry of ice*, 627–634. RSC Publishing, Cambridge, 2007. Proceedings of the 11th International Conference on the Physics and Chemistry of Ice.

[243] Y J Lü, W J Xie and B Wei, 'Observation of ice nucleation in acoustically levitated water drops', *Appl Phys Lett*, **87**, 184107, 2005.

[244] S Bauerecker, P Ulbig, V Buch, L Vrbka and P Jungwirth, 'Monitoring ice nucleation in pure and salty water via high-speed imaging and computer simulations', *J Phys Chem C*, **112**, 7631–7636, 2008.

[245] D G Thomas and L A K Staveley, 'A study of the supercooling of drops of some molecular liquids', *J Chem Soc*, 4569–4577, 1952.

[246] K Sassen and G C Dodd, 'Homogeneous nucleation rate for highly supercooled cirrus cloud droplets', *J Atmos Sci*, **45**, 1357–1369, 1988.

[247] P J DeMott and D C Rogers, 'Freezing nucleation rates of dilute solution droplets measured between -30° and $-40\,^{\circ}$C in laboratory simulations of natural clouds', *J Atmos Sci*, **47**, 1056–1064, 1990.

[248] T Koop and B Zobrist, 'Parameterizations for ice nucleation in biological and atmospheric systems', *Phys Chem Chem Phys*, **11**, 10839–10850, 2009.

[249] D A Knopf and M D Lopez, 'Homogeneous ice freezing temperatures and ice nucleation rates of aqueous ammonium sulfate and aqueous levoglucosan particles for relevant atmospheric conditions', *Phys Chem Chem Phys*, **11**, 8056–8068, 2009.

[250] W Cantrell and A Heymsfield, 'Production of ice in tropospheric clouds: A review', *B Am Meteorol Soc*, **86**, 795–807, 2005.

[251] C Vega, J L F Abascal, M M Conde and J L Aragones, 'What ice can teach us about water interactions: A critical comparison of the performance of different water models', *Faraday Discuss*, **141**, 251–276, 2009.

[252] C Vega and J L F Abascal, 'Simulating water with rigid non-polarizable models: A general perspective', *Phys Chem Chem Phys*, **13**, 19663–19688, 2011.

[253] P T Kiss and A Baranyai, 'Sources of the deficiencies in the popular SPC/E and TIP3P models of water', *J Chem Phys*, **134**, 054106, 2011.

[254] E G Noya, C Menduiña, J L Aragones and C Vega, 'Equation of state, thermal expansion coefficient, and isothermal compressibility for ices I_h, II, III, V, and VI, as obtained from computer simulation', *J Phys Chem C*, **111**, 15877–15888, 2007.

[255] H L Pi, J L Aragones, C Vega, E G Noya, J L Abascal, M A Gonzalez and C McBride, 'Anomalies in water as obtained from computer simulations of the TIP4P/2005 model: Density maxima, and density, isothermal compressibility and heat capacity minima', *Mol Phys*, **107**, 365–374, 2009.

[256] J L F Abascal, E Sanz and C Vega, 'Triple points and coexistence properties of the dense phases of water calculated using computer simulation', *Phys Chem Chem Phys*, **11**, 556–562, 2009.

[257] M Agarwal, M P Alam and C Chakravarty, 'Thermodynamic, diffusional, and structural anomalies in rigid-body water models', *J Phys Chem B*, **115**, 6935–6945, 2011.

[258] D Rozmanov and P G Kusalik, 'Transport coefficients of the TIP4P-2005 water model', *J Chem Phys*, **136**, 044507, 2012.

[259] S Tazi, A Boţan, M Salanne, V Marry, P Turq and B Rotenberg, 'Diffusion coefficient and shear viscosity of rigid water models', *J Phys: Condens Mat*, **24**, 284117, 2012.

[260] U Essmann, L Perera, M L Berkowitz, T Darden, H Lee and L G Pedersen, 'A smooth particle mesh Ewald method', *J Chem Phys*, **103**, 8577–8593, 1995.

[261] D Paschek and A Geiger, *MOSCITO 4*. Department of Physical Chemistry, University of Dortmund, Dortmund, 2002.

[262] H Nada and J P J M van der Eerden, 'An intermolecular potential model for the simulation of ice and water near the melting point: A six-site model of H_2O', *J Chem Phys*, **118**, 7401–7413, 2003.

[263] M W Mahoney and W L Jorgensen, 'A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions', *J Chem Phys*, **112**, 8910–8922, 2000.

[264] S W Rick, 'A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums', *J Chem Phys*, **120**, 6085–6093, 2004.

[265] H Nada and Y Furukawa, 'Anisotropy in growth kinetics at interfaces between proton-disordered hexagonal ice and water: A molecular dynamics study using the six-site model of H_2O', *J Cryst Growth*, **283**, 242–256, 2005.

[266] H J C Berendsen, J R Grigera and T P Straatsma, 'The missing term in effective pair potentials', *J Phys Chem*, **91**, 6269–6271, 1987.

[267] J L F Abascal, E Sanz, R García Fernández and C Vega, 'A potential model for the study of ices and amorphous water: TIP4P/Ice', *J Chem Phys*, **122**, 234511, 2005.

[268] N Bernstein, C Varnai, I Solt, S A Winfield, M C Payne, I Simon, M Fuxreiter and G Csanyi, 'QM/MM simulation of liquid water with an adaptive quantum region', *Phys Chem Chem Phys*, **14**, 646–656, 2012.

[269] D J Huggins, 'Correlations in liquid water for the TIP3P-Ewald, TIP4P-2005, TIP5P-Ewald, and SWM4-NDP models', *J Chem Phys*, **136**, 064518, 2012.

[270] G Lamoureux, E Harder, I V Vorobyov, B Roux and A D MacKerell Jr, 'A polarizable model of water for molecular dynamics simulations of biomolecules', *Chem Phys Lett*, **418**, 245–249, 2006.

[271] S Habershon and D E Manolopoulos, 'Free energy calculations for a flexible water model', *Phys Chem Chem Phys*, **13**, 19714–19727, 2011.

[272] I G Tironi, R M Brunne and W F van Gunsteren, 'On the relative merits of flexible versus rigid models for use in computer simulations of molecular liquids', *Chem Phys Lett*, **250**, 19–24, 1996.

[273] M A González and J L F Abascal, 'A flexible model for water based on TIP4P/2005', *J Chem Phys*, **135**, 224516, 2011.

[274] S Habershon, T E Markland and D E Manolopoulos, 'Competing quantum effects in the dynamics of a flexible water model', *J Chem Phys*, **131**, 024501, 2009.

[275] S Habershon and D E Manolopoulos, 'Thermodynamic integration from classical to quantum mechanics', *J Chem Phys*, **135**, 224111, 2011.

[276] C McBride, C Vega, E G Noya, R Ramírez and L M Sesé, 'Quantum contributions in the ice phases: The path to a new empirical model for water—TIP4PQ/2005', *J Chem Phys*, **131**, 024506, 2009.

[277] C McBride, E G Noya, J L Aragones, M M Conde and C Vega, 'The phase diagram of water from quantum simulations', *Phys Chem Chem Phys*, **14**, 10140–10146, 2012.

[278] D R Nutt and J C Smith, 'Choosing an appropriate water model for use in biomolecular simulations', in W F Kuhs, editor, *Physics and chemistry of ice*, 451–458. RSC Publishing, Cambridge, 2007. Proceedings of the 11th International Conference on the Physics and Chemistry of Ice.

[279] C J Tainter, P A Pieniazek, Y-S Lin and J L Skinner, 'Robust three-body water simulation model', *J Chem Phys*, **134**, 184501, 2011.

[280] F H Stillinger and T A Weber, 'Computer simulation of local order in condensed phases of silicon', *Phys Rev B*, **31**, 5262–5271, 1985.

[281] V Molinero, S Sastry and C A Angell, 'Tuning of tetrahedrality in a silicon potential yields a series of monatomic (metal-like) glass formers of very high fragility', *Phys Rev Lett*, **97**, 075701, 2006.

[282] E B Moore and V Molinero, 'Ice crystallization in water's "no-man's land"', *J Chem Phys*, **132**, 244504, 2010.

[283] K Ramasesha, S T Roberts, R A Nicodemus, A Mandal and A Tokmakoff, 'Ultrafast 2D IR anisotropy of water reveals reorientation during hydrogen-bond switching', *J Chem Phys*, **135**, 054509, 2011.

[284] D Laage, G Stirnemann, F Sterpone and J T Hynes, 'Water jump reorientation: From theoretical prediction to experimental observation', *Acc Chem Res*, **45**, 53–62, 2012.

[285] K Koga, H Tanaka and X C Zeng, 'First-order transition in confined water between high-density liquid and low-density amorphous phases', *Nature*, **408**, 564–567, 2000.

[286] M Matsumoto, S Saito and I Ohmine, 'Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing', *Nature*, **416**, 409–413, 2002.

[287] R Radhakrishnan and B L Trout, 'Nucleation of hexagonal ice (I_h) in liquid water', *J Am Chem Soc*, **125**, 7743–7747, 2003.

[288] R Radhakrishnan and B L Trout, 'Nucleation of crystalline phases of water in homogeneous and inhomogeneous environments', *Phys Rev Lett*, **90**, 158301, 2003.

[289] M A Carignano, P B Shepson and I Szleifer, 'Molecular dynamics simulations of ice growth from supercooled water', *Mol Phys*, **103**, 2957–2967, 2005.

[290] L Vrbka and P Jungwirth, 'Molecular dynamics simulations of freezing of water and salt solutions', *J Mol Liq*, **134**, 64–70, 2007.

[291] A V Brukhno, J Anwar, R Davidchack and R Handel, 'Challenges in molecular simulation of homogeneous ice nucleation', *J Phys: Condens Mat*, **20**, 494243, 2008.

[292] R G Pereyra, I Szleifer and M A Carignano, 'Temperature dependence of ice critical nucleus size', *J Chem Phys*, **135**, 034508, 2011.

[293] V C Weiss, M Rullich, C Köhler and T Frauenheim, 'Kinetic aspects of the thermostatted growth of ice from supercooled water in simulations', *J Chem Phys*, **135**, 034701, 2011.

[294] H Nada, 'Analysis of ice crystal growth shape under high pressure using molecular dynamics simulation', *Cryst Growth Des*, **11**, 3130–3136, 2011.

[295] P Pirzadeh, E N Beaudoin and P G Kusalik, 'Structural evolution during water crystallization: Insights from ring analysis', *Chem Phys Lett*, **517**, 117–125, 2011.

[296] D Rozmanov and P G Kusalik, 'Temperature dependence of crystal growth of hexagonal ice (I_h)', *Phys Chem Chem Phys*, **13**, 15501–15511, 2011.

[297] D Rozmanov and P Kusalik, 'Isoconfigurational molecular dynamics study of the kinetics of ice crystal growth', *Phys Chem Chem Phys*, **14**, 13010–13018, 2012.

[298] D Rozmanov and P G Kusalik, 'Anisotropy in the crystal growth of hexagonal ice, I_h', *J Chem Phys*, **137**, 094702, 2012.

[299] M Seo, E Jang, K Kim, S Choi and J S Kim, 'Understanding anisotropic growth behavior of hexagonal ice on a molecular scale: A molecular dynamics simulation study', *J Chem Phys*, **137**, 154503, 2012.

[300] A Okabe, B Boots, K Sugihara and S N Chiu, *Spatial tessellations: Concepts and applications of Voronoi diagrams*. Wiley, Chichester, 2nd edition, 1999.

[301] J A van Meel, L Filion, C Valeriani and D Frenkel, 'A parameter-free, solid-angle based, nearest-neighbor algorithm', *J Chem Phys*, **136**, 234107, 2012.

[302] W Mickel, S C Kapfer, G E Schröder-Turk and K Mecke, 'Shortcomings of the bond orientational order parameters for the analysis of disordered particulate matter', *J Chem Phys*, **138**, 044501, 2013.

[303] G B Arfken and H J Weber, *Mathematical methods for physicists*. Elsevier Academic, London, 6th edition, 2005.

[304] C Jekeli, 'Spherical harmonic analysis, aliasing, and filtering', *J Geodesy*, **70**, 214–223, 1996.

[305] E Hivon, K M Górski, C B Netterfield, B P Crill, S Prunet and F Hansen, 'MASTER of the cosmic microwave background anisotropy power spectrum: A fast method for statistical analysis of large and

complex cosmic microwave background data sets', *Astrophys J*, **567**, 2–17, 2002.

[306] A Kudlicki, M Rowicka, M Gilski and Z Otwinowski, 'An efficient routine for computing symmetric real spherical harmonics for high orders of expansion', *J Appl Cryst*, **38**, 501–504, 2005.

[307] R J Morris, R J Najmanovich, A Kahraman and J M Thornton, 'Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons', *Bioinformatics*, **21**, 2347–2355, 2005.

[308] K Khairy and J Howard, 'Minimum-energy vesicle and cell shapes calculated using spherical harmonics parameterization', *Soft Matter*, **7**, 2138–2143, 2011.

[309] M A Blanco, M Flórez and M Bermejo, 'Evaluation of the rotation matrices in the basis of real spherical harmonics', *J Mol Struc-Theochem*, **419**, 19–27, 1997.

[310] J Wedekind and D Reguera, 'What is the best definition of a liquid cluster at the molecular scale?', *J Chem Phys*, **127**, 154516, 2007.

[311] W Lechner and C Dellago, 'Accurate determination of crystal structures based on averaged local bond order parameters', *J Chem Phys*, **129**, 114707, 2008.

[312] S Jungblut and C Dellago, 'Crystallization of a binary Lennard-Jones mixture', *J Chem Phys*, **134**, 104501, 2011.

[313] P-L Chau and A J Hardwick, 'A new order parameter for tetrahedral configurations', *Mol Phys*, **93**, 511–518, 1998.

[314] J R Errington and P G Debenedetti, 'Relationship between structural order and the anomalies of liquid water', *Nature*, **409**, 318–321, 2001.

[315] E E Santiso and B L Trout, 'A general set of order parameters for molecular crystals', *J Chem Phys*, **134**, 064109, 2011.

[316] B Senger, P Schaaf, D S Corti, R Bowles, J-C Voegel and H Reiss, 'A molecular theory of the homogeneous nucleation rate. I. Formulation and fundamental issues', *J Chem Phys*, **110**, 6421–6437, 1999.

[317] C Vega, E Sanz and J L F Abascal, 'The melting temperature of the most common models of water', *J Chem Phys*, **122**, 114507, 2005.

[318] R Handel, R L Davidchack, J Anwar and A Brukhno, 'Direct calculation of solid-liquid interfacial free energy for molecular systems: TIP4P ice-water interface', *Phys Rev Lett*, **100**, 036104, 2008.

[319] R L Davidchack, R Handel, J Anwar and A V Brukhno, 'Ice I_h–water interfacial free energy of simple water models with full electrostatic interactions', *J Chem Theory Comput*, **8**, 2383–2390, 2012.

[320] E K Bigg, 'The supercooling of water', *Proc Phys Soc B*, **66**, 688–694, 1953.

[321] D Pan, L-M Liu, B Slater, A Michaelides and E Wang, 'Melting the ice: On the relation between melting temperature and size for nanoscale ice crystals', *ACS Nano*, **5**, 4562–4569, 2011.

[322] L M Ghiringhelli, C Valeriani, E J Meijer and D Frenkel, 'Local structure of liquid carbon controls diamond nucleation', *Phys Rev Lett*, **99**, 055702, 2007.

[323] W Lechner, C Dellago and P G Bolhuis, 'Role of the prestructured surface cloud in crystal nucleation', *Phys Rev Lett*, **106**, 085701, 2011.

[324] E B Moore and V Molinero, 'Structural transformation in supercooled water controls the crystallization rate of ice', *Nature*, **479**, 506–508, 2011.

[325] W Lechner, C Dellago and P G Bolhuis, 'Reaction coordinates for the crystal nucleation of colloidal suspensions extracted from the reweighted path ensemble', *J Chem Phys*, **135**, 154110, 2011.

[326] E B Moore and V Molinero, 'Growing correlation length in supercooled water', *J Chem Phys*, **130**, 244505, 2009.

[327] N Kastelowitz, J C Johnston and V Molinero, 'The anomalously high melting temperature of bilayer ice', *J Chem Phys*, **132**, 124511, 2010.

[328] E González Solveyra, E de la Llave, D A Scherlis and V Molinero, 'Melting and crystallization of ice in partially filled nanopores', *J Phys Chem B*, **115**, 14196–14204, 2011.

[329] E B Moore, J T Allen and V Molinero, 'Liquid-ice coexistence below the melting temperature for water confined in hydrophilic and hydrophobic nanopores', *J Phys Chem C*, **116**, 7507–7514, 2012.

[330] J C Johnston and V Molinero, 'Crystallization, melting, and structure of water nanoparticles at atmospherically relevant temperatures', *J Am Chem Soc*, **134**, 6650–6659, 2012.

[331] T D Shepherd, M A Koc and V Molinero, 'The quasi-liquid layer of ice under conditions of methane clathrate formation', *J Phys Chem C*, **116**, 12172–12180, 2012.

[332] S Plimpton, 'Fast parallel algorithms for short-range molecular dynamics', *J Comp Phys*, **117**, 1–19, 1995.

[333] J Russo and H Tanaka, 'The microscopic pathway to crystallization in supercooled liquids', *Sci Rep*, **2**, 2012.

[334] E Sanz, C Valeriani, D Frenkel and M Dijkstra, 'Evidence for out-of-equilibrium crystal nucleation in suspensions of oppositely charged colloids', *Phys Rev Lett*, **99**, 055501, 2007.

[335] L C Jacobson, W Hujo and V Molinero, 'Thermodynamic stability and growth of guest-free clathrate hydrates: A low-density crystal phase of water', *J Phys Chem B*, **113**, 10298–10307, 2009.

[336] P Beaucage and N Mousseau, 'Nucleation and crystallization process of silicon using the Stillinger–Weber potential', *Phys Rev B*, **71**, 094102, 2005.

[337] J Rudnick and G Gaspari, 'The asphericity of random walks', *J Phys A: Math Gen*, **19**, L191–L193, 1986.

[338] J-M Leyssale, J Delhommelle and C Millot, 'Atomistic simulation of the homogeneous nucleation and of the growth of N_2 crystallites', *J Chem Phys*, **122**, 104510, 2005.

[339] Z-J Wang, C Valeriani and D Frenkel, 'Homogeneous bubble nucleation driven by local hot spots: A molecular dynamics study', *J Phys Chem B*, **113**, 3776–3784, 2009.

[340] T Li, D Donadio and G Galli, 'Nucleation of tetrahedral solids: A molecular dynamics study of supercooled liquid silicon', *J Chem Phys*, **131**, 224519, 2009.

[341] R P Sear, 'Nucleation in the presence of slow microscopic dynamics', *J Chem Phys*, **128**, 214513, 2008.

[342] J Wang, Y W Tang and X C Zeng, 'Solid-liquid interfacial free energy of water: A molecular dynamics simulation study', *J Chem Theory Comput*, **3**, 1494–1498, 2007.

[343] P Pirzadeh, E N Beaudoin and P G Kusalik, 'Interfacial free energy: An entropy portent to energy changes', *Cryst Growth Des*, **12**, 124–128, 2012.

[344] J L F Abascal and C Vega, 'Widom line and the liquid–liquid critical point for the TIP4P/2005 water model', *J Chem Phys*, **133**, 234502, 2010.

[345] J L F Abascal and C Vega, 'Note: Equation of state and compressibility of supercooled water: Simulations and experiment', *J Chem Phys*, **134**, 186101, 2011.

[346] G T Gao, X C Zeng and H Tanaka, 'The melting temperature of proton-disordered hexagonal ice: A computer simulation of 4-site transferable intermolecular potential model of water', *J Chem Phys*, **112**, 8534–8538, 2000.

[347] P H Hünenberger and W F van Gunsteren, 'Alternative schemes for the inclusion of a reaction-field correction into molecular dynamics simulations: Influence on the simulated energetic, structural, and dielectric properties of liquid water', *J Chem Phys*, **108**, 6117–6134, 1998.

[348] X Wei, S Xiao and J Ni, 'Studies of ice melting using molecular dynamics', *Mol Sim*, **36**, 823–830, 2010.

[349] R G Pereyra, A B di Lorenzo, D C Malaspina and M A Carignano, 'On the relation between hydrogen bonds, tetrahedral order and molecular mobility in model water', *Chem Phys Lett*, **538**, 35–38, 2012.

[350] M Hermanns, *Parallel programming in Fortran 95 using OpenMP*. School of Aeronautical Engineering, Universidad Politécnica de Madrid, 2002.

[351] M R Wilson and J M Ilnytskyi, 'Parallel computer simulation techniques for the study of macromolecules', *NATO Science Series*, **177**, 335–359, 2005.

[352] A K Soper, 'The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa', *Chem Phys*, **258**, 121–137, 2000.

[353] T Strässle, A M Saitta, Y L Godec, G Hamel, S Klotz, J S Loveday and R J Nelmes, 'Structure of dense liquid water by neutron scattering to 6.5 GPa and 670 K', *Phys Rev Lett*, **96**, 067801, 2006.

[354] R M Lynden-Bell and P G Debenedetti, 'Computational investigation of order, structure, and dynamics in modified water models', *J Phys Chem B*, **109**, 6527–6534, 2005.

[355] G Stirnemann and D Laage, 'Communication: On the origin of the non-Arrhenius behavior in water reorientation dynamics', *J Chem Phys*, **137**, 031101, 2012.

[356] A Pérez and A Rubio, 'A molecular dynamics study of water nucleation using the TIP4P/2005 model', *J Chem Phys*, **135**, 244505, 2011.

[357] G M Whitesides and M Boncheva, 'Beyond molecules: Self-assembly of mesoscopic and macroscopic components', *Proc Natl Acad Sci USA*, **99**, 4769–4774, 2002.

[358] D S Goodsell, *Bionanotechnology: Lessons from Nature*. Wiley-Liss, Hoboken, 2004.

[359] A Zlotnick, 'Theoretical aspects of virus capsid assembly', *J Mol Recognit*, **18**, 479–490, 2005.

[360] L A Amos, 'Microtubule structure and its stabilisation', *Org Biomol Chem*, **2**, 2153–2160, 2004.

[361] M Antonietti and S Förster, 'Vesicles and liposomes: A self-assembly principle beyond lipids', *Adv Mater*, **15**, 1323–1333, 2003.

[362] H N Miras, G J T Cooper, D-L Long, H Bögge, A Müller, C Streb and L Cronin, 'Unveiling the transient template in the self-assembly of a molecular oxide nanowheel', *Science*, **327**, 72–74, 2010.

[363] N C Seeman, 'DNA in a material world', *Nature*, **421**, 427–431, 2003.

[364] J Bath and A J Turberfield, 'DNA nanomachines', *Nat Nanotechnol*, **2**, 275–284, 2007.

[365] S C Glotzer and M J Solomon, 'Anisotropy of building blocks and their assembly into complex structures', *Nat Mater*, **6**, 557–562, 2007.

[366] A B Pawar and I Kretzschmar, 'Fabrication, assembly, and application of patchy particles', *Macromol Rapid Comm*, **31**, 150–168, 2010.

[367] L Hong, S Jiang and S Granick, 'Simple method to produce Janus colloidal particles in large quantity', *Langmuir*, **22**, 9495–9499, 2006.

[368] Y-S Cho, G-R Yi, S-H Kim, S-J Jeon, M T Elsesser, H K Yu, S-M Yang and D J Pine, 'Particles with

coordinated patches or windows from oil-in-water emulsions', *Chem Mater*, **19**, 3183–3193, 2007.

[369] E W Edwards, D Wang and H Möhwald, 'Hierarchical organization of colloidal particles: From colloidal crystallization to supraparticle chemistry', *Macromol Chem Physic*, **208**, 439–445, 2007.

[370] G A DeVries, M Brunnbauer, Y Hu, A M Jackson, B Long, B T Neltner, O Uzun, B H Wunsch and F Stellacci, 'Divalent metal nanoparticles', *Science*, **315**, 358–361, 2007.

[371] S-M Yang, S-H Kim, J-M Lim and G-R Yi, 'Synthesis and assembly of structured colloidal particles', *J Mater Chem*, **18**, 2177–2190, 2008.

[372] L Hong, A Cacciuto, E Luijten and S Granick, 'Clusters of amphiphilic colloidal spheres', *Langmuir*, **24**, 621–625, 2008.

[373] D J Kraft, W S Vlug, C M van Kats, A van Blaaderen, A Imhof and W K Kegel, 'Self-assembly of colloids with liquid protrusions', *J Am Chem Soc*, **131**, 1182–1186, 2009.

[374] D J Kraft, J Groenewold and W K Kegel, 'Colloidal molecules with well-controlled bond angles', *Soft Matter*, **5**, 3823–3826, 2009.

[375] L Wang, L Xia, G Li, S Ravaine and X S Zhao, 'Patterning the surface of colloidal microspheres and fabrication of nonspherical particles', *Angew Chem, Int Ed*, **47**, 4725–4728, 2008.

[376] Z Mao, H Xu and D Wang, 'Molecular mimetic self-assembly of colloidal particles', *Adv Funct Mater*, **20**, 1053–1074, 2010.

[377] S Sacanna, W T M Irvine, P M Chaikin and D J Pine, 'Lock and key colloids', *Nature*, **464**, 575–578, 2010.

[378] M F Hagan and D Chandler, 'Dynamic pathways for viral capsid assembly', *Biophys J*, **91**, 42–54, 2006.

[379] G Villar, A W Wilber, A J Williamson, P Thiara, J P K Doye, A A Louis, M N Jochum, A C F Lewis and E D Levy, 'Self-assembly and evolution of homomeric protein complexes', *Phys Rev Lett*, **102**, 118106, 2009.

[380] A W Wilber, J P K Doye, A A Louis and A C F Lewis, 'Monodisperse self-assembly in a model with protein-like interactions', *J Chem Phys*, **131**, 175102, 2009.

[381] R P Sear, 'Phase behavior of a simple model of globular proteins', *J Chem Phys*, **111**, 4800–4806, 1999.

[382] N Kern and D Frenkel, 'Fluid-fluid coexistence in colloidal systems with short-ranged strongly directional attraction', *J Chem Phys*, **118**, 9882–9889, 2003.

[383] Z Zhang, A S Keys, T Chen and S C Glotzer, 'Self-assembly of patchy particles into diamond structures through molecular mimicry', *Langmuir*, **21**, 11547–11551, 2005.

[384] E Bianchi, J Largo, P Tartaglia, E Zaccarelli and F Sciortino, 'Phase diagram of patchy colloids: Towards empty liquids', *Phys Rev Lett*, **97**, 168301, 2006.

[385] A W Wilber, J P K Doye, A A Louis, E G Noya, M A Miller and P Wong, 'Reversible self-assembly of patchy particles into monodisperse icosahedral clusters', *J Chem Phys*, **127**, 085106, 2007.

[386] J P K Doye, A A Louis, I C Lin, L R Allen, E G Noya, A W Wilber, H C Kok and R Lyus, 'Controlling crystallization and its absence: Proteins, colloids and patchy models', *Phys Chem Chem Phys*, **9**, 2197–2205, 2007.

[387] E G Noya, C Vega, J P K Doye and A A Louis, 'Phase diagram of model anisotropic particles with octahedral symmetry', *J Chem Phys*, **127**, 054501, 2007.

[388] F Sciortino, A Giacometti and G Pastore, 'Phase diagram of Janus particles', *Phys Rev Lett*, **103**, 237801, 2009.

[389] F Romano, E Sanz and F Sciortino, 'Role of the range in the fluid–crystal coexistence for a patchy particle model', *J Phys Chem B*, **113**, 15133, 2009.

[390] F Romano, E Sanz and F Sciortino, 'Phase diagram of a tetrahedral patchy particle model for different interaction ranges', *J Chem Phys*, **132**, 184501, 2010.

[391] E G Noya, C Vega, J P K Doye and A A Louis, 'The stability of a crystal with diamond structure for patchy particles with tetrahedral symmetry', *J Chem Phys*, **132**, 234511, 2010.

[392] F Sciortino, A Giacometti and G Pastore, 'A numerical study of one-patch colloidal particles: From square-well to Janus', *Phys Chem Chem Phys*, **12**, 11869–11877, 2010.

[393] W L Miller and A Cacciuto, 'Hierarchical self-assembly of asymmetric amphiphatic spherical colloidal particles', *Phys Rev E*, **80**, 021404, 2009.

[394] S Whitelam and S A F Bon, 'Self-assembly of amphiphilic peanut-shaped nanoparticles', *J Chem Phys*, **132**, 074901, 2010.

[395] T Narayanan and A Kumar, 'Reentrant phase transitions in multicomponent liquid mixtures', *Phys Rep*, **249**, 135–218, 1994.

[396] S Whitelam and P L Geissler, 'Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles', *J Chem Phys*, **127**, 154101, 2007.

[397] L A Davies, G Jackson and L F Rull, 'Simulation study of the link between molecular association and reentrant miscibility for a mixture of molecules with directional interactions', *Phys Rev Lett*, **82**, 5285–5288, 1999.

[398] G Jackson, 'Theory of closed-loop liquid-liquid immiscibility in mixtures of molecules with directional attractive forces', *Mol Phys*, **72**, 1365–1385, 1991.

[399] L A Davies, G Jackson and L F Rull, 'Closed-loop phase equilibria of a symmetrical associating mixture of square-well molecules examined by Gibbs ensemble Monte Carlo simulation', *Phys Rev E*, **61**, 2245–2256, 2000.

[400] F F Nord, M Bier and S N Timasheff, 'Investigations on proteins and polymers. IV. Critical phenomena in polyvinyl alcohol-acetate copolymer solutions', *J Am Chem Soc*, **73**, 289–293, 1951.

[401] M Corti, C Minero and V Degiorgio, 'Cloud point transition in nonionic micellar solutions', *J Phys Chem*, **88**, 309–317, 1984.

[402] P E Cladis, 'A one hundred year perspective of the reentrant nematic phase', *Mol Cryst Liq Cryst*, **165**, 85–121, 1988.

[403] Y Levin and M E Fisher, 'Criticality in the hard-sphere ionic fluid', *Physica A*, **225**, 164–220, 1996.

[404] M E Fisher and Y Levin, 'Criticality in ionic fluids: Debye-Hückel theory, Bjerrum, and beyond', *Phys Rev Lett*, **71**, 3826–3829, 1993.

[405] E Sanz, C Vega, J L F Abascal and L G MacDowell, 'Phase diagram of water from computer simulation', *Phys Rev Lett*, **92**, 255701, 2004.

[406] M S Wertheim, 'Fluids with highly directional attractive forces. I. Statistical thermodynamics', *J Stat Phys*, **35**, 19–34, 1984.

[407] M S Wertheim, 'Fluids with highly directional attractive forces. II. Thermodynamic perturbation theory and integral equations', *J Stat Phys*, **35**, 35–47, 1984.

[408] G Villar, J P K Doye, A A Louis, A W Wilber and A J Williamson, 'Analysis and simulation of the self-assembly of tetrameric protein complexes', in preparation.

[409] R Laghaei, A E Nasrabad and B C Eu, 'Excluded volume in the generic van der Waals equation of state and the self-diffusion coefficient of the Lennard-Jones fluid', *J Chem Phys*, **124**, 154502, 2006.

[410] P H V Konynenburg and R L Scott, 'Critical lines and phase equilibria in binary van der Waals mixtures', *Philos T R Soc A*, **298**, 495–540, 1980.

[411] S Torquato, T M Truskett and P G Debenedetti, 'Is random close packing of spheres well defined?', *Phys Rev Lett*, **84**, 2064–2067, 2000.

[412] D Zwillinger, editor, *CRC standard mathematical tables and formulae*. CRC Press, Boca Raton, 31st edition, 2003.

[413] J P Sethna, *Statistical mechanics: Entropy, order parameters, and complexity*. Oxford University Press, Oxford, 2006.

[414] M Hillert, *Phase equilibria, phase diagrams and phase transformations: Their thermodynamic basis*. Cambridge University Press, New York, 2nd edition, 2008.

[415] A Giacometti, F Lado, J Largo, G Pastore and F Sciortino, 'Effects of patch size and number within a simple model of patchy colloids', *J Chem Phys*, **132**, 174110, 2010.

[416] S Torquato and Y Jiao, 'Dense packings of the Platonic and Archimedean solids', *Nature*, **460**, 876–879, 2009.

[417] J P K Doye and W C K Poon, 'Protein crystallization in vivo', *Curr Opin Colloid In*, **11**, 40–46, 2006.

[418] V J Anderson and H N W Lekkerkerker, 'Insights into phase transition kinetics from colloid science', *Nature*, **416**, 811–815, 2002.

[419] A J Williamson, *Methods, rules and limits of successful self-assembly*. Thesis, University of Oxford, Oxford, 2011.

[420] W Steurer, 'Twenty years of structure research on quasicrystals. Part I. Pentagonal, octagonal, decagonal and dodecagonal quasicrystals', *Z Kristallogr*, **219**, 391–446, 2004.

[421] W Steurer, 'Why are quasicrystals quasiperiodic?', *Chem Soc Rev*, **41**, 6719–6729, 2012.

[422] D Shechtman, I Blech, D Gratias and J W Cahn, 'Metallic phase with long-range orientational order and no translational symmetry', *Phys Rev Lett*, **53**, 1951–1953, 1984.

[423] X Zeng, 'Liquid quasicrystals', *Curr Opin Colloid In*, **9**, 384–389, 2005.

[424] S Fischer, A Exner, K Zielske, J Perlich, S Deloudi, W Steurer, P Lindner and S Förster, 'Colloidal quasicrystals with 12-fold and 18-fold diffraction symmetry', *Proc Natl Acad Sci USA*, **108**, 1810–1814, 2011.

[425] T Dotera, 'Quasicrystals in soft matter', *Isr J Chem*, **51**, 1197–1205, 2011.

[426] M Widom, K J Strandburg and R H Swendsen, 'Quasicrystal equilibrium state', *Phys Rev Lett*, **58**, 706–709, 1987.

[427] P W Leung, C L Henley and G V Chester, 'Dodecagonal order in a two-dimensional Lennard-Jones system', *Phys Rev B*, **39**, 446–458, 1989.

[428] M Dzugutov, 'Formation of a dodecagonal quasicrystalline phase in a simple monatomic liquid', *Phys Rev Lett*, **70**, 2924–2927, 1993.

[429] A Skibinsky, S V Buldyrev, A Scala, S Havlin and H E Stanley, 'Quasicrystals in a monodisperse system', *Phys Rev E*, **60**, 2664–2669, 1999.

[430] M Engel and H-R Trebin, 'Self-assembly of monatomic complex crystals and quasicrystals with a double-well interaction potential', *Phys Rev Lett*, **98**, 225505, 2007.

[431] A S Keys and S C Glotzer, 'How do quasicrystals grow?', *Phys Rev Lett*, **99**, 235503, 2007.

[432] A Haji-Akbari, M Engel, A S Keys, X Zheng, R G Petschek, P Palffy-Muhoray and S C Glotzer, 'Dis-

ordered, quasicrystalline and crystalline phases of densely packed tetrahedra', *Nature*, **462**, 773–777, 2009.

[433] J C Johnston, S Phippen and V Molinero, 'A single-component silicon quasicrystal', *J Phys Chem Lett*, **2**, 384–388, 2011.

[434] J C Johnston, N Kastelowitz and V Molinero, 'Liquid to quasicrystal transition in bilayer water', *J Chem Phys*, **133**, 154516, 2010.

[435] C R Iacovella, A S Keys and S C Glotzer, 'Self-assembly of soft-matter quasicrystals and their approximants', *Proc Natl Acad Sci USA*, **108**, 20935–20940, 2011.

[436] A Haji-Akbari, M Engel and S C Glotzer, 'Degenerate quasicrystal of hard triangular bipyramids', *Phys Rev Lett*, **107**, 215702, 2011.

[437] C Zhi-Wei and F Xiu-Jun, 'Quasicrystalline phase with eighteen-fold diffraction symmetry in molecular dynamics simulations', *Chinese Phys Lett*, **29**, 050204, 2012.

[438] T Dotera, 'Toward the discovery of new soft quasicrystals: From a numerical study viewpoint', *J Polym Sci: Pol Phys*, **50**, 155–167, 2012.

[439] M N van der Linden, J P K Doye and A A Louis, 'Formation of dodecagonal quasicrystals in two-dimensional systems of patchy particles', *J Chem Phys*, **136**, 054904, 2012.

[440] A-P Tsai, '"Back to the Future"–An account discovery of stable quasicrystals', *Acc Chem Res*, **36**, 31–38, 2003.

[441] D Frenkel and A J C Ladd, 'New Monte Carlo method to compute the free energy of arbitrary solids. Application to the fcc and hcp phases of hard spheres', *J Chem Phys*, **81**, 3188–3193, 1984.

[442] A Haji-Akbari, M Engel and S C Glotzer, 'Phase diagram of hard tetrahedra', *J Chem Phys*, **135**, 194101, 2011.

[443] A Kiselev, M Engel and H-R Trebin, 'Confirmation of the random tiling hypothesis for a decagonal quasicrystal', *Phys Rev Lett*, **109**, 225502, 2012.

[444] G Doppelbauer, E Bianchi and G Kahl, 'Self-assembly scenarios of patchy colloidal particles in two dimensions', *J Phys: Condens Mat*, **22**, 104105, 2010.

[445] A J Williamson, A W Wilber, J P K Doye and A A Louis, 'Templated self-assembly of patchy particles', *Soft Matter*, **7**, 3423–3431, 2011.

[446] M Antlanger, G Doppelbauer and G Kahl, 'On the stability of Archimedean tilings formed by patchy particles', *J Phys: Condens Mat*, **23**, 404206, 2011.

[447] G Doppelbauer, E G Noya, E Bianchi and G Kahl, 'Self-assembly scenarios of patchy colloidal particles', *Soft Matter*, **8**, 7768–7772, 2012.

[448] G Doppelbauer, E G Noya, E Bianchi and G Kahl, 'Competing ordered structures formed by particles with a regular tetrahedral patch decoration', *J Phys: Condens Mat*, **24**, 284124, 2012.

[449] E Duguet, A Desert, A Perro and S Ravaine, 'Design and elaboration of colloidal molecules: An overview', *Chem Soc Rev*, **40**, 941–960, 2011.

[450] Y Wang, Y Wang, D R Breed, V N Manoharan, L Feng, A D Hollingsworth, M Weck and D J Pine, 'Colloids with valence and specific directional bonding', *Nature*, **491**, 51–55, 2012.

[451] Q Chen, S C Bae and S Granick, 'Directed self-assembly of a colloidal kagome lattice', *Nature*, **469**, 381–384, 2011.

[452] Z Zhang and S C Glotzer, 'Self-assembly of patchy particles', *Nano Lett*, **4**, 1407–1413, 2004.

[453] F Romano and F Sciortino, 'Two dimensional assembly of triblock Janus particles into crystal phases in the two bond per patch limit', *Soft Matter*, **7**, 5799–5804, 2011.

[454] E Bianchi, R Blaak and C N Likos, 'Patchy colloids: State of the art and perspectives', *Phys Chem Chem Phys*, **13**, 6397–6410, 2011.

[455] F C Frank and J S Kasper, 'Complex alloy structures regarded as sphere packings. I. Definitions and basic principles', *Acta Crystallogr*, **11**, 184–190, 1958.

[456] F C Frank and J S Kasper, 'Complex alloy structures regarded as sphere packings. II. Analysis and classification of representative structures', *Acta Crystallogr*, **12**, 483–499, 1959.

[457] M Oxborrow and C L Henley, 'Random square-triangle tilings: A model for twelvefold-symmetric quasicrystals', *Phys Rev B*, **48**, 6966–6998, 1993.

[458] D A Kofke, 'Gibbs–Duhem integration: A new method for direct evaluation of phase coexistence by molecular simulation', *Mol Phys*, **78**, 1331–1336, 1993.

[459] D A Kofke, 'Direct evaluation of phase coexistence by molecular simulation via integration along the saturation line', *J Chem Phys*, **98**, 4149–4162, 1993.

[460] M H J Hagen and D Frenkel, 'Determination of phase diagrams for the hard-core attractive Yukawa system', *J Chem Phys*, **101**, 4093–4097, 1994.

[461] G A Vliegenthart, J F M Lodge and H N W Lekkerkerker, 'Strong, weak and metastable liquids: Structural and dynamical aspects of the liquid state', *Physica A*, **263**, 378–388, 1999.

[462] J P K Doye and D J Wales, 'The structure and stability of atomic liquids: From clusters to bulk', *Science*, **271**, 484–487, 1996.

[463] H Yan, S H Park, G Finkelstein, J H Reif and T H LaBean, 'DNA-templated self-assembly of protein arrays and highly conductive nanowires', *Science*, **301**, 1882–1884, 2003.

[464] Y He, Y Chen, H Liu, A E Ribbe and C Mao, 'Self-assembly of hexagonal DNA two-dimensional (2D) arrays', *J Am Chem Soc*, **127**, 12202–12203, 2005.

[465] Y He, Y Tian, A E Ribbe and C Mao, 'Highly connected two-dimensional crystals of DNA six-point-stars', *J Am Chem Soc*, **128**, 15978–15979, 2006.

[466] C Zhang, M Su, Y He, X Zhao, P-a Fang, A E Ribbe, W Jiang and C Mao, 'Conformational flexibility facilitates self-assembly of complex DNA nanostructures', *Proc Natl Acad Sci USA*, **105**, 10665–10669, 2008.

[467] J L Katz, 'Homogeneous nucleation theory and experiment: A survey', *Pure & Appl Chem*, **64**, 1661–1666, 1992.

[468] E R Buckle, 'Studies on the freezing of pure liquids. II. The kinetics of homogeneous nucleation in supercooled liquids', *Proc R Soc Lond A*, **261**, 189–196, 1961.

[469] E W Weisstein, 'Euler angles', accessed 06-02-2013, URL http://mathworld.wolfram.com/EulerAngles.html.

[470] D J Evans, 'On the representation of orientation space', *Mol Phys*, **34**, 317–325, 1977.

[471] E W Weisstein, 'Quaternion', accessed 06-02-2013, URL http://mathworld.wolfram.com/Quaternion.html.

[472] F J Vesely, 'Angular Monte Carlo integration using quaternion parameters: A spherical reference potential for CCl_4', *J Comput Phys*, **47**, 291–296, 1982.

[473] L Verlet, 'Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules', *Phys Rev*, **159**, 98–103, 1967.

[474] R W Hockney and J W Eastwood, *Computer simulation using particles*. A. Hilger, Bristol, 1988.

[475] W C Swope, H C Andersen, P H Berens and K R Wilson, 'A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters', *J Chem Phys*, **76**, 637–649, 1982.

[476] B J Leimkuhler and R D Skeel, 'Symplectic numerical integrators in constrained Hamiltonian systems', *J Comput Phys*, **112**, 117–125, 1994.

[477] D J Evans and S Murad, 'Singularity free algorithm for molecular dynamics simulation of rigid polyatomics', *Mol Phys*, **34**, 327–331, 1977.

[478] D Rapaport, 'Molecular dynamics simulation using quaternions', *J Comput Phys*, **60**, 306–314, 1985.

[479] T S van Erp, D Moroni and P G Bolhuis, 'A novel path sampling method for the calculation of rate constants', *J Chem Phys*, **118**, 7762–7774, 2003.

[480] T H Cormen, C E Leiserson, R L Rivest and C Stein, *Introduction to algorithms*. MIT Press, London, 2nd edition, 2001.

[481] P Atkins and R Friedman, *Molecular quantum mechanics*. Oxford University Press, Oxford, 4th edition, 2005.

Appendices

The following chapters contain information that either does not fit neatly into the main framework of the thesis (in particular, appendices A and B present research that was undertaken during my DPhil project, but which does not relate to the thesis title) or presents certain details or background information that supplements the material presented.

In appendix A, we study the phase behaviour of a system which exhibits competition between self-assembly and phase separation using classical statistical thermodynamics. In appendix B, we study the thermodynamics of a system of patchy particles in which a quasicrystalline phase is thermodynamically stable. In appendix C, we derive the classical nucleation theory expression for the rate of nucleation. Finally, in appendix D, we look at certain aspects of simulation methods in more detail than in chapter 3.

A Re-entrant phase behaviour for systems with competition between phase separation and self-assembly

This appendix chapter presents work undertaken during the course of my DPhil project, but which does not fit into the main structure of the thesis. The work was undertaken in collaboration with Alexander Williamson in the Doye group. We performed the calculations together, sharing the work approximately equally. The contents of much of this appendix have been published,[1] although some aspects are slightly extended here.

A.1 Summary

In patchy particle systems where there is competition between the self-assembly of finite clusters and liquid-vapour phase separation, re-entrant phase behaviour can be observed, with the system passing from a monomeric vapour phase to a region of liquid-vapour phase coexistence and then to a vapour phase of clusters as the temperature is decreased at constant density. Here, we present a classical statistical mechanical approach to the determination of the complete phase diagram of such a system. We model the system as a Van der Waals fluid, but one where the monomers can assemble into monodisperse clusters that have no attractive interactions with any of the other species. The resulting phase diagrams show a clear region of re-entrance. However, for the most physically reasonable parameter values of the model, this behaviour is restricted to a certain range of density, with phase separation still persisting at high densities.

A.2 Introduction

Self-assembly is a process by which a material, be it molecular, colloidal or macroscopic, spontaneously forms a well-defined structured aggregate.[357] There are many examples of such self-assembling systems, be they biological[358] (*e.g.* virus capsids,[359] protein fibres[360] and motors) or synthetic (*e.g.* surfactant micelles[361] and supramolecular complexes[362]), although the latter are not yet able to replicate the exquisite control of self-assembly exhibited by biological systems. Indeed, some of the most remarkable synthetic self-assembling systems borrow from biology; for example, the use of DNA and RNA to form complex nanostructures[363] and devices.[364] From the materials perspective, the advantages of self-assembly are clear, potentially enabling new materials and devices to be fabricated, particularly those with features on the nanoscale.

For colloids and nanoparticles, there is a particular interest in developing particles that have 'patchy' anisotropic interactions,[365;366] with a view to extending the range of structures that can be formed by self-assembly. Experimental methods for synthesising such particles, although in their early stages, are advancing rapidly.[367–377] Similarly, there is an increasing amount of computational work on patchy particles, both as simple models for understanding biological self-assembly,[378–380] and to explore what might be possible with synthetic patchy particles and to learn the design principles for their successful self-assembly.[109;381–392]

One particular focus in the simulations of patchy particles has been the formation of finite aggregates, be they monodisperse[109;383;385] (similar to virus capsids) or polydisperse[388;392–394] (micelle-like structures). One intriguing result for these systems is that for certain parameter ranges, the competition between self-assembly and liquid-vapour phase separation can lead to 're-entrant' phase behaviour.[109;385;388;392] A phase transition is said to be re-entrant if it involves the transformation of a system from one state into a macroscopically similar (or identical) state via at least two phase transitions through the variation of a single thermodynamic

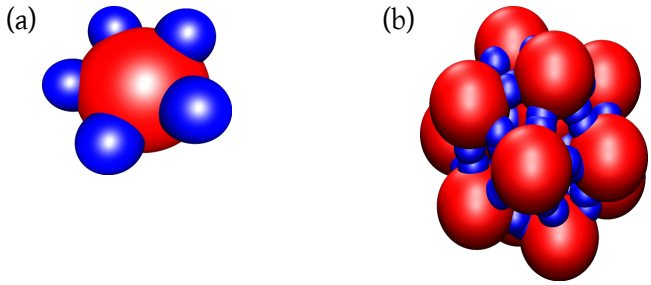


Figure A.1: The icosahedron-forming system: in (a), the monomer is depicted, and in (b), the formed icosahedral cluster is shown. Red spheres represent the Lennard-Jones site of each particle, and blue spheres represent the patches. The patches do not occupy any volume themselves.

parameter (such as the temperature).[395]

For a system that is designed to form monodisperse 12-particle icosahedra (figure A.1),[385] such re-entrant behaviour is illustrated in figure A.2. This figure shows the average

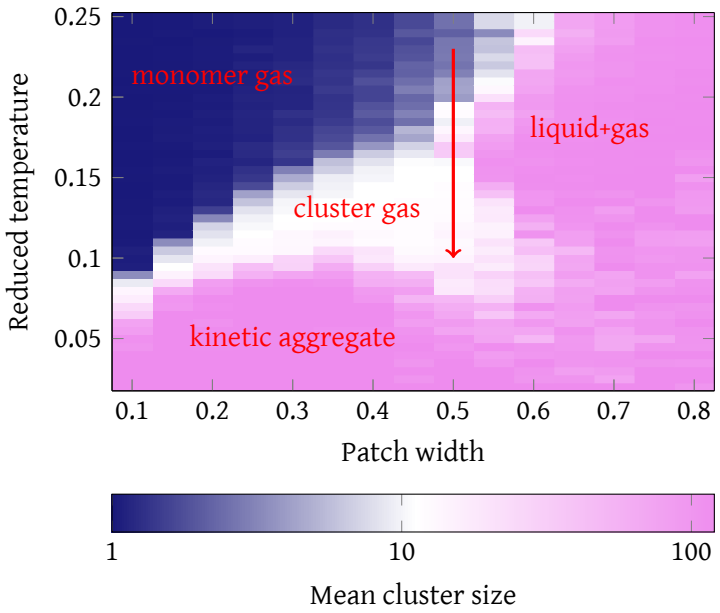


Figure A.2: The mean size of clusters formed as a function of the reduced temperature and the patch width for patchy particles designed to form 12-particle icosahedral clusters. These Monte Carlo simulations involved 120 particles simulated for 1.2×10^8 Monte Carlo steps at a constant number density of $\rho\sigma^3 = 0.15$ and used the virtual move Monte Carlo algorithm.[396] Results were averaged over 5 independent simulations. Decreasing the temperature at a patch width of approximately 0.5 radians involves two phase transitions; from a monomer gas through a liquid-gas coexistence region to a cluster gas.

size of clusters as a function of the temperature and the angular width of the patches at the end of dynamical simulations that initially started from a monomeric gas. Values close to 12 indicate the formation of icosahedra, whereas very large cluster sizes indicate either a kinetic aggregate or the formation of a liquid droplet within a background vapour (and thus correspond to a liquid-vapour coexistence region). At very broad patch widths, the patches are relatively non-specific and the behaviour tends to that of a Lennard-Jones fluid with phase separation between a vapour of monomers and a liquid of monomers. Conversely, at very narrow patch widths, the potential is so angularly dependent that the only thermodynamically stable phases are a vapour of monomers at higher temperatures and a vapour of clusters at lower temperatures (note that in these dynamical simulations, large aggregates form at low temperatures due to the slow kinetics, but this state is not a thermodynamically stable phase). At intermediate patch widths, we expect competition between self-assembly and phase separation to occur. For example, simulation results suggest that decreasing the temperature at a patch width of approximately 0.5 radians and at a constant density involves two phase transitions: first going from a *gas* of monomers to a *liquid-gas* mixture and then to a *gas* of icosahedral clusters. This re-entrance is driven by the lower energy of the cluster gas phase (due to the strong internal bonding within the icosahedra) compared to the liquid phase, which has a greater entropy.

Clear evidence of re-entrant phase behaviour has also been observed by Sciortino and co-workers in simulations of Janus particles with a single attractive patch. Certain parameter ranges lead to a competition between liquid-vapour phase separation and micelle formation.[388;392] In the phase diagrams that they computed, the standard behaviour of an ever-larger two-phase region in the T-ρ phase diagram as the temperature is decreased below the critical point is perturbed by the formation of micelles, shrinking the coexistence region and shifting it to higher density (figure A.3). However, what happens to this two-phase region at low temperature (does it end at a critical point, or persist to zero temperature?) is unclear,

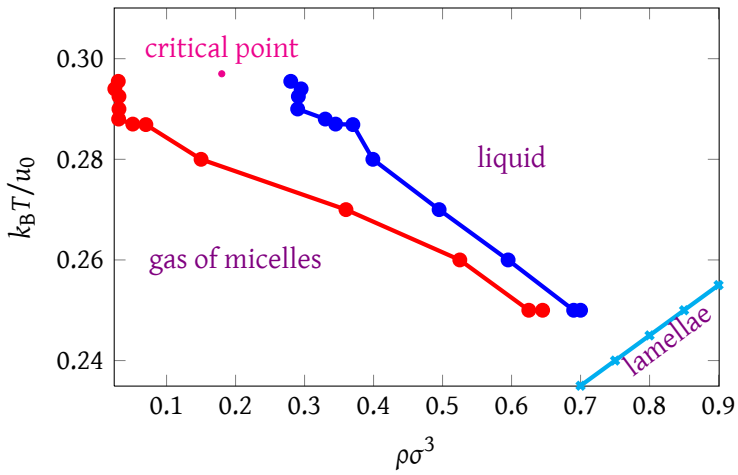


Figure A.3: Approximate phase diagram of a Janus particle system exhibiting re-entrant phase behaviour as observed in simulations by Sciortino and co-workers.[388] (This figure was redrawn from the data points given in reference 388.)

because the system instead formed a lamellar phase. In this work, our aim is to address such questions by calculating complete phase diagrams for the fluid using a simple theoretical model of a system that can both self-assemble into clusters and exhibit liquid-vapour phase separation.

Re-entrant phase behaviour is of course not limited to patchy particle systems. Indeed, the classic example of such behaviour is the nicotine-water mixture, which exhibits a closed loop in its temperature-composition phase diagram with both an upper and a lower critical solution temperature.[40] Reducing the temperature whilst keeping the mole fraction of nicotine constant first leads to phase separation, and then to remixing driven by nicotine-water association. Liquid-liquid immiscibility leading to such phase diagrams is often attributed to hydrogen bonding.[397–399] Complete mixing will occur if the temperature is sufficiently high so as to overcome the enthalpic incompatibility of mixing. However, at sufficiently low temperatures, miscibility results from hydrogen bonding between different components in the system. Similar considerations have been used to explain re-entrant phase behaviour seen in polymer aggregation in water,[400] in non-ionic micellar solutions[401] and in nematic re-entrant phases in liquid crystals.[402]

Re-entrant phase behaviour as a function of temperature can also be seen in other

systems, such as polymer aggregation.[400] In polymer aggregation of polyvinyl alcohol-acetate copolymer solutions, it is the competition between hydrophilic hydroxyl groups and hydrophobic acetyl groups and hydrocarbon chains that can account for the lower critical point. At low temperatures, hydration of the hydrophilic groups is a strong enthalpic force for solvation; as the temperature increases, hydration (as an exothermic process) becomes less pronounced by Le Chatelier's principle, and aggregation ensues. At higher temperatures still, it is naturally the entropy that drives polymer solvation.[400] Micelles can also be considered to be similar to polymers: hydrophilic interactions ensure separation at low temperatures, whilst above the lower critical solution temperature, attractive interactions between the aggregates separate the micellar solution into two isotropic micellar solutions.[401] Nematic re-entrant phases are similarly commonplace in the field of liquid crystals, where molecular and ionic association are commonly invoked to rationalise this behaviour.[402]

Although the systems of interest in this work are not two-component mixtures, unlike most systems where re-entrant phase behaviour has previously been observed, the behaviour can nevertheless in some ways be analogous to such systems. The similarities arise due to the mixing of two states of the same component (monomers and associated clusters), rather than two components proper. An example of re-entrant phase behaviour as a function of temperature in a system with a single effective density is ionic association.[403;404] When dielectric constants are large, phase separation is driven by the usual enthalpic incompatibility. With low dielectric constants, coulombic-style phase separation between a low density (neutral cluster) structured gas phase and a high density (free ion) liquid phase occurs due to the long-range ionic interactions between the ionic components, rather than between the ions and the solvent. This strong ionic association of ion pairs can be very significant compared to the thermal energy, and an equilibrium is established between neutral dipolar pairs and potentially larger clusters on the one and dissociated free ions on the other hand. The formation of ionic

clusters is somewhat similar to the cluster formation in the self-assembling patchy particle system that we focus upon, and so the re-entrant behaviour (in simple theories of these ionic systems, 'banana'-shaped coexistence curves are found[403]) may have similar origins.

Re-entrant behaviour as a function of variables other than the temperature is also relatively well-known. For example, it can be observed as a function of pressure in p-T phase diagrams of common water models such as TIP4P and SPC.[405] This is attributed to the low compressibility of ice phases compared to liquid water: water becomes denser at high pressures than the low-density ices. According to the Clapeyron equation, the slope of the p-T curve will depend on the change in entropy over the change in volume between the delimited phases;[40] since the entropy is always positive for melting, it is the density ordering changeover that allows the p-T curve to change slope and thus to exhibit re-entrant behaviour. Patchy particle systems, where bonds are typically similarly directional as hydrogen bonds are in water systems, and thus readily form open crystal structures, can also exhibit re-entrant phase behaviour in the p-T plane.[387;390] Nevertheless, although patchy particles can exhibit such re-entrance, this is not the kind of re-entrance phase behaviour that we consider here.

In this work, we investigate a theoretical model which can be used to rationalise our earlier simulation results of self-assembling patchy colloidal particles and one which yields a phase diagram reminiscent of the work of Sciortino and co-workers.[388;392] In particular, we are interested in constructing a phase diagram for single-component systems that can potentially form a vapour of monomers, a liquid of monomers and a gas of monodisperse clusters. Here, these clusters are for the most part chosen to be 12-particle icosahedra,[385] but the approach can easily be generalised to other clusters, as in subsection A.4.4. We have chosen to model the chemical equilibrium between clusters and monomers with explicit molecular partition functions, and to treat the vapour-liquid equilibrium as a familiar Van der Waals fluid; this is a more straightforward approach compared to using Wertheim theory,[406;407] which has

already been used in the study of patchy particles.[384] We first look at just the monomer-cluster equilibrium in section A.3, and then determine the necessary changes to the pure Van der Waals fluid model in order to be able to form the three-state 'self-assembling Van der Waals fluid' and calculate appropriate phase diagrams for this system in section A.4.

A.3 Monomer-cluster equilibrium

A.3.1 Partition functions

We first construct the molecular partition function of an icosahedral cluster in a manner similar to that done previously for a simple model of tetrameric protein complexes.[379;408]

The total molecular partition function of the icosahedral cluster is given by

$$q_{\mathrm{clus}} = q_{\mathrm{tr}} q_{\mathrm{rot}} q_{\mathrm{vib}} \exp\left[-E_{\mathrm{c}}/k_{\mathrm{B}}T\right], \tag{A.1}$$

where E_{c} is the ground state energy of the cluster. As we are comparing results to classical simulations, and since we are modelling colloidal particles, it is reasonable to use classical expressions for the various partition functions. The translational partition function is given by

$$\begin{aligned} q_{\mathrm{tr}} &= \frac{V - BN_{\mathrm{clus}} - bN_{\mathrm{mon}}}{\Lambda^3} \\ &= \frac{48\sqrt{6}(\pi m k_{\mathrm{B}}T)^{3/2}}{h^3}(V - BN_{\mathrm{clus}} - bN_{\mathrm{mon}}), \end{aligned} \tag{A.2}$$

where Λ is the de Broglie thermal wavelength, the mass of the icosahedron is $12m$ (with m being the mass of a monomer), and N_{mon} and N_{clus} refer to the number of monomers and the number of clusters in the system, respectively. For the volume term, we choose to use the free volume approximation,[403] which is one generalisation of the Van der Waals expression to multiple components. The volume available to clusters is thus $V - BN_{\mathrm{clus}} - bN_{\mathrm{mon}}$, taking into account the reduction in the total volume due to the presence of other clusters and monomers. Here, B and b are free volume coefficients (rather than hard sphere second virial coefficients) for

clusters and monomers, respectively. The choice of values we assign to B and b is discussed in subsection A.3.2.

The moment of inertia about the principal axes of an icosahedron is given by $I_c = 80m\sigma^2/(5-\sqrt{5})$, where m is the monomer mass and σ is the diameter of the monomers. The symmetry number of an icosahedral cluster is 60. We can therefore calculate the rotational partition function as

$$\begin{aligned} q_{\text{rot}} &= \frac{\sqrt{I_c^3\pi}}{60} \times \left(\frac{8\pi^2 k_B T}{h^2}\right)^{3/2} \\ &= \frac{128\pi^{7/2}}{3}\sqrt{1+\frac{2}{\sqrt{5}}}\left(\frac{mk_B T\sigma^2}{h^2}\right)^{3/2}. \end{aligned} \tag{A.3}$$

There will be $(6\times 12-6)$ overall vibrational modes of the cluster.[1] For simplicity, we assume that $(3\times 12-6)$ probe the radial part of the potential, and 3×12 probe the orientational degrees of freedom, and that all the modes of each type have the same spring constant. Hence, we can write the vibrational partition function as

$$q_{\text{vib}} = \prod_i q_{\text{vib},i} \approx \left(\frac{k_B T}{\hbar\omega_{\text{rad}}}\right)^{30}\left(\frac{k_B T}{\hbar\omega_{\text{ang}}}\right)^{36}. \tag{A.4}$$

To obtain expressions for ω_{rad} and ω_{ang}, we need to assume some form for the interparticle potential. We choose the potential that we used in our simulations of the self-assembly of icosahedra.[109;385] The potential is based on the Lennard-Jones form, but where the attractive region is modulated by an angular term that measures how well patches point at each other, and is given by

$$U_{ij}(\boldsymbol{r}_{ij}, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j) = \begin{cases} U^{\text{LJ}}(r_{ij}) & r_{ij} < \sigma, \\ U^{\text{LJ}}(r_{ij}) V^{\text{ang}}(\hat{\boldsymbol{r}}_{ij}, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j) & \sigma \le r_{ij}, \end{cases} \tag{A.5}$$

where $\boldsymbol{r}_{ij}$ is the interparticle vector connecting the centres of the two particles i and j, r_{ij} is its

[1] This can be deduced by noting that each of the 12 monomers in the cluster has three translational and three rotational degrees of freedom. This gives us 6×12 overall degrees of freedom; however, three translational and three rotational degrees of freedom associated with the cluster as a whole must be removed, leading to the expression given.

magnitude, $\boldsymbol{\Omega}_i$ and $\boldsymbol{\Omega}_j$ are the orientations of the particles i and j, the Lennard-Jones potential is

$$U^{\mathrm{LJ}}(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \tag{A.6}$$

and

$$U^{\mathrm{ang}}(\hat{\boldsymbol{r}}_{ij}, \boldsymbol{\Omega}_i, \boldsymbol{\Omega}_j) = \exp\left[\frac{-\theta^2_{k_{\min}ij}}{2\,\sigma^2_{\mathrm{pw}}}\right]\exp\left[\frac{-\theta^2_{l_{\min}ji}}{2\,\sigma^2_{\mathrm{pw}}}\right], \tag{A.7}$$

where σ_{pw} is the *patch width* (we measure σ_{pw} in radians), θ_{kij} is the angle between the patch vector of patch k on particle i and the interparticle vector $\boldsymbol{r}_{ij}$, and $k_{\min}$ is the patch that minimises the magnitude of this angle. Hence, two particles interact only through a single pair of patches.

We assume that ω_{rad} and ω_{ang} are simply related to the second derivative of the radial and angular parts of the above anisotropic modified Lennard-Jones potential.[2] Namely, they are given by

$$\omega^2_{\mathrm{ang}} = -\frac{\varepsilon}{I_{\mathrm{mon}}}\left(\frac{\partial^2 U^{\mathrm{ang}}}{\partial\theta^2_{kij}}\right)_{\substack{\theta_{kij}\\ \theta_{lji}}\}=0} = \frac{\varepsilon}{I_{\mathrm{mon}}\sigma^2_{\mathrm{pw}}}, \tag{A.8}$$

where I_{mon} is the moment of inertia of the monomer (given by $I_{\mathrm{mon}} = (2/5)m\sigma^2$ for a spherical monomer), and

$$\omega^2_{\mathrm{rad}} = \frac{2}{m}\left(\frac{\partial^2 U^{\mathrm{LJ}}}{\partial r^2}\right)_{r=2^{1/6}\sigma} = 72\times 2^{2/3}\times\frac{\varepsilon}{m\sigma^2}. \tag{A.9}$$

Although these approximations appear somewhat crude, this level of description is reasonable given some of the other approximations used elsewhere. Furthermore, using the exact form of the vibrational partition function would be unlikely to have a significant effect on the overall phase behaviour.

Finally, the Boltzmann factor in equation (A.1) reflects the additional energy obtained as a result of clustering – that is to say, the energy of bonding between the patches of the monomers when arranged in an icosahedron. Ignoring next-neighbour interactions, the ground state energy of the icosahedron is $E_{\mathrm{c}} = -30\varepsilon$, since each of the 12 monomers in an icosahedron

[2] The angular frequencies are taken to be related to the harmonic spring constant in the standard way, $\omega = \sqrt{k/\mu}$, where μ is the relevant effective mass associated with the motion.

is bonded to five other monomers with bond energy ε.

Overall, the icosahedral cluster molecular partition function is given by

$$q_{\text{clus}} = \left[\frac{k_{\text{B}}^{69} m^{36} T^{69} \sigma^{69} \sigma_{\text{pw}}^{36} \exp\left[30\varepsilon/k_{\text{B}}T\right]}{h^{72}\varepsilon^{33}}\right] \times \left[V - BN_{\text{clus}} - bN_{\text{mon}}\right] \times 9.4 \times 10^{20}. \quad \text{(A.10)}$$

The monomers have both rotational and translational degrees of freedom, but not generic attractions at this stage. Therefore, the molecular partition function of the monomer is given by $q_{\text{mon}} = q_{\text{tr, mon}} \times q_{\text{rot, mon}}$, where

$$q_{\text{tr, mon}} = \frac{V - BN_{\text{clus}} - bN_{\text{mon}}}{\Lambda^3} \quad \text{(A.11)}$$

and

$$q_{\text{rot, mon}} = \frac{16\sqrt{2}\pi^{7/2}\left(I_{\text{mon}}k_{\text{B}}T\right)^{3/2}}{sh^3} = \frac{64\pi^{7/2}\sigma^3(mk_{\text{B}}T)^{3/2}}{25\sqrt{5}h^3}. \quad \text{(A.12)}$$

Note that the symmetry number s of a monomer is 5, since there are five identical patches on each monomer, and the same volume term occurs as in the cluster translational partition function because we are using the free volume approximation.[403] The overall monomer molecular partition function is therefore given by

$$q_{\text{mon}} = \frac{128\sqrt{\frac{2}{5}}\pi^5 k_{\text{B}}^3 m^3 \sigma^3 T^3}{25h^6} \times \left[V - BN_{\text{clus}} - bN_{\text{mon}}\right]. \quad \text{(A.13)}$$

A.3.2 Estimation of free volume coefficients

We choose to use the free volume approximation[403] summation to take into account the effects of excluded volume. Although there are more accurate functions which describe such effects better,[409] we deliberately wish to keep the partition functions as simple as possible, as we are interested in the *generic* properties and phase behaviour resulting from the competition between self-assembly and phase separation, rather than the detailed behaviour of a specific system. Moreover, we want a form consistent with the Van der Waals description of the fluid, and one that works reasonably well both at high and, to a degree, at low system densities.

Consider again the translational partition function, given by

$$q_{\mathrm{tr}} = \frac{48\sqrt{6}(\pi m k_{\mathrm{B}} T)^{3/2}}{h^3}(V - B_{\mathrm{clus}}N_{\mathrm{clus}} - B_{\mathrm{mon}}N_{\mathrm{mon}}), \tag{A.14}$$

where the term $V - B_{\mathrm{clus}}N_{\mathrm{clus}} - B_{\mathrm{mon}}N_{\mathrm{mon}}$ is the volume available to clusters, taking into account the reduction in the total volume due to the presence of other clusters and monomers, where B_{clus} and B_{mon} are new designations of the effective hard-sphere second virial coefficients for clusters and monomers, respectively. We should note a number of features concerning this expression. Firstly, we have chosen the form of this term to be consistent with that used for the Van der Waals fluid, which will be important in section A.4 when we consider a modified Van der Waals fluid that can also self-assemble into clusters. Hence, we use a term $B_{\mathrm{clus}}N_{\mathrm{clus}}$ to take into account the cluster-cluster excluded volume interactions. Secondly, we could match the B_{clus} (and B_{mon}) values to the exact hard-sphere second virial coefficients; however, this choice only works well at low densities where higher-order virial coefficients can be neglected and indeed leads to an unphysical range of possible densities. Instead, we choose B_{clus} and B_{mon} values to correspond to the reciprocals of the maximum densities of clusters and monomers, respectively. Thirdly, to take into account the reduction of the volume available to the clusters due to monomers, we use the free volume approximation,[403] and so a further term $B_{\mathrm{mon}}N_{\mathrm{mon}}$ is taken away from the volume, rather than a term $B_{\mathrm{mon\text{-}clus}}N_{\mathrm{mon}}$. The latter form would be appropriate if we were matching to exact hard-sphere second virial coefficients, and in this case $B_{\mathrm{mon\text{-}clus}}$ could be related to B_{clus} and B_{mon} by the expression $B_{\mathrm{mon\text{-}clus}} = [(B_{\mathrm{clus}}^{1/3} + B_{\mathrm{mon}}^{1/3})/2]^3$, which is sometimes known as the Lorentz expression.[410] However, this form again leads to unphysical behaviour at higher densities, where it overestimates the volume excluded. We refer to B_{clus} and B_{mon} as B and b, respectively, throughout the remainder of the chapter.

Within the free volume approximation,[403] the maximum density in the system will equal the reciprocal of the free volume coefficient b, $\rho_{\mathrm{max}} = 1/b$. We choose to set the maximum

density achievable in the system to $1\,\sigma^{-3}$, and hence $b = \sigma^3$. This choice is reasonable both in terms of the comparison of the maximum packing fraction with that for random close packing, and of the position of the Van der Waals critical point with that of the Lennard-Jones fluid. If we consider the maximum density of the system to equal the maximum density of a random close-packed system, we can establish the packing fraction

$$\phi = \frac{N_{\mathrm{mon}} V_{\mathrm{mon}}}{V} = \rho \times \frac{4\pi}{3}\left(\frac{\sigma}{2}\right)^3 = \frac{\rho\pi\sigma^3}{6}, \tag{A.15}$$

which, using $\phi_{\mathrm{RCP}} \approx 0.64$,[411] gives

$$b = \frac{1}{\rho_{\max}} = \frac{\pi\sigma^3}{6\phi_{\mathrm{RCP}}} = 0.82\sigma^3 \approx \sigma^3. \tag{A.16}$$

This suggests that our choice for b is a reasonable approximation.

The relative values of B and b should be proportional to the cluster and monomer sizes. To determine the size of the cluster relative to that of the monomer, we assume that clusters are spherical objects, but such that we take their icosahedral character into account. The circumradius of an icosahedron[412] with edge length a is $r = (a/4)\sqrt{10 + 2\sqrt{5}} \approx 0.95\,a$; if the icosahedron is built of monomeric spheres of diameter σ, then $a = \sigma$. However, we also need to take into account the fact that these monomers protrude out of the icosahedron itself, and so another hard-core radius of these constituent spheres, $\sigma/2$, must be added to the circumradius of the icosahedron to estimate the overall cluster radius R, giving $R = 1.45\,\sigma$ (figure A.4).

The volume occupied by this sphere relative to the volume occupied by a single particle is $R^3/(\sigma/2)^3 = 24.4$. This is therefore a suitable measure of how much bigger the icosahedral cluster is compared to a single patchy particle, and suggests that an appropriate ratio of cluster and monomer coefficients is $B/b \approx 24$.[3] We consider this case most comprehensively, and also a

[3] Alternatively, instead of considering the cluster edge length to be at the changeover from an attractive to a repulsive Lennard-Jones potential energy ($a = \sigma$), we can consider the minimum of the potential to be the appropriate edge length, namely $a = 2^{1/6}\sigma$. In this case, $B/b \approx 31$ is a suitable ratio of cluster and monomer free volume coefficients. In practice, apart from the numerical values being appropriately changed, there is little qualitative difference in the resulting phase diagrams when this alternative ratio is used.

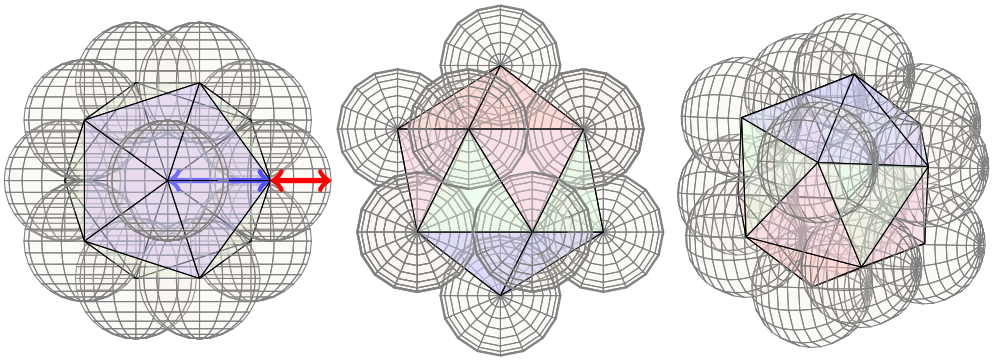


Figure A.4: In an icosahedron, provided the system is incompressible, we calculate the radius of the circumscribed sphere and then add a further radius of the particle, as shown in red in the figure. This gives a physically relevant value of B/b to be approximately 24, as used throughout most of this work.

few other values in order to characterise more fully the types of phase diagram that may arise. This ratio of B/b is specific to the icosahedral system described, and would take values different from the ones estimated here if the system were allowed, for example, to adopt a different cluster shape (for the same number of monomers) or to assemble into a different oligomeric state.

A.3.3 Clustering transition

We assume that the only relevant reaction in the system is $12\mathrm{A} \rightleftharpoons \mathrm{A}_{12}$, which is to say that there are no partially constructed clusters whatsoever. We justify this assumption from simulations, which show that the free energy of a 12-particle cluster is considerably lower than any intermediate size bigger than a monomer.[385] At equilibrium, the chemical potentials of the individual species, appropriately weighted by their stoichiometric coefficients, will be equal; in this case, $12\mu_{\mathrm{mon}} = \mu_{\mathrm{clus}}$, where $\mu_i = -k_{\mathrm{B}}T\ln(q_i/N_i)$. We choose to define the centre of this equilibrium as the point at which the probabilities of a particle being in a cluster or a monomer are equal, which imposes the condition $\rho_{\mathrm{mon}} = 12\rho_{\mathrm{clus}}$, where $\rho_{\mathrm{mon}} = N_{\mathrm{mon}}/V$ and $\rho_{\mathrm{clus}} = N_{\mathrm{clus}}/V$. We denote the temperature at which both conditions are fulfilled as the clustering temperature, T_{clus}.

A viable alternative definition of the equilibrium centre would be where the equilibrium constant is unity (subject to a reasonable definition of a 'standard' pressure); in practice, the difference in T_{clus} would be slight, because near to the midpoint of the equilibrium, the probability of being in the different states changes very rapidly with density and temperature.

We are not able to solve this system of equations analytically, and so we use numerical minimisation of the difference in chemical potential using the density constraint to obtain T_{clus} as a function of density, as plotted in figure A.5. The transformation between the two states is continuous (figure A.6) and the $T_{\text{clus}}(\rho)$ line is not a phase coexistence curve, but simply the centre of a chemical equilibrium that indicates where clusters and monomers are equally probable states.

We first consider the ideal limit ($b = 0$ and $B = 0$) and see from figure A.5 that at low temperatures, the cluster gas is more stable than the monomer gas. Although the clusters have a lower entropy as a consequence of the lower number of translational degrees of freedom compared to the monomer gas, at sufficiently low temperatures, the clusters will have a lower free energy, because their lower entropy is more than overcome by their lower energy due to bonding between the patches. Also, the equilibrium favours the cluster state as the density increases (*i.e.* T_{clus} increases with ρ), since the effective number of particles in the gas phase is lower.

Although the above are the most important basic features of the monomer-cluster equilibrium, there are a few more features that we should consider. For example, the model predicts that at high temperatures, the cluster gas becomes favoured (figure A.5(b)). This behaviour is an unphysical consequence of the harmonic approximation. For vibrations, the molecular partition function per degree of freedom rises with temperature as $q \propto T$, whereas for both rotations and translations, $q \propto T^{1/2}$. As the cluster state has more vibrational degrees of freedom than the monomer states, it will invariably have a lower free energy in our model

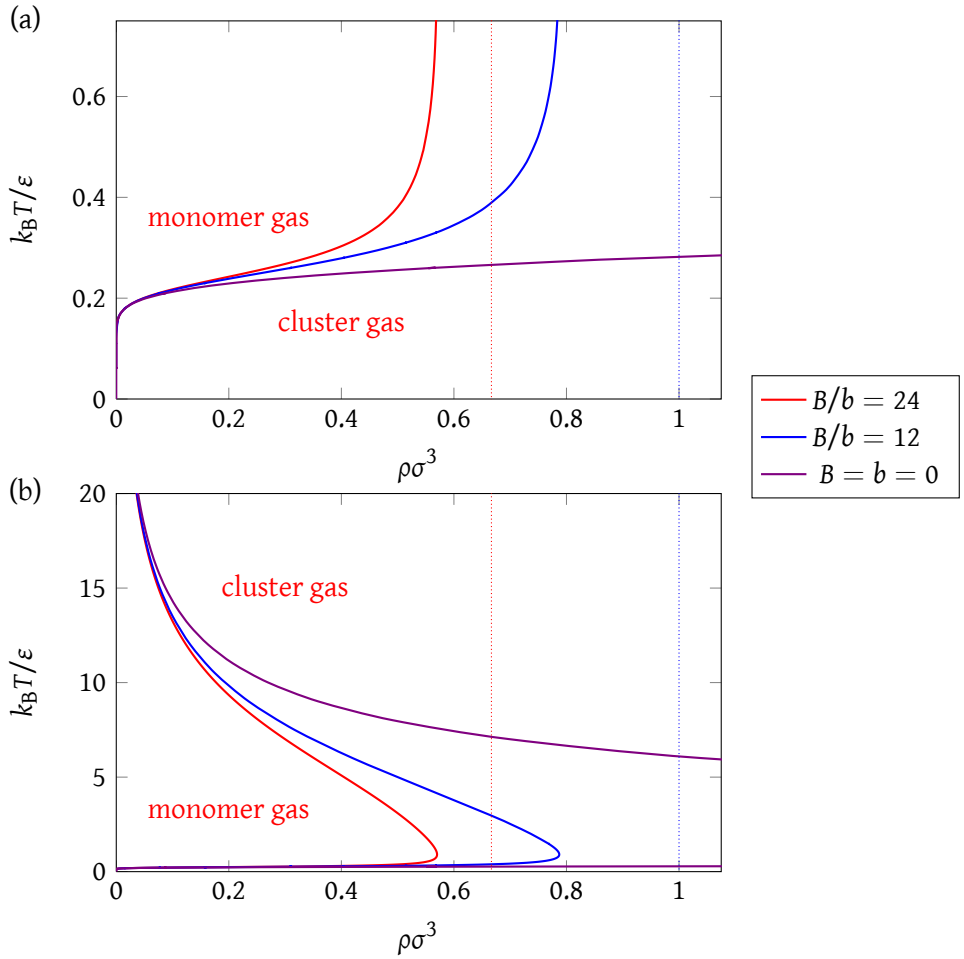


Figure A.5: $T_{\text{clus}}(\rho)$ for the transition from monomers to icosahedral clusters for different ratios of the free volume coefficients, B/b. For the non-ideal systems, $b/\sigma^3 = 1$, and dotted lines show $\rho^{\text{eql}}_{\text{max}}\sigma^3$, the maximum density at which particles can have an equal probability of being monomeric or in a cluster. Although both (a) and (b) depict the same systems, the temperature ranges differ. In (a), the section of the T-ρ plane that will be relevant when we consider the competition between phase separation and self-assembly is shown, whereas (b) illustrates the unphysical behaviour of our model that occurs at high temperatures. $\sigma_{\text{pw}} = 0.5$.

at sufficiently high temperatures. However, at such high temperatures ($k_BT \gg \varepsilon$), it is not appropriate to apply the harmonic approximation, as the thermal energy will be sufficient to dissociate all the bonds in the cluster. Nevertheless, this deficiency in the model is not a problem for the current study, as we are interested in reduced temperatures that are well below where this phenomenon occurs.

Figure A.5 also shows the effect of introducing excluded volume interactions on the

monomer-cluster equilibrium. There is no maximum density in principle for the ideal system ($b = 0, B = 0$), and clusters are always stable to the right of the clustering transition line. The first effect of the excluded volume is to limit the maximum density. For the $B/b = 24$ case, the maximum density of monomers is $\rho_{\mathrm{mon,\,max}}\sigma^3 = 1$, and the maximum particle density associated with the clusters is $12\rho_{\mathrm{clus,\,max}}\sigma^3 = 1/2$. For the $B/b = 12$ case, both the monomers and the clusters have a maximum particle density of 1.

At low densities, the excluded volume interactions have little effect and the $T_{\mathrm{clus}}(\rho)$ lines initially follow that for the ideal case. However, as the density increases, the lines are displaced upwards with respect to the ideal case because the volume term $V - BN_{\mathrm{clus}} - bN_{\mathrm{mon}}$ in the translational partition functions of both monomers and clusters decreases, and this destabilises the monomeric state more because it has a greater number of translational degrees of freedom.

A typical dependence of the fluid composition on the density is shown in figure A.6 for a system in which $B/b > 12$. The fraction of particles in clusters increases rapidly as the system passes through the clustering transition on increasing the density, and as $\rho = 12\rho_{\mathrm{clus,\,max}}$ is approached, this fraction is usually close to unity. However, when clusters exclude more volume than 12 monomers (which is the usual physical situation), *i.e.* when $B > 12b$, achieving densities beyond $12\rho_{\mathrm{clus,\,max}}$ is only possible if the monomeric state again becomes populated, because monomers pack more efficiently (for example, when $B/b = 24$, 12 monomers take up only half the volume of a single icosahedral cluster). Hence we see in figure A.6 that as the density increases further, the fraction of monomers increases, becoming the dominant species beyond $\rho_{\mathrm{max}}^{\mathrm{eql}}$, the maximum density for which particles can have an equal probability of being monomeric or in a cluster. This maximum density is set by the constraint that the available volume (and hence the translational partition functions) must always be non-negative. In general, $\rho_{\mathrm{max}}^{\mathrm{eql}} = 24/(B + 12b)$, where this expression is the simultaneous solution of both $V - BN_{\mathrm{clus}} - bN_{\mathrm{mon}} = 0$ and $12\rho_{\mathrm{clus}} = \rho_{\mathrm{mon}}$. For $B/b = 24$, $\rho_{\mathrm{max}}^{\mathrm{eql}}\sigma^3 = 2/3$; for $B/b = 12$,

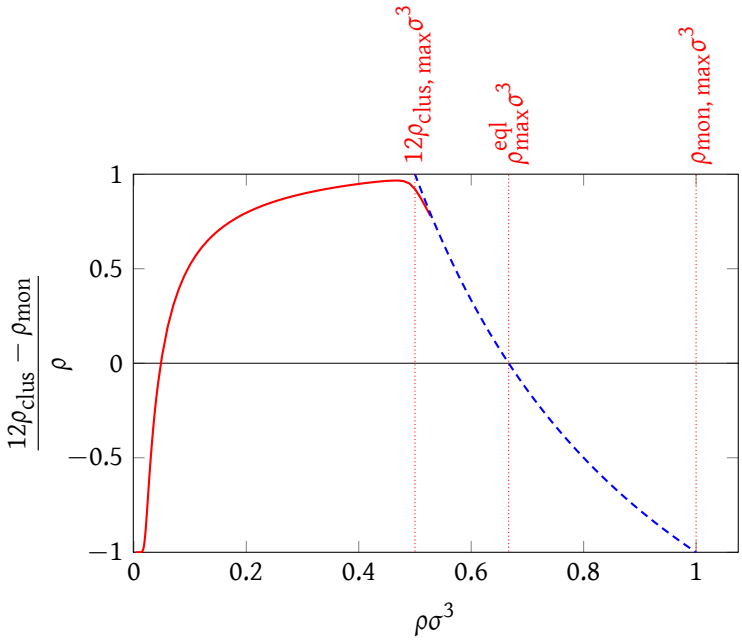


Figure A.6: The relative proportion of monomers and clusters in the equilibrium mixture is shown as a function of the density at $k_{\text{B}}T/\varepsilon = 0.2$ for $B/b = 24$, $\sigma_{\text{pw}} = 0.5$. The function $(12\rho_{\text{clus}} - \rho_{\text{mon}})/\rho$ takes the value -1 when only monomers are present, $+1$ when only clusters are in the system, and 0 at the centre of the clustering transition (*i.e.* on the $T_{\text{clus}}(\rho)$ line). Vertical dotted lines indicate the maximum particle density associated with clusters $12\rho_{\text{clus, max}}$, the maximum monomer density $\rho_{\text{mon, max}}$ and the maximum density at which the monomers and clusters can have an equal probability $\rho^{\text{eql}}_{\text{max}}$, as labelled. The solid line is calculated from the partition functions, whilst the dashed line is the extrapolation in the regime where the available volume is held at zero.

$\rho^{\text{eql}}_{\text{max}}\sigma^3 = 1$.

In figure A.5, the dotted vertical lines correspond to $\rho^{\text{eql}}_{\text{max}}$, and so for $B/b = 24$, monomers dominate to the right of the line. We also note that, in practice, it becomes numerically extremely difficult to calculate the equilibrium densities of monomers and clusters beyond $\rho = 12\rho_{\text{clus, max}}$, because the $V - BN_{\text{clus}} - bN_{\text{mon}}$ term in the translational partition function becomes very close to zero and the chemical potentials of the clusters and the monomers both become extremely large. However, it is relatively easy to extrapolate beyond $12\rho_{\text{clus, max}}$ using the approximation that the available volume $V - BN_{\text{clus}} - bN_{\text{mon}}$ is zero, as illustrated in figure A.6.

We note here that when considering the combined self-assembling Van der Waals fluid in section A.4, for the most part we need not worry about some of these phenomena that occur when $\rho > 12\rho_{\mathrm{clus,\,max}}$, because phase separation intervenes before this density is reached.

A.4 Self-assembling Van der Waals fluid

Having developed an approach to calculate the equilibrium between a vapour of monomers and clusters, we now wish to calculate the phase diagram of a 'self-assembling' Van der Waals fluid, where monomers can either assemble into inert clusters (with no attractive interactions) or condense to form a liquid. We have chosen to use a Van der Waals fluid as a starting point because, although we could use a more sophisticated description of the monomer and liquid states, we are mainly interested in the underlying fundamental physical behaviour of the system, rather than in trying to describe the behaviour of any one particular experimental system.

A.4.1 Partition functions

To model clusters, we use the cluster molecular partition function from equation (A.10) above. The canonical partition function for a Van der Waals (VdW) monatomic fluid is given by

$$Q = \frac{1}{N!}\left(\frac{2\pi m k_{\mathrm{B}} T}{h^2}\right)^{(3N/2)} (V - Nb)^N \exp\left[\frac{aN^2}{V k_{\mathrm{B}} T}\right], \tag{A.17}$$

where a and b are positive constants.[68] This expression is analogous to that for an ideal gas, but with an excluded volume term (Nb) and an attractive exponential term. To model the vapour and the liquid, we use the molecular partition function for a Van der Waals fluid derived from equation (A.17), except for two modifications. First, we include an additional term in the volume expression in order to take into account the exclusion of monomers by clusters as well as by monomers themselves. Using the free volume approximation[403] leads to the same available volume term as in equation (A.10), namely $V - BN_{\mathrm{clus}} - bN_{\mathrm{mon}}$. Furthermore, we multiply this

molecular partition function by q_{rot} of the same form as in equation (A.12) to take into account the fact that the monomers are now particles with orientational degrees of freedom. There are no attractive interactions between clusters, or between clusters and monomers, to reflect the fact that in our patchy particle model, no patches are exposed on the surface of the clusters (we ignore partially formed clusters); all patches are instead used in the internal bonding.

A.4.2 Coexistence curve

The coexistence curve, *i.e.* the curve where the difference in the free energy between two phases is zero, is often calculated using a Maxwell construction, whereby a straight line is constructed in the p-V phase diagram over the mechanically unstable region such that the areas of the regions defined by the Van der Waals loop and this line are equal.[413] This is to say, a line is constructed at some pressure $p_1 = p_2$ such that

$$p_1(V_2 - V_1) = \int_{V_1}^{V_2} p \, \mathrm{d}V, \tag{A.18}$$

so that the area under the pressure curve matches the area of the rectangle defined by V_1 and V_2. By evaluating the areas near coexistence and varying the pressure in the appropriate direction in ever smaller steps, we are able to obtain the coexistence points numerically. A representative Maxwell construction is shown in figure A.7(a). Performing this at a variety of temperatures allows us to calculate the binodal curve (figure A.8).

However, in many interesting cases, Van der Waals loops do not exist,[403;404] and it would be useful to have a more generic (and, indeed, easier) approach. We can calculate the chemical potential of species j using $\mu_j = (\partial A/\partial N_j)_{V,T,N'}$, where N' refers to all components other than j, and the pressure as $p = -(\partial A/\partial V)_{T,N}$. Plotting the chemical potential against the pressure parametrically allows us to determine binodal points where the curves cross. Whenever the chemical potential defines two (or more) different pressures, the stable phase invariably has

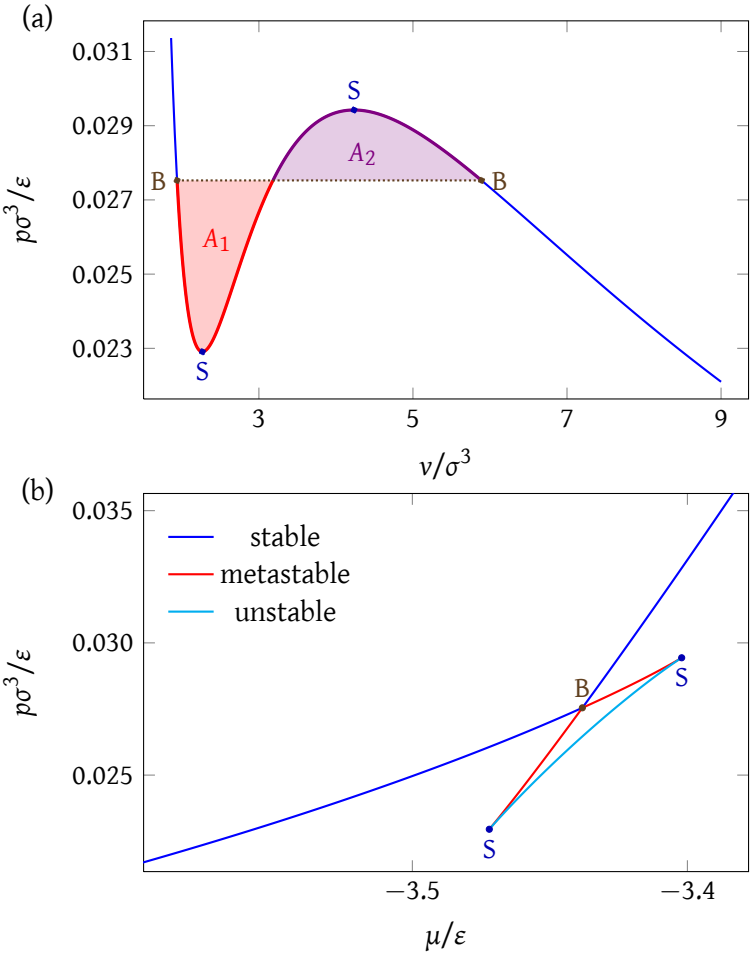


Figure A.7: Alternative methods of calculating binodals. In (a), a Maxwell construction on a Van der Waals loop is shown on the p-v diagram, where $v = V/N = 1/\rho$. In (b), a parametric plot of p versus μ whilst varying ρ is shown. Binodal (B) and spinodal (S) points are marked. $k_\mathrm{B}T/\varepsilon = 0.276$, $a\sigma^{-3}/\varepsilon = 1$ and $b\sigma^{-3} = 1$.

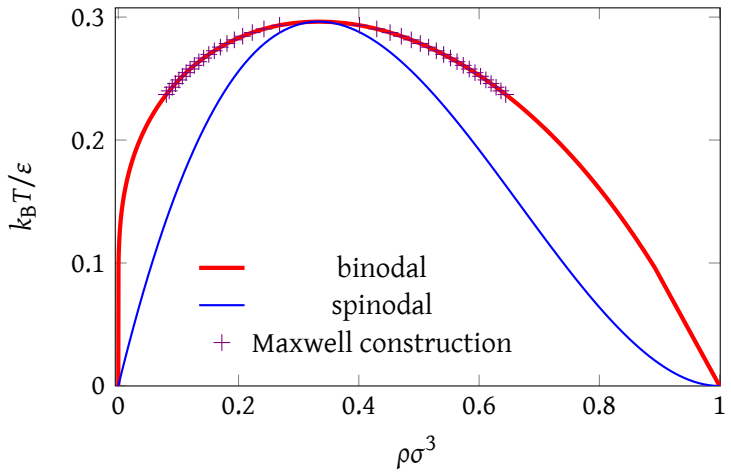


Figure A.8: The T-ρ phase diagram of a Van der Waals fluid with $a\sigma^{-3}/\varepsilon = 1$ and $b\sigma^{-3} = 1$, with spinodals and binodals depicted.

the *largest pressure*, which maximises the entropy and is consequently favoured by the second law of thermodynamics,[403;414] although this may perhaps more readily be appreciated by inverting the axes and minimising the chemical potential at any given pressure. An illustration of this method is shown in figure A.7(b).

As in section A.3, we assume that the only relevant reaction in the system is $12\mathrm{A} \rightleftharpoons \mathrm{A}_{12}$. We then seek to find the coexistence curves for the system where Van der Waals monomers and clusters are in equilibrium. This is achieved by setting $\mu_{\mathrm{clus}} = 12\mu_{\mathrm{mon}}$ and constructing p-μ curves to calculate binodals. Since the density of clusters is a function of the density of Van der Waals monomers, and vice versa, we use numerical techniques for solving the system of equations. To find p as a function of μ, we impose a chemical potential μ and numerically minimise both $|\mu_{\mathrm{clus}} - 12\mu_{\mathrm{mon}}|$ and $|\mu_{\mathrm{clus}}/12 + \mu_{\mathrm{mon}} - 2\mu|$ as functions of ρ_{mon} and ρ_{clus}, all under the constraints that densities are positive and that the available volume $(V - BN_{\mathrm{clus}} - bN_{\mathrm{clus}})$ is non-negative. The first minimisation ensures that the system is at equilibrium ($\mu_{\mathrm{clus}} \approx 12\mu_{\mathrm{mon}}$), whereas the second fixes the individual chemical potentials to the imposed value ($\mu_{\mathrm{clus}} \approx 12\mu$ and $\mu_{\mathrm{mon}} \approx \mu$). We only admit answers which have a combined difference equalling zero to within a small error term of $10^{-6}\varepsilon$, although in practice the error term is generally several orders of magnitude smaller. Having now calculated ρ_{clus} and ρ_{mon}, we can calculate the pressure. Example p-v and p-μ curves are given in figure A.9. By performing this analysis at a large number of relevant temperatures, the full phase diagram can be calculated.

A.4.3 Stability criteria

For a normal two-component system, where species 1 and 2 are independent components (such as in a mixture), the standard stability criteria are[414]

$$\left(\frac{\partial p}{\partial V}\right)_{T,N_1,N_2} < 0, \tag{A.19}$$

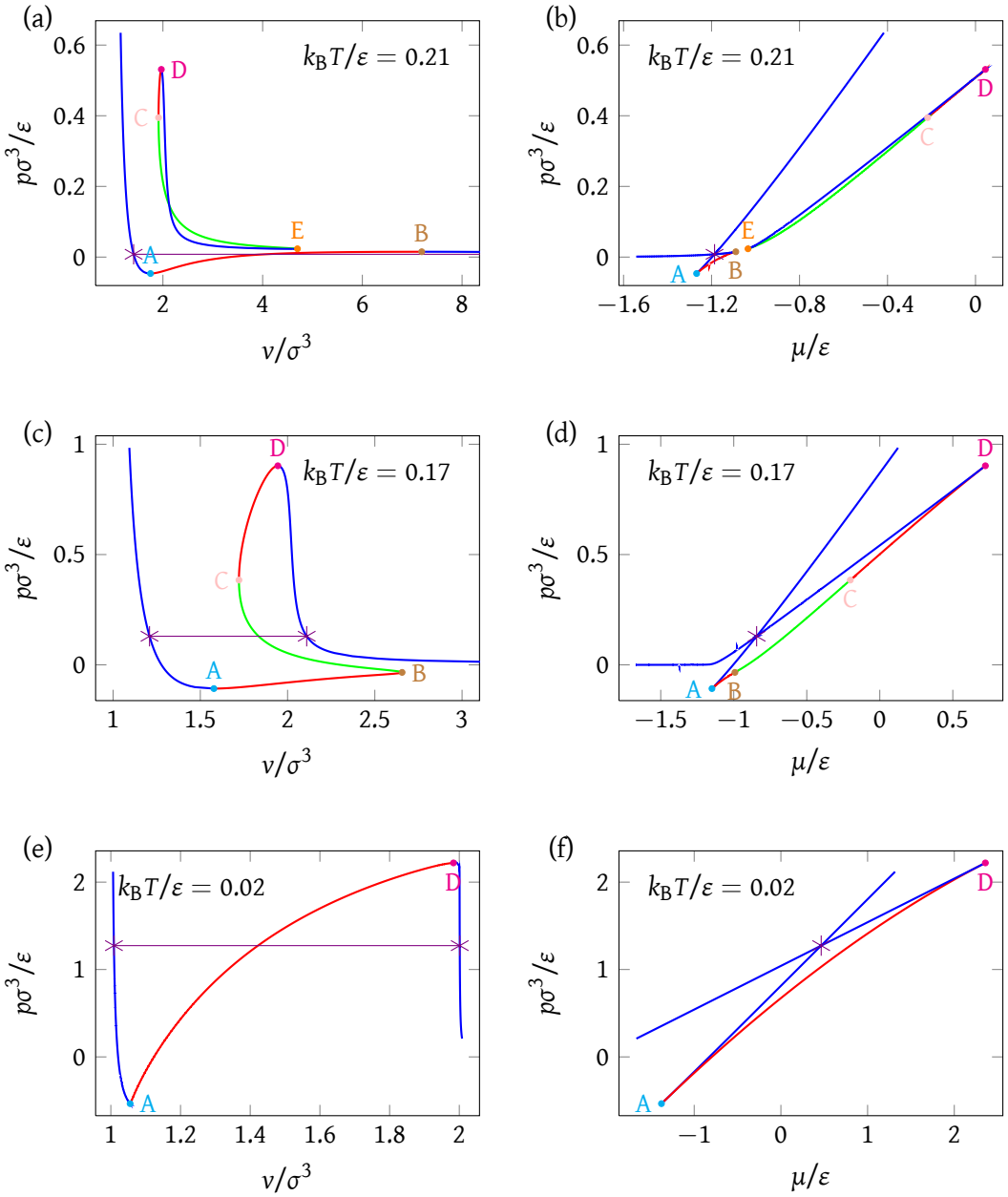


Figure A.9: The p-v ((a), (c) and (e)) and p-μ ((b), (d) and (f)) curves for the self-assembling Van der Waals fluid at reduced temperatures of (a) and (b) $k_BT/\varepsilon = 0.21$, (c) and (d) $k_BT/\varepsilon = 0.17$ and (e) and (f) $k_BT/\varepsilon = 0.02$. Regions of mechanical stability are coloured blue, those of mechanical instability red, and the mechanically stable, but compositionally unstable, region of back-bending is coloured green. Binodal points are marked by violet asterisks connected by a tie line. Spinodal points, where the derivative of the pressure with respect to the volume is either zero or infinity, are also shown with the labelling and colour-coding matching that used in figure A.10. $v = V/N$, $a/\varepsilon\sigma^3 = 1$, $b/\sigma^3 = 1$, $\sigma_{pw} = 0.5$, $B/b = 24$.

$$\left(\frac{\partial \mu_1}{\partial N_1}\right)_{T,p,N_2} > 0 \qquad \text{and} \tag{A.20}$$

$$\left(\frac{\partial \mu_2}{\partial N_2}\right)_{T,p,\mu_1} > 0. \tag{A.21}$$

However, in the system under investigation, the numbers of monomers and clusters are not independent. In the equilibrium $12\mathrm{A}_1 \rightleftharpoons \mathrm{A}_{12}$, we can write the total number of particles as $N = 12N_{\mathrm{clus}} + N_{\mathrm{mon}}$ and the differential as $\mathrm{d}N = 12\,\mathrm{d}N_{\mathrm{clus}} + \mathrm{d}N_{\mathrm{mon}}$. The fundamental equation for the Helmholtz energy can be written as

$$\begin{aligned} \mathrm{d}A &= -p\,\mathrm{d}V - S\,\mathrm{d}T + \mu_{\mathrm{mon}}\,\mathrm{d}N_{\mathrm{mon}} + \mu_{\mathrm{clus}}\,\mathrm{d}N_{\mathrm{clus}} \\ &= -p\,\mathrm{d}V - S\,\mathrm{d}T + \mu_{\mathrm{mon}}\,\mathrm{d}N + \left(\mu_{\mathrm{clus}} - 12\mu_{\mathrm{mon}}\right)\,\mathrm{d}N_{\mathrm{clus}}, \end{aligned} \tag{A.22}$$

and we are thus able to write a new stability criterion by requiring that the second derivatives be positive (and so that the Helmholtz energy is a minimum), including

$$\left(\frac{\partial^2 A}{\partial V^2}\right) = -\left(\frac{\partial p}{\partial V}\right)_{T,N,N_{\mathrm{clus}}} > 0. \tag{A.23}$$

This is very similar to the standard definition of spinodal points. However, since we will already have calculated equilibrium points, it would be convenient to establish a stability criterion based on these data. Since $\mu_{\mathrm{clus}} - 12\mu_{\mathrm{mon}} = 0$ at coexistence, and

$$\left(\frac{\partial(\mu_{\mathrm{clus}} - 12\mu_{\mathrm{mon}})}{\partial V}\right)_{N,T,\mathrm{eql}} = 0 \tag{A.24}$$

since the difference in chemical potential is constant (identically nought) along the p-V equilibrium curve, we write

$$\left(\frac{\partial A}{\partial V}\right)_{N,T,\mathrm{eql}} = -p, \tag{A.25}$$

and consequently

$$\left(\frac{\partial^2 A}{\partial V^2}\right)_{\mathrm{eql}} = -\left(\frac{\partial p}{\partial V}\right)_{T,N,\mathrm{eql}} > 0 \tag{A.26}$$

is also a stability criterion reflecting mechanical stability. This may readily be evaluated numerically from the equilibrium curve in the p-V phase diagram (such as those represented in

figure A.9). To include spinodals on the phase diagram, we can bracket regions of mechanical stability by finding where the gradients of the p-v curves change sign.

Notice from figure A.9(c) that there are parameter ranges where there are two regions of mechanical instability (AB and CD), separated by a region of mechanical stability (BC). In this system, however, we also have to check for compositional stability, *i.e.* whether fluctuations away from the equilibrium composition of monomers and clusters are stable. To derive a condition for compositional stability, we again consider the fundamental equation for Helmholtz energy given by equation (A.22); it follows that

$$\left(\frac{\partial A}{\partial N_{\mathrm{clus}}}\right)_{N,V,T} = \mu_{\mathrm{clus}} - 12\mu_{\mathrm{mon}}, \tag{A.27}$$

which is of course zero at equilibrium. The condition for compositional stability is hence that

$$\left(\frac{\partial^2 A}{\partial N_{\mathrm{clus}}^2}\right) = \left(\frac{\partial(\mu_{\mathrm{clus}} - 12\mu_{\mathrm{mon}})}{\partial N_{\mathrm{clus}}}\right)_{N,V,T} > 0. \tag{A.28}$$

This stability criterion can be evaluated at every equilibrium point of the combined system. The system is compositionally unstable exactly along the line BC.

A.4.4 Phase diagrams

We are most interested in computing phase diagrams for our system in the region of parameter space where re-entrant behaviour might be expected to occur. We choose a value for the patch width from the region where simulations (figure A.2) suggest that this might be the case, namely $\sigma_{\mathrm{pw}} = 0.5$. We choose the free volume coefficients B and b in line with subsection A.3.2, such that $b/\sigma^3 = 1$, and use the case $B/b = 24$ as our main example. We then choose the Van der Waals parameter a such that the critical Van der Waals temperature $k_{\mathrm{B}}T_{\mathrm{c}} = 8a/27b$ is somewhat greater than the clustering temperature at the critical density. This is true for our choice $a/\varepsilon\sigma^3 = 1$, for which $k_{\mathrm{B}}T_{\mathrm{crit}}/\varepsilon \approx 0.30$ for the pure Van der Waals fluid and $k_{\mathrm{B}}T_{\mathrm{clus}}(\rho_{\mathrm{crit}})/\varepsilon \approx 0.23$ in the absence of liquid formation (*i.e.* $a = 0$).

As in subsection A.3.3, we can also calculate the clustering transition temperature, $T_{\text{clus}}(\rho)$, although we should note that where $T_{\text{clus}}(\rho)$ lies in the two-phase region, phase separation of course occurs instead of clustering. With liquid formation being favourable at high densities, however, the clustering transition curve does not simply increase in temperature as the density increases, but beyond $\rho\sigma^3 \approx 0.15$ instead gradually decreases, such that it reaches $\rho^{\text{eql}}_{\text{max}}$ at $k_{\text{B}}T/\varepsilon = 0$, as shown in figure A.10.

The full phase diagram for $B/b = 24$ is depicted in figure A.10. Re-entrant phase behaviour is readily observed in this T-ρ plane for $0.05 \lessapprox \rho\sigma^3 \lessapprox 0.5$. As the temperature decreases at constant density within this range, the system first undergoes a transition from a monomeric fluid to a two-phase liquid-vapour coexistence region, and then as the temperature is further decreased, to a fluid of icosahedral clusters.

At high temperatures, the self-assembling Van der Waals fluid exhibits liquid-vapour

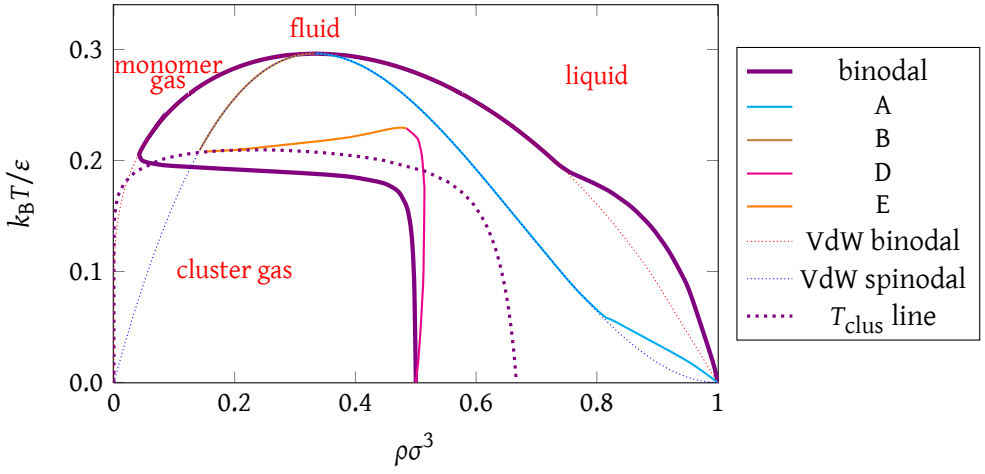


Figure A.10: The T-ρ phase diagram for the self-assembling Van der Waals fluid for $B/b = 24$. The thick lines give the binodals for the system. Also represented are lines which correspond to points A, B, D and E in the p-v plots in figure A.9, where the derivative of the pressure with respect to the volume is zero or $\pm\infty$, where they are true limits of stability. The binodals and spinodals for the standard, pure Van der Waals fluid are plotted for comparison (dotted lines). Finally, the clustering temperature (T_{clus}) is also plotted (thick dotted line). $a/\varepsilon\sigma^3 = 1$, $b/\sigma^3 = 1$, $\sigma_{\text{pw}} = 0.5$.

coexistence behaviour exactly analogous to the pure Van der Waals fluid, and the binodals of the self-assembling fluid follow those of the Van der Waals fluid. This liquid-vapour phase separation is driven by the attractions between monomers. In this region, the phase boundary exhibits a positive gradient in the p-T phase diagram (figure A.11). The Clapeyron equation[40] relates this gradient to

$$\frac{\mathrm{d}p}{\mathrm{d}T} = \frac{\Delta_{\mathrm{trs}}H}{T_{\mathrm{trs}}\Delta_{\mathrm{trs}}V} = \frac{\Delta_{\mathrm{trs}}S}{\Delta_{\mathrm{trs}}V}, \tag{A.29}$$

and since the entropy, enthalpy and volume changes are all negative for the vapour-liquid transition, the resulting slope is naturally positive.

By contrast, at sufficiently low temperatures, $T_{\mathrm{clus}}(\rho)$ lies at lower density than the Van der Waals binodal. Therefore, as the density increases, the vapour forms clusters before it reaches the point at which it would otherwise have become phase separated. As clusters have no attractions between them, the vapour is now stable and does not phase separate.

As the density of this system approaches the maximum possible density of clusters, packing constraints become important. If the density is to move beyond $12\rho_{\mathrm{clus,\,max}}$ and remain homogeneous, the system must form a monomer-cluster mixture with the proportion

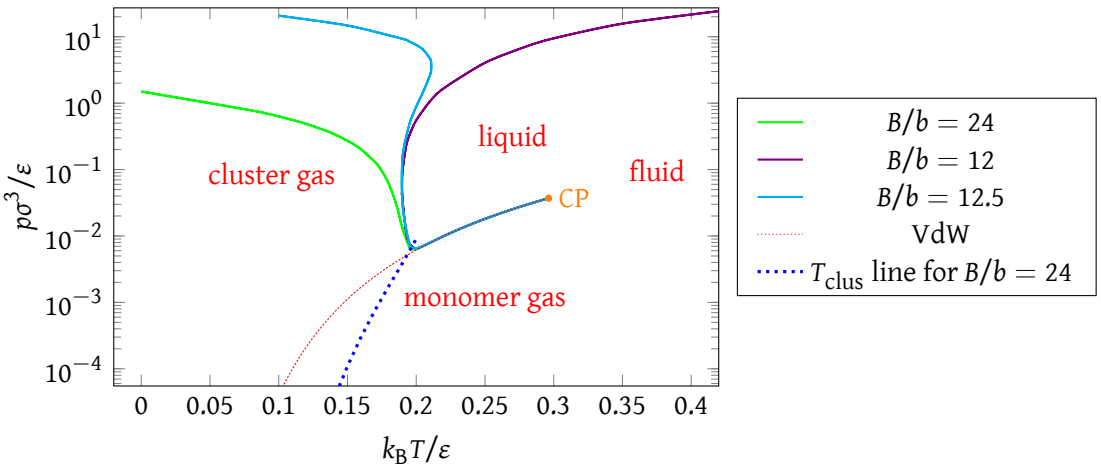


Figure A.11: The p-T phase diagram for the self-assembling Van der Waals fluid at different values of B/b. The critical point is labelled 'CP'. The clustering temperature (T_{clus}) is also plotted for $B/b = 24$ (thick dotted line) in the monomer-cluster region only. $a/\varepsilon\sigma^3 = 1$, $b/\sigma^3 = 1$, $\sigma_{\mathrm{pw}} = 0.5$.

of monomers increasing with ρ. However, unlike in subsection A.3.3, the system now prefers to de-mix into a vapour of clusters and a monomeric liquid because of a lack of attractions between clusters and monomers. Consequently, both the solubility of monomers in the cluster fluid and of clusters in the monomeric liquid is very low. For example, at $k_B T/\varepsilon = 0.08$, $\rho_{mon}\sigma^3 = 3.44 \times 10^{-7}$ at the lower-density binodal ($\rho\sigma^3 = 0.498$), and $\rho_{clus}\sigma^3 = 7.28 \times 10^{-120}$ at the higher-density binodal ($\rho\sigma^3 = 0.955$). This de-mixing is a density-driven transition, and in this regime, the position of the lower-density binodal varies little with temperature. As the vapour phase is now lower in energy than the liquid due to the strong internal bonding in the clusters, the enthalpy change associated with the vapour-liquid transition is now positive, while the volume change is of course still negative. Hence, the cluster-liquid transition has a negative slope in the p-T diagram (figure A.11).

The crossover between these two regimes occurs at temperatures near the point at which $T_{clus}(\rho)$ crosses the pure Van der Waals binodal and leads to a rapid increase in the density of the lower-density binodal with decreasing temperature until $12\rho_{clus,\,max}$ is approached. The width of this crossover regime reflects the width of the monomer-cluster equilibrium. The change in slope of the binodal occurs slightly above T_{clus}, as the vapour first begins to cluster. Initially, when monomers are still in the majority at the binodal, the transition is still driven by the attractions between the monomers, and the binodal occurs when the monomer density reaches the density of the binodal for the pure Van der Waals system. However, as clusters begin to predominate in the coexisting mixture, the phase transition is driven more by the lack of attractions between monomers and clusters, and occurs when the monomer density reaches a critical solubility.

These changes in the phase diagram are also reflected in the p-v and p-μ diagrams (figure A.9). For $0.23\,\varepsilon/k \lessapprox T < T_{crit}$, we observe Van der Waals-like p-v and p-μ curves. However, below this temperature, even though not yet in the region of re-entrance, we can observe

additional metastable closed curves in both p-v and p-μ plots. As the temperature decreases, the cluster-rich loop (as seen at $k_\mathrm{B}T/\varepsilon = 0.21$) merges with the Van der Waals curve and thus exhibits two additional spinodal points compared to the situation in the pure Van der Waals case, due to a back-bending of the p-v curve. This merging occurs when $T_\mathrm{clus}(\rho)$ meets the spinodal density, as shown in figure A.10. Below this temperature, because the cluster predominates in the vapour, the vapour is no longer destabilised by the presence of attractions and instead only becomes unstable when the density goes beyond $12\rho_{\mathrm{clus,\,max}}$.

To obtain spinodals, we numerically calculated derivatives obtained directly from p-v curves. As already mentioned in subsection A.4.3, the unusual sections of back-bending in the p-v isotherms (*e.g.* curve BC in figure A.9(c)) are, although mechanically stable, compositionally unstable, and therefore the points at which $\partial p/\partial v = \infty$ are not 'true' borders of (in)stability. The physically relevant spinodals are those of type A and B in the Van der Waals-like region of the phase diagram, and then D instead of B in lower regions; these are shown in figure A.10.

For $B/b = 24$, despite the re-entrance, at all temperatures below T_crit, there is always a density range for which phase separation occurs. We will now consider how the topology of the phase diagram can be modified by varying the parameters of the model. In particular, we want to investigate whether the phase separation at low temperature can be removed, and if so, whether the phase diagram might exhibit a closed loop similar to the nicotine-water mixture.

First, we investigate the effect of decreasing B, as this parameter determines $\rho_{\mathrm{clus,\,max}}$ and hence the onset of de-mixing in the low-temperature regime. As B/b decreases, the clusters are able to pack more efficiently and the lower-density binodal moves to higher density. Initially, this does not lead to any change in the topology of the phase diagram. However, the situation becomes more interesting as the binodal associated with the cluster fluid meets and crosses that associated with the monomeric liquid. Figure A.12(a) illustrates the type of phase diagram that results. We note that this scenario does not lead to a lower critical point; rather, at point A

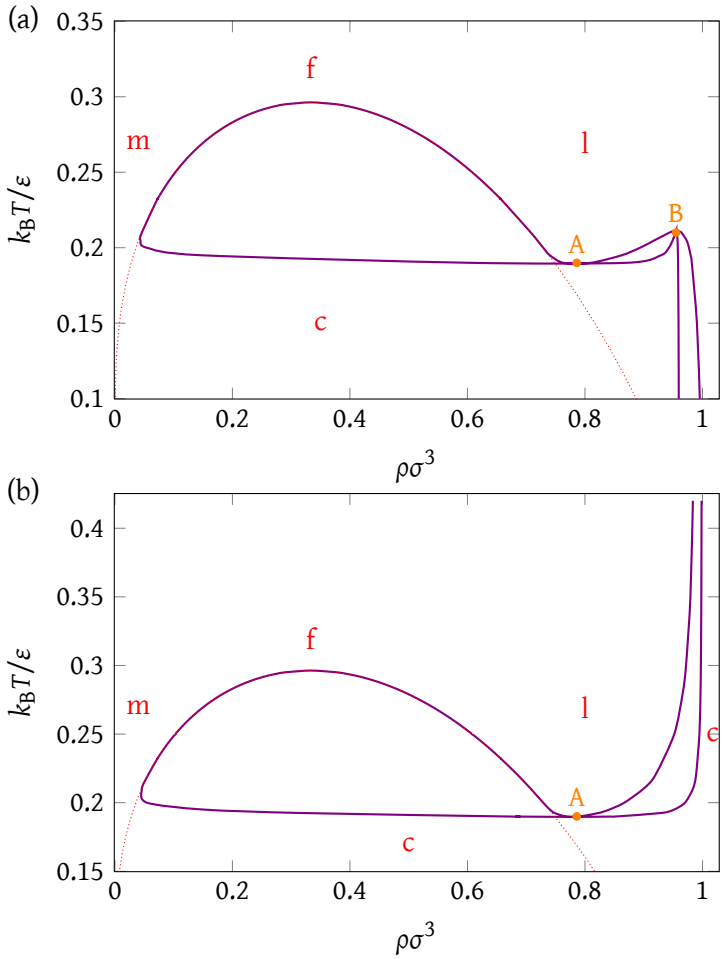


Figure A.12: Binodal lines for (a) $B/b = 12.5$ and (b) $B/b = 12$ are shown in violet, compared with the dotted red line representing the pure Van der Waals binodal. The labels 'c', 'm', 'l' and 'f' refer to cluster gas, monomer gas, liquid and fluid states, respectively. $a/\varepsilon\sigma^3 = 1$, $b/\sigma^3 = 1$, $\sigma_{\mathrm{pw}} = 0.5$.

in the phase diagram, there are two coexisting fluids with the same density, but which differ significantly in other characteristics. One fluid is a monomeric liquid with high entropy, while the other is a dense cluster fluid with low energy. These two fluids do not evolve continuously into each other as point A is approached (as would need to be the case for there to be a critical point), but instead retain their separate identities.

As $\Delta V = 0$ at point A in figure A.12, $\mathrm{d}p/\mathrm{d}T \to \infty$ and there is a change in sign of the slope of the phase boundary in the p-T phase diagram (figure A.11). Initially, beyond this point, the coexisting cluster fluid is more dense, and so $\mathrm{d}p/\mathrm{d}T$ is again positive. However, for

$B/b > 12$, the clusters still pack less efficiently, and so at low temperature and densities beyond $12\rho_{\mathrm{clus,\,max}}$, there must again be a region of de-mixing because of the lack of attractions between clusters and monomers. The slope of the p-T phase boundary corresponding to this de-mixing must again be negative and so there must be a second point (B) where the slope of the phase boundary changes and where the coexisting fluids have the same density. Curiously, for this phase diagram topology, there is a temperature range ($0.19 \lessapprox k_{\mathrm{B}}T/\varepsilon \lessapprox 0.21$ for $B/b = 12.5$) where there are three fluid-fluid phase transitions as the system is pressurised (figure A.11).

Finally, at $B/b = 12$, there must be a further change in the phase diagram topology, because at this point monomers no longer pack more efficiently than clusters, and at low temperature, the cluster fluid will be more stable for any feasible density. As B/b approaches 12 from above, point B moves to higher and higher temperatures, and the density range for the cluster fluid/monomeric liquid de-mixing becomes smaller and smaller until it disappears at $B/b = 12$. The resulting phase diagram is illustrated in figure A.12(b).

The second effect we wish to consider is changing the position of the liquid-vapour critical point with respect to the clustering transition. For a given b, T_{crit} is determined by the value of a, and T_{clus} is determined by σ_{pw}. In contrast to our model, for the simulated patchy particle systems, the two parameters are not independent, but instead the position of the critical point varies with the patch width (and the number of patches).[384;415] For example, it can be seen from figure A.2 that the temperature range associated with liquid-vapour coexistence decreases as σ_{pw} decreases. Here, we therefore choose to vary one of the parameters (a) whilst keeping the other one (σ_{pw}) fixed.

The results for a variety of values of a are shown in figure A.13. When the Van der Waals critical temperature is above the clustering temperature (for example, at $a = 1\,\varepsilon\sigma^3$ and $a = 0.8\,\varepsilon\sigma^3$, which correspond to critical temperatures of $k_{\mathrm{B}}T_{\mathrm{crit}}/\varepsilon = 0.296$ and $k_{\mathrm{B}}T_{\mathrm{crit}}/\varepsilon = 0.237$, respectively), Van der Waals behaviour is observed at high temperatures. When the critical

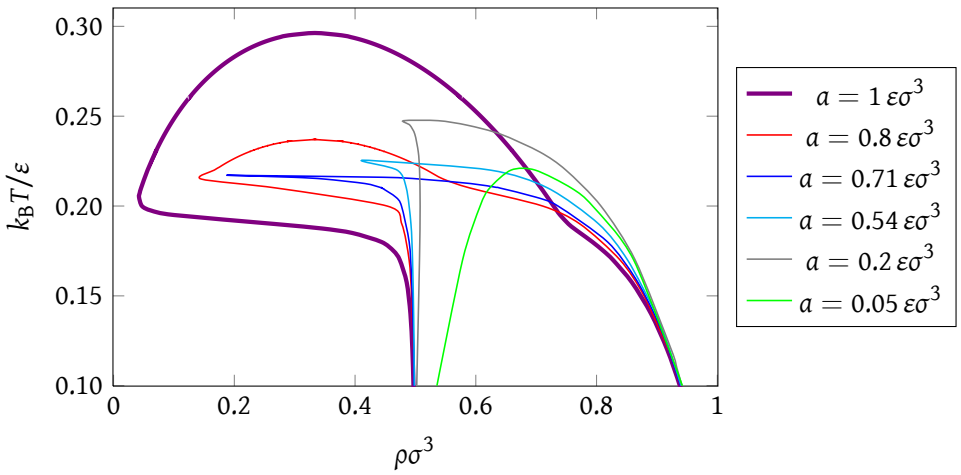


Figure A.13: A T-ρ phase diagram for the self-assembling Van der Waals fluid with a selection of values for the Van der Waals attractive parameter a. $b/\sigma^3 = 1$, $b/B = 24$, $\sigma_{pw} = 0.5$.

temperature descends below the clustering temperature, we nevertheless continue to observe a bending of the binodal curve to lower densities in the region where liquid-vapour coexistence previously took place. However, this feature is not some remnant of the liquid-vapour transition, but is instead associated with de-mixing in the vicinity of $T_{clus}(\rho)$. At T_{clus}, there is an equal probability of particles being monomeric or as part of clusters; however, if ρ and a are sufficiently large, the system will not form a homogeneous mixture, but will instead de-mix. The narrowness in temperature of this feature is associated with the sharpness with which the chemical equilibrium shifts between being monomer dominated to being cluster dominated as the temperature decreases. As a decreases, there is a progressively smaller de-mixing driving force and so the size of this feature decreases. Note that these changes lead to a non-monotonic dependence of T_{crit} on a, because T_{crit} initially increases as ρ_{crit} increases, mirroring the dependence of T_{clus} on ρ (figure A.5). As a becomes smaller still, the clusters and monomers begin to be able to mix at high temperatures because less entropy of mixing is required to overcome the enthalpic mixing incompatibility. Hence, T_{crit} again begins to decrease. However, even at a small value of a, *e.g.* $a = 0.05\,\varepsilon\sigma^3$ in figure A.13, there is still a substantial region of de-mixing. As a tends to zero, the de-mixing regime disappears, recovering the monomer-cluster

equilibrium studied in section A.3.

So far, we have considered exclusively particles that can self-assemble into icosahedral clusters, but our approach can easily be generalised to the formation of other types of cluster. Here, we illustrate this by calculating the phase diagram for particles able to form tetrahedra.[109] The requisite changes in the model from the icosahedral system discussed above are as follows. The moment of inertia of a tetrahedron about the principal axes is $I = 4m\sigma^2$, and the mass of the cluster is $4m$. The symmetry number of a tetrahedron is 12, and that of an individual monomer is 3. The ground state energy of the tetrahedron is -6ε. The number of particles in the cluster is $N = 4$, and this correspondingly affects the number of radial and orientational vibrational modes. Finally, the circumradius of a tetrahedron with edge length a is $r = \sqrt{3/8}\,a$,[412] and so the physical value of B/b which we have used is $B/b = 11$. This gives a maximum particle density associated with clusters of $4\rho_{\mathrm{clus,\,max}}\sigma^3 = 4/11 \approx 0.36$. We have assumed that tetrahedra will pack approximately like spheres; this approximation is not so far-fetched as it might seem at first glance, and recent research confirms that the packing fractions are very similar.[416] We

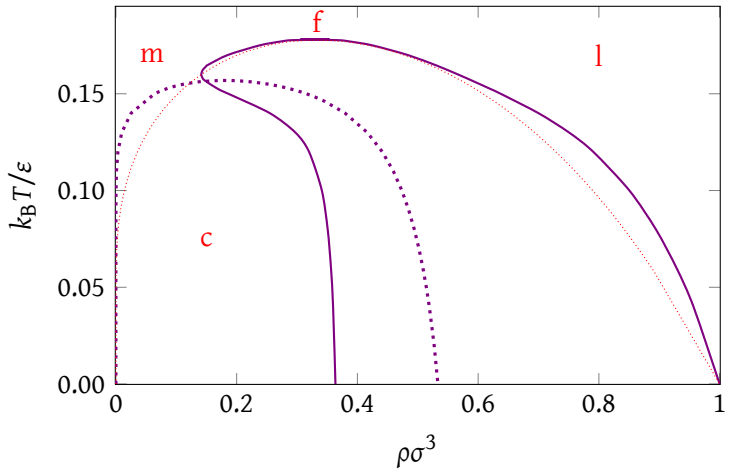


Figure A.14: A T-ρ phase diagram for the self-assembling Van der Waals fluid forming tetrahedral clusters. The solid line is the binodal curve, whilst the thin dotted line represents the pure Van der Waals system binodal. The thick dashed line is the clustering temperature, $T_{\mathrm{clus}}(\rho)$, for this system. The labels 'c', 'm', 'l' and 'f' refer to cluster gas, monomer gas, liquid and fluid states, respectively. $a/\varepsilon\sigma^3 = 0.6$, $B/b = 11$, $b/\sigma^3 = 1$, $\sigma_{\mathrm{pw}} = 0.5$.

also scale the value of the parameter a by 3/5 compared to that for the icosahedron in order to take into account that there are only three rather than five patches that give rise to the interactions that drive the liquid-vapour phase separation. The phase diagram of this system is depicted in figure A.14. The behaviour of the system is similar to the icosahedral system, with the main differences being easy to rationalise. For example, the clustering temperature is significantly lower due to the lower energetic driving force to form clusters. Similarly, the crossover between the liquid-vapour and de-mixing regimes occurs over a wider temperature range due to the greater width of the monomer-cluster equilibrium, which is in turn caused by the smaller size of the clusters.

A.5 Discussion and conclusions

We have calculated the complete phase diagram for a system exhibiting competition between phase separation and self-assembly into monodisperse clusters. For nearly all of the parameterisations of the model that we have considered, decreasing the temperature at constant density results, for a limited range of density, in the transition from a gas of monomers, to a liquid-gas phase-separated mixture, and then to a gas of clusters. Therefore, the existence of re-entrant phase behaviour, driven by the lower energy of clusters, is clear within this model and is a robust feature of such systems. In this aspect, the model agrees with our group's simulation results for patchy colloidal particles[109;385] and the Janus particle simulations of Sciortino and co-workers,[388;392] both of which showed evidence of re-entrant phase behaviour.

In particular, a comparison of the final phase diagram with the yield plot shown in figure A.2 shows that decreasing the temperature at a constant density of $\rho\sigma^3 = 0.15$ and a patch width of $\sigma_{\mathrm{pw}} = 0.5$ involves a re-entrant phase transition in both cases. Even though we have not chosen to match the parameter a precisely to the icosahedral system studied in simulations, there is surprisingly good agreement not only in the underlying physical behaviour, but also in

the numerical values.

While the phase behaviour of the fluid at low temperature is not yet fully clear in the simulated systems, our model allows the complete phase diagram to be calculated. We have established that phase separation persists at low temperatures within the framework of our model. The driving force for it, however, is not the attractions between monomers as in the pure Van der Waals fluid, but rather cluster-liquid immiscibility. For this reason, re-entrance is observed over only a limited range of density in all physically relevant phase diagrams; at higher densities, de-mixing between the cluster and the liquid states ensures continued phase separation.

In our model, the change in topology of the phase diagram as the model parameters are varied is easy to calculate. The free volume parameter ratio B/b changes the maximum density of the cluster, and therefore affects the position of the lower-density binodal of the T-ρ phase diagram in the de-mixing regime. The change in the Van der Waals parameter a, on the other hand, changes the Van der Waals critical temperature, and, provided that it is greater than the clustering temperature, affects the temperature range of the phase-separated region.

Of course, in our theoretical model, there are a significant number of approximations, the effect of which we should try to understand, in particular in relation to the comparison of our results with phase diagrams obtained from simulations. Firstly, our focus has just been on the fluid behaviour of these self-assembling systems, and, unlike in simulations, we can explore the low temperature form of the fluid phase behaviour without having to worry about other phases obscuring the fundamental behaviour. However, crystallisation (of both monomers and clusters) is likely to have a major effect on the overall form of the phase diagram. There are precedents for such cluster crystals; for example, icosahedral virus particles (monodisperse aggregates of virus capsid proteins, perhaps with a nucleic acid genome inside) are often observed to form crystals when they are densely packed in cells.[417]

We can construct a schematic phase diagram illustrating the potential effects of crystallisation on it by assuming hard-sphere crystallisation of both components. This approximation is probably quite reasonable for the clusters, as all the attractive patches are involved in the internal bonding of the cluster, but probably less so for the monomers, where the effect of the patches on the crystal stability and form is less clear. For hard spheres, the fluid-crystal coexistence limits occur at packing fractions $\phi = 0.494$ and $\phi = 0.545$,[418] and by associating the maximum densities of the fluids with the packing fraction in random close packing, $\phi = 0.64$,[411] we can obtain estimates for the densities of the fluid-crystal binodals.

We do this by calculating an effective unit of length such that the maximum density of clusters corresponds to random close packing. This is to say, for $B/b = 24$, $12\rho_{\mathrm{clus}}^{\mathrm{max}}\sigma^3 = 0.5$ and hence

$$\phi = \frac{4\pi r_{\mathrm{eff}}^3 \rho_{\mathrm{clus}}^{\mathrm{max}}}{3} = 0.64. \tag{A.30}$$

We then use the resulting r_{eff} in the calculation of densities corresponding to $\phi = 0.494$ and $\phi = 0.545$. We repeat the procedure to obtain the effective radius for the liquid by matching random close packing to $\rho_{\mathrm{mon}}\sigma^3 = 1$.

Further, the maximum hard-sphere crystalline packing fraction, $\phi = 0.74$, provides a means for obtaining an upper density limit for the stability of the cluster crystal. Finally, the clustering temperature provides an upper temperature stability limit for the cluster crystal. The resulting schematic phase diagram is shown in figure A.15. Note that this phase diagram has a second triple point associated with the disappearance of the cluster crystal phase, as well as the usual one associated with the disappearance of the liquid phase. We have not attempted to consider the effect of other possible phases on the phase diagram, such as the lamellar phase observed for Janus particles[388;392] and the two-dimensional crystalline lamellae[419] seen in simulations of octahedron-forming patchy particles.[109]

Secondly, there is also a series of approximations associated with our description of the

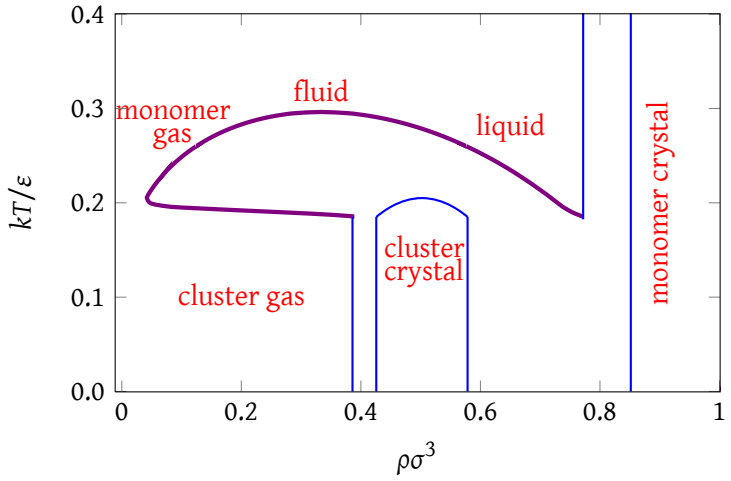


Figure A.15: A schematic T-ρ phase diagram to illustrate the potential effects of crystallisation on the phase diagram in figure A.10.

fluid. For example, the free volume approximation[403] is not flexible enough to capture fully both the high- and the low-density behaviour (we choose B and b values appropriate to the high-density limit) and the form of the available volume term differs from that for the exact hard-sphere second virial coefficient for interactions between monomers and clusters. There is also an obvious problem when the density of clusters approaches the maximum allowed, as the free volume approximation predicts that the translational partition function of both clusters and monomers should tend to zero. However, because of the large differences in size between monomers and icosahedral clusters, even when the clusters are close-packed, there is still space available for the monomers in the gaps between the clusters. In other words, the translational partition function of the monomers should not tend to zero as the translational partition function of the clusters does so. However, such correlation effects associated with the positions of the particles are not so straightforward to take into account.

In our model, we also effectively assume that the clusters pack, in essence, like hard spheres, without any change of shape or compression, regardless of the overall density of the system. This assumption, which is responsible for the near vertical slope of the lower-density binodal in the low-temperature de-mixing regime, may be reasonable for monodisperse clusters

that have a well-defined shape and are held together by specific bonds (although for our patchy particle potential, the clusters will be slightly compressible). However, this assumed cluster incompressibility may be a factor in the mismatch in the detailed shape of our phase diagrams compared to those obtained in the simulations of Sciortino and co-workers,[388;392] where the shapes of clusters are much more deformable, as well as not being monodisperse. In their computed phase diagrams, they observe the binodals moving to higher densities with decreasing temperature considerably more gradually than we do.

Another approximation we have made is that we can model the monomeric liquid and vapour phases as a Van der Waals fluid. In this model, the critical density has a fixed value (depending only on the free volume parameter b), whereas we know that in patchy particle colloids, the critical point varies with both the patch width and the patch number.[384;415] Nevertheless, this deficiency is probably not so important for the qualitative shape of the phase diagram, especially as we are most concerned here with the behaviour at lower temperatures associated with the onset of clustering.

Finally, we have assumed that clusters form either completely or not at all, *i.e.* only a single size of clusters can be formed. Although this is an approximation, the probability of observing partly-formed clusters *at equilibrium* is very small, and so the effect on the phase diagrams will be negligible.[385]

Although these approximations will affect the quantitative comparison between the phase diagrams for the current model and those computed in patchy particle simulations,[109;385] and hopefully at some future point those found experimentally for self-assembled patchy colloids, we believe that the fundamental insights the current model gives will help to provide a theoretical framework for understanding the fluid phase behaviour of these kinds of system.

B Computing phase diagrams for a quasicrystal-forming patchy-particle system

This appendix chapter presents work undertaken during the course of my DPhil project, but which does not fit into the main structure of the thesis. The work was undertaken in collaboration with Flavio Romano in the Doye group. The contents of this appendix have recently been accepted for publication.[4]

B.1 Summary

We introduce an approach to computing the free energy of quasicrystals, which we use to calculate equilibrium phase diagrams for systems of two-dimensional patchy particles with five regularly arranged patches that have previously been shown to form dodecagonal quasicrystals. We find that the quasicrystal is a thermodynamically stable phase for a wide range of conditions and remains a robust feature of the system as the potential's parameters are varied. We also demonstrate that the quasicrystal is entropically stabilised over its crystalline approximants.

B.2 Introduction

Quasicrystals are a type of aperiodic crystal structure: they have well-ordered structures, but cannot be described by a periodic lattice, not even one with incommensurate lattice parameters.[420;421] They were first reported by Shechtman and co-workers,[422] in whose work a rapidly cooled Al-Mn alloy was found to result in a metastable quasicrystal. Most quasicrystals discovered so far are binary or ternary metallic alloys;[420;421] however, recently, an increasing number of examples have been reported in the field of soft condensed matter.[423–425]

Quasicrystals have also been observed to form in computer simulations, in which not only binary, but also one-component quasicrystals have been seen.[426–439] While certain quasicrystals have been shown to be stable in experiment,[440] in simulations, the relatively short timescales accessible mean that it is not necessarily clear whether quasicrystals are truly the stable phase, or rather a metastable phase that is more kinetically accessible. In this work, we address this important question of assessing the stability of quasicrystals and consider a model soft matter system that has the potential to be realised experimentally.

To prove that quasicrystals are thermodynamically stable, their free energy must be computed and compared with that of the competing phases. However, such a calculation is not straightforward: the principal problem is that there is no obvious reference state for which the free energy is known and from which thermodynamic integration[190;441] could be used to calculate the free energy of the quasicrystalline phase. A phase transition intervenes when integrating from an ideal gas, which is normally used as a reference state for fluid phases, and an Einstein crystal, used as a reference state for crystalline phases, would also not be a suitable reference state because it would fail to capture the configurational entropy associated with the quasicrystal's many possible structures.

Several attempts to assess the stability of quasicrystals have been undertaken thus far. One approach is to compute the free energy of an approximant crystalline phase. However, it is not immediately obvious whether quasicrystals are more or less stable than their approximants;[432;436;442] whilst the enthalpy and vibrational properties of the quasicrystal and its approximant phases are likely to be similar, the free energy of a quasicrystal has a contribution from its configurational entropy that is not present for the approximant.[435] Nevertheless, the calculation of the free energy of quasicrystal approximants allows for some insight to be gained into the phase behaviour of such systems. For example, Iacovella and co-workers calculated free energies for several approximant phases for a spherical micelle model and found

that such phases are thermodynamically stable under certain conditions of polydispersity,[435] while Haji-Akbari and co-workers calculated free energies of approximants in systems of hard tetrahedra and bipyramids.[436;442] By contrast, Kiselev and co-workers recently estimated the free energy of the quasicrystal itself by combining a phonon contribution for a particular quasicrystal configuration from thermodynamic integration and a configurational contribution based on an approximation of uncorrelated phason flips.[443] They confirmed that, within their approximation, the free energy of the quasicrystal is in certain conditions lower than that of its approximant.

Our solution to this conundrum is to note that thermodynamic integration is not the only way the melting point of a solid can be computed. Another method is to simulate the direct coexistence of two phases with an interface.[190] By performing such simulations at a range of temperatures, we can bracket the regions where the quasicrystal melts ($T > T_{\text{fus}}$) and grows ($T < T_{\text{fus}}$). At $T = T_{\text{fus}}$, the chemical potential of the quasicrystal is equal to that of the fluid, which we can calculate using thermodynamic integration. Such an equilibrium approach by construction accounts for the quasicrystal's configurational entropy. Once the quasicrystal's free energy is known at one point, we can use thermodynamic integration to reach other points of interest on the phase diagram.

B.3 Model

Here, we apply this approach to a two-dimensional system of patchy particles[365;366] with five regularly arranged attractive patches, which we model using a simple potential analogous to that introduced in section A.3, which has previously been used in simulations of self-assembly and crystallisation.[385–387;391;439;444–448] The potential is based on the Lennard-Jones form, but where the attractive region is modulated by an angular term that measures how well patches

point at each other, and is given by

$$U_{ij}(\boldsymbol{r}_{ij},\,\varphi_i,\,\varphi_j) = \begin{cases} U^{\mathrm{LJ}}(r_{ij}) & r_{ij} < \sigma_{\mathrm{LJ}}, \\ U^{\mathrm{LJ}}(r_{ij})U^{\mathrm{ang}}(\hat{\boldsymbol{r}}_{ij},\,\varphi_i,\,\varphi_j) & \sigma_{\mathrm{LJ}} \leq r_{ij}, \end{cases} \tag{B.1}$$

where $\boldsymbol{r}_{ij}$ is the interparticle vector connecting the centres of the two particles i and j, r_{ij} is its magnitude, φ_i and φ_j are the orientations of the particles i and j, the generalised Lennard-Jones potential is

$$U^{\mathrm{LJ}}(r_{ij}) = 4\varepsilon\left[\left(\frac{\sigma_{\mathrm{LJ}}}{r_{ij}}\right)^{2n} - \left(\frac{\sigma_{\mathrm{LJ}}}{r_{ij}}\right)^{n}\right], \tag{B.2}$$

where $n = 6$ for the standard Lennard-Jones potential, and

$$U^{\mathrm{ang}}(\hat{\boldsymbol{r}}_{ij},\,\varphi_i,\,\varphi_j) = \exp\left[\frac{-\theta^2_{k_{\mathrm{min}}ij}}{2\sigma^2_{\mathrm{pw}}}\right]\exp\left[\frac{-\theta^2_{l_{\mathrm{min}}ji}}{2\sigma^2_{\mathrm{pw}}}\right], \tag{B.3}$$

where σ_{pw} is the 'patch width' (measured in radians), θ_{kij} is the angle between the patch vector of patch k on particle i and the interparticle vector $\boldsymbol{r}_{ij}$, and k_{min} is the patch that minimises the magnitude of this angle. Two particles therefore interact only through a single pair of patches. In the simulations reported here, we use a potential cutoff of $r_{\mathrm{cut}} = 3\sigma_{\mathrm{LJ}}$, and the potential in equation (B.2) is shifted so that it equals zero at $r_{ij} = r_{\mathrm{cut}}$. The crossover to including angular modulation in the potential (equation (B.1)) is likewise adjusted so that it still occurs when the potential energy is identically zero.

This interaction potential provides a simple model for the patchy colloidal particles that many experimental groups have, with progressively increasing success, been seeking to develop.[365;366;368;370;371;373;375;376;449–451] The enhanced range of structural behaviour that such patchy interactions could facilitate has been extensively studied in a variety of computer simulations.[382;384–388;391;439;444;445;448;452–454]

B.4 Results and discussion

Dodecagonal quasicrystals were previously found to form on cooling systems of these five-patch particles in certain conditions, and structures with the lowest enthalpy were also identified.[439] Some of the relevant phases studied are shown in figure B.1; particles are classified based on a common neighbour analysis[439] into σ ({21111}), H ({22110}) and Z ({222222}) environments (shown in figure B.2(a)), by analogy to the Frank–Kasper phases,[455;456] with the same cutoff radius ($r_{\mathrm{neigh}} = 1.38\sigma_{\mathrm{LJ}}$) as used by Van der Linden and co-workers. The hexagonal (Z) phase (figure B.1(c)) was found to form at high pressures (as it is densest) and for wide patches (as the potential is then closer to being isotropic). At low pressures and at reasonably narrow patch widths, the structure which best satisfies the attractive patches is the enthalpically favoured phase. A crystal cannot exist with five-fold symmetry, and so no crystal with five

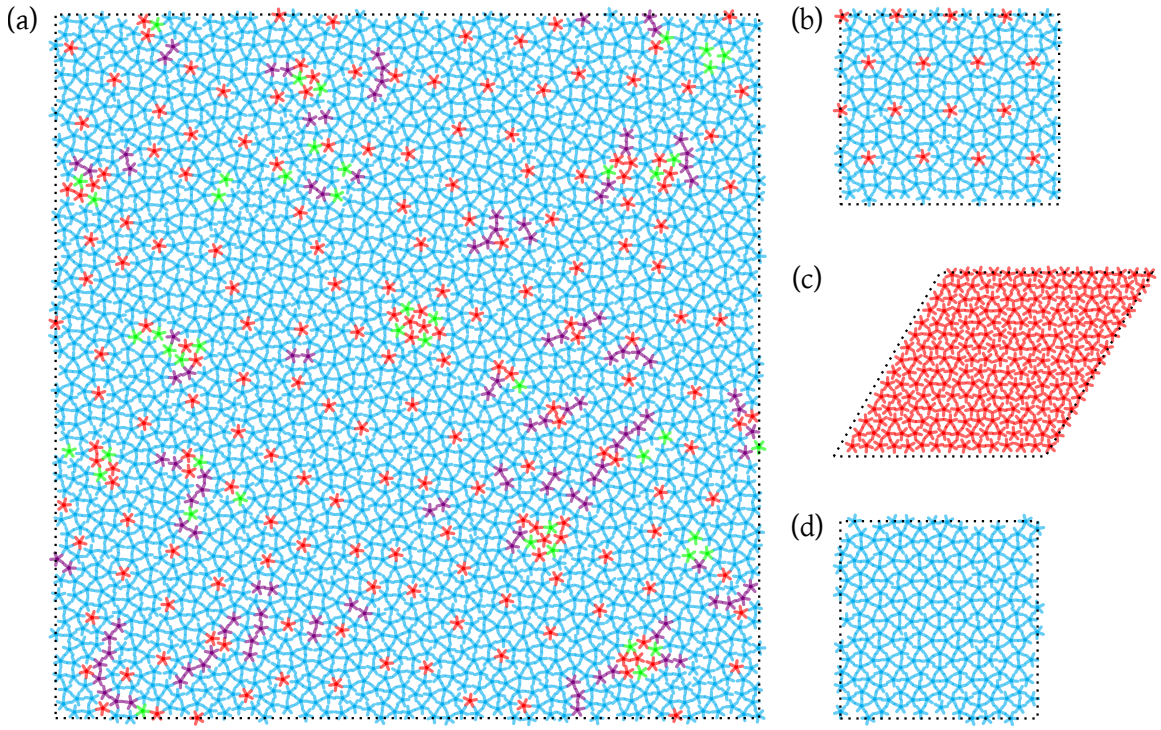


Figure B.1: Examples of the main phases studied in this work. Particle colours correspond to their classification: σ particles are coloured cyan, H particles violet, Z particles red and any unclassified particles green. (a) Quasicrystal, $N = 2500$. (b) Approximant crystal. (c) Plastic hexagonal (Z) phase. (d) σ phase.

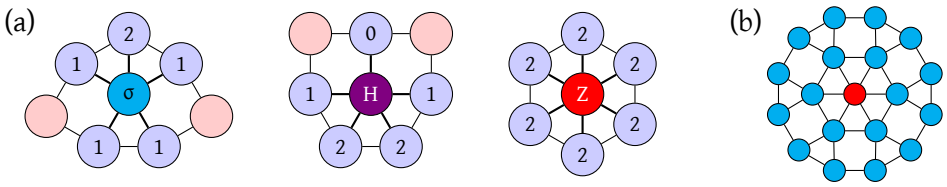


Figure B.2: (a) Neighbour classification of σ, H and Z environments. The numbers of common neighbours are given for each of the central particle's neighbours. (b) A dodecagonal motif characteristic of the quasicrystal.

perfectly aligned patches exists. As a compromise, each particle has five neighbours, but the angles between the neighbours do not perfectly match the five-fold symmetry of the particle itself. Local environments satisfying this requirement are the σ and H environments shown in figure B.2(a); the lowest enthalpy structure at low pressure and reasonably narrow patch widths is the σ crystal shown in figure B.1(d).[439]

Quasicrystals were observed in some cooling runs in the region where the σ phase (figure B.1(d)) is the lowest in enthalpy, but near to the hexagonal 'boundary'. The structure of a typical quasicrystal is shown in figure B.1(a); its diffraction pattern (figure B.3(a)) exhibits a characteristic twelve-fold symmetry. The local environments in the quasicrystal are predominantly of the σ type, but a significant fraction of Z environments can be seen: these are typically located at the centre of a dodecagonal motif (figure B.2(b)). Much of the structure can be analysed in terms of packing of such dodecagons into triangular, square and rectangular arrangements.[439] which can readily be seen in figure B.1(a); the various possible ways of arranging these dodecagons are likely to lead to substantial entropy. There is no translational order, but bonds can be orientated in any one of twelve directions, resulting in long-range orientational order of dodecagonal character.

We perform Monte Carlo simulations in the *NpT* ensemble using a rectangular box with periodic boundaries. As a starting point in the determination of phase diagrams for this system, we chose a patch width, as well as temperature and pressure, at which the quasicrystal is known to form on cooling and thus locate the fluid-quasicrystal (F-QC) equilibrium transition.

This transition is mostly rapid and facile, with essentially no hysteresis, which allows us to determine the coexistence points directly by performing simulations in relatively large boxes. Whilst we initially performed direct coexistence simulations to determine the melting point of the quasicrystal, these simulations mostly did not prove to be any more efficient than direct brute-force simulations, in which no initial interface was introduced, as the growth of the phases was not restricted to the initial interface. This lack of hysteresis is most likely indicative of a particularly low F-QC interfacial free energy. Phase transitions can be observed from a kink in the potential energy or density plotted against the variable we are changing (such as the temperature or the pressure). However, we can further test for the nature of the phase by calculating the appropriate diffraction pattern,

$$\mathrm{Re}(S(\boldsymbol{q})) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\cos\left[2\pi\boldsymbol{q}\cdot(\boldsymbol{r}_i - \boldsymbol{r}_j)\right], \tag{B.4}$$

where $\boldsymbol{q}$ is the reciprocal space vector. The quasicrystal exhibits a distinctive twelve-fold diffraction pattern (figure B.3(a)), the Z phase exhibits a six-fold pattern (figure B.3(c)), and the fluid phase does not exhibit a well-defined pattern.

Once a F-QC coexistence point is known, we can equate the chemical potential of the fluid (as calculated by thermodynamic integration) with that of the quasicrystal. From here,

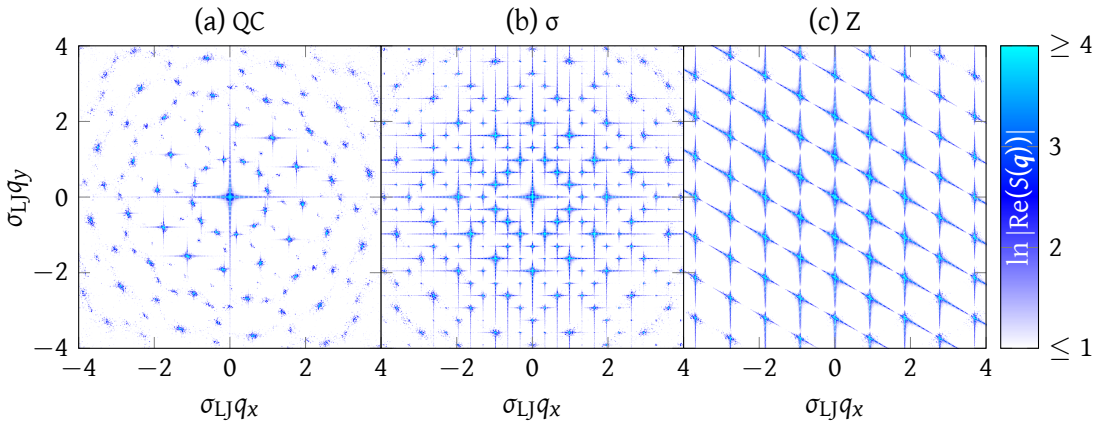


Figure B.3: Typical diffraction patterns corresponding to (a) the quasicrystal, (b) the σ phase and (c) the Z phase. Each system contains approximately 2500 particles.

we can use thermodynamic integration to calculate the chemical potential of the quasicrystal at other temperatures and pressures. By also performing thermodynamic integration to find the chemical potential of the σ phase,[1] we can locate the putative σ-QC transition. We plot in figure B.4 the free energies of the σ, QC and F phases as a function of T at a constant βp. From this figure, we can identify two clear transitions (σ-QC and QC-F), and a temperature window in which the quasicrystal is thermodynamically stable. In the same figure, we also show the

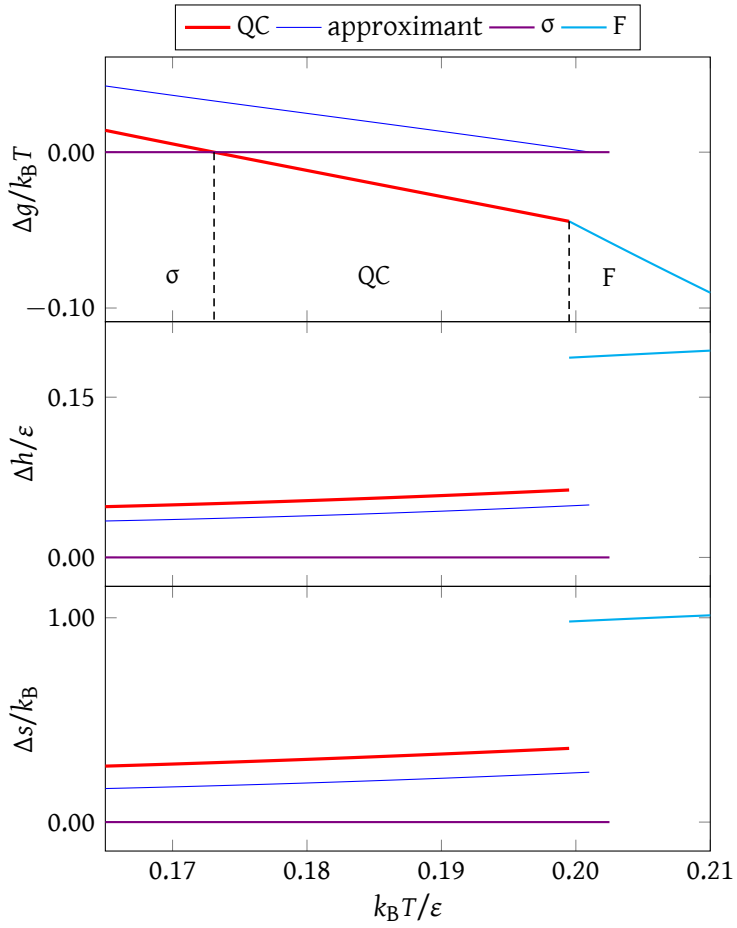


Figure B.4: Per-particle Gibbs energies, enthalpies and entropies, relative to the σ phase, as a function of T at $\sigma_{LJ}^2 \beta p = 1.5$, $\sigma_{pw} = 0.49$ and $n = 6$. Lines end at the limit of (meta)stability of the phases. Dashed lines denote coexistence points, $k_B T_{\sigma \leftrightarrow QC}/\varepsilon = 0.173$ and $k_B T_{QC \leftrightarrow F}/\varepsilon = 0.1995$. The stable phase in each region is explicitly marked. The error in $\Delta g/k_B T$ is estimated to be $< 0.012\, k_B T$.

[1] In the Frenkel–Ladd scheme,[190] we take the rotational energy to be $u_{\text{Ein, or, } i} = \Lambda_{\text{rot}} \sin^2\left[\frac{p}{2}\left(\varphi_i - \varphi_{i,\text{orig}}\right)\right]$, where Λ_{rot} is a constant that we vary and $p = 5$ is the number of patches. We thus ensure that the particle's symmetry is accounted for: rotations into degenerate positions give the same Frenkel–Ladd energy.

free energy of an approximant crystal phase that is based on the most common packing of dodecagons in the quasicrystal; it is never thermodynamically stable. To rationalise why the quasicrystal is stable, we also calculate the enthalpy (obtained directly from simulations) and hence obtain the entropy as a function of the temperature. These functions are also depicted in figure B.4, and it is clear that in the region in which the quasicrystal is stable, it does indeed have a higher enthalpy than the σ phase, but its significantly higher entropy nevertheless allows it to become more stable. The origin of this entropy is a combination of the differing vibrational properties and the configurational disorder present in the quasicrystalline phase. If we assume the vibrational properties of the quasicrystal and its approximant are identical, the configurational component of this entropy can be estimated from the entropy difference between the quasicrystal and its approximant, $\Delta s/k_\mathrm{B} \approx 0.113$ at $T_{\sigma\leftrightarrow\mathrm{QC}}$.[2]

From the free energies in figure B.4, we can determine two coexistence points on the phase diagram. We construct the remainder of the phase diagram partly through the same procedure as above at, say, other pressures, but also through the use of Gibbs–Duhem integration.[190;458;459] This method allows us to integrate Clapeyron equation analogues in order to obtain new coexistence points from already known coexistence points, but requires metastability and hysteresis in the transition so that the simulated phases retain their identity when simulated at and around the coexistence point. As a result, the method cannot be applied to the QC-F and QC-Z transitions for much of the phase diagram due to their relative reversibility.

The resulting phase diagram is shown in figure B.5. We note that the quasicrystal is stable for a wide range of pressures. At very high pressures, the Z phase becomes stable due to its greater density. The Z phase is a plastic crystal, meaning that particles are able to rotate relatively freely: this is understandable, as there is no obvious preferred way to orientate a

[2] This value of Δs is significantly larger than the tiling entropy of a random Stampfli tiling, but somewhat smaller than that of a maximally random square-triangle tiling, where both have been evaluated at zero phason strain.[457]

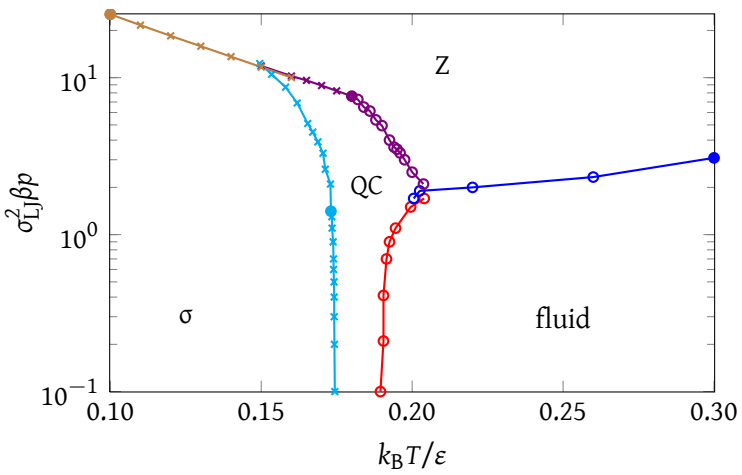


Figure B.5: The βp-T phase diagram. Markers show the technique used to obtain each point: Frenkel–Ladd and thermodynamic integration (filled circles), Gibbs–Duhem integration (crosses) and direct simulation (open circles). $\sigma_{pw} = 0.49$, $n = 6$.

five-fold particle in a six-fold environment.[3] No orientational ordering ensues in the Z phase even at low temperatures. Whether it would form an orientationally-ordered crystal or an orientational glass at sufficiently low T is not clear. It is noteworthy that for the QC-Z transition, $\mathrm{d}p/\mathrm{d}T < 0$; since $\rho_Z > \rho_{QC}$, the Clapeyron equation implies that $S_Z > S_{QC}$: this is likely to stem from the orientational entropy of the plastic crystal. In addition to the two triple points (σ-QC-Z and QC-Z-F), it is interesting to note that there is a range of pressures where the stable thermodynamic phase changes from F to Z to QC to σ as the system is cooled.

We can also look at the effect of modifying potential parameters on the phase diagram, thus allowing us to assess the extent of quasicrystal stability and how finely-tuned the particle properties would need to be in order to observe a quasicrystalline phase in experiment. To determine the phase diagram as the potential parameters change, we additionally use hamiltonian Gibbs–Duhem integration, a technique analogous to Gibbs–Duhem integration, except that we numerically integrate a Clapeyron equation that has been generalised to allow for changes to the potential itself.[190]

[3] We have also considered the additional non-plastic hexagonal crystal structures suggested by Doppelbauer and co-workers for pentavalent patchy particles.[444] but these were found quickly to lose their rotational specificity at the temperatures used in our simulations.

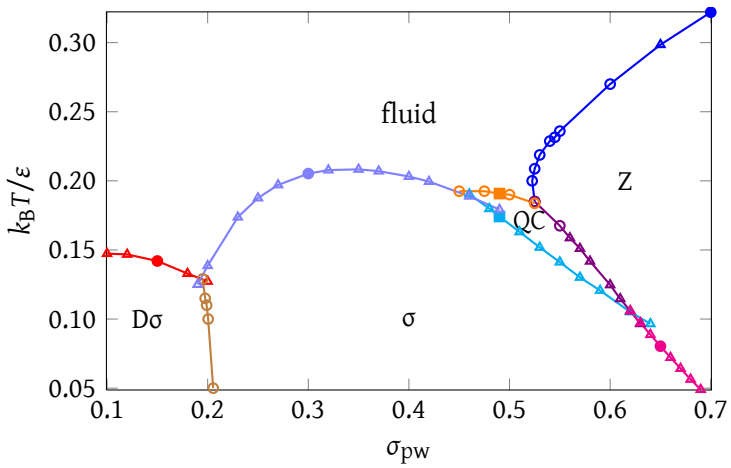


Figure B.6: The T-σ_{pw} phase diagram. Markers show the technique used to obtain each point: Frenkel–Ladd and thermodynamic integration (filled circles), hamiltonian Gibbs–Duhem integration (triangles) and direct simulation (open circles). Points obtained from the previous phase diagram are shown as squares. $\sigma_{LJ}^2 \beta p = 0.5, n = 6$.

We first investigate the behaviour of the system as the patch width is varied. From the phase diagram depicted in figure B.6, we see that the quasicrystal is only stable for a limited range of σ_{pw}, although the range may perhaps be wider at other pressures. As the patch width narrows, the quasicrystal becomes increasingly enthalpically destabilised with respect to the σ phase, as the many six-fold environments cannot satisfactorily fulfil patch-patch interactions. By contrast, as the patch width increases, the quasicrystal becomes more stable with respect to the σ phase, but the hexagonal phase is stabilised even more: as the patches are wider and closer to the isotropic case, six-fold environments are enthalpically preferred.

Another noteworthy feature of this phase diagram is that below $\sigma_{pw} \approx 0.35$, $T_{\sigma \leftrightarrow F}$ begins to decrease: as the patches become narrower, there is an increasing energetic penalty for patches that do not point directly at each other. For very narrow patches, the σ phase transforms into a distorted σ phase (Dσ).[444] in which three of the five patches point directly at the patches of neighbouring particles, thus breaking the square symmetry of the lattice. The reason for this behaviour is that at narrow patch widths, only 'perfect' connections count, as a slight misalignment of the patches rapidly reduces the interparticle attraction, and three

perfect interactions become favoured over five imperfect ones.

Because colloidal interactions are often quite short-ranged, we also wish to investigate what happens to the system when we change the Lennard-Jones exponent (n), which changes the range of the potential and the width of the potential well. We show the phase diagram as a function of n in figure B.7. The effects of increasing n on isotropic potentials are well understood:[460;461] the liquid phase is energetically destabilised because the intrinsic disorder in interparticle neighbour distances is penalised by the narrower potential.[462] In figure B.7, the range of stability of the quasicrystal initially increases. The enthalpies of the QC and σ phases increase only slightly as n increases, because their relative order means that most interparticle neighbour distances can be close to optimal, but do so more for the quasicrystal, as some disorder is typically present in the quasicrystal configurations (for example, particles with unclassified environments in figure B.1(a)); the σ-QC coexistence temperature does therefore increase slightly. The quasicrystal's enthalpy increases slightly more than that of the σ phase, so the QC-σ coexistence temperature does therefore increase slightly. By contrast, the fluid is initially much more enthalpically penalised: the fluid enthalpy increases rapidly, and the

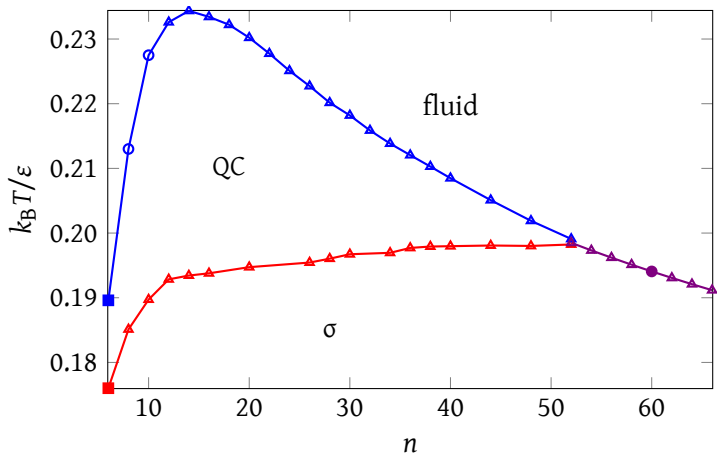


Figure B.7: The T-n phase diagram. Markers show the technique used to obtain each point: Frenkel–Ladd and thermodynamic integration, verified by direct coexistence (filled circles), hamiltonian Gibbs–Duhem integration (triangles) and direct simulation (open circles). Points obtained from the previous phase diagram are shown as squares. $\sigma_{LJ}^2 \beta p = 0.5$, $\sigma_{pw} = 0.49$.

fluid density decreases as n increases. At the maximum in $T_{\mathrm{QC}\leftrightarrow\mathrm{F}}$ ($n \approx 14$), the size of the temperature window of quasicrystal stability is three times that of the $n = 6$ Lennard-Jones potential. However, a narrower potential well associated with increasing n also means that there is less vibrational entropy associated with the ordered structures, and beyond $n \approx 14$, this effect becomes dominant. The QC-F coexistence temperature decreases until the quasicrystal loses stability at $n \approx 53$, and the σ-F coexistence temperature continues to decrease with n beyond the point at which the quasicrystal disappears.

The potential at the higher values of n shown in figure B.7 is very short-ranged indeed. The fact that the quasicrystal maintains and even increases its window of stability for ranges of attraction typical of colloidal particles suggests that this is a robust, general result and that stable quasicrystals can be expected to form in experimental realisations of the system of study.

There are several potential approaches to such an experimental realisation. If patchy colloidal particles similar to the ones studied here could be created, it would be reasonable to expect that a quasicrystal may form when they are confined in two dimensions. For example, Chen and co-workers observed the formation of a 2D kagomé lattice from tri-block Janus particles by introducing a density mismatch with the solvent to confine their colloidal system into two dimensions.[451] A possible alternative to using colloidal patchy particles might involve the use of DNA multi-arm motifs.[463–466] for which a wide variety of 2D crystalline arrays have been observed for motifs with different numbers of arms, including a σ phase for five-arm motifs that is analogous to what we see in the current model when patches are narrow. However, such DNA motifs have a well-defined 'valence', and there is no equivalent of the patch width that could be tuned to make two co-ordination numbers compete. Therefore, a possible approach to forming a DNA quasicrystal might be to use a two-component mixture of five-arm and six-arm motifs of the appropriate composition.

B.5 Conclusion

In summary, we have shown how we can compute if and where a quasicrystalline phase is thermodynamically stable. To the best of our knowledge, this is the first such calculation of the chemical potential of a quasicrystal and the associated phase diagrams that has been obtained directly from simulations with no approximations. For our patchy particle system, we found that the quasicrystalline phase is stable over a significant portion of the phase diagram and is stabilised primarily by its configurational entropy. The quasicrystal we studied involves both five and six co-ordinate sites, and is found to be stable for intermediate patch widths where both environments are competitive. The quasicrystal disappears at narrower patch widths, where the five-co-ordinate environment dominates, and at larger patch widths, where the six-co-ordinate environment dominates. It is robust to parameter changes in the model, which inspires confidence that such a thermodynamically stable quasicrystal might be experimentally realised.

C Rate of nucleation in classical nucleation theory

In classical nucleation theory, clusters are assumed to grow stepwise by the addition of single molecules, rather than by merging clusters. This is a sensible assumption, because the formation of any reasonably sized cluster involves a significant free energy barrier: the formation of two clusters which can interact is thus less probable still. By contrast, the (liquid) monomers are in huge excess. The converse, namely that clusters shrink by stepwise removal of single molecules, is also true from the principle of microscopic reversibility.[467] A cluster comprising n molecules can thus be formed in the following two ways (assuming a lack of memory and constant temperature[12]):

$$\mathrm{C}_{n-1} + \mathrm{C}_1 \underset{k'_n}{\overset{k_{n-1}}{\rightleftharpoons}} \mathrm{C}_n, \tag{C.1}$$

$$\mathrm{C}_n + \mathrm{C}_1 \underset{k'_{n+1}}{\overset{k_n}{\rightleftharpoons}} \mathrm{C}_{n+1}. \tag{C.2}$$

Here, C_i refers to a cluster comprising i molecules. The number of monomers, C_1, is in huge excess, so we can use pseudo-first order rate constants. The rate of change in the number of clusters[1] composed of i molecules, N_i, is therefore easily expressed as[12]

$$\frac{\mathrm{d}N_n(t)}{\mathrm{d}t} = k_{n-1}N_{n-1}(t) - k'_n N_n(t) - k_n N_n(t) + k'_{n+1}N_{n+1}(t), \tag{C.3}$$

[1] Although the derivation presented here is given in terms of particle numbers, the procedure is unchanged if we use concentrations instead; indeed, experimental rates are typically reported per unit volume.

and the net nucleation rate at which clusters of size n become clusters of size $n+1$ is given by the total flux J at the fictional position $n+\frac{1}{2}$, yielding[12]

$$J_{n+\frac{1}{2}}(t) = N_n(t)k_n - N_{n+1}(t)k'_{n+1}, \tag{C.4}$$

so that

$$\frac{\mathrm{d}N_n(t)}{\mathrm{d}t} = J_{n-\frac{1}{2}}(t) - J_{n+\frac{1}{2}}(t). \tag{C.5}$$

To obtain the values of $N_n(t)$, it is reasonable to assume that the Boltzmann distribution holds for clusters of various sizes being formed.[48] It is then possible to derive, based on a perfect gas of clusters assumption,[47;50] the equilibrium size distribution

$$N_n = N \exp\left[-\frac{\Delta G_n}{k_B T}\right], \tag{C.6}$$

where N refers to the total number of molecules in the liquid phase and ΔG_n refers to the Gibbs energy change on the formation of a cluster of size n of the stable phase. As this is the equilibrium distribution, it does not depend on time. It is important to note that, although the Boltzmann distribution predicts the number of clusters to begin rising after the maximum has been reached, this is not physically sound; the equilibrium distribution can only hold up to the critical nucleus size at best.[46]

It is possible to derive an expression for the rate of nucleation using this approach by assuming that there is no back flux once the critical nucleus size has been reached, following Volmer and Weber;[42] however, we can also consider a kinetic approach based on the steady-state assumption; in fact, the resulting formula for vapour to liquid condensation is identical to the one developed using a quasi-equilibrium approach due to the various assumptions that must be employed later in the derivation, so in a sense adopting such a kinetic approach to nucleation is often simply a matter of aesthetics.

In the steady-state approach of Becker and Döring,[43] the distribution of cluster sizes is

also not time-dependent; it is assumed that after some initial relaxation time, during which the distribution rapidly changes, a steady state emerges (for as long as the depletion of monomers is not significant).[12] In the steady state, all fluxes are equal to some constant flux J at all times, and this is the flux that we call the rate of nucleation and that we can measure experimentally.

For future convenience, we define a function $f(n)$ by the difference equation[12]

$$f(n+1) = \frac{k_n}{k'_{n+1}} f(n), \tag{C.7}$$

and $f(1) = 1$. By simply rewriting it as

$$f(n+1) = \frac{k_n}{k'_{n+1}} \cdot \frac{k_{n-1}}{k'_n} \cdot \cdots \cdot \frac{k_1}{k'_2} \cdot 1, \tag{C.8}$$

we can immediately write the solution to this recurrence,

$$f(n) = \prod_{i=1}^{n-1} \frac{k_i}{k'_{i+1}}. \tag{C.9}$$

During nucleation, the forward rate should dominate for sufficiently high values of n (past the maximum in the free energy barrier). Therefore we expect exponential growth of $f(n)$ for large n, as each consecutive factor will be greater than unity. In fact, an alternative definition of the critical cluster is simply that this ratio is one.[467]

Dividing equation (C.4) (with constant J) by $k_n f(n) = k'_{n+1} f(n+1)$ gives[12]

$$\frac{J}{k_n f(n)} = \frac{N_n}{f(n)} - \frac{N_{n+1}}{f(n+1)}. \tag{C.10}$$

Let us now take the sum of this equation from $n = 1$ to $n = n_{\max}$. Since nearly all terms on the right-hand side cancel out, the result is particularly simple,[12]

$$J \sum_{n=1}^{n_{\max}} \frac{1}{k_n f(n)} = N_1 - \frac{N_{n_{\max}+1}}{f(n_{\max}+1)}. \tag{C.11}$$

We note that $N_{n_{\max}+1}$ must be much less than N_1 in order to ensure that we are still in the steady state where the monomer concentration is overwhelmingly high. Since $f(n)$ grows exponentially

for large n, the final term in equation (C.11) can be taken to be nought, and the sum can be extended to infinity. Therefore the nucleation rate in terms of the individual forward and reverse rate constants in classical nucleation theory is[12]

$$J = N_1 \left(\sum_{n=1}^{\infty} \frac{1}{k_n f(n)} \right)^{-1}. \tag{C.12}$$

This is the solution for any system, provided that we know the forward and reverse rate constants for each step in the above scheme. For nucleation of a liquid from a vapour, or crystallisation from a dilute solution, it is possible to derive an elegant final result on the simple assumption that the forward rate is proportional to the gas density whilst the reverse rate is independent of it.[12;43;46;468] This is not feasible when considering crystallisation from the melt.[12]

Instead, we follow the approach of Turnbull and Fisher,[45] where we effectively calculate the individual rate constants using transition state theory. In the generalised reaction step

$$\mathrm{C}_n + \mathrm{C}_1 \rightleftharpoons \mathrm{C}_{n+1}, \tag{C.13}$$

the Gibbs energy of the reactants is ΔG_n and that of the products is ΔG_{n+1} (relative to the Gibbs energy of the metastable liquid). We can consider that the addition of a molecule to a cluster, and the analogous removal of a molecule from it, has with it an associated free energy barrier (a 'transition state'). The notation for the energies is depicted in figure C.1.

The rate of the forward reaction (equation (C.13)) is then given by the Eyring equation,[40]

$$J^+ = N_n \cdot a n^{2/3} \frac{k_\mathrm{B} T}{h} \exp\left[-\frac{\Delta^\ddagger g_1}{k_\mathrm{B} T} \right], \tag{C.14}$$

where $an^{2/3}$ is the number of C_1 particles in contact with each C_n nucleus.[45] The reverse rate is similarly given by

$$J^- = N_{n+1} \cdot a n^{2/3} \frac{k_\mathrm{B} T}{h} \exp\left[-\frac{\Delta^\ddagger g_2}{k_\mathrm{B} T} \right], \tag{C.15}$$

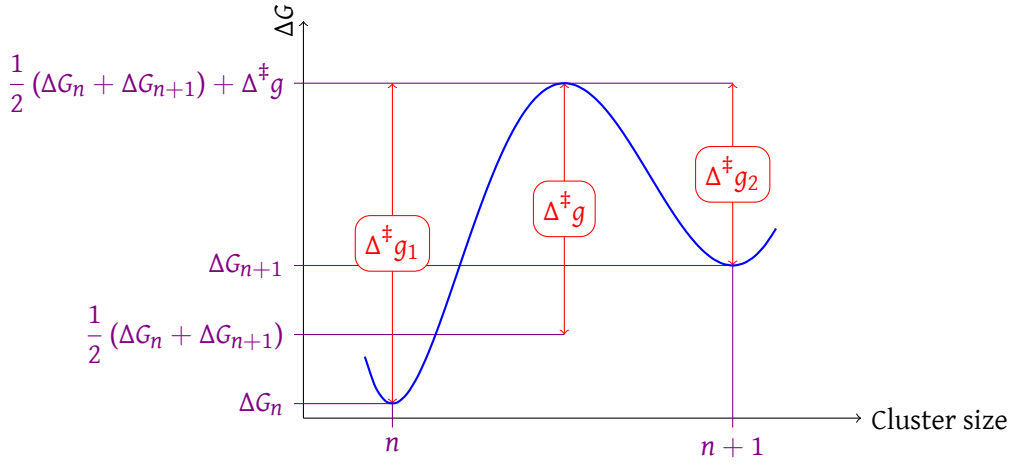


Figure C.1: The notation used in the text in the derivation of the Turnbull–Fisher nucleation rate is presented in this diagram.[45] The free energy maximum represents the activated complex (transition state).

where $an^{2/3}$ is the number of particles within each C_n cluster in contact with C_1 particles.[45] We have assumed that the two surface coefficients (a) are equal, which is reasonable except for very small clusters.[45] The net forward reaction rate is[45]

$$J^{\ddagger} = J^{+} - J^{-} = \frac{an^{2/3} k_B T}{h} \left(N_n \exp\left[-\frac{\Delta^{\ddagger} g_1}{k_B T} \right] - N_{n+1} \exp\left[-\frac{\Delta^{\ddagger} g_2}{k_B T} \right] \right). \tag{C.16}$$

We can treat N_n and ΔG_n as continuous functions except for very small clusters. We can write changes in N_n as n changes by one unit as $\frac{\partial N_n}{\partial n}$, and likewise, $\frac{\partial \Delta G_n}{\partial n}$ is a good approximation for the change in Gibbs energy when n varies by one unit.[45] We also assume that[45]

$$\frac{1}{k_B T} \frac{\partial \Delta G_n}{\partial n} \ll 1 \qquad \text{and} \qquad \frac{1}{N_n} \frac{\partial N_n}{\partial n} \ll 1. \tag{C.17}$$

We can thus rewrite

$$\Delta^{\ddagger} g_1 = \Delta^{\ddagger} g + \frac{1}{2}(\Delta G_n + \Delta G_{n+1}) - \Delta G_n \approx \Delta^{\ddagger} g + \frac{1}{2} \frac{\partial \Delta G_n}{\partial n} \tag{C.18}$$

and analogously

$$\Delta^{\ddagger} g_2 \approx \Delta^{\ddagger} g - \frac{1}{2} \frac{\partial \Delta G_n}{\partial n}. \tag{C.19}$$

Therefore the total rate can be rewritten as

$$J^{\ddagger} = \frac{an^{2/3}k_{\mathrm{B}}T}{h}\left(N_n \exp\left[-\frac{\Delta^{\ddagger}g}{k_{\mathrm{B}}T}\right]\exp\left[-\frac{1}{2k_{\mathrm{B}}T}\frac{\partial\Delta G_n}{\partial n}\right] - N_{n+1}\exp\left[-\frac{\Delta^{\ddagger}g}{k_{\mathrm{B}}T}\right]\exp\left[+\frac{1}{2k_{\mathrm{B}}T}\frac{\partial\Delta G_n}{\partial n}\right]\right). \tag{C.20}$$

By performing a series expansion ($e^x \approx 1 + x$ for small x), we can express the exponential terms (assuming equation (C.17) holds) as

$$\exp\left[\pm\frac{1}{2k_{\mathrm{B}}T}\frac{\partial\Delta G_n}{\partial n}\right] \approx 1 \pm \frac{1}{2k_{\mathrm{B}}T}\frac{\partial\Delta G_n}{\partial n}. \tag{C.21}$$

By also rewriting $N_n + N_{n+1} \approx 2N_n$, we write

$$J^{\ddagger} = \frac{an^{2/3}k_{\mathrm{B}}T}{h}\exp\left[-\frac{\Delta^{\ddagger}g}{k_{\mathrm{B}}T}\right]\left(N_n - N_{n+1} - 2N_n\left(\frac{1}{2k_{\mathrm{B}}T}\frac{\partial\Delta G_n}{\partial n}\right)\right), \tag{C.22}$$

and rewriting the change in N as a differential,

$$J^{\ddagger} = \frac{an^{2/3}k_{\mathrm{B}}T}{h}\exp\left[-\frac{\Delta^{\ddagger}g}{k_{\mathrm{B}}T}\right]\left(-\frac{\partial N_n}{\partial n} - N_n\left(\frac{1}{k_{\mathrm{B}}T}\frac{\partial\Delta G_n}{\partial n}\right)\right). \tag{C.23}$$

Let $\mathcal{R}$ be defined as

$$\mathcal{R} = \frac{J^{\ddagger}h}{ak_{\mathrm{B}}T\exp\left[-\frac{\Delta^{\ddagger}g}{k_{\mathrm{B}}T}\right]}, \tag{C.24}$$

whereby equation (C.23) reduces[2] into the differential equation[45]

$$\frac{\mathrm{d}N_n}{\mathrm{d}n} + \frac{1}{k_{\mathrm{B}}T}\frac{\mathrm{d}\Delta G_n}{\mathrm{d}n}N_n = -\mathcal{R}n^{-2/3}. \tag{C.25}$$

This is a linear inhomogeneous first order differential equation, whose solution can be obtained by the integrating factor method, and is given by

$$N_n = \exp\left[-\frac{\Delta G_n}{k_{\mathrm{B}}T}\right] \times \left(N_{n_0}\exp\left[\frac{\Delta G_{n_0}}{k_{\mathrm{B}}T}\right] - \mathcal{R}\int_{n_0}^{n}\exp\left[\frac{\Delta G_\eta}{k_{\mathrm{B}}T}\right]\eta^{-2/3}\,\mathrm{d}\eta\right). \tag{C.26}$$

Here, N_{n_0} is the steady state number of clusters comprising n_0 particles, and η is a dummy variable of integration. The steady state number of nuclei sufficiently below the critical nucleus

[2] We assume that we are working at constant temperature, so that we can convert partial derivatives into total derivatives.

is not considerably different from the equilibrium concentration. We can assume that, for example, even at $n_0 = n_{\text{crit}}/3$, the equilibrium number (equation (C.6)) holds,[45] and write

$$N_{n_0} = N \exp\left[-\frac{\Delta G_{n_0}}{k_B T}\right], \tag{C.27}$$

where N is the total number of particles in the liquid phase. We can therefore rewrite the steady state equation as

$$N_n = \exp\left[-\frac{\Delta G_n}{k_B T}\right] \times \left(N - \mathcal{R}\int_{n_0}^{n} \exp\left[\frac{\Delta G_\eta}{k_B T}\right] \eta^{-2/3}\, \mathrm{d}\eta\right). \tag{C.28}$$

In the limit of infinitely large clusters, $\lim_{n\to\infty} \exp\left[-\frac{\Delta G_n}{k_B T}\right] > 1$, since ΔG_n will be favourable (*i.e.* negative). Furthermore, the number of clusters of this size must vanish in this limit (as every cluster will quickly grow; *i.e.* the quasi-equilibrium equation does not hold for clusters larger than the critical cluster). Therefore it must be the case that

$$\lim_{n\to\infty}\left(N - \mathcal{R}\int_{n_0}^{n} \exp\left[\frac{\Delta G_\eta}{k_B T}\right] \eta^{-2/3}\, \mathrm{d}\eta\right) = 0, \tag{C.29}$$

and therefore

$$\mathcal{R} = N\left(\int_{n_0}^{\infty} \exp\left[\frac{\Delta G_\eta}{k_B T}\right] \eta^{-2/3}\, \mathrm{d}\eta\right)^{-1} \equiv N/I, \tag{C.30}$$

where we define I as the integral in this expression for future convenience.

Since ΔG is largest for clusters near the critical cluster size, it dominates the integral due to the exponential dependence; the lower bound is therefore unimportant (provided it is not too close to the critical size). We can also approximate the Gibbs energy term with the first two non-zero Taylor expansion terms about n_{crit} (noting that the first derivative is zero by construction),

$$\Delta G_\eta \approx \Delta G_{n_{\text{crit}}} + \frac{1}{2}\Delta G''_{n_{\text{crit}}}\left(\eta - n_{\text{crit}}\right)^2. \tag{C.31}$$

Furthermore, since the exponential has a very sharp maximum at $\eta = n_{\text{crit}}$ (see figure C.2), we can replace the $\eta^{-2/3}$ term by its value at $\eta = n_{\text{crit}}$ without introducing significant error.[46]

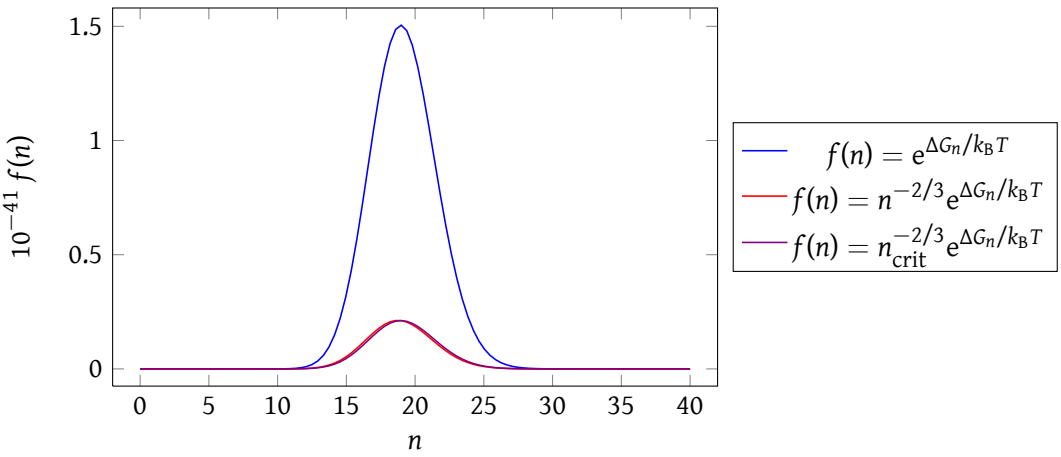


Figure C.2: This shows the sharp maximum of the function $e^{\Delta G_n/k_B T}$, where the Gibbs energy function is a generalised form of equation (2.9), $\Delta G_n/k_B T = An^{2/3} - Bn$. In this particular example, $A = 40$ and $B = 10$, so that $n_{crit} = \left(2A/3B\right)^3 \approx 19$. There is essentially no difference between the second curve and the equivalent when using a fixed scaling factor of $n_{crit}^{-2/3}$.

This allows us to rewrite the integral I as[3]

$$I = n_{crit}^{-2/3} \exp\left[\frac{\Delta G_{n_{crit}}}{k_B T}\right] \left(\int_{n_0}^{\infty} \exp\left[\frac{1}{2k_B T} \Delta G''_{n_{crit}} (\eta - n_{crit})^2\right] \mathrm{d}\eta\right). \tag{C.32}$$

It is readily seen from figure C.2 that the lower bound of the integral can be recast as $-\infty$ for convenience, and the integral can be evaluated as[4]

$$I = n_{crit}^{-2/3} \exp\left[\frac{\Delta G_{n_{crit}}}{k_B T}\right] \sqrt{\frac{2\pi k_B T}{-\Delta G''_{n_{crit}}}}. \tag{C.33}$$

Therefore

$$\mathcal{R} = \frac{J^{\ddagger} h}{a k_B T \exp\left[-\frac{\Delta^{\ddagger} g}{k_B T}\right]} = N \left(n_{crit}^{-2/3} \exp\left[\frac{\Delta G_{n_{crit}}}{k_B T}\right] \sqrt{\frac{2\pi k_B T}{-\Delta G''_{n_{crit}}}}\right)^{-1}, \tag{C.34}$$

and the rate of nucleation can be written as

$$J^{\ddagger} = \frac{aNk_B T n_{crit}^{2/3}}{h} \sqrt{-\frac{\Delta G''_{n_{crit}}}{2\pi k_B T}} \exp\left[-\frac{1}{k_B T}\left(\Delta^{\ddagger} g + \Delta G_{n_{crit}}\right)\right]. \tag{C.35}$$

[3] Note that this procedure is different from the one used by Turnbull and Fisher,[45] since the integral in equation (C.30) was misevaluated in that paper.[468] A procedure analogous to the Zeldovich derivation[44;46] is adopted instead.

[4] To evaluate a gaussian integral, we use the standard result $\int_{-\infty}^{\infty} \exp\left[-ax^2\right] \mathrm{d}x = \sqrt{\frac{\pi}{a}}$.

We can recast this as

$$J^{\ddagger} = Nk_{\text{crit}}Z\exp\left[-\frac{\Delta G_{n_{\text{crit}}}}{k_{\text{B}}T}\right], \tag{C.36}$$

where

$$Z = \sqrt{-\frac{\Delta G''_{n_{\text{crit}}}}{2\pi k_{\text{B}}T}} \tag{C.37}$$

is the Zeldovich factor[44;52] and k_{crit} has been recovered from the Eyring formulation in equation (C.14). Equation (C.36), and its concentration-based analogue (where the density replaces the number of particles), are especially useful forms of the rate equation for use in simulations.

Alternatively, we can insert the classical nucleation result for the second derivative at the critical nucleus (calculated from equation (2.9)),

$$\Delta G''_{n_{\text{crit}}} = -\frac{2}{3}\gamma\left(\frac{4\pi}{3}\cdot\frac{1}{n_{\text{crit}}^4\rho^2}\right)^{1/3}, \tag{C.38}$$

we can rewrite the rate equation as

$$J^{\ddagger} = a\frac{Nk_{\text{B}}T}{h}\left(\frac{2}{9\pi\rho}\right)^{1/3}\left(\frac{\gamma}{k_{\text{B}}T}\right)^{1/2}\exp\left[-\frac{1}{k_{\text{B}}T}\left(\Delta^{\ddagger}g + \Delta G_{n_{\text{crit}}}\right)\right], \tag{C.39}$$

which can for many systems of interest be approximated by

$$J^{\ddagger} \approx \frac{Nk_{\text{B}}T}{h}\exp\left[-\frac{1}{k_{\text{B}}T}\left(\Delta^{\ddagger}g + \Delta G_{n_{\text{crit}}}\right)\right], \tag{C.40}$$

since the remaining terms often yield a constant within an order of magnitude from unity and can thus be neglected.[45]

The free energy of activation associated with attaching a new particle to a cluster is typically thought of as the short-range diffusion of particles moving some fraction of the interparticle distance across the interface to join the newly-forming lattice.[45] The self-diffusion

coefficient is often given by the Arrhenius-type equation[5]

$$D = D_0 \exp\left[-\frac{\Delta^{\ddagger} g}{k_B T}\right] \tag{C.41}$$

so that $\ln D = -\frac{\Delta^{\ddagger} g}{k_B T} + \text{constant}$, and the activation energy can then be determined by considering the gradient of the logarithm of the self-diffusion coefficient plotted against the inverse temperature,

$$\Delta^{\ddagger} g = -k_B \frac{\mathrm{d} \ln D}{\mathrm{d} T^{-1}}. \tag{C.42}$$

This is not strictly true, however, because D_0 is a (weak) function of temperature as well, and, more importantly, there is an implicit assumption that the temperature dependence in the exponential is greater than that of the activation energy itself. Unfortunately, in the supercooled region, $\Delta^{\ddagger} g$ itself is a strongly varying function of temperature, so this assumption is not necessarily justified. Alternative approaches to the determination of $\Delta^{\ddagger} g$ exist.[5] In the majority of nucleation work on ice, the nucleation free energy barrier dominates the process, in particular the variation of the rate with the temperature;[174] many studies therefore ignore the diffusion activation free energy altogether.

D More on simulation methods

D.1 Rigid body rotations and quaternions

When thinking about how to specify the rotation of a rigid molecule, perhaps the most intuitive way is to specify an axis of rotation and an angle of rotation (in the 'laboratory' co-ordinate system). This poses some problems in practice because generating a random change in rotation that is uniformly distributed is very difficult in this representation, making it inconvenient to use in simulations. Another way of achieving rotations is to specify a set of three angles called Euler angles. In the z-x-z convention, where $(x,\ y,\ z)$ are the laboratory axes and $(X,\ Y,\ Z)$ are the internal axes, the Euler angles are[303;469]

- ϕ, the angle between the x-axis and the vector $\hat{\boldsymbol{z}} \times \hat{\boldsymbol{Z}}$;
- θ, the angle between the z-axis and the Z-axis; and
- ψ, the angle between the vector $\hat{\boldsymbol{z}} \times \hat{\boldsymbol{Z}}$ and the X-axis.

The difficulties arising in using this method include the problem of gimbal lock. This occurs when, for example, the angle θ is zero or π: in that case, only $\phi + \psi$ is defined, but not each of them individually. In other words, when we happen to rotate a molecule such that it is centred on one of the axes, we lose one degree of freedom. Since the Euler angles are effectively equivalent to three rotations about each of the axes in turn, consider for example a rotation in the y-axis rotating a vector parallel to the x-axis so that the vector is parallel to the z-axis. Now any rotations about the z-axis will have no effect on this final vector. A further difficulty in using Euler angles is that when generating positions – which are needed when calculating

the energy, for example – a series of computationally expensive and occasionally numerically unstable trigonometric functions need to be evaluated.

An elegant solution to this problem is to use unit quaternions,[470] which effectively represent a rotation on a four-dimensional unit sphere. A quaternion is represented by a four-dimensional vector $\boldsymbol{q} = (w \quad x \quad y \quad z)^{\mathrm{T}}$, which we can also consider to be an extended complex number, $\boldsymbol{q} = w + \boldsymbol{i}x + \boldsymbol{j}y + \boldsymbol{k}z$. The individual values are easily related to the Euler angles above by[66]

$$w = \cos\left(\frac{\theta}{2}\right)\cos\left(\frac{\phi+\psi}{2}\right), \tag{D.1}$$

$$x = \sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\phi-\psi}{2}\right), \tag{D.2}$$

$$y = \sin\left(\frac{\theta}{2}\right)\sin\left(\frac{\phi-\psi}{2}\right), \tag{D.3}$$

$$z = \cos\left(\frac{\theta}{2}\right)\sin\left(\frac{\phi+\psi}{2}\right). \tag{D.4}$$

A rotation of a vector $\boldsymbol{v}$ (which can, for example, represent the position of each atom within the molecule whose centre of mass lies at the internal co-ordinate origin) by the quaternion $\boldsymbol{q}$ to give the rotated vector $\boldsymbol{v}'$ is given by $\boldsymbol{q}\cdot\boldsymbol{v}\cdot\boldsymbol{q}^{-1} = \boldsymbol{v}'$. The inverse rotation is therefore simply $\boldsymbol{q}^{-1}\boldsymbol{q}\cdot\boldsymbol{v}\cdot\boldsymbol{q}^{-1}\boldsymbol{q} \equiv v \equiv \boldsymbol{q}^{-1}\cdot\boldsymbol{v}'\cdot\boldsymbol{q}$.

The multiplication of two quaternions $\boldsymbol{q}_a$ and $\boldsymbol{q}_b$ (where a 'normal' vector is simply a quaternion with a real value of nought) is given by[470;471]

$$\boldsymbol{q}_a\cdot\boldsymbol{q}_b \equiv \begin{pmatrix} a_w \\ a_i \\ a_j \\ a_k \end{pmatrix} \cdot \begin{pmatrix} b_w \\ b_i \\ b_j \\ b_k \end{pmatrix} = \begin{pmatrix} a_w b_w - a_i b_i - a_j b_j - a_k b_k \\ a_w b_i + a_i b_w + a_j b_k - a_k b_j \\ a_w b_j - a_i b_k + a_j b_w + a_k b_i \\ a_w b_k + a_i b_j - a_j b_i + a_k b_w \end{pmatrix}. \tag{D.5}$$

The inverse of a quaternion is therefore very simply its complex conjugate (as long as the quaternion has been normalised), $\boldsymbol{q}^{\star} \equiv (w \quad -x \quad -y \quad -z)^{\mathrm{T}}$.

The rotation matrix representing a rotation by a quaternion $\boldsymbol{q} = (w \quad x \quad y \quad z)^{\mathrm{T}}$ is given by[66]

$$\boldsymbol{R} = \begin{pmatrix} w^2 + x^2 - y^2 - z^2 & 2xy - 2wz & 2wy + 2xz \\ 2wz + 2xy & w^2 - x^2 + y^2 - z^2 & 2yz - 2wx \\ 2xz - 2wy & 2wx + 2yz & w^2 - x^2 - y^2 + z^2 \end{pmatrix}. \tag{D.6}$$

It is worth noting that the quaternion $-\boldsymbol{q}$ results in the same rotation as the quaternion $\boldsymbol{q}$.

The major advantage of using this method is that there are no trigonometric functions to evaluate; furthermore, it is easier to store a quaternion than a full rotation matrix. In a Monte Carlo simulation, we generate a random move by constructing a random quaternion on a sphere[472] and adding it to the original quaternion, and finally normalising the resulting quaternion.

D.2 Detailed balance in hybrid Monte Carlo simulations

Here, we more formally show that hybrid Monte Carlo simulations with time-reversible molecular dynamics schemes satisfy detailed balance. In section 3.1, we assumed that the transition probability α was symmetric and it thus cancelled out in the acceptance criterion. This is not fulfilled in hybrid Monte Carlo, and so we need to include it explicitly. The standard detailed balance condition in a Monte Carlo simulation is thus

$$P_{\text{old}}\alpha(\text{old} \to \text{new})\Theta(\text{old} \to \text{new}) = P_{\text{new}}\alpha(\text{new} \to \text{old})\Theta(\text{new} \to \text{old}) \tag{D.7}$$

and the Metropolis acceptance criterion associated with it is

$$\Theta(\text{old} \to \text{new}) = \min\left\{1, \frac{P_{\text{new}}\alpha(\text{new} \to \text{old})}{P_{\text{old}}\alpha(\text{old} \to \text{new})}\right\}. \tag{D.8}$$

For convenience, the 'old' and the 'new' state can be written as pairs of generalised position and momentum co-ordinates (x, p) and (x', p'), respectively.[119;120] Each MD step is deterministic,

so α(old $\to$ new) depends only on the choice of momenta,[119;120]

$$\alpha(p) \propto \exp\left[-\beta \sum_{i=1}^{N} p_j^2/2m_j\right] = \exp\left[-\beta K\right]. \tag{D.9}$$

Let g be a function describing the change in position and momentum following an MD move (sometimes, this function is known as the MD propagator[119;120]), such that $g^{\delta t}(x, p) = (x', p')$ for a time step δt. The acceptance criterion can thus be recast as

$$\Theta((x, p) \to g^{\delta t}(x, p)) = \min\left\{1, \exp\left[-\beta\Delta(U + K)\right]\right\} \tag{D.10}$$

$$= \min\left\{1, \exp\left[-\beta\Delta H\right]\right\}, \tag{D.11}$$

where H is the 'hamiltonian' giving the total potential and kinetic energy. For later use, we derive the following.[39;120] We take the factor $e^{-\beta H(g^{\delta t}(x,p))}$ out of $P(x, p)\Theta((x, p) \to g^{\delta t}(x, p))$ and then rearrange the expression in terms of the acceptance criterion for the reverse move,

$$P(x, p)\Theta((x, p) \to g^{\delta t}(x, p)) = \exp[-\beta H(x, p)] \min\left\{1, \exp\left[-\beta\Delta H\right]\right\} \tag{D.12}$$

$$= \exp[-\beta H(g^{\delta t}(x, p))] \exp[-\beta H(x, p)]$$

$$\times \min\left\{\exp[+\beta H(g^{\delta t}(x, p))], \exp\left[\beta H(x, p)\right]\right\} \tag{D.13}$$

$$= \exp[-\beta H(g^{\delta t}(x, p))] \times \min\left\{\exp[+\beta\Delta H], 1\right\} \tag{D.14}$$

$$= P(x', p')\Theta(g^{\delta t}(x, p) \to (x, p)). \tag{D.15}$$

Finally,[120]

$$P_{\text{old}}\alpha(\text{old} \to \text{new})\Theta(\text{old} \to \text{new}) = P(x)\alpha(p)\Theta((x, p) \to g^{\delta t}(x, p)) \qquad \text{use (D.15)} \tag{D.16}$$

$$= P(x')\alpha(p')\Theta(g^{\delta t}(x, p) \to (x, p)) \tag{D.17}$$

$$= P(x')\alpha(p')\Theta((x', p') \to g^{-\delta t}(x', p')) \tag{D.18}$$

$$= P_{\text{new}}\alpha(\text{new} \to \text{old})\Theta(\text{new} \to \text{old}), \tag{D.19}$$

which demonstrates that detailed balance is obeyed. To proceed from equation (D.17) to equation (D.18), we require that the MD propagator is time reversible,[120] such that $(x, p) = g^{-\delta t}(x', p')$.

The MD propagator must also be symplectic, or area-preserving.[39;120] Technically, what we called probabilities for simplicity above are actually probability densities, and we should in principle have written $P(x, p)\,\mathrm{d}x\,\mathrm{d}p$ for the probability: we require that $\mathrm{d}x\,\mathrm{d}p = \mathrm{d}x'\,\mathrm{d}p'$ for the above derivation to hold.

D.3 Verlet and leapfrog MD integration schemes

In an MD simulation, we compute forces between particles; these are given by the derivatives of the potential energy,[40] $\boldsymbol{f}(r) = -\operatorname{grad} U(r)$, and we can split the vector into its component directions,[66] *e.g.*

$$f_x(r) = -\left(\frac{\partial U(r)}{\partial x}\right)_{y,z} = -\frac{\mathrm{d}U(r)}{\mathrm{d}r}\left(\frac{\partial r}{\partial x}\right)_{y,z} = -\frac{x}{r}\frac{\mathrm{d}U(r)}{\mathrm{d}r}, \tag{D.20}$$

where $r^2 = x^2 + y^2 + z^2$. We can often evaluate the derivative of the potential energy with respect to the particle separation r analytically. In the following, we restrain ourselves to a single dimension, x, and forces, velocities, accelerations and similar are assumed to refer to this dimension only. Analogous expressions hold for y and z dimensions.

Once we have the forces, we can integrate the equations of motion. Following Verlet,[473] we can expand a particle's position at time $t + \delta t$ in a Taylor series about t as[66]

$$x(t+\delta t) = x(t) + \frac{\mathrm{d}x(t)}{\mathrm{d}t}\delta t + \frac{1}{2!}\frac{\mathrm{d}^2x(t)}{\mathrm{d}t^2}\delta t^2 + \frac{1}{3!}\frac{\mathrm{d}^3x(t)}{\mathrm{d}t^3}\delta t^3 + \mathcal{O}(\delta t^4). \tag{D.21}$$

From Newton's second law, $f = m\dfrac{\mathrm{d}^2r}{\mathrm{d}t^2}$, and we can use $\dfrac{\mathrm{d}r}{\mathrm{d}t} = v$. Hence,

$$x(t+\delta t) = x(t) + v(t)\delta t + \frac{f(t)}{2m}\delta t^2 + \frac{1}{3!}\frac{\mathrm{d}^3x(t)}{\mathrm{d}t^3}\delta t^3 + \mathcal{O}(\delta t^4). \tag{D.22}$$

Performing the same expansion at time $t - \delta t$, we find

$$x(t - \delta t) = x(t) - v(t)\delta t + \frac{f(t)}{2m}\delta t^2 - \frac{1}{3!}\frac{\mathrm{d}^3 x(t)}{\mathrm{d}t^3}\delta t^3 + \mathcal{O}(\delta t^4), \tag{D.23}$$

and summing equations (D.22) and (D.23), we obtain an expression for the position of a particle at time $t + \delta t$,[473]

$$x(t + \delta t) = -x(t - \delta t) + 2x(t) + \frac{f(t)}{m}\delta t^2 + \mathcal{O}(\delta t^4), \tag{D.24}$$

which is accurate to third order in time.[66] This is the Verlet integration scheme. Velocities can be calculated by subtracting equation (D.23) from (D.22), giving

$$v(t) = \frac{x(t + \delta t) - x(t - \delta t)}{2\delta t} + \mathcal{O}(\delta t^3). \tag{D.25}$$

It will be helpful to derive the leapfrog velocity and position integration algorithm.[66;474] Note that in the leapfrog algorithm, we do not know the positions and the velocities of particles at the same time; the velocities are half a step out-of-sync with the positions. If we expand $x(t + \delta t/2)$ and $x(t - \delta t/2)$ in a Taylor series about t, we obtain

$$v(t) = \frac{x(t + \delta t/2) - x(t - \delta t/2)}{2\delta t}, \tag{D.26}$$

or equivalently

$$v(t + \delta t/2) = \frac{x(t + \delta t) - x(t)}{\delta t} \tag{D.27}$$

and

$$v(t - \delta t/2) = \frac{x(t) - x(t - \delta t)}{\delta t}. \tag{D.28}$$

Rearranging gives

$$x(t + \delta t) = x(t) + v(t + \delta t/2)\delta t, \tag{D.29}$$

while inserting the Verlet definition of $x(t + \delta t)$ (equation (D.24)) into equation (D.27) gives[66]

$$v(t + \delta t/2) = v(t - \delta t/2) + \frac{f(t)}{m}\delta t. \tag{D.30}$$

Equations (D.29) and (D.30) are the leapfrog algorithm.

There are several considerations that need to be borne in mind when applying the leapfrog algorithm to hybrid Monte Carlo simulations. First of all, an important question is what velocities we must use for the first step of the integration. If we load the velocities or if we generate the positions and the velocities at $t = 0$, then we have $x(0)$ and $v(0)$ for each particle; however, to find $x(\delta t)$, we require the knowledge of $v(\delta t/2)$. Although this is a subtle point, it is nevertheless required in order to ensure we are measuring the kinetic and potential energies at the same position in time. To calculate the half-step velocity, we can simply evaluate

$$v(\delta t/2) = v(0) + \frac{f(0)}{2m}\delta t. \tag{D.31}$$

(Alternatively, we could generate $v(\delta t/2)$ and calculate $v(0)$ by subtraction, as the algorithm is time-reversible.) Furthermore, we do not know the potential and the kinetic energy at the same time in the leapfrog algorithm. Should we require the kinetic energy of the 'current' step, for example in thermostatting, we could likewise calculate

$$v(t) = v(t - \delta t/2) + \frac{f(t)}{2m}\delta t. \tag{D.32}$$

This involves no significant computational cost (except in the final step), as the forces will already have been calculated in order to establish the velocities at step $t + \delta t/2$.

The velocity Verlet algorithm[475] is analogous to the leapfrog algorithm, except that we calculate the velocity explicitly in every step using equation (D.32). In the usual implementation, the calculation is performed in two stages; first, we calculate

$$v(t + \delta t/2) = v(t) + \frac{f(t)}{2m}\delta t. \tag{D.33}$$

From this, we calculate the position at time $t + \delta t$,

$$x(t + \delta t) = x(t) + v(t + \delta t/2)\delta t. \tag{D.34}$$

We now evaluate the forces using these new positions, $f(t + \delta t)$, and in the second stage of the scheme, we calculate the velocity at this new timestep,

$$v(t + \delta t) = v(t + \delta t/2) + \frac{f(t + \delta t)}{2m}\delta t. \tag{D.35}$$

The equations are ultimately the same as in the leapfrog approach, since if we incorporate equation (D.35) into equation (D.33), we recover the leapfrog algorithm; however, as mentioned above, there is little computational overhead, as the forces are required to be calculated in any case.

D.4 Rigid body rotations in MD simulations

Although it is possible to perform simulations of rigid bodies by first performing a non-constrained move and then applying a constraints algorithm, we have encountered significant difficulties doing so. Common constraints algorithms such as RATTLE and SHAKE are symplectic and time reversible if they are iterated to full convergence,[476] and so the analytic analogue SETTLE[131] used for constrained triangular objects (such as water molecules) is also, in principle, symplectic and time reversible.[129] The requirements for use in hybrid Monte Carlo simulations therefore appear to be met; however, in practice, although reasonably long MD simulations result in the correct behaviour, short MD simulations of only a few steps within a hybrid Monte Carlo scheme invariably led to a (catastrophic) decrease in the potential energy of the system. A much more intuitive approach that does not result in such unusual behaviour is to use actual rigid body dynamics, in which atoms only ever move within the rigid framework of the molecule. To achieve this, we use the symplectic and time reversible algorithm of Miller III and co-workers.[129]

We wish to evolve the rotational motion in the same way as we do the translational motion. To do this, we move the centre of mass of every molecule to take into account the

translational motion, and explicitly evaluate the rotational motion by considering not only the forces, but also the torques acting on the system, ignoring those components of the forces that would distort the rigid body structure.[1] The scheme of the previous section can be used for translational motion of the centre of mass, where the force acting on the centre of mass is the sum of the forces on each of the atoms within the molecule,

$$\boldsymbol{F}_{\mathrm{cm}} = \sum_{i=1}^{N_{\mathrm{atoms}}} \boldsymbol{f}_i. \tag{D.36}$$

For rotational motion, we can calculate the torque acting on the rigid molecule, which is given by

$$\boldsymbol{\tau} = \sum_{i=1}^{N_{\mathrm{atoms}}} \boldsymbol{r}_i \times \boldsymbol{f}_i, \tag{D.37}$$

where $\boldsymbol{r}_i$ is the vector from the centre of mass to atom i. The equation of motion for rotational motion is

$$\boldsymbol{\tau} = \frac{\mathrm{d}(\boldsymbol{I}\boldsymbol{\omega})}{\mathrm{d}t}, \tag{D.38}$$

where $\boldsymbol{I}$ is the moment of inertia tensor and $\boldsymbol{\omega}$ is the angular velocity vector. Generally, we will consider motion such that the angular velocities are described in the molecular (principal) frame, which is to say the co-ordinate system in which the moment of inertia tensor is diagonal. We need to account for the fact that the principal frame is rotating;[2] within the principal frame, Euler's equations of motion are[129]

$$\frac{\mathrm{d}\omega_x}{\mathrm{d}t} = \frac{\tau_x}{I_{xx}} + \frac{I_{yy} - I_{zz}}{I_{xx}}\omega_y\omega_z, \tag{D.39}$$

$$\frac{\mathrm{d}\omega_y}{\mathrm{d}t} = \frac{\tau_y}{I_{yy}} + \frac{I_{zz} - I_{xx}}{I_{yy}}\omega_x\omega_z, \tag{D.40}$$

$$\frac{\mathrm{d}\omega_z}{\mathrm{d}t} = \frac{\tau_z}{I_{zz}} + \frac{I_{xx} - I_{yy}}{I_{zz}}\omega_x\omega_y, \tag{D.41}$$

[1] In other words, we are invoking the idea that some unspecified internal force acts to ensure that the body remains rigid.[129]

[2] This means simply taking the relevant time derivatives of the unit vectors themselves as well using a product rule.

where all the angular velocities are expressed in the principal frame. The principal frame angular velocity is related to the quaternion $\boldsymbol{q} = (q_0 \quad q_1 \quad q_2 \quad q_3)^{\mathrm{T}}$ through[129;470;477;478]

$$\frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}t} = \frac{1}{2}\begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix}\begin{pmatrix} 0 \\ \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}, \tag{D.42}$$

where an explicit expression for $\boldsymbol{\omega}$ can be obtained by noting that the 4×4 matrix above is orthogonal (since $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$) and thus easily invertible. However, because these equations relate to the principal frame, they are only valid instantaneously.

We require a symplectic, time reversible algorithm to use in hybrid Monte Carlo simulations (appendix D.2). We use the algorithm proposed by Miller III and co-workers,[129] and in particular the recast version of Kamberaj and co-workers,[130] to achieve this. In this algorithm, the hamiltonian of the system is decomposed exactly into four separate hamiltonians, and then a numerical integration scheme for the hamiltonian dynamics[66] is introduced by using the Trotter identity for the Liouville time propagation operator.[66] Because it uses hamiltonian dynamics, the algorithm is guaranteed to be symplectic,[66] and this is shown explicitly by Miller III and co-workers.[129] In the following, we describe the algorithm as implemented in our code.

We define auxiliary quaternion momenta,[3]

$$\boldsymbol{P} = \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} = 2\begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix}\begin{pmatrix} 0 \\ I_{xx}\omega_x \\ I_{yy}\omega_y \\ I_{zz}\omega_z \end{pmatrix}, \tag{D.43}$$

[3] We follow the nomenclature of Kamberaj and co-workers; this is not, however, the same quaternion as the one denoted $\boldsymbol{P}$ by Miller III and co-workers.

and auxiliary quaternion torques,[4]

$$\boldsymbol{m} = \begin{pmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \end{pmatrix} = 2 \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \begin{pmatrix} 0 \\ \tau_x \\ \tau_y \\ \tau_z \end{pmatrix}. \tag{D.44}$$

We first calculate the auxiliary momenta following a half-step,

$$\boldsymbol{P}' = \boldsymbol{P}(t) + \frac{\delta t}{2}\boldsymbol{m}(t) \tag{D.45}$$

and set $\boldsymbol{q}' = \boldsymbol{q}(t)$. We then apply a sequence of time evolution operators to the quaternions $\boldsymbol{q}$ and the quaternion momenta $\boldsymbol{P}$ of the form[130]

$$\boldsymbol{P}' \longleftarrow \cos(h\phi_i)\,\boldsymbol{P}' + \sin(h\phi_i)\boldsymbol{D}_i\boldsymbol{P}', \tag{D.46}$$

$$\boldsymbol{q}' \longleftarrow \cos(h\phi_i)\,\boldsymbol{q}' + \sin(h\phi_i)\boldsymbol{D}_i\boldsymbol{q}', \tag{D.47}$$

where[130]

$$\phi_i = \frac{1}{4I_i}\left(\boldsymbol{P}'\right)^{\mathrm{T}}\boldsymbol{D}_i\boldsymbol{q}' \tag{D.48}$$

(and is calculated appropriately before each $(\boldsymbol{P}',\ \boldsymbol{q}')$ pair of time evolution operator applications[129]), I_i is the moment of inertia ($I_1 = I_{xx}$, $I_2 = I_{yy}$ and $I_3 = I_{zz}$). $\boldsymbol{D}_i$ is a vector operator given by

$$\boldsymbol{D}_i = \begin{cases} (1 \quad \boldsymbol{i} \quad \boldsymbol{j} \quad \boldsymbol{k}) & \text{if } i = 0, \\ (\boldsymbol{i} \quad -1 \quad -\boldsymbol{k} \quad \boldsymbol{j}) & \text{if } i = 1, \\ (\boldsymbol{j} \quad \boldsymbol{k} \quad -1 \quad -\boldsymbol{i}) & \text{if } i = 2, \\ (\boldsymbol{k} \quad -\boldsymbol{j} \quad \boldsymbol{i} \quad -1) & \text{if } i = 3, \end{cases} \tag{D.49}$$

where $\boldsymbol{i}, \boldsymbol{j}$ and $\boldsymbol{k}$ are unit vectors (see appendix D.1). The time evolution operators are applied in the order $i = 3, i = 2, i = 1$ (with double the timestep), $i = 2, i = 3$, where h is the integration step such that δt is an integer multiple of h (we typically use $\delta t = 2h$). After the relevant

[4] Auxiliary quaternion torques are denoted $\boldsymbol{F}^{(4)}$ by Kamberaj and co-workers; to avoid any confusion with forces, we use the symbol $\boldsymbol{m}$, standing for 'moment'.

calculations have been iterated the required number of times, we evaluate[130]

$$\boldsymbol{P}(t+\delta t) = \boldsymbol{P}' + \frac{\delta t}{2}\boldsymbol{m}(t+\delta t), \tag{D.50}$$

where $\boldsymbol{m}(t+\delta t)$ is evaluated using equation (D.44) at time $t+\delta t$. This is the velocity Verlet analogue for rigid body rotations. The angular velocities can be determined from the final $\boldsymbol{P}$ using equation (D.43), noting that the inverse of the matrix involving elements of $\boldsymbol{q}$ is just the transpose of that matrix, as the matrix is orthogonal.

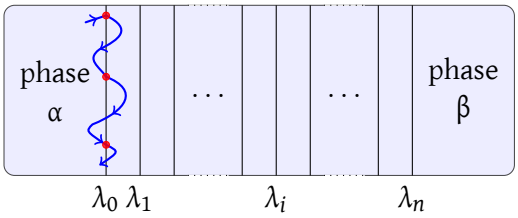
Note that it is no longer possible to construct a leapfrog analogue of the velocity Verlet scheme.[129] Although we can again calculate only half-step auxiliary momenta by combining equations (D.45) and (D.50), and equation (D.45) can be written as

$$\boldsymbol{P}'(t+\delta t/2) = \boldsymbol{P}'(t-\delta t/2) + \delta t\boldsymbol{m}(t), \tag{D.51}$$

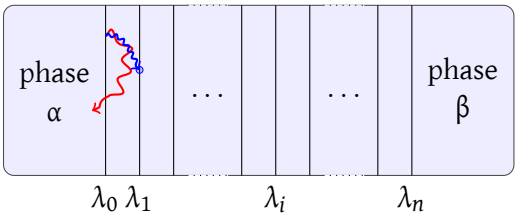
which is exactly analogous to equation (D.30), this is not especially helpful, because $\boldsymbol{Q}$ depends on the evolution of $\boldsymbol{P}'$ in the algorithm. We need to know the angular velocity (or its auxiliary momentum) and the current rotation *at the same instantaneous time* when performing the Trotter factorisation.[129]

D.5 Forward flux sampling

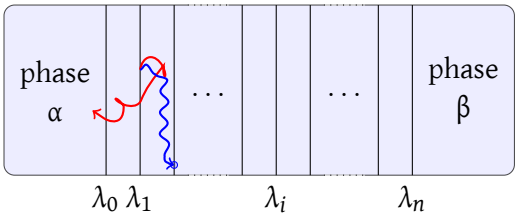
Conceptually, forward flux sampling[50;94–97] is fairly simple and can readily be understood by considering figure D.1. We split the energy landscape between the two phases of interest into individual interfaces (λ_i) based on some sort of order parameter. In our case, phase α might be liquid water and phase β might be ice. After we have generated a variety of starting configurations at the interface separating phase α and the 'interface' region (λ_0), we start some trial number of simulations for each of these interfaces (figure D.1(b)). Any trajectories reaching the next interface are also stored and new simulations are fired off from these saved



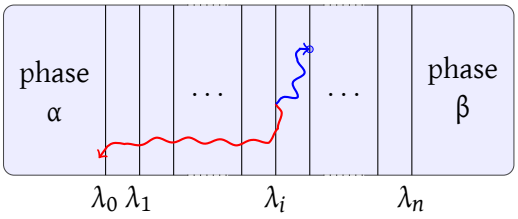
(a) The first stage in the forward flux sampling simulation: generating initial points. A brute force MC simulation is run from phase α and the 'forward' direction crossings into the first interface region are recorded (red circles) and taken into account when calculating the flux.



(b) Simulations are fired off from the first interface (λ_0). We stop if the trajectory returns to the initial state (λ_0), or whenever the next barrier (λ_1) is crossed and save this configuration.



(c) Successful simulations from the previous step are then used to launch new trials from the second interface (λ_1).



(d) The procedure is then repeated until interface λ_n is reached. For a generalised interface λ_i, successful trials reach the interface λ_{i+1}, whilst unsuccessful ones go all the way back to phase α.

Figure D.1: The stages of a forward flux sampling simulation.

configurations (figure D.1(c)). This is then repeated all the way across the interface region until interface λ_n is reached. Forward flux sampling is only efficient in stochastic systems (such as in Monte Carlo simulations or molecular dynamics with randomly assigned initial particle velocities); in a deterministic simulation, forward flux sampling offers no advantage over brute force simulations.[94]

The reason we employ a series of interfaces is that the probability of transition from one phase to the other is very small indeed; but by splitting it up into more reasonable components, we are able to measure successive conditional probabilities that each trajectory coming from phase α will reach the next interface (λ_{i+1}) given that it has reached interface λ_i. Together, these make up the overall probability of crossing. They are given by

$$P(\lambda_{i+1}|\lambda_i) \equiv P_{i\to i+1} = \frac{N_{i\to i+1}^{\text{successful}}}{N_{i\to i+1}^{\text{total}}}. \tag{D.52}$$

This works provided that λ is a monotonically increasing function, and, in particular, the interfaces must be sufficiently widely spaced such that each 'success' proceeds to the immediately adjacent interface λ_{i+1}, and not directly to, say, λ_{i+2}.[95]

The rate constant is then given by[100;479] $k_{\alpha\beta} = \Phi_{\alpha\to\beta}$, where $\Phi_{\alpha\to\beta}$ is the flux of trajectories coming from phase α that arrive in phase β. This reduces to

$$k_{\alpha\beta} = \Phi_{\alpha\to 0} P_{0\to\beta} = \Phi_{\alpha\to 0} \prod_{i=0}^{n-1} P_{i\to i+1}, \tag{D.53}$$

where $\Phi_{\alpha\to 0} = \dfrac{N_{\alpha\to 0}}{tV}$ is the flux of trajectories crossing the phase α boundary and reaching λ_0 for the first time. Here, t refers to the number of steps spent in phase α and $N_{\alpha\to 0}$ is the number of trajectories reaching λ_0 from phase α (*i.e.* the red circles in figure D.1(a)).

Transition interface sampling is similar to forward flux sampling, and indeed the rate equation is identical. However, the sampling itself is different in the forward flux sampling scheme, as 'shooting' does not take place with a Metropolis Monte Carlo probability,[100] as it does in transition interface sampling. New trajectories are instead created from the distribution achieved at the previous interface. This means that no previous knowledge of the (steady state) phase space density is required, which is a significant advantage of forward flux sampling over other similar sampling techniques.[94;97]

The stationary distribution and hence the Gibbs energy barrier to nucleation (in the isobaric-isothermal ensemble) can be obtained in the forward flux sampling simulation with some additional effort.[50;96] Namely, this requires both a forward and a reverse simulation.

D.6 Cells, neighbour lists, heaps and sorting

In simulations, we generally wish to avoid calculating interactions of particles which are so far away from each other that they do not interact. In many simulations, we achieve this by splitting the simulation box into cells[66] at least as large as the cutoff radius, and construct

linked lists connecting all the particles in a given cell. Whenever we need to calculate a particle's interactions with other particles, we loop through these linked lists within its own cell and the 26 cells adjacent to it (*i.e.* $3 \times 3 \times 3$ cells in total). Provided the system is large enough, using cells significantly reduces the required computational effort: the system must contain more than 27 cells at a minimum, but usually the overhead involved makes the use of cells worthwhile only for systems considerably larger than this.

Alternatively, we can construct neighbour lists,[473] where a particle's neighbours (within some radius larger than the cutoff radius, $r = r_c + x$) are explicitly tracked, and neighbour lists are recalculated once a particle has moved by more than $x/2$ from its original position. We only recalculate the moved particle's neighbour lists and insert or delete the moved particle from the neighbour lists of former and new neighbours in the moved particle's list. Furthermore, any particles whose 'original positions' were in the new shell are included in the new neighbour list even though its actual distance is greater than our initial requirement. This ensures that the particle is not missed should it happen to return to its original place, as its neighbour list would not then be updated. So as to ease the process of adding and removing particles from neighbour lists, and in particular to allow us to terminate calculations when particles are outside the cutoff radius, we implement neighbour lists as heaps, in which particles are sorted by distance.

A graph is called a *tree* if it is connected and contains no cycles. A *rooted tree* is a directed tree that contains a vertex from which there is a path to every other vertex. Such a vertex is called a *root*. A *binary rooted tree* is a rooted tree where each vertex has at most two children.[480] A *binary heap* is a binary tree where each node is less than (or equal to) each of its children, and the binary tree is complete (except in the lowest level). ('Less than' need not be interpreted in a mathematical way; it can refer to any mathematical comparison, including a 'greater than' function. However, if it is in fact 'less than' in mathematical terms, this is known as a *min-heap*).

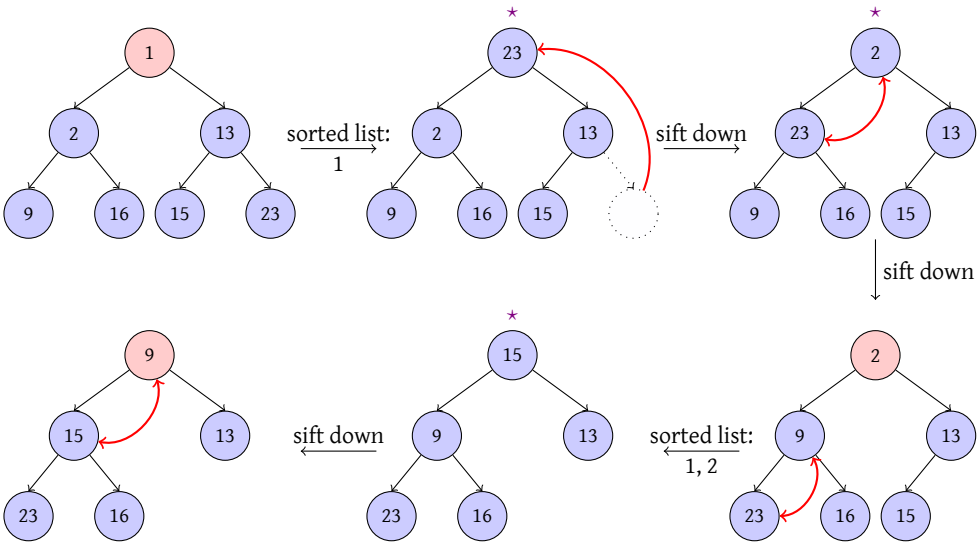


Figure D.2: Heapsort. Structures marked with a star are not proper binary heaps. The smallest element in a heap (marked in pale red) is removed whenever we have recovered the heap structure.

A heap is thus sorted on any 'upward' path, but there is no ordering (except by chance) between elements horizontally across.

Once we have a data structure that implements a heap, sorting it is straightforward. We take the root, which is the smallest element, and remove it from the heap. We then promote the final element of the maximum depth level into the root position; this is now no longer a heap. What we then do is to swap this element with the smaller of its children, and repeat until the binary tree again satisfies the heap property. By repeating the process until the heap is empty, we can therefore obtain a fully sorted list.[480] An example of a heapsort is shown in figure D.2.

Heaps and (almost) complete binary trees can very easily be stored in a compact array. For an array whose first element is labelled 1 (as opposed to 0), for each index i, that node's children are given by indices $2i$ and $2i + 1$, while its parent is given by $i \div 2$, where '$\div$' denotes an integer division (*i.e.* $a \div b = \lfloor a/b \rfloor$ for non-negative integers). In practice, we can store the ordered list in the same array as a heap by keeping track of the size of the heap and simply placing the minimum element in the place from which we have moved a node to the root position. However, in our implementation, we prefer to have both a sorted list and a heap

structure for every particle in the box, and so we will use two arrays. A sorted list is, of course, also a heap, but by placing the final element to the top position in any potential new sorts, this will make sorting slower; therefore, it is best to keep the structures separate.

Although there is no real need for us to sort neighbour lists, it turns out – perhaps surprisingly – that keeping neighbour lists in heaps is actually marginally faster than not doing so, since we can terminate the sorting algorithm as soon as we are past the cutoff point.

D.7 Spherical harmonics

Spherical harmonics are solutions of the angular part of the Laplace equation (in spherical polar co-ordinates), $\nabla^2 f = 0$, where ∇^2 is the laplacian operator. A derivation of their form can be found in the standard texts;[303;481] however, it is helpful to be aware of some of their properties that can allow us to implement them recursively in a computer code.

Spherical harmonics are defined for non-negative integers l and integers $m \in [-l, l] \cap \mathbb{Z}$, and are calculated using

$$Y_{lm}(\theta, \varphi) = N \exp[\mathrm{i}m\varphi] P_l^m(\cos\theta), \tag{D.54}$$

where N is the normalisation constant given by

$$N = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} \tag{D.55}$$

and $P_l^m(x)$ are the associated Legendre functions of the first kind. These satisfy a few important analytic results which, with the help of several recurrence relations, we can use to calculate their value recursively.

Associated Legendre functions are defined in the domain $-1 \leq x \leq 1$ and for $-l \leq m \leq l$. For negative values of m,[303]

$$P_l^{-m}(x) = (-1)^m \frac{(l-m)!}{(l+m)!} P_l^m(x). \tag{D.56}$$

This includes the Condon–Shortley phase factor, $(-1)^m$, as is common in quantum mechanics.[309]

When $m = l$ (and $m \geq 0$), we can write the analytic expression,[303]

$$P_m^m(x) = \begin{cases} (-1)^m(2m-1)!!(1-x^2)^{m/2} & \text{if } m > 0, \\ 1 & \text{if } m = 0. \end{cases} \tag{D.57}$$

The double factorial is given for a non-negative integer n by

$$n!! = \begin{cases} n \cdot (n-2) \cdot (n-4) \cdot \cdots \cdot 1 & \text{if } n \text{ is odd}, \\ n \cdot (n-2) \cdot (n-4) \cdot \cdots \cdot 2 & \text{if } n \text{ is even}, \\ 1 & \text{if } n = 0. \end{cases} \tag{D.58}$$

We can also calculate P_{m+1}^m from the analytic result

$$P_{m+1}^m(x) = x(2m+1)P_m^m(x), \tag{D.59}$$

and any other associated Legendre function of the first kind can be calculated by recursion on

$$(l-m+1)P_{l+1}^m(x) = (2l+1)xP_l^m(x) - (l+m)P_{l-1}^m(x). \tag{D.60}$$

It is important to note that the factorials of the form $\dfrac{(l-m)!}{(l+m)!}$, which occur in several places in this section, can become enormous when calculated for the numerator and the denominator separately for even relatively small values of l and m. This can result in a severe loss of numerical precision. If, however, the factorials are evaluated at the same time by recursion (*i.e.* the largest factor of the numerator divided by the largest factor of the denominator and so forth), no number is likely to become too large or too small to handle using ordinary double precision, and thus numerical stability can be maintained for longer.[5]

In our work, we do not generally use complex spherical harmonics as presented above. Instead, we use their real analogues,[309] which are given by

$$S_{lm}(\theta, \varphi) = \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}} P_l^{|m|}(\cos\theta) f(m), \tag{D.61}$$

[5] In our applications, where we only require one value of l, we can in fact pre-calculate the required factorial fractions and use a lookup table when the results are required.

where

$$f(m) = \begin{cases} \sqrt{2}\sin(|m|\,\varphi) & \text{if } m < 0, \\ \sqrt{2}\cos(|m|\,\varphi) & \text{if } m > 0, \\ 1 & \text{if } m = 0. \end{cases} \tag{D.62}$$

The use of real spherical harmonics is advantageous due to a very significant increase in computation speed. The reason for this is that the complex spherical harmonics have mutually dependent values for the quantum numbers $+m$ and $-m$: the real and imaginary parts of such spherical harmonics for the same solid angle are equal in magnitude and differ only in sign, but yet are typically evaluated explicitly if using complex spherical harmonics, and thus effectively doubling the amount of calculations required, whilst gaining no new information. The use of complex numbers furthermore doubles the amount of memory required, as well as doubling the number of mathematical operations performed, since they need to be performed on the real and on the imaginary part separately.

Real spherical harmonics are perhaps less commonly used due to our familiarity with quantum problems, where the use of complex versions is advantageous due to their conjugation properties, but in spherical harmonic transforms, such properties are not required and the use of complex spherical harmonics is not therefore beneficial. In our tests with Perl and Fortran implementations of spectral analysis on the sphere, an up to five-fold decrease in computation time was observed when using real spherical harmonics as opposed to their complex analogues.

Finally, it is convenient, in order to avoid repeating calculations, to note that when we consider the spherical harmonic calculation for a particular vector between two particles i and j, we will again encounter the same vector, except for the sign, when calculating the spherical harmonics for the same pair in the other direction. This is to say that $\boldsymbol{v}_{ij} = -\boldsymbol{v}_{ji}$, and we can also note that $\theta_{ij} = \pi - \theta_{ji}$ (or $\cos\theta_{ij} = -\cos\theta_{ji}$) and $\varphi_{ij} = \varphi_{ji} \pm \pi$ (where the plus or minus sign depends on the value of φ_{ji} in order to give the principal value). We can calculate the spherical

harmonics especially easily for this pair of vectors, since[303]

$$S_{lm}(\theta, \varphi) = (-1)^l S_{lm}(\pi - \theta, \varphi \pm \pi). \tag{D.63}$$