A numerical study on the enhancement and suppression of crystal nucleation

A numerical study on the enhancement and suppression of crystal nucleation

Academisch Proefschrift

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR AAN DE UNIVERSITEIT VAN AMSTERDAM, OP GEZAG VAN DE RECTOR MAGNIFICUS PROF. DR. D. C. VAN DEN BOOM TEN OVERSTAAN VAN EEN DOOR HET COLLEGE VOOR PROMOTIES INGESTELDE COMMISSIE, IN HET OPENBAAR TE VERDEDIGEN IN DE AGNIETENKAPEL OP

DINSDAG, 20 OKTOBER 2009 TE 10:00 UUR

DOOR

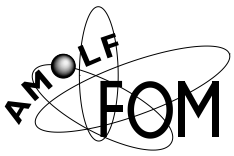
Jacobus Antoon van Meel

GEBOREN TE AMSTERDAM

Promotor: Prof. dr. D. Frenkel

Overige leden: Prof. dr. P. Bolhuis
Prof. dr. W. Briels
Prof. dr. B. Mulder
Prof. dr. S. Portegies Zwart
Dr. P. Schall
Dr. R. Sear

Faculteit: Natuurwetenschappen, Wiskunde en Informatica.



Nederlandse titel: *Een numerieke studie over de bevordering en suppressie van kristal nucleatie.*

The work described in this thesis was performed at the FOM Institute for Atomic and Molecular Physics (AMOLF) in Amsterdam, The Netherlands. This work is part of the research programme of the 'Stichting voor Fundamenteel Onderzoek der Materie (FOM)', which is financially supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)'.

ISBN: 978-90-6464-360-6

A digital version of this thesis can be obtained from `http://www.amolf.nl`. Printed copies can be obtained by request via email to `library@amolf.nl` or by addressing the library at the FOM Institute for Atomic and Molecular Physics (AMOLF), Science Park 113, 1098 XG, Amsterdam, The Netherlands.

This thesis is based in part on the following articles:

Chapter 3:
J.A. van Meel, A. Axel, D. Frenkel, S.F. Portegies Zwart, and R.G. Belleman.
Harvesting graphics power for MD simulations.
Mol. Sim. **34**, Issue 3, page 259 - 266 (2007).

Chapter 4:
J.A. van Meel, A. Page, R. Sear, and D. Frenkel.
Two-step vapor-crystal nucleation close below triple point.
J. Chem. Phys. **129**, 204505 (2008).

Chapter 5:
J.A. van Meel and D. Frenkel.
Liquid wetting can enhance vapor-crystal nucleation in a pore.
Submitted to J. Chem. Phys.

Chapter 6:
J.A. van Meel and D. Frenkel. (in prep.)
Disordered substrates favor heterogeneous crystal nucleation of model globular proteins.

Chapter 7:
J.A. van Meel, D. Frenkel, and P. Charbonneau.
Geometrical frustration: a study of four-dimensional hard spheres.
Phys. Rev. E **79**, 030201(R) (2009).

J.A. van Meel, B. Charbonneau, A. Fortini, and P. Charbonneau.
How rare is hard-sphere crystallization?
Submitted to Phys. Rev. E.

Appendix A:
J.A. van Meel and D. Frenkel. (in prep.)
A parameter-free nearest-neighbor algorithm.

Other articles by the same author:
J.A. van Meel and H.A. Posch. Phys. Rev. E **80**, 016206 (2009).
P. Charbonneau, A. Ikeda, J.A. van Meel, and K. Miyazaki. (Submitted to Phys. Rev. Lett.)

Contents

Chapter 1

Introduction

What is crystal nucleation? To answer this question we need to recall that most substances that surround us can exists in two or more different phases. Think for instance of water, a substance we deal with on a daily basis. At room temperature it is a liquid, but we all know that cooling water below $0\,^{\circ}C$ will turn it into ice, and that it evaporates when heated above $100\,^{\circ}C$. We refer to these two phases as crystal and vapor, respectively. Sometimes, the crystal forms directly out of the vapor, a process called deposition (or de-sublimation). This is how snow-flakes grow in clouds, or hoar frost on a shrub (see Figure 1.1).

But not only temperature determines the phase, pressure plays an important role, too. You might not realize it, but your precious diamond is not stable at ambient conditions. Heating it up to extreme temperatures, well above those of molten lava, will turn it into a carbon liquid. But on subsequent cooling at ambient pressure you will not regain your diamond. Instead you will find yourself with a piece of graphite, the common crystalline form of carbon. Diamonds only grow at the enormous pressures that exist deep inside a planet, and without these pressures they eventually turn to graphite. Fortunately, the natural transition from diamond to graphite takes place on the time scale of a billion years.

A slightly different crystallization process is neither driven by temperature nor by pressure, but by concentration. For instance, the two crystal products with the world's largest industrial production are table salt and sugar. Both are obtained from solution by increasing the concentration beyond the solubility threshold, at which crystals start to form. This can be achieved by reducing the temperature or - at constant temperature - by evaporation of the solvent. The latter method accounts for more than 50% of the world's production of crystals [1].

If we combine all the information about the different phases of a substance, we arrive at what is called a phase diagram. It shows, as a function of temperature, pres-

Figure 1.1: Hoar frost on a shrub in Akureyri, Iceland. At high humidity and low temperatures the ice crystals grow directly by deposition from the vapor. Picture by Ævar Arnfjörð Bjarmason (released into public domain).

sure, or concentration, which phase is stable at the given conditions. The boundaries between these phases are the coexistence lines - crossing them eventually leads to the transformation from one phase into the other. It is important to note that a phase diagram shows the stability of bulk phases in thermodynamic equilibrium. In other words, it only holds for a substance without confinements (e.g. far away from a wall) after sufficiently long waiting times. Figure 1.2 depicts a schematic phase diagram of a simple mono-atomic substance, such as the noble gas Argon, both in its temperature/pressure and temperature/density representations.

Phase transitions on the molecular scale

Although the phase diagram is known for many substances, it is not enough to describe the *kinetics* of a phase transition. As an example imagine the following experiment: Provided you have some very pure water, kept at room temperature in a jar with a very smooth surface (or, if you prefer, floating in space as a huge spherical drop). If you now cool it down very slowly - and make sure you don't disturb the water in the slightest (which is a challenge) - you will witness that the water temperature can drop to $-20\,^{\circ}C$ without freezing. The water can stay liquid for days, but the tiniest perturbation, a piece of dust or some small vibrations, will turn it to ice. You

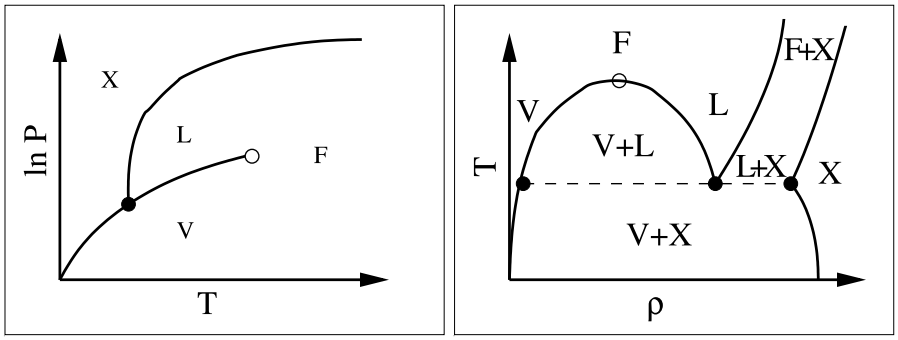


Figure 1.2: Schematic phase diagram similar to that of the noble gas Argon. It shows the vapor (V), liquid (L), crystal (X), and fluid (F) phases, and both the triple point (filled circle) and the critical point (open circle), which is where vapor and liquid become indistinguishable and are called a fluid. Left: temperature-pressure (T-P) representation showing the coexistence lines. Right: temperature-density (ρ-T) representation indicating the vapor-liquid (V+L), liquid-crystal (L+X), and vapor-crystal (V+X) coexistence regions.

might argue that this sounds a bit academic, but in facts its not. Many liquids can be undercooled and kept for a long time without crystallizing, but eventually the liquid will freeze by itself. As a consequence, the temperature where a crystal melts and the temperature where the corresponding liquid freezes do not necessarily coincide for one and the same substance, a phenomenon known as hysteresis. So what prevents a material from freezing as soon as it is brought to a state where the crystal phase is the most stable?

Another effect that is not described by a phase diagram concerns an observation reported by Turnbull half a century ago [2, 3]. Investigating the characteristic undercooling temperature of liquid metals, he found that the freezing temperature depends on the temperature of the previous melting cycle. In particular, metals that were prepared at temperatures only slightly above the melting point froze much sooner compared to those prepared at a very high temperature. He concluded that upon heating some microscopic pores in the container wall remained crystalline and acted as *seeds* inducing bulk crystallization much sooner. Once the liquid was heated far above the melting point, even the material inside these pores melted, and the original undercooling temperature was recovered again. But how is it possible that pores remain crystalline above the melting point?

Both examples show that knowing the thermodynamic properties of a substance is not enough to describe the kinetics of phase transitions in detail. All we can observe macroscopically is the *growth* of a new phase, but not how it starts. To explain what happens at the initial step of a phase transition, the so-called *nucleation* process, we

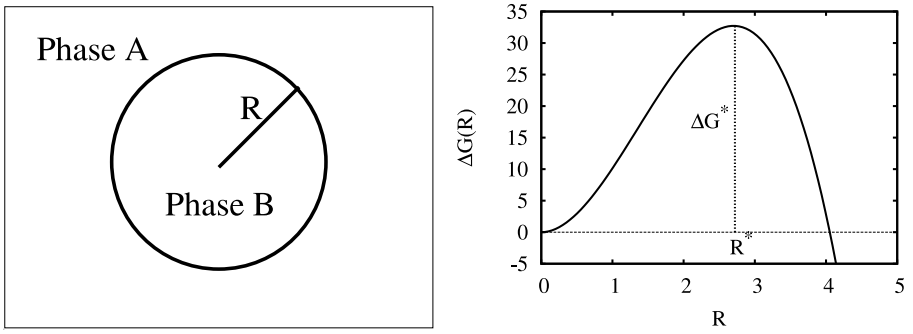


Figure 1.3: Left: schematics of a spherical nucleus of radius R of a new phase B embedded in a meta-stable parent phase A. Right: CNT prediction for the free energy change due to the presence of a spherical nucleus as function of its radius R.

need to understand what happens on a microscopic scale.

Classical Nucleation Theory

The first microscopic theory of crystal nucleation was presented early last century by Volmer and Weber [4] and by Becker and Döring [5]. Based on the conjecture of Gibbs [6] that a phase' stability is related to the work of formation of small nuclei of the new phase, they developed a theoretical framework for the study of nucleation, now commonly known as Classical Nucleation Theory (CNT). They assumed that a nucleus of the new phase is approximately spherical and has the same properties as its final bulk phase (i.e. same structure and density). They argued that the driving force for the phase transition, the difference in chemical potential $|\Delta\mu|$ between the two phases, is countered by the work to form an interface with surface tension γ. From this they deducted an expression for the (Gibbs) free energy change of a system due to the presence of a nucleus of the new phase,

$$\Delta G(R) = \gamma A(R) - |\Delta\mu|\rho V(R),$$

where $A(R)$ and $V(R)$ are the surface area and volume of a sphere with radius R, respectively, and ρ denotes the bulk (number) density of the nucleating phase. This curve is plotted in Figure 1.3. For small radii $\Delta G(R)$ increases steadily, but once the radius exceeds a critical size R^*, the gain in chemical potential outweighs the interfacial cost, and the nucleus is likely to grow to a macroscopic size. Small nuclei form and dissolve all the time in a supercooled liquid, caused by thermal fluctuations. Statistical thermodynamics tells us that the average number of fluctuations of a given size that can be observed in a volume V is related to the free energy cost associated

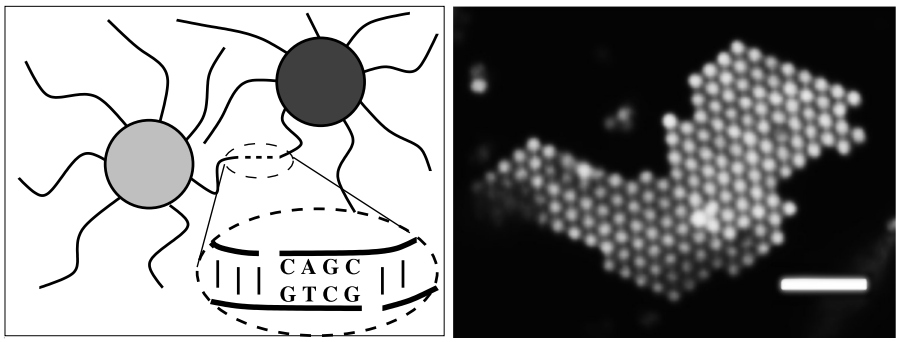


Figure 1.4: Left: schematics of the binding of two DNA-coated colloidal particles. The DNA strings have a short single-stranded end which binds specifically to DNA with a complementary sequence. Right: confocal-microscopy image of DNA-coated colloids forming a two-dimensional "flying carpet". The scale bar indicates $5\mu m$, and the colloids have a diameter of $1\mu m$. Picture courtesy of N. Geerts, with kind permission.

with this fluctuation:

$$N(\Delta G) = Ve^{-\beta\Delta G},$$

which scales linearly with the system volume V, and $\beta = 1/(k_B T)$, where T is the temperature and k_B is Boltzmann's constant. From this expression one can construct an approximate expression for J, the *rate* at which nuclei form in a system. All we need is the work to form a critical nucleus, $\Delta G(R^*) = \Delta G^*$, and a constant taking into account kinetic effects, such as the rate of particle adsorption to the critical nucleus and the probability of barrier re-crossings.

$$J = J_0 Ve^{-\beta\Delta G^*}.$$

This equation provides the starting point for many theoretical studies on nucleation. Although the simplicity of CNT limits its applicability for quantitative predictions, it has great explanatory power and often predicts the qualitative behavior correctly.

Novel materials, self-assembly, and protein crystals

It should be stressed that understanding the process of nucleation and growth on a microscopic scale is not only of academic interest. For many industrial processes it is crucial to optimize the crystallization process in order to increase output, improve the quality of the resulting materials, or lower the costs of manufacturing. Just think of computers: their chips are made from large silicon single crystals called wavers. Improving their quality, size, or growth speed enhances the rate at which high-quality chips can be produced, which is worth billions.

Another example comes from nano-technology. One of the holy grails of this field of research is to arrive at novel materials with pre-specified properties and phase behavior, not by a random process of trial and error but by rational design. Such rational materials design would be greatly facilitated if we could understand how (macro-)molecules interact and how this interaction affects the spatial organization of these molecules into larger structures. Using the techniques of modern molecular biology, we can functionalize micron-sized particles ("colloids") in such a way that we can impose which colloid will be specifically attracted to which other colloid as the DNA ensures very specific binding between the different units. Careful design of "intelligent" building blocks might make them self-assemble into increasingly complex macroscopic structures. Figure 1.4 shows both a schematics for a DNA-coated colloid and a confocal-microscopy image [1] of an experimental colloidal "flying carpet".

A final example comes from drug discovery: here, proteins are in the center of interest. Once in solution, these long chains of amino-acids fold into a very specific structure giving them their ability to function as building blocks of life. Detailed knowledge of a protein's 3D structure is crucial for efficient drug design, and protein crystal X-ray diffraction has provided us with most of the protein structures that we know so far. The main challenge here is the production of high-quality crystals, as many proteins are poor crystallizers. In fact, evolution has "selected" proteins such that they tend to stay in solution. However, under conditions where the solubility is decreased, many proteins do not order in bulk crystals but form fibrillar aggregates. Such "crystalline" aggregates have been implicated in numerous diseases such as Creutzfeld-Jacob or Altzheimer. Clearly, control of undesirable fibrillar crystallization of proteins is of considerable interest.

1.1 Current research on nucleation phenomena

Nucleation phenomena have been approached from many different perspectives. Ever since Volmer and Weber laid the foundation of what is now known as CNT, other scientists have proposed extensions to compensate for its short-comings. There exist corrections for surface curvatures effects, compressibility, or line tension, and the discussion is still on-going. But CNT is not the only theoretical approach: originated from quantum mechanical calculations of electron densities in condensed matter physics and chemistry, density functional theory is now also used in combination with classical statistical mechanics to assess nucleation phenomena. It expands the grand potential in terms of a local density and takes into account the effect of thermal density fluctuations. Minimizing this functional numerically with respect to the

[1] Courtesy of N. Geerts, with kind permission.

local density leads to the system's equilibrium properties. With DFT it is possible to predict nucleation rates, barrier heights, and provide an estimate for the size of the critical nucleus without having to assume, as is done in CNT, that a microscopic nucleus has the same bulk and surface properties as a macroscopic crystal.

In addition to theoretical advances, recent years have witnessed a spectacular expansion of experimental techniques to probe the details of crystal nucleation. A particularly interesting development concerns colloidal systems. These micron-sized particles are dispersed in solution, and with buoyancy-matching to eliminate gravitational forces, they can diffuse freely in 3D space. Using light from the visible spectrum they can be observed in "real space" using confocal microscopy. Due to the rather slow movement of colloids, experimentalists can follow the motion of individual particles or extract the 3D position of all particles at any given time. Moreover, colloidal interactions can be tuned almost at will by changing the solvent composition or adding external fields. And with optical tweezers created by focused laser beams individual particles can be trapped and moved around. With all these properties colloidal systems make for a generic experimental model system for testing theories. In fact, during the past decade, the first direct "microscopic" observations of crystal nucleation have been reported.

As the third pillar of modern physics, computational physics provides an excellent tool for the study of molecular systems, both for testing theoretical predictions and for investigating the molecular mechanism. The rapid pace at which computing power increases makes it possible to calculate the properties of systems with many thousands of particles on a home computer. But in the context of phase transitions it is the recent advances in rare-event algorithms that allow for an efficient simulation of nucleation events, particularly at conditions close to coexistence. It is here where this Thesis attempts to advance on the topic of nucleation. Some of the more technical chapters may at first sight appear unrelated to nucleation. However, as I will briefly explain below, it is indeed the study of nucleation events that provides the logical connection between all chapters.

Massively parallel computing

Rather than solving analytically the complex equations of statistical mechanics for a system as a whole, numerical many-particle simulations pursue a different course. Molecular Dynamics (MD) refers to one class of simulations that computes the forces acting on all particles and solves iteratively Newton's equation of motion to yield dynamical trajectories. Equilibrium properties of the system are then obtained through time averages. Metropolis Monte Carlo (MC) simulations form another class of simulations. In contrast to MD it computes ensemble averages by sampling over possible micro-states of a system. A new state is typically constructed from an existing state

by randomly changing the microscopic state of the system. This change can be either a particle displacement/insertion/removal, a volume change, or another physical or unphysical move. A new state is then accepted or rejected according to its weight in the statistical ensemble.

Although both methods seem rather different, in fact they are equivalent in many ways. From a numerical perspective they share the principle that all the particle interactions are evaluated many times over. Even with the rapid increase of available computational power, for many systems the performance of a single processor (CPU) is not sufficient. Large-scale compute clusters offer the option to run these simulations in parallel on many processors. But in order to work simultaneously on the same problem, all processors have to exchange information. Not only does this communication cost time, which lowers the efficiency of such a parallel computer well below its nominal performance, but the required low-latency high-throughput network hardware often exceeds the price of the processing units themselves.

In Chapter 3 we explore an alternative hardware for parallelism, put forward by a field very distinct from natural sciences: the gaming industry. To accommodate the growing demand for 3D video graphics in computer games, the graphics processing unit (GPU) was developed. Within the last two decades it matured from a graphical co-processor to a fully programmable one-chip supercomputer, capable of a (theoretical) peak performance exceeding that of a conventional CPU by two orders of magnitude. With its massively parallel arrangement of floating point units and an on-board memory working at ten times the speed of conventional memory the GPU makes an ideal compute platform for MD simulations. Using nVidia's CUDA language we demonstrate that a 25- to 80-fold speed increase can be achieved. And instead of moving only the computationally most expensive parts on the GPU, we show that in fact the entire simulation can run on a GPU.

Ostwald's step rule

Chapter 4 is the first to deal with the physical aspects of nucleation. Phase transitions between two phases, like condensation or crystallization, have been studied before for many systems. But already in 1897 Ostwald conjectured that the phase that forms during nucleation is not necessarily the phase that is thermodynamically most stable [7]. Instead, another metastable phase may nucleate, provided that it is more stable than the parent phase and that the transition to that phase is kinetically favored. A phase transition to a thermodynamically stable phase may therefore proceed via a two step process. In the first nucleation event a metastable intermediate phase is formed; this then goes on to form the final phase, possibly via a second nucleation event.

There is experimental evidence that such a two-step nucleation scenario may be

relevant to protein crystallization. For instance, lysozyme is sometimes found to separate into liquid-like aggregates before crystallization [8, 9]. This was also observed in simulation [10] and several theoretical models provide a rationale for the behavior observed [11–14].

In Chapter 4 we study the vapor of the Lennard-Jones system at temperatures below its triple point at chemical potentials where the crystal phase is the stable phase and surface melting is expected. We find that the phase transition does not occur directly to the crystal phase but via a liquid intermediate phase. Both transitions act on different time scales and correspond to distinct independent nucleation events. The surface tension of the liquid-vapor interface is less than the surface free-energy density of the crystal-vapor interface, making the free-energy barrier to liquid droplet formation lower than to crystal formation. Recently, this behavior has been confirmed experimentally in a quasi-2D colloidal system [15].

Heterogeneous nucleation and wetting

Nucleation can proceed either homogeneously, that is in the bulk of the parent phase, far away from walls or other external influences, or heterogeneously at the surface of a "seed". In most cases of practical interest, heterogeneous nucleation dominates the rate of crystal formation [16]. For this reason, heterogeneous crystal nucleation has been studied extensively in theory, experiment, and simulation, investigating a variety of geometries, such as walls, templates, and pores. In particular, micro-porous media has received much attention as experiments by Chayen and co-workers [17] suggest that such materials may act as universal nucleation agents for protein crystallization.

Based on their theoretical study of the 2D Ising model Page and Sear [18] predicted that a pore of suitable size can indeed enhance nucleation. However, such a lattice model cannot capture one of the most interesting features of protein crystallization, namely that it is most effective in the presence of a meta-stable high-density disordered phase. The Ising model does not distinguish between the liquid and the crystal state. Another limitation that will turn out to be crucial is that a pore in an Ising model is by definition always commensurable with the lattice spacing.

In Chapter 5 we therefore study pore nucleation in an off-lattice model, where the liquid and crystalline states are distinct, and variations in the pore diameter may influence the free-energy barrier to crystallization. As in the case of homogeneous nucleation, we show that crystallization proceeds via an intermediate meta-stable liquid phase. As is to be expected, we observe that the wetting properties of a pore have a strong influence on the nucleation free-energy barrier. In addition, for a narrow range of pore radii we find that the liquid nucleation itself becomes a two-step process, where in a first nucleation event the liquid fills the pore, and in a second independent event the liquid breaks out into the bulk. Interestingly, the crystal nucle-

ation remains largely unaffected. Only for the largest pore featuring two-step liquid nucleation is the crystal formed before the liquid grows into the bulk. But since the liquid nucleation is the rate-limiting step the overall nucleation rate is still enhanced significantly by the presence of a pore with suitable wetting properties.

Surface roughness

So far we have provided two examples that nucleation via an intermediate meta-stable phase can facilitate the transition to the thermodynamically stable phase, both for homogeneous and heterogeneous nucleation. But the latter scenario holds more options. A judicious choice of the wetting properties of the nucleation-inducing material may lead to the adsorption of a thin microscopic layer of an intermediate phase even at conditions where its bulk is unstable. Moreover, the structure and geometry of the nucleant surface may have a strong influence on the phase that wets it, and in the case of crystallization, on the structure of the emerging nucleus.

These options are investigated in Chapter 6 for a model system with very short-range interactions, which is known to reproduce a phase diagram characteristic of colloids and proteins. These phase diagrams feature only the fluid and the crystal as thermodynamically stable phases, with the (meta-stable) vapor-liquid coexistence curve hidden within the fluid-crystal two-phase region. Close to this fluid-fluid de-mixing, where protein crystallization is observed experimentally, we explore the effects of surface structure on the nucleation pathway. We find that surface disorder leads to the adsorption of a thin layer of a high-density liquid phase. Both the amount and the isotropy of the adsorbed phase is maximal if the surface roughness is of the order of the radius of gyration of the protein. Moreover, for deep cylindrical pores cut from a disordered substrate we observe pore filling if the cylinder falls within a narrow size range. In such a filled pore the crystal emerges in its center, i.e. away from the rough surface, and the nucleation rate is enhanced tremendously. Moreover, in contrast to periodically structured surfaces, the liquid in the disordered pore appears not to impose stress on the growing crystal, and may therefore lead to high-quality crystals.

Geometrical frustration and higher dimensions

Most of the discussion so far was focused on systems with very little undercooling. But for some systems a quench deep into the meta-stable or even unstable regime does not enhance crystallization, but instead promotes the formation of a glass. In this case the fluid remains disordered and becomes steadily more viscous, until the microscopic relaxation processes become slower than the experimental or simulation timescales.

Geometrical frustration is one of the factors thought to slow the formation of ordered phases and thus assist glass formation [19]. The earliest theories of geometrical frustration conjectured that it is related to the optimal way to pack "kissing" spheres around a central particle [20]. For hard spheres, the icosahedron is the optimally packed arrangement of kissing-spheres. In contrast to the fcc lattice with the same number of kissing-spheres, the icosahedron has a higher entropy since its spheres can still "wiggle around". But the five-fold symmetry of an icosahedral arrangement does not allow to fill space, leading to a competition between local and global order. More recent formulations of geometrical frustration focus on the concept that tetrahedra or, more generally, simplexes are the smallest building blocks in a Delaunay decomposition of space [21, 22]. This poly-tetrahedral scenario then ascribes the presence of icosahedra to their singularly easy assembly from quasi-regular tetrahedra. But again, 3D space cannot be tiled with tetrahedra. Hence, also from this perspective there appears to be a conflict between local and global optimal packing.

The last part of this Thesis attempts to resolve the ambiguity of the origin of geometrical frustration. For this task we leave our familiar three-dimensional world and enter the space of four-dimensional hard hyper-spheres, because in such a system the optimal arrangement for kissing spheres *is* compatible with the lattice with the highest density, the D_4 crystal. Thus, if geometrical frustration has its origin in the optimal way of arranging kissing spheres, crystal nucleation should encounter only a very small free-energy barrier in $4D$. With the simulation of crystal nucleation in Chapter 7, we find that in 4D the free-energy barrier to nucleation is considerably higher than for its 3D analog at comparable conditions, suggesting that tetrahedra are the origin of frustration. In fact, using higher-dimensional bond-order analysis, we show that the 4D optimal kissing cluster plays no identifiable role in the supercooled fluid.

In the appendix of Chapter 7, we extend this line of thought into even higher dimensions. We assess the validity of higher-dimensional theoretical predictions on the equations of state, the random close packing, and the growing fluid-crystal dissimilarity. From the behavior in higher dimensions we deduct some consequences of our 3D world, for instance that conventional hard spheres are remarkably little frustrated.

Love thy neighbor[2]

We conclude this Thesis with an appendix on a particular aspect of numerical simulations: the challenge to handle the vast amount of data produced in a single run. Sometimes this data is analyzed after the simulation has finished, but at other times this has to happen on-the-fly during a simulation. Appendix A deals specifically with

[2] Title of an oral presentation by S. Tindemans.

a key ingredient for structural analysis: the identification of the nearest neighbors of a given particle. Two algorithms are widely used. The first algorithm simply considers all particles neighbors which are within a certain distance. This fixed-distance cutoff algorithm is generally applied in simulations with particles interacting through a short-range potential, where a neighbor corresponds to an interaction partner, and the cutoff to the range of the interaction potential. For a structural analysis, for instance to determine whether a particle is part of a crystalline or liquid environment, the choice of cutoff is not so clear. The first minimum of the pair correlation function $g(r)$ forms arguably a natural choice, as it relates to the neighbors in the first coordination shell. However, this quantity depends on temperature, pressure, and the systems thermodynamic phase, and therefore needs to be computed for each system. In contrast, the so-called Voronoi construction is based on purely geometric constraints and requires no parameters. But this advantage comes at the cost of significant computational effort. This is why the Voronoi construction is generally applied only in post-analysis and not on-the-fly in simulations.

In Appendix A we propose a novel algorithm for the identification of nearest neighbors. Like a Voronoi construction it is parameter-free (scale-free), but performs almost an order of magnitude faster. However, some of the properties of the Voronoi construction are not preserved, such as neighbor symmetry and local volume conservation. We compare our algorithm to both a well-tuned fixed-distance cutoff and the Voronoi construction (in fact to its dual, the Delaunay triangulation) for the analysis of both Lennard-Jones and Carbon liquid and crystal phases. We find that our algorithm yields results very similar to a well-tuned distance cutoff, performs well at interfaces, and is robust against thermal fluctuations. In particular, we suggest that our algorithm may be suitable for protein structure analysis, for which the currently applied Voronoi construction only partially meets those criteria.

Chapter 2

Methodology

This chapter reviews the key ingredients required for a computational approach to nucleation phenomena. It starts with a brief introduction to Classical Nucleation Theory (CNT), a phenomenological theory based on macroscopic quantities, which will then be extended to heterogeneous nucleation and to a higher-dimensional space. CNT provides a tool to assess the efficiency of different nucleation pathways and allows for a qualitative comparison with experiment. Next, an overview is presented over recent simulation techniques for the study of phase transitions and rare events. In particular, both Umbrella Sampling and Forward-Flux Sampling are discussed in more detail, which are the two algorithms used in this Thesis. In the last part of this chapter the focus is on the role of the order parameter, also known as reaction coordinate, in the study of rare events. The chapter concludes with a detailed description of two common order parameters, one for condensation and one for crystallization,

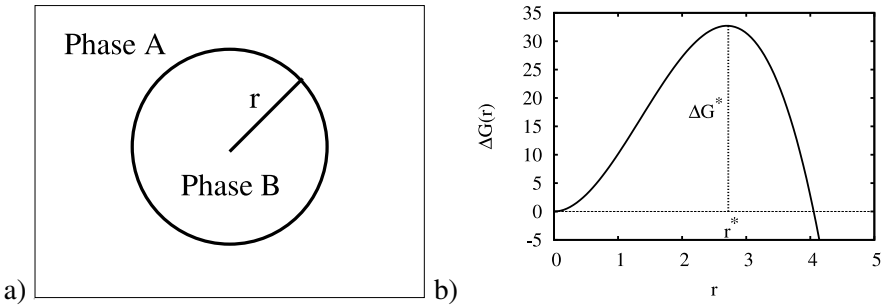


Figure 2.1: a) Phase A with a spherical nucleus of phase B as assumed in Classical Nucleation Theory (CNT). b) Free-energy barrier for nucleation according to CNT, along with the critical nucleus size r^* and the critical barrier height ΔG^*. The parameters correspond to a vapor-liquid nucleation of a Lennard-Jones system below the triple point, as studied in Chapter 4.

2.1 Classical Nucleation Theory

In the early part of the last century Volmer and Weber [4], and subsequently Becker and Döring [5] developed a phenomenological theory of nucleation based on macroscopic thermodynamic properties. This theory is nowadays called *Classical Nucleation Theory* (CNT). It relates the rate of nucleation of a new phase from a metastable parent phase to probability that a spontaneous fluctuation will create a nucleus of the new phase that can continue to grow spontaneously. This probability is related to the reversible work required to form such a nucleus. In particular, CNT yields an expression for the nucleation free-energy barrier height that depends only on *macroscopic* quantities *at coexistence*, i.e. quantities that are in principle accessible both in simulation and experiment.

CNT is based on several assumptions: First, the nucleus of the new phase is assumed to be perfectly spherical. This assumption may be reasonable for droplet nucleation from a saturated vapor, but need not be correct for nucleation of a strongly-facetted crystal [23]. Second, the nucleus is taken to be incompressible. While this is reasonable for the nucleation of a high-density phase like a liquid or a crystal, this is definitively not a good approximation in the case of vapor nucleation in an overheated liquid. The next simplification is that the nucleus is assumed to behave as its bulk phase, e.g. with the same density and structure. This is a serious approximation, considering that the Laplace pressure inside a droplet scales with the inverse droplet radius, or taking account of the fact that a small crystal nucleus consist mainly of interface, which is expected to be less ordered than a bulk crystal. Finally: the surface tension (or, more precisely, the "surface free-energy density") is taken to be that of a planar surface at coexistence. On the one hand, this assumption is crucial, as

surface tension is only defined unambiguously at coexistence: the surface tension of a microscopic nucleus is a rather ill-defined quantity. On the other hand, nucleation can only occur away from coexistence, and surface tension often depends strongly on curvature [24, 25].

Despite these approximations, CNT offers a reasonable theoretical framework for description of nucleation phenomena. Although its rate prediction is typically off by several orders of magnitude [26], when it comes to a qualiitative comparison with experiment it helps to understand the underlying processes and is still often able to predict the trends. Therefore, we now review its essentials.

Consider a system prepared in a phase A under conditions where A is meta-stable with respect to the more stable, nucleating phase B. Figure 2.1a schematically depicts such a system in phase A with a nucleus of phase B. We assume that we know, either from calculations or from experiments, the precise coexistence conditions, the difference in chemical potential between A and B, and the surface tension at coexistence. Using the approximations mentioned above, CNT arrives at the following expression for the change in (Gibbs) free energy due to the formation of a spherical nucleus of phase B with radius r,

$$\Delta G(r) = A(r)\gamma - |\Delta\mu|\rho V(r). \tag{2.1}$$

Here, $\Delta\mu$ is the difference in chemical potential between both phases, γ is the planar surface free-energy density at coexistence, ρ the bulk density of phase B, and $A(r)$ and $V(r)$ are the sphere's surface and volume, respectively. Physically, $\Delta G(r)$ represents the change in free energy due to the cost to form an interface between the two phases, which scales with the sphere's surface, and the gain in bulk Gibbs free energy of the new phase, which scales with the volume. This equation can also be derived from thermodynamical considerations as is shown in Section 2.1.1. As $A \propto r^2$ and $V \propto r^3$, there exists a radius r^*, called the *critical radius*, from which on further growth results in a decrease of the free energy and therefore becomes beneficial. This radius corresponds to the maximum of $\Delta G(r)$, and is commonly referred to as the nucleation free-energy barrier height. The maximum of Eqn. 2.1 is easily located and yields

$$r^* = \frac{2\gamma}{\rho\,|\Delta\mu|}, \tag{2.2}$$

with the corresponding barrier height

$$\Delta G^* = \frac{16\pi}{3}\frac{\gamma^3}{\rho^2\,|\Delta\mu|^2}. \tag{2.3}$$

Equation 2.3 shows that the free-energy barrier for nucleation diverges when approaching coexistence where $|\Delta\mu| \to 0$. Figure 2.1b shows the CNT free-energy barrier (Eqn. 2.1) as well as the critical nucleus size r^* and the corresponding barrier

height ΔG^* for a Lennard-Jones vapor-liquid transition under the conditions considered in Chapter 4. Note that often it is more practical to rewrite the critical nucleus size in terms of the number of particles n in a nucleus, rather than its radius. Using $n = \rho V$ we can rewrite Eqn. 2.2 to obtain

$$n^* = \frac{32\pi}{3} \frac{\gamma^3}{\rho^2 |\Delta\mu|^3}. \tag{2.4}$$

2.1.1 Thermodynamic derivation of CNT

The expression for the CNT free-energy barrier, Eqn. 2.1, can be also derived from thermodynamic considerations. The starting point is the thermodynamic expression for the internal energy of a system in phase A,

$$U = TS - PV + \mu N, \tag{2.5}$$

with T the temperature, S the entropy, P the pressure, V the system volume, μ the chemical potential of phase A under these conditions, and N the number of particles in the system. We now change the system to transform a small amount of particles into phase B, surrounded by an interface with area A and surface tension γ. Assuming that the interface is infinitesimally thin, the internal energy reads

$$\begin{aligned} U_{A+B} &= TS_{A+B} - P_A(V - V_B) + P_B V_B + \mu_A(N - N_B) + \mu_B N_B + A\gamma \\ &= TS - PV + \mu N + T\Delta S - P\Delta V + N_B\Delta\mu + A\gamma \\ &= U + T\Delta S - P\Delta V + N_B\Delta\mu + A\gamma. \end{aligned}$$

In the second step we have assumed that the transition from the pure-phase system to the two-phase system is isothermal-isobaric, and that $N_A \gg N_B$ such that $\mu_A \approx \mu$. The work of formation for a new phase B in A is then the change in (Gibbs) free energy, $\Delta G = \Delta U + P\Delta V - T\Delta S$. If we rewrite the (negative) difference in chemical potential using its absolute value, $\Delta\mu = -|\Delta\mu|$, we recover the CNT expression Eqn. 2.1,

$$\Delta G = A\gamma - N_B|\Delta\mu|. \tag{2.6}$$

2.1.2 Extension to higher dimensions

From a theoretical perspective, equation 2.1 is not limited to a three-dimensional space. In fact it can easily extended to a d-dimensional space,

$$\Delta G(r) = \gamma A_d r^{(d-1)} - |\Delta\mu|\rho V_d r^d, \tag{2.7}$$

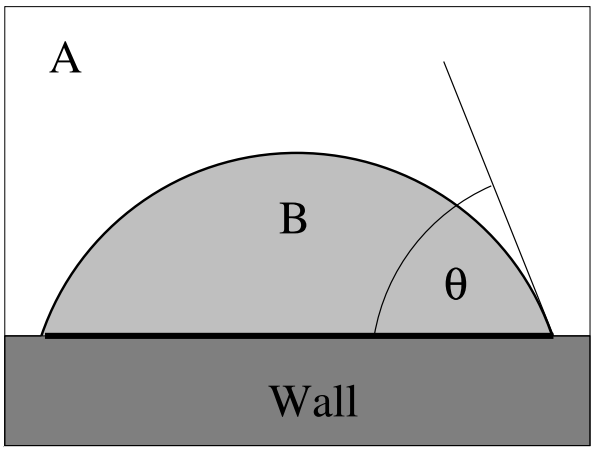


Figure 2.2: Schematic representation of a nucleating droplet of a phase B in contact with an adsorbing wall embedded in a meta-stable parent phase A. The symbol θ indicates the macroscopic contact angle.

with the surface area A_d and volume V_d of a d-dimensional unit sphere given by

$$A_d = dV_d \qquad \text{and} \qquad V_d = \frac{\pi^{n/2}}{\Gamma\Big(n/2+1\Big)}, \tag{2.8}$$

where $\Gamma(x)$ is the gamma function. Consequently, the critical radius then yields

$$r^* = (d-1)\frac{\gamma}{\rho\,|\Delta\mu}, \tag{2.9}$$

the number of particles in the critical nucleus is

$$n^* = (d-1)^d V_d \frac{\gamma^d}{\rho^{(d-1)}\,|\Delta\mu|^d}, \tag{2.10}$$

and the free-energy barrier height results in

$$\Delta G^* = (d-1)^{(d-1)} V_d \frac{\gamma^d}{\rho^{(d-1)}\,|\Delta\mu|^{(d-1)}}. \tag{2.11}$$

This extension is used in Chapter 7 in the analysis of crystal nucleation of four-dimensional hard hyper-spheres.

2.1.3 Heterogeneous nucleation

In most systems of practical interest homogeneous nucleation is not the dominating pathway. The default route to nucleation starts at a pre-existing “seed” crystal, or at

impurities (e.g. dust particles) or even at the container walls. Nucleation that starts at a pre-existing interface is called heterogeneous nucleation. CNT can easily be extended to describe this phenomenon. Imagine a meta-stable phase A that is in contact with an impurity, for instance a wall. Whether or not a nucleus of the nucleating phase B forms at the wall depends on the wetting properties of the wall, which are related to the surface tensions between the various phases involved. Following the procedure for homogeneous nucleation, we can write down the work of formation for a small nucleus in contact with a wall,

$$\Delta G(r) = \gamma_{AB} A_{AB}(r) - |\Delta\mu|\rho V(r) + A_{BW}(r)\gamma_{BW} - A_{AW}(r)\gamma_{AW}. \tag{2.12}$$

Compared to Eqn. 2.1 this expression does not only contain a contribution from the interface between phase A and B, but also from the interfaces between each phase and the wall, as is denoted by the subscript. When the nucleating phase grows, its interface area with the wall A_{BW} increases. At the same time the interfacial area of the parent phase with the wall A_{AW} is reduced, since the parent phase is expelled from the wall. Using Young's equation for the contact angle and the information that A_{AW} is equal to A_{BW} for symmetry reasons, the latter equation becomes

$$\Delta G(r) = \gamma_{AB} A_{AB}(r) - |\Delta\mu|\rho V(r) - A_W(r)\gamma_{AB}\cos\Theta, \tag{2.13}$$

where Θ is the (macroscopic) contact angle, and $A_W(r) = A_{AW}(r) = A_{BW}(r)$ is the contact area of A with the wall. Such a system is schematically depicted in Figure 2.2, showing a nucleating droplet of the phase B in contact with an adsorbing wall embedded in a meta-stable parent phase A. The expressions for the barrier height and critical nucleus size are derived in the appendix of Chapter 5 for the nucleation on a planar wall, a planar circular patch, and a hemi-spherical pore. However, note that the current CNT expression neglects effects of line tension, which is known to affect the barrier [27].

2.2 Rare events sampling

First-order phase transitions are inherently difficult to simulate. The often large free-energy barrier separating the meta-stable from the stable phase causes nucleation to become a rare event. As a consequence, the system will usually fluctuate around its meta-stable free-energy minimum in the course of a normal ("brute-force") simulation. For this reason, the study of nucleation requires advanced sampling schemes. In fact, observing a nucleation event is not the only difficulty for studying phase transitions. Often, it is unclear which pathway will be followed, or even which phase will nucleate. Moreover, even if both the destination and route are known, many algorithms require a parameter that specifies precisely how far the transition has evolved.

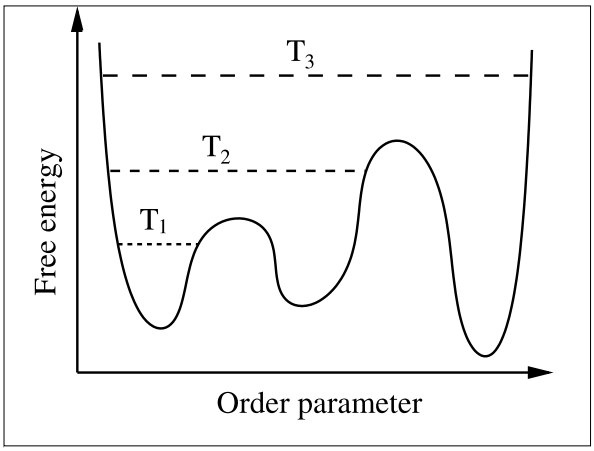


Figure 2.3: Schematic representation of the Parallel Tempering method. The region that is accessible to the system depends on the temperature, $T_1 < T_2 < T_3$. A higher temperature, indicated by the horizontal dashed lines, enables the simulation to overcome the barriers separating neighboring free-energy minima.

And finally, we need to reduce the complexity to get some physical insight into the underlying mechanism. Often it is convenient to quantify the progress of the system along the transition pathway with the help of a so-called order parameter or reaction-coordinate. Ideally, this quantity provides an intuitive measure of the degree of nucleation. However, what is intuitively simple need not be computationally straightforward and indeed, order parameter are often complex mathematical functions. We will come back to this point in detail in Section 2.3. In what follows, we assume that we are equipped with a set of order parameters capable of describing the transition of interest, and attempt to group the available algorithms according to the just-mentioned challenges they address.

The first class of algorithms that we discuss are so-called one-ended methods. They aim to explore possible transitions in a system where only the initial state is known. The key to this challenge is to escape the initial basin of attraction and move the system to other parts in phase-space. One method that works without the need of an order parameter is parallel tempering [28–30]. In this method, one runs simultaneously several simulations of the same system at slightly different temperatures (or more generally, with slightly different Hamiltonians). Periodically, one attempts to swap the states of two neighboring temperatures. This is schematically depicted in Figure 2.3, which indicates the different temperatures as horizontal lines indicating the accesible region in the free-energy landscape. Note that the order parameter in the schematic is only used for visualization, but not actually used in the method. At a higher temperature the system can escape shallow free-energy minima and there-

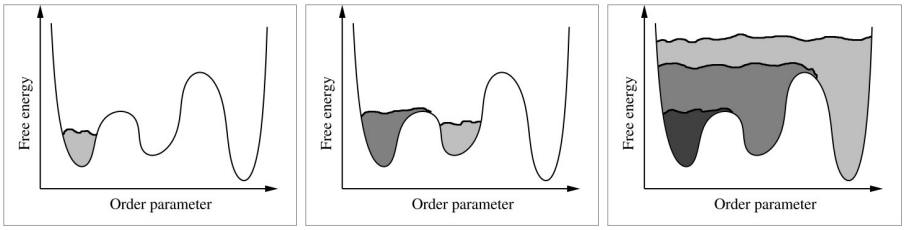


Figure 2.4: Schematic representation of the Meta Dynamics algorithms, which leaves a trail of entropic ”sand” indicated by the shaded area to fill up free-energy minima. In this picture the sand that was deposited most recently has the lightest shade of gray. The panels correspond to various stages of the simulation: sampling the initial basin of attraction (left), exploring the second minimum (center), and the final stage of the simulation, when the free-energy landscape becomes effectively flat (right). Once all minima are filled up, the free-energy landscape can be extracted from the (negative) sand profile.

fore explore the phase-space. At a lower temperature the system settles down into the free-energy minima revealing the (meta-)stable states of interest. The use of both high and low temperatures in combination with the state swapping leads to an effective sampling. Of course, this method is not only limited to temperature; other quantities like the chemical potential can be used, or more than one quantity simultaneously [31, 32]. Another algorithm that does not involve an order parameter is based on dynamical instabilities.This Lyapunov-weighted sampling [33, 34] evolves simultaneously several copies, so-called walkers, of the same system, too. But in contrast to parallel tempering, each walker evolves at the same conditions (i.e. temperature, pressure, etc.). New walkers are created and existing ones destroyed according to a probability related to their (largest) local Lyapunov exponent, a measure for chaotic instability. In this way unstable attractors are approached, which for a phase transition correspond to the saddle points between two (free-)energy minima. A different approach to explore the free-energy landscape was proposed by Wang and Landau [35, 36]. The idea behind this method is to sample phase-space whilst, on the fly, biasing the sampling in such a way that regions that have already been visited become less favored. This is achieved by means of a histogram sampling the density of states, which affects the acceptance probability for the generation of new states. The computed density of states is slowly refined until the overall sampling becomes flat, in which case the underlying free-energy landscape can be extracted. A closely related method is called Meta-Dynamics [37–39]. Rather than computing a histogram, it leaves behind a trail of Gaussians in order-parameter space. An intuitive description of this process might be that it is like walking through a mountain landscape and filling up the low-lying areas that one visits. Using such an approach

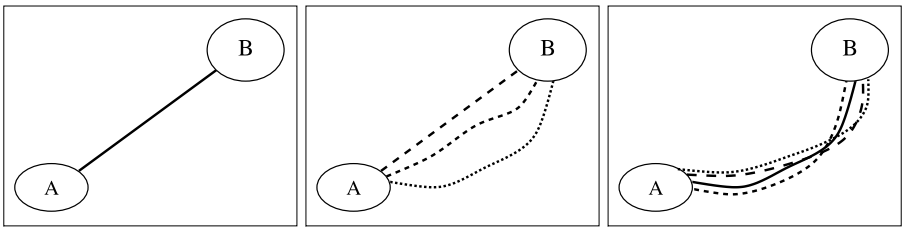


Figure 2.5: Schematics of transition path sampling: left) initial path between A and B. center) the path is evolved in a Monte Carlo manner. c) The path has converged and only minor fluctuations are observed.

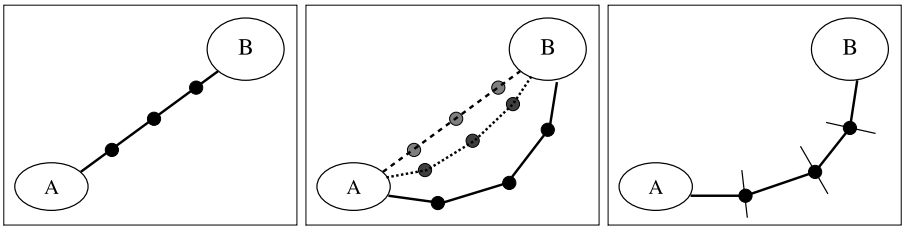


Figure 2.6: Schematics of the string method: a) initial path between A and B. b) Piece-wise evolution of the path.. c) The final path has evolved.

in a simulation ensures that the sampling moves increasingly to regions that would not be visited in a normal Boltzmann sampling procedure. As time progress, the free-energy minimum where the simulation is started fills up completely and the system escapes to another minimum. Once the sampling landscape becomes flat, the underlying free energy landscape can be deduced from the biasing potential. The state of the free-energy landscape at different times during the course of a simulation is presented in Figure 2.4, which shows the potential energy bias as shaded area filling up the free-energy minima. A similar idea is pursued in temperature accelerated molecular dynamics [40–42] (TAMD). Here, the system dynamics is restrained to a reference point in order-parameter space like a ball-and-chain. In short runs the local free-energy landscape is sampled providing a potential of mean-force for a Langevin-type evolution of the reference point. Because the movement of the reference point is unrelated to the actual dynamics in phase-space, a much higher effective temperature can be applied to the dynamics in order-parameter space, leading to an efficient sampling.

The second class of algorithms are two-ended methods. They are deployed if the starting and final state are known, but the actual pathway from A to B is unclear. The most prominent algorithm is transition path sampling [43, 44] (TPS). It is schematically depicted in Figure 2.5. It starts with an estimate for the transition path, which

is then subsequently evolved using a Monte Carlo path-sampling scheme. To generate a new path from an existing one, a point along the path is chosen at random and slightly perturbed. The perturbed path is then evolved both forward and backward in time until it ends up in one of the two basins of attraction. If the new path connects A with B it is added to the transition path ensemble, else it is rejected. This procedure is repeated until no significantly different pathways are found. The set of well-equilibrated paths generated in this way are referred to as the transition path ensemble. Although TPS needs an order parameter that distinguishes products from reactants, the order parameter only has to determine whether the system is in A or B, or in neither of both, but not where it is in-between. Since start and end are generally quite distinct the requirements for a "good" order parameter are less demanding. Another algorithm is the string method [45–49]. Like TPS it starts with an estimate for the transition path, but then it evolves the transition path directly in order parameter space, as shown in Figure 2.6. For this task the transition path is discretized, and the system is sampled at each of the points X_i along the paths to compute the potential of mean-force. To keep the system at or around its point X_i it has to be either constrained, e.g. by means of a blue moon ensemble [50, 51], or restrained, adding a bias on the order parameter [52]. Each point along the path is then evolved ccording to the mean force perpendicular to the path, and the string re-parametrized. Once the path has converged the minimum free-energy path is obtained.

The last class of algorithms concern the calculation of free-energy barriers or reaction rates. This is already possible with some of the algorithms presented in the last paragraph or its derivatives, such as transition interface sampling [53–55] or the finite-temperature string method [48], but a few methods have been designed specifically with this task in mind. Already in 1977 Torrie and co-workers developed the so-called umbrella sampling [52] (US), which has later been applied to the study of nucleation phenomena [56, 57] and recently extended to obtain the steady-state probability distribution of non-equilibrium ergodic processes [58, 59]. The concept is to split the sampling range into several independent windows, which can then be merged afterwards to yield the entire free-energy curve. How this can be applied to compute the free-energy barrier for nucleation will be explained in detail in the next section. A method that calculates reaction rates rather than free-energy barriers is forward-flux sampling (FFS) [60–62]. It splits the path into so-called interfaces and divides the rate calculation into two parts: the computation of the flux through the first interface and the computation of the probability to reach the final state B before returning to state A starting at the first interface. The latter probability is rewritten as a product of probabilities to reach interface $(i+1)$ before A starting at interface i, which is sampled sequentially. A detailed description of this method is presented in Section 2.2.2. As a final example for the calculation of transition rates we refer to the computation of mean first passage times [63, 64]. Based on theoretical arguments,

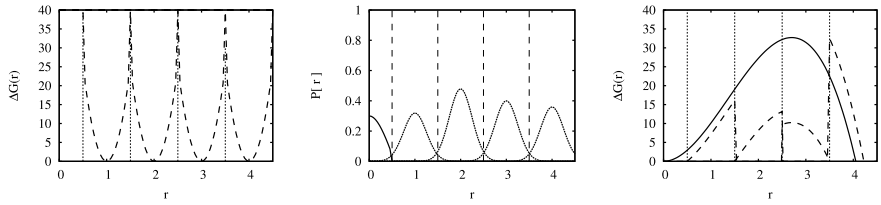


Figure 2.7: Various stages during umbrella sampling. Left: definition of the sampling window borders (thin dotted lines) and the bias potentials (dashed lines). Center: the sampled distribution for each window. Note that the first window ($r < 0.5$) is sampled without bias. Right: Un-biased distributions and the merged final free-energy curve.

the mean time τ_i to first observe the system at position i along the path can be fitted to an expression that then yields both the nucleation rates and the committor, the point at which the probability to grow to the final phase is 0.5. Although it was shown to yield results close to the experimental nucleation rates for Argon condensation, it is limited to highly supersaturated systems, where the free-energy barrier for nucleation is sufficiently low to be overcome in a un-biased simulation.

From this overview it is evident that there exist a large variety of algorithms to explore all aspects of a phase transition, and it sounds appealing to use them all: start with a one-ended method to find the relevant free-energy minima of the system, then apply a two-ended method to optimized the transition path, and finish with computing both the free-energy barrier and rate for nucleation along this path. However, in practice the limited computing facilities force us to focus on one aspect only, because although significantly more efficient than brute-force sampling, the computational cost of each individual algorithm is still considerable.

2.2.1 Umbrella Sampling

Free-energy difference can be computed as an ensemble average. However, traditional Metropolis Monte Carlo simulations are extremely inefficient for such computations because of the main contributions to the average often come from parts of configuration space that are hardly sample in a normal MC simulation. To address this problem, Torrie and Valleau [52] proposed a more efficient Monte-Carlo sampling scheme to compute free-energy differences. The authors proposed a method to use arbitrary sampling distributions, which can be chosen to facilitate the sampling of free-energy differences. In essence, one adds a weighting function w, also known as biasing function, to the phase-space coordinate $\mathbf{q}^\mathbf{N}$, to obtain a distribution $\pi(\mathbf{q}^\mathbf{N})$,

$$\pi(q^N) = \frac{w(\mathbf{q}^\mathbf{N})\exp\{-\beta U(\mathbf{q}^\mathbf{N})\}}{\int d\bar{\mathbf{q}}^\mathbf{N} w(\bar{\mathbf{q}}^\mathbf{N})\exp\{-\beta U(\bar{\mathbf{q}}^\mathbf{N})\}}, \tag{2.14}$$

which is then applied for sampling. The weighting function $w(\mathbf{q}^{\mathbf{N}})$ is chosen to favor those configurations $\mathbf{q}^{\mathbf{N}}$ that are important to the ensemble average of interest, and optimally results in a flat sampling histogram. The unbiased ensemble average of any function $\theta(\mathbf{q}^{\mathbf{N}})$ can be recovered according to

$$\langle\theta\rangle_0 = \frac{\int d\mathbf{q}^{\mathbf{N}}(\theta/w)w\exp\{-\beta U\}}{\int d\mathbf{q}^{\mathbf{N}}(1/w)w\exp\{-\beta U\}} = \frac{\langle\theta/w\rangle_\pi}{\langle 1/w\rangle_\pi}, \tag{2.15}$$

where $\langle\ldots\rangle_\pi$ denotes the average over the distribution $\pi(\mathbf{q}^{\mathbf{N}})$ of Eqn. 2.14. Because this sampling scheme should span all the relevant parts of the free-energy landscape, the authors named the method Umbrella Sampling (US).

In the context of phase transitions US enables a simulation to sample the behavior of the critical nucleus at the top of the free-energy barrier, a point in phase-space that is both highly unlikely and unstable. The biasing function then acts on the order parameter. Many implementations use a quadratic biasing function, restraining the system to fluctuate around a given value. More advanced implementations steadily improve the biasing function with a fit to the sampled distribution after unbiasing. In this manner the biasing function slowly converges to the (negative of the) free energy, resulting in a flat histogram.

Moreover, Torrie et al. argued that for a wide-ranging distribution the choice of a suitable bias function becomes tedious, and that it is more convenient to use several overlapping windows with different bias functions. In fact, following this line of thought, for many systems using multiple windows leads to faster convergence. This can be understood assuming a one-dimensional random walk, for which the time to explore the full range of a window scales quadratically with its width $\Delta = L/N$, and hence the total sampling time for all N windows scales as $T \sim L^2/N$. Although $T \to 0$ for $N \to \infty$, in practice numerical aspects limit the number of windows. Note that using multiple windows yields the free energy locally except for a constant offset, so a slight overlap between neighboring windows facilitates merging the local free-energy curves.

Figure 2.7 depicts schematically the application of US to sample a free-energy barrier for nucleation. The bias function acts quadratically on the radius of the largest cluster r and the calculation is split into 5 windows. Note that the (meta-stable) free-energy minimum at $r < 0.5$ is sampled without the use of a biasing function. The reason for this lies with the choice of using a the order parameter, as will be explained below. Panel (a) shows the quadratic biasing functions for all windows. Panel (b) depicts the obtained sampling histogram, and panel (c) presents both the individual unbiased window results and the merged free-energy barrier for nucleation.

From my experience, the best ratio of accuracy versus computational cost is achieved using many windows ($10 \leq N \leq 50$). For each window I start with a

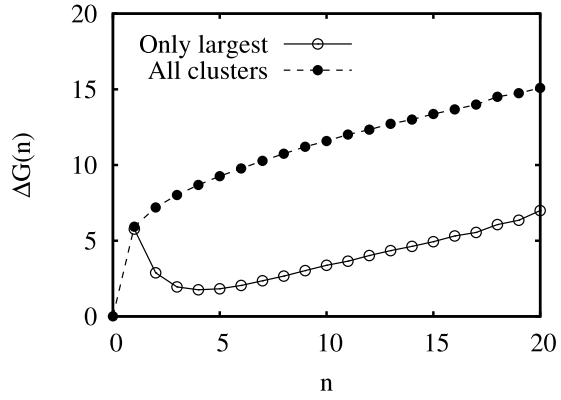


Figure 2.8: Free-energy curve sampled with umbrella sampling in the basin of attraction. The probability distribution obtained from a histogram sampling only the largest cluster yields a spurious minimum (open circles), because it neglects the presence of other clusters. The correct free-energy curve (filled circles) is calculated from a histogram including all clusters including monomers or, in the case of crystallization, liquid particles ($n = 0$). As a consequence, in the case of crystallization the correct free-energy curve includes a cost associated with the formation of a crystal monomer, since the majority of particles will be identified as liquid-like.

quadratic potential, and after some sampling I fit the results using a low-order polynomial (linear or quadratic) to improve the efficiency of my biasing function.

Spurious artifacts of US with a local order parameter

In the context of phase transitions, Umbrella Sampling requires an order parameter that monitors the transition progress. Order parameters suitable for this task can be classified into two groups: *local* and *global*. Order parameter of the latter group refer to the system as a whole. Examples are the system density, e.g. for a vapor-liquid transition, or the fraction of particles in a crystalline environment. In contrast, a local order parameter refers to a small part of the system, for instance the number of particles in the largest high-density cluster. The properties of such a cluster may be very different from the bulk and therefore does not represent the system as a whole. Although both groups may be suitable to compute the free-energy barrier for nucleation, their use in US comes with some subtle differences [57].

Umbrella sampling using a global order parameter x is fairly straightforward. Sampling the basin of attraction with a bias function yields, after un-biasing, the same probability distribution as sampling without a bias, and hence the same free-energy curve. The subtlety becomes evident at a later stage in the path analysis.

Imagine a vapor-liquid transition: we are interested in how a liquid droplet nucleates and grows. However, if the bias involves a global order parameter x, small values of x will not result in one small droplet, but instead in many tiny droplets, a process called entropic break-up [57]. Once x exceeds a certain threshold it becomes favorable to form just one big droplet in the system, which means that all the tiny droplets have to coalescent. In a condensation transition this might be possible, because the liquid droplets diffuse almost freely in the vapor, but for crystallization coalescence becomes unlikely during the run-time of a simulation. Hence, a global order parameter does not only lead to numerical problems, but also to a physically questionable nucleation pathway.

Local order parameters, such as the number of particles in the largest cluster of a new phase, offer a solution to avoid entropic break-up. They follow the evolution of a single cluster rather than the whole system, which compares well with CNT's picture of a growing nucleus. Umbrella sampling with such an order parameter produces a nucleation pathway that is physically more appealing. The free-energy barrier for nucleation is again calculated from the probability histogram, $\beta G(x) = -\ln P(x)$, but here the subtlety lies within the definition of $P(x)$. In contrast to a global order parameter, which describes the whole system, it is possible to have many clusters in the system, all with a different x. If the sampled cluster is very large, the probability to observe a second cluster of this size or larger is negligible due the high free-energy cost associated with its formation. In this situation it is justified to sample only the largest cluster using a bias function. But for small clusters the situation is different. Due to the presence of other small clusters, a bias forcing the largest cluster to fluctuate around a certain size will prevent any other cluster to exceed this size and hence constrain the system. This forces the system into an artificial state with suppressed cluster sizes, which leads to an increase in free energy at small cluster sizes, and hence create a spurious free-energy minimum. Such a behavior is depicted in Figure 2.8 (open circles). The solution to this problem is to sample, for small x, the entire cluster size distribution without a bias, and compute the free-energy barrier from the histogram including all cluster sizes. Figure 2.8 also shows the resulting correct behavior of the free-energy barrier (filled circles).

2.2.2 Forward Flux Sampling

Forward-flux sampling (FFS) was first introduced by Allen and co-workers in the context of modeling genetic switches [60, 61]. The method was later applied to the study of crystallization in binary-charged colloids [65]. The idea behind this method is to divide the path along a reaction coordinate x by a set of interfaces $\{\lambda_i\}$ with $A \leq \lambda_A < \lambda_1 < \cdots < \lambda_N \leq B$. The rate calculation is then split into two parts: sampling the flux Φ through the first interface λ_1 and computing the probability to reach B

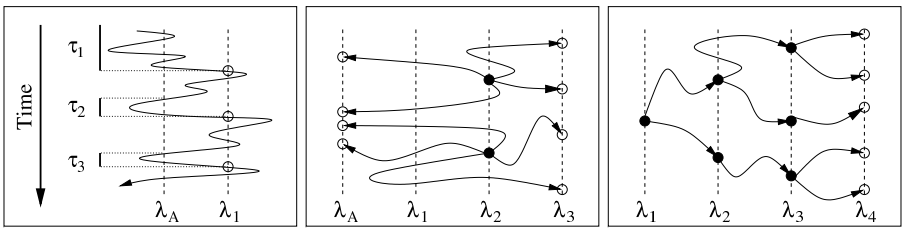


Figure 2.9: Schematic representation of the various stages during forward-flux sampling. Left: sampling of the flux through the first interface. Indicated are both the crossing times τ_i and the future starting points (open circles) for the shooting phase. Center: Shooting of trial trajectories from interface λ_2. Starting points are indicated by filled circles and end points by open circles. Trajectories end successfully if they reach λ_3, or unsuccessfully if they go all the way back to λ_A. Right: Final path reconstruction connecting trajectories across interfaces.

before A starting at λ_1. The latter quantity is replaced by a product of probabilities $P[\lambda_i \rightarrow \lambda_{(i+1)}]$ to reach interface $\lambda_{(i+1)}$ before returning to A (by crossing λ_A) starting at interface λ_i. The overall transition rate from A to B (i.e. the forward rate) is then given by

$$J_{A\rightarrow B} = \Phi \, \Pi_{i=1}^{N-1} P\Big[\lambda_i \rightarrow \lambda_{(i+1)}\Big]. \tag{2.16}$$

It is worth noting that this method does not involve the calculation of a free-energy barrier (although it is reported in Ref. [62] that it can be obtained with a minor extension), which makes it suitable for the study of both equilibrium and non-equilibrium processes.

The flux provides a measure for the rate at which a system in state A crosses the barrier for nucleation. For its computation a simulation evolves the system in the basin of attraction A according to its normal dynamics. During the run the order parameter x is monitored, and every time the first interface λ_1 is crossed in the forward direction (i.e. coming from A), the state is saved for future use and the time τ is sampled. The next time measurement τ starts the moment the system returns to state A (i.e. passing through λ_A coming from λ_1) , such that the time to relax to the basin of attraction is not included. After sufficient sampling this part of the algorithm provides the flux as the inverse of the average crossing time, $\Phi = 1/\langle\tau\rangle$, and the collection of states at the first interface forms the starting points for the next step in the procedure.

The second part of the method concerns the calculation of the probabilities. For this task a series of trajectories, so-called *shots*, are fired from the set of points at the first interface. Such a short run is stopped once the trajectory returns all the way to λ_A, or alternatively if it reaches λ_2, the next interface, at which point it is called *successful* and the state point is saved again. The fraction of successful shots provides

an estimate for the probability to advance from λ_1 to λ_2 before returning to A. In this manner all the interfaces are sampled sequentially, and the product of probabilities is obtained. Note that, to improve statistics, one and the same state point can be a starting point many times over, provided every generated trajectory is different, which can be ensured in chaotic many-particle simulations by a slight random perturbation or in stochastic systems by a change in the random number generator seed.

There exist several options for the implementation of the second part of FFS differing in the way the state points are sampled at each interface [61]. Examples are the branched growth, the Rosenbluth sampling, and the FFS sampling. In the branched growth sampling, a fixed number of shots are fired from every starting point of a given interface. Every successful shot is saved and forms a starting point at the next interface. If more than one trajectory arrives at the next interface, the path branches and forms a tree. If no trajectory arrives at the next interface, that branch dies out. The problem with this scheme is that one has little control over the number of trajectories at each interface. In particular, once the system reaches the top of the barrier, almost every shot will be successful, leading to an exponential increase. In the Rosenbluth scheme an interface is sampled until sufficient statistics are obtained. Then one successful shot is chosen at random, which is then taken to be the starting point for the next interface. After the sampling is complete, the path has to be re-weighted to correct for picking paths randomly, and not according to their weight in the transition path ensemble. The last scheme offers control over the sampling accuracy and the number of trajectories without the need of path re-weighting. In this "classical" scheme a fixed number of shots is performed at each interface, picking a starting point at random, and all successful paths are stored for the next interface. In the appendix of Ref. [61] it is demonstrated that this method selects paths according to their correct weight in the transition path ensemble eliminating the need of re-weighting. In practice, it is advantageous to continue sampling until either a minimum number of shots or a minimum number of successful paths are generated. This leads to a good control of accuracy for both a negative and a positive free-energy slope, respectively.

2.3 Order parameters

Section 2.2 reviewed algorithms for the study of rare events. As mentioned, many of these techniques require an order parameter, and some even a whole order-parameter space. Order parameters, also referred to as reaction coordinates in a chemical context or collective variables in a mathematical context, capture the essential physical aspects of a phase transition, but remove redundancies, e.g. due to translational or rotational symmetries. Often a single order parameter is not enough to characterize the

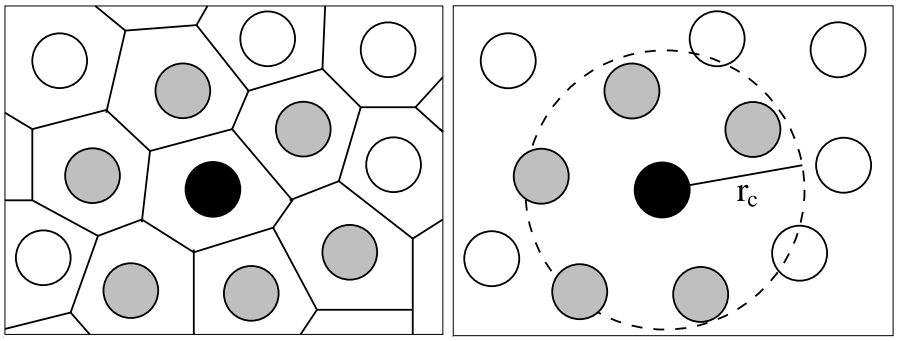


Figure 2.10: Two examples of a neighborhood definition. In both cases the center particle is shown in black, and the identified neighbors in gray. Left: two-dimensional Voronoi construction: particles that share a line are considered neighbors. Right: fixed-distance cutoff: all particles within a circle of radius r_c are neighbors.

system, and a set of them is deployed, which then span the so-called order-parameter space. The criteria for a good order parameter differ by method. The authors of the string method claim that every order parameter may be included into the set of collective variables, because even if it does not contribute actively to the path, it will not cause any harm to the method. For some methods, such as transition path sampling, it is sufficient to identify the (meta-)stable start and end points of a transition. For other methods, such as FFS or US, it is essential that the order parameter correlates closely with the transition progress, and that it increases monotonously moving along the path from A to B.

From the rare-event methods' dependence on the order parameter it is evident that a judicious choice is crucial for the success and the numerical efficiency of any algorithm. This is why the remainder of this Chapter is dedicated to the order parameters used in this Thesis. As was already discussed in Section 2.2.1 of this chapter, both a local and a global approach is possible. However, a local order parameter yields a physically more plausible pathway, which is the reason why we consider it here. Moreover, in practice a local order parameter based on the number of particles in the largest cluster of the new phase has proven itself many times over. This leaves the question how to compute the largest cluster and how to identify particles of the new phase. Since the latter question depends on details of the phases involved, we will address it separately for condensation and crystallization in Sections 2.3.1 and 2.3.2, and first focus on the determination of the largest cluster.

Assume that all particles of a new phase have been identified. To calculate the number of particles in the largest cluster we first need to define a cluster. A natural criterion that proved itself useful considers a cluster to consist of particles of the same phase that are all connected, and two particles are connected if they are nearest

neighbors. Again, there exist more than one way to define nearest neighbors. The Voronoi construction considers particles to be neighbors if their Wigner-Seitz cells share a face (or a line in 2D) as depicted in Figure 2.10a. But since the Voronoi construction is computationally rather expensive it is rarely used on-the-fly during simulation. A simpler, and faster, definition is based only on the particles' distance. This fixed-distance cutoff criterion considers all particles neighbors who's center-of-mass distance is smaller than the radius corresponding to the first minimum of the radial distribution function $g(r)$. This algorithm is schematically depicted for a 2D system in Figure 2.10b. Which neighborhood definition is used depends on the particular requirements. In Appendix A of this Thesis the advantages and disadvantages of these two neighborhood criteria are explained in more detail, and an alternative definition is proposed.

2.3.1 Liquid vs. vapor

The process of condensation is investigated in several chapters of this Thesis. Order parameters for condensation are somewhat simpler than those for crystallization discussed in Section 2.3.2, because both the vapor and the liquid are disordered structures, and the transition is easily identified by a large change in density. However, as detailed in Section 2.2.1, the overall system density is not the first choice. Instead, we deploy a local order parameter, which requires to identify particles of the new phase. Since density is the key quantity, we need a definition for a local density on a per-particle basis. This allows to distinguish between particles in a liquid-like environment and particles in a vapor-like environment by comparing their local density to a well-tuned threshold. Such a threshold can be obtained from computing the local-density distribution for both phases in the bulk and minimizing their overlap.

As for the definition of neighboring particles, the Voronoi construction provides a natural measure for the local density as it constructs the Wigner-Seitz cell around each particle, and the inverse of the cell's volume corresponds to the local density. Although this definition sounds intriguing, it involves significant computational cost, and is prone to strong fluctuations.

An alternative is a criterion based on the inter-particle distance. This fixed-distance criterion, also known as Stillinger's overlapping spheres criterion [66], defines all particles within a sphere of certain radius centered around a particle of interest to be its neighbors. Although it is possible to construct a local volume based on the number of neighbors and the sphere's volume, it is equivalent to impose the threshold directly on the number of neighbors. In a study on Lennard-Jones condensation Wedekind et al. [67] argued that, using a distance cutoff corresponding to the first minimum of the radial distribution function, a threshold of $n = 5$ particles yielded a critical nucleus size that correlated best with the experimental observations

for Argon. But the physical justification for this observation is, at this stage, unclear.

2.3.2 Crystal vs. liquid

In the process of crystallization the system undergoes a density change, too, but it is significantly smaller than for condensation. In fact, on the basis of individual particles a local density-based order parameter cannot distinguish between an emerging crystalline nucleus and a local density fluctuation of the liquid. But another difference lies in the local structural arrangement of particles, since the liquid is disordered and the crystal ordered. For instance, with the Voronoi construction the local arrangement of a particle can be classified with a set of polyhedra [68–70]. The so-called Common Neighbor Analysis decomposes the local environment into pairs, triples, and higher-order n-tuples of neighbors shared between two particles [71]. Steinhardt et al. [72] proposed to expand the local environment of particles into a set of spherical harmonics capturing the symmetries of the emerging lattice. This approach was later on extended by ten Wolde and co-workers [57] to yield reliable results on a per-particle basis. In what follows this method will be reviewed.

The first step in the procedure is to identify the nearest neighbors $\{j\}$ for each particle i. This is achieved commonly by means of a fixed-distance cutoff corresponding to the first minimum of the radial distribution function of the crystal phase. Each particle is then assigned a $(2l+1)$-dimensional complex vector based on the spherical harmonics Y_{lm} of order l capturing the local environment of particle i,

$$\Big(q_l(i)\Big)_m = \frac{1}{N_b(i)} \sum_j Y_{lm}(\hat{\mathbf{r}}_{ij}), \tag{2.17}$$

where components m of the spherical harmonic range from $-l \leq m \leq l$, $N_b(i)$ is the number of nearest neighbors of particle i, and $\hat{\mathbf{r}}_{ij}$ denotes a unit vector connecting the centers of mass of particles i and j. The order l depends on the symmetry of the lattice and must be chosen accordingly. For instance, both a fcc and bcc lattice are well-represented using $l = 6$, but for a diamond lattice, which is anti-symmetric, order $l = 3$ is better suited. However, if the second neighbor shell of the diamond lattice is included, the 6^{th}-order spherical harmonics perform well again (see Appendix A).

From this we can construct a quantity that is invariant versus both translation and rotation, and does not depend on the frame of reference. This is an important property, since a crystal nucleus can form anywhere in the system with any orientation,

$$\bar{q}_l(i) = \left(\frac{4\pi}{2l+1} \left| \mathbf{q}_l(i) \cdot \mathbf{q}_l^*(i) \right| \right)^{1/2}, \tag{2.18}$$

with the superscript star (*) indicating the complex conjugate. A distribution of $\bar{q}_6$ is presented in Figure 2.11 for both a meta-stable liquid and a fcc crystal of a Lennard-

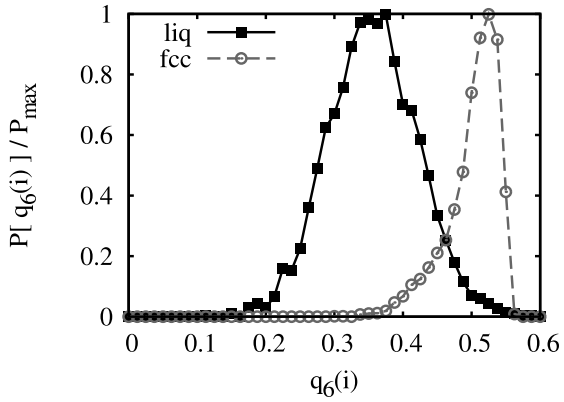


Figure 2.11: Distribution $P[\bar{q}_6(i)]$ for the order parameter of Eqn. 2.18 for both a Lennard-Jones liquid and fcc crystal bulk phase. Note that there exists a significant overlap between the two phases preventing a unique classification. Both curves are normalized by their maximum P_{max}.

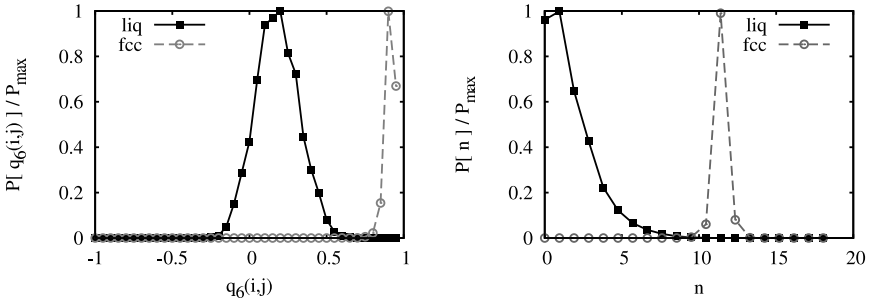


Figure 2.12: a) Local bond-order correlator distributions $P[q_6(i,j)]$ for both a Lennard-Jones liquid and fcc crystal bulk phase. b) Distribution of the number of links per particle $n(i)$ resulting in a unique classification. All curves are normalized by their maximum P_{max}.

Jones system ($T = 0.92$, $P = 5.68$). Although the distributions are distinct, there exists a significant overlap preventing a unique classification on a per-particle basis.

To improve the classification, Ref. [57] proposed an invariant that does not correlate the neighbor arrangement of one particle with itself, but with that of a neighboring particle,

$$q_l(i,j) = \frac{\mathbf{q}_l(i) \cdot \mathbf{q}_l^*(j)}{\sqrt{|\mathbf{q}_l(i) \cdot \mathbf{q}_l^*(i)|}\sqrt{|\mathbf{q}_l(j) \cdot \mathbf{q}_l^*(j)|}}. \tag{2.19}$$

This quantity, which we will refer to as "bond-order correlator", is invariant versus both translation and rotation, and does not depend on the frame of reference. But the main difference is that it compares effectively the neighborhoods of two particles, and that each particle has as many $q_l(i,j)$ as it has neighbors. Figure 2.12a plots the bond-order correlator distribution $P[q_6(i,j)]$ for the same system and conditions as in Figure 2.11. Although the distributions are already quite distinct, Ref. [57] goes one step further. Whenever $q_l(i,j)$ exceeds θ, the two particles i and j are considered to form a "link". Figure 2.12b presents the distribution of the number of links per particle $n(i)$. With this final step both the liquid and crystal distributions are well-separated, and a unique classification is possible by imposing a cutoff n_c on the number of links. At this point we like to mention that these distributions are different for each system, and that all parameters should be tuned to achieve a good classification of both bulk phases before an attempt is made to identify a growing nucleus.

Particles at surfaces

A simple extension to the procedure outlined in Ref. [57] concerns particles at a surface or at an interface with a low-density phase: they tend to have fewer nearest neighbors than particles in the bulk. Imposing a threshold on the absolute number of links prevents those particles from being identified correctly, even if they are positioned in a perfectly crystalline arrangement. To avoid this problem, Mendez-Villuendas and Bowles [73] proposed to change the threshold on the number of links to a threshold on the *fraction* of neighbors that form a link. Using this extension they could study surface nucleation in the freezing of gold nano-particles. In Chapter 4 we apply this modification to study vapor-crystal nucleation in a Lennard-Jones system at conditions below the triple point to check whether crystal nucleation in liquid droplets starts at the vapor-liquid interface or within the "bulk" of the droplet. And in Chapter 6 this modification enables us to grow a crystal cluster directly from the low-density fluid phase without the need of a second order parameter to first obtain a small cluster of a high-density disordered phase [10].

Imposing compactness

In Chapter 7 we present another extension to the local bond-order parameter which improves the compactness of crystalline clusters of four-dimensional hard hyperspheres. In this system, the fluid phase is very distinct from the crystal phase, and crystalline particles are rarely observed in an over-compressed meta-stable fluid. In order to grow an initial nucleus, the parameters had to be chosen to identify even weakest ordering. As a consequence, the growing crystal had many defects and the tendency to grow crystalline "arms" or branches into the fluid. This is an unphysical entropic effect induced by the choice of order parameter settings. To facilitate the formation of an initial nucleus without this artifact we modified the order parameter such that it returns the total number of links in the largest cluster rather than its number of particles. With 24 nearest neighbors in 4D a cluster can therefore increase its order parameter without growing in size by means of optimizing its inner order. This leads to the formation of very compact clusters, as was confirmed by computing the cluster's radius of gyration. For the final sampling of the free-energy barrier for crystallization and the comparison with Classical Nucleation Theory, we used these compact clusters only as input, but sampled with a bias on the conventional order parameter definition. Although the compactness was not encouraged anymore, the clusters remained dense compact objects - an indication that the initial branching was indeed unphysical.

Chapter 3

Harnessing graphics power for MD simulations

with A. Axel, S. Portegies Zwart, and R. Belleman

In this chapter we discuss an implementation of molecular dynamics (MD) simulations on a graphic processing unit (GPU) in the NVIDIA CUDA language. We tested our code on a modern GPU, the NVIDIA GeForce 8800 GTX. Results for two MD algorithms suitable for short-ranged and long-ranged interactions, and a congruential shift random number generator are presented. The performance of the GPU's is compared to their main processor counterpart. We achieve speedups of up to 40, 80 and 150 fold, respectively. With the latest generation of GPU's one can run standard MD simulations at 10^7 *flops/$.*

Over the last 30 years computer simulations have become an important tool in materials science, often bridging the gap between theory and experiment. Simulations can be used both to predict the outcome of experiments and to test the assumptions of theories. The basic idea of most classical simulations is to calculate the forces acting on all particles, and then integrate Newton's equations of motion using these forces. This approach is not limited to single atoms or molecules, the same approach can be used to model the motion of stars within galaxies.

With the rapid increase of available computational power, more systems become tractable for simulations. Nowadays it is possible to simulate the time evolution of simple molecules over microseconds with atomistic detail on a conventional personal computer. However, for many systems, the computational power of a single processor (CPU) is not sufficient. In this case, simulations are run in parallel on many processors, which allows us to simulate hundreds of thousands of molecules over time-spans of milliseconds. The increase of computational power comes at a price: the different processors have to exchange information on the simulated system continuously. This communication costs time, reducing the effective performance of a parallel system to typically less than 80% [74, 75] of the total performance of all its individual CPUs. And although there are standardised software tools for the implementation of this communication, such as PVM [76], MPI [77] or OpenMP [78], writing a code for parallel execution is not trivial. Moreover, the necessary very low latency, high throughput communication hardware often costs as much as the processing units themselves.

An alternative approach speeds up the simulations by using special purpose hardware. For example, in simulations of stars or charged molecules, more than 90% of the computation time is typically spent on the calculation of the gravitational or electrostatic interaction. Most prominently, the GRAPE board [79] is a special purpose hardware designed to calculate such interactions; recently, a variant called MD-GRAPE [80] has been put forward to calculate the interactions of more general pair potentials. Due to their specificity, these boards can achieve several orders of magnitude higher throughput compared to conventional CPUs, but are only of interest for a limited community of researchers. This makes these boards relatively expensive and their development cycle long.

Since 2003, a new route to gain additional computational power has opened: the graphics processors (GPUs) of recent PC hardware have become general purpose processors, which can be programmed using C–like programming environments such as the GL shader language (GLSL) [81], C for graphics (Cg) [82] or the NVIDIA compute unified device architecture (CUDA) [83]. Their computational power exceeds that of the CPU by orders of magnitude: while a conventional CPU has a peak performance of around 20 Gigaflops, a NVIDIA GeForce 8800 Ultra reaches theoretically 500 Gigaflops. This means, that 4 graphics cards can replace a complete 64 proces-

sor PC cluster, saving space and reducing the necessary power supply from 15kW to around 2kW. Moreover, graphics processors follow a Moore–law with a computational power doubling every 9 months, in contrast to 18 months for conventional CPUs. For the end of 2007, the first Teraflops-cards are expected.

There have been early attempts to harvest this computational power for various applications, including fast Fourier transforms [84], matrix operations [85], lattice Boltzmann simulations [86] or Monte Carlo (MC) simulations of the 2D Ising model [87]. Recently, Portegies Zwart et al. [88, 89] presented a N-body simulation with gravitational interactions, where the force calculation was performed on the GPU. For the latter application, the graphics cards are in direct competition with the GRAPE boards, and achieve similar performances at much lower costs and higher reliability. Yang et al. already presented a proof–of–concept molecular dynamics simulation for the thermal conductivity of solid argon [90]; their implementation is however limited to the simulation of defect–free solids.

In this chapter, we aim to assess the portability of classical molecular simulation systems onto GPUs using NVIDIA's CUDA [83]. Unlike the previous attempts of putting only the computationally most expensive parts of the simulation onto the graphics cards, we demonstrate that in fact the entire simulation can be ported to the graphics cards. The resulting program reproduces all data obtained from a standard single–processor simulation. We report benchmarks of three codes: two simulating the classical "work–horse" of coarse–grained molecular simulation, the Lennard-Jones system, and a classical rand48 random number generator [91]. We tested these codes on a system consisting of an Intel Xeon CPU running at 3.2 GHz and a NVIDIA GeForce 8800 GTX (16 multiprocessors, running at 675 MHz each). For both the simulation and the calculation of random numbers, we achieve an about 25– to 150–times speedup using the GPU compared to the CPU.

3.0.3 GPU architecture

To facilitate the discussion on the technical implementations, it is necessary to briefly summarise the key aspects of the GPUs hardware architecture and its nomenclature (see also Figure 3.1). We use the NVIDIA CUDA system for programming the GPU, which allows to write functions for the GPU, so–called *kernels*, in a C–like language. For detailed information we refer to NVIDIA CUDA programming guide [83].

The NVIDIA GeForce 8800 GTX consists of 16 multiprocessors (MPs). Each MP has a single–instruction–multiple–data (SIMD) architecture and is capable of performing 32 times the same operation on different data per two clock cycles. Many copies of a kernel, so-called *threads*, are executed in parallel on all available MPs on the GPU. To fit the SIMD architecture, groups of 32 threads form a *warp* which is executed on the same MP. If a kernel contains a branch and threads of the same

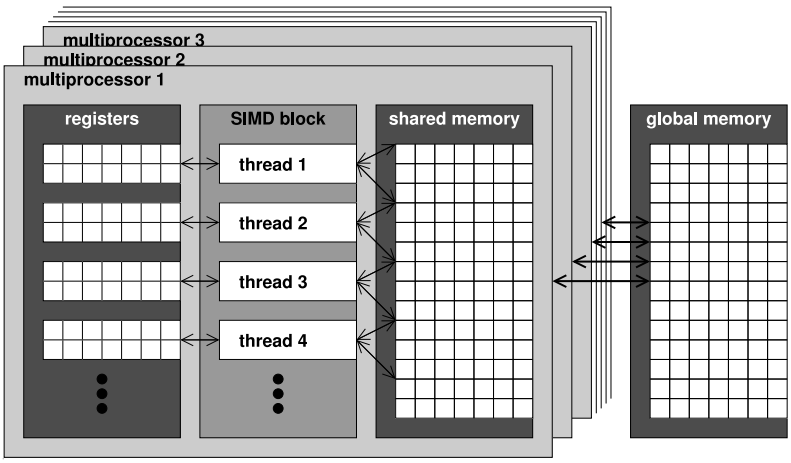


Figure 3.1: Schema of the multiprocessor and memory organisation of current NVIDIA GPUs. While registers are bound to a thread, all threads of the same SIMD block have access to a common shared memory, and all threads on all MPs have access to the global memory.

warp take different routes, then both routes are executed sequentially and the total run time is the sum of both branches. This *warp divergence* can have a serious impact on performance.

Threads can store data in 8192 32–bit registers per MP, and a high-speed *shared memory* of 16 KB per MP is available to share data among threads running on the same MP. For this, threads are grouped into *blocks* of up to 512 threads which are forced to run on the same MP. A slower *global memory* of 768 MB is also available that is shared among all MPs. To hide register read-write latencies of one to two clock cycles, it is recommended to use block sizes of 192 or more threads, and more than one block per MP should be scheduled in order to hide the much larger global memory read latencies of 200 to 400 clock cycles. Note however, that GPU global memory is still ten times faster than the main memory of recent PCs.

As a final remark we point out that nowadays graphics hardware only supports single precision floating point arithmetic. This might not suffice for systems where energy conservation is crucial. But for systems in thermal equilibrium, i.e. with a stochastic thermostat, this forms no limitation.

3.1 N-squared MD

We start with the most simple molecular dynamics algorithm in which each particle interacts with all other particles. Therefore, the total force calculation scales

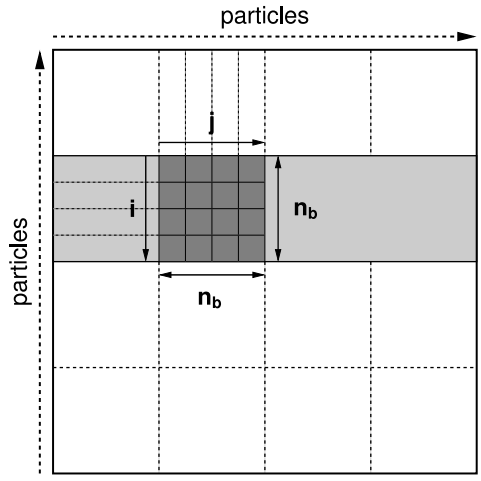


Figure 3.2: Schema of the splitting of the force calculation into blocks. Each of the n_b threads is dedicated to calculating all interactions of particle i. All particles j are loaded in blocks of again n_b particles, which are then shared among all threads.

quadratic with the particle number N. The force $\mathbf{f}_i$ on a particle i is given by

$$\mathbf{f}_i = -\sum_{\substack{j=1 \\ j\neq i}}^{N} f(|\mathbf{r}_j - \mathbf{r}_i|)\frac{\mathbf{r}_j - \mathbf{r}_i}{|\mathbf{r}_j - \mathbf{r}_i|}, \tag{3.1}$$

where $f(r)$ is the well-known Lennard-Jones pair force, truncated at a distance $r_c = 2.5\sigma$ and shifted such that the force at the cutoff distance was zero. The full Lennard-Jones pair force is given by

$$f_{\mathrm{LJ}}(r) = 24\varepsilon\left[2\left(\frac{\sigma}{r}\right)^{13} - \left(\frac{\sigma}{r}\right)^{7}\right], \tag{3.2}$$

our truncated and shifted force by

$$f(r) = \begin{cases} f_{\mathrm{LJ}}(r) - f_{\mathrm{LJ}}(r_c) & r < r_c \\ 0 & r \geq r_c\,. \end{cases} \tag{3.3}$$

The Velocity Verlet algorithm was applied to integrate Newton's equations of motion [92].

3.1.1 Implementation details

A Molecular Dynamics simulation is naturally suited for a SIMD architecture, because it performs the same set of operations on each particle. The most simple way

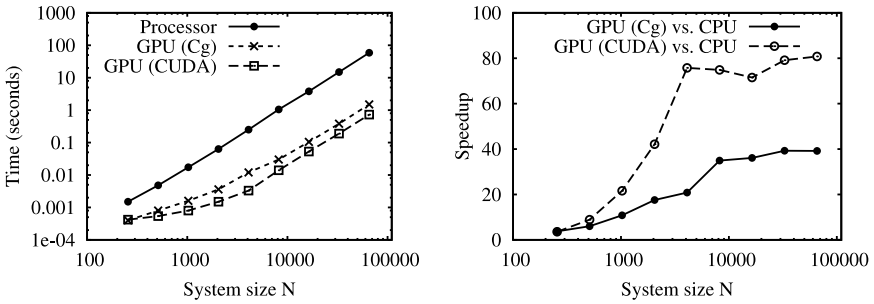


Figure 3.3: Left: Time (in seconds) required to integrate a single MD time step as a function of system size N. Right: Speedup factor for both GPU implementations, Cg and CUDA. The speedup saturates once both the processor and GPU versions have reached the quadratic scaling regime.

to parallelise this algorithm is to have one independent thread per particle. However, naively implementing Equation 3.1 turns out to be far from efficient. The reason for this is that every thread loads all particle positions from global memory, which is not cached. Each read access comes with some latency causing the processor to idle until the data arrives. A huge improvement can be achieved by taking advantage of the fact that all threads need the same data. By grouping threads into blocks, data can be shared among them, effectively reducing memory bandwidth and idle times.

Our implementation works as follows: each thread loads one different particle from global memory and stores it into shared memory. Then all threads of a block are synchronised to ensure loading has finished. Now the data of all threads are accessible through high–speed shared memory, and each thread can calculate the interactions of its dedicated particle with all other particles in shared memory (see figure 3.2). For a block of n_B threads, this reduces memory bandwidth by a factor $1/n_B$. In addition, each thread can now compute more interactions per memory read, allowing the thread scheduler to more efficiently hide global memory latencies. The optimal block size depends on the resources used by the kernel: number of registers and shared memory size. A block size of $n_B = 64$ turned out to be the optimal choice for our program. For details about the interplay between register usage, shared memory usage, block size and number of blocks per multiprocessor, we refer to NVIDIA's CUDA programming guide [83].

3.1.2 Results

To compare the performance of the GPU and CPU implementations on our test system, we measured the time required to integrate a single MD time step as a function

of system size N. The speedup factor x is defined as

$$x = \frac{T_{CPU}}{T_{GPU}}, \tag{3.4}$$

where T_{CPU} is the time used by the CPU implementation and T_{GPU} the time used by the GPU implementation. In addition to the GPU implementation with CUDA, we also present results from a GPU implementation in Cg [82], a language designed for graphic processing. It is supported by the majority of current graphics hardware, but does not provide the flexibility required for more complex MD algorithms.

The left graph in Figure 3.3 shows the average time used to integrate a single MD time step. The quadratic scaling of run time with system size is clearly visible for the CPU version. For the GPU code at small system sizes, the overhead of invoking the graphic program is comparable to the actual computation time. Therefore, the quadratic scaling regime is reached when this overhead becomes negligible, which corresponds to a system size of approximately 4000 particles.

The speedup factor for the GPU implementation is depicted in the right graph in Figure 3.3. Although the GPU version is faster for all our system sizes, it requires a system size larger than 4000 particles to reach its full speedup of around 80.

3.2 Cell-lists MD

If the pair interaction is short-ranged, the simulation box is typically decomposed into smaller domains, so-called cells, with a side length equal to or greater than the maximum interaction range. For a given particle, all interaction partners are then located in the same and directly neighbouring cells. Therefore, the algorithm scales linearly with the number of particles, but suffers some penalty due to the overhead associated with maintaining the cell structure. For small systems, this might be disadvantageous compared to the N-squared algorithm, but for large systems it generally results in a huge performance gain. The system size at which both algorithms perform equally well is called *break–even point*.

Another way of optimisation are the so-called Verlet lists. For each particle, a list holds all neighbour particles within a sphere of $r_V = r_c + \Delta r$, the Verlet radius. The Verlet lists *skin* with width Δr prevents particles to move into interaction range unnoticed and generally acts as an invalidation criterion. Every time a particles list is updated, the particles current position is stored as Verlet list centre. If this particle moved further than $\Delta r/2$ away from its Verlet list centre, the list has expired and needs to be rebuild. Obviously, the larger Δr, the less frequent the lists have to be updated, but the more unnecessary interactions with $r > r_c$ have to be computed. Updating Verlet lists is rather expensive and scales like $O(N^2)$. Therefore, cell lists

are often used to reduce its costs to $O(N)$. Compared to cell lists, Verlet lists further reduce the number of possible interaction partners and result in a theoretical seven fold speedup.

Yang et al. [90] used the Verlet lists approach to compute the thermal conductivities in solid argon on a GPU. However, their Verlet lists were computed only once (on the CPU) and never updated, which restricts its use to defect–free solids. To fit the SIMD architecture, they added virtual particles to obtain the same list size for all particles. Moreover, to avoid inner-loop branching which deteriorates the performance, the interaction cutoff distance was set to the Verlet list radius r_V. In doing so, they removed the essential skin from their Verlet lists allowing interactions due to fluctuations to be ignored. This shows that Verlet lists are not particularly suited for a SIMD architecture. The MD code of Ref. [90] is therefore useful as a proof of concept, but cannot be used for production runs.

In our program we applied only cell lists. They seem more suitable for the hardwares architecture and could be implemented to run entirely on the GPU. Care was taken not to neglect any interactions and to include cell list updates.

3.2.1 Implementation details

There are plenty of schemes to implement the cell lists technique [92]. One approach uses one linked list per cell to store the identities of the particles located in it. The advantage is that this scheme works well for all densities without parameter modifications, because there are no size limitations on a linked list. The disadvantage is that memory access is random, not sequential, and therefore a linked list cannot be loaded in parallel.

Another way is to assign a fixed sized array of placeholders (AOP) to every cell and physically copy particles position into this array. The advantage of this scheme is that interacting particles are physically close together in memory allowing for fast parallel loading. The disadvantage is that it generally requires more memory, because each AOP has to provide space enough to store particles at the highest possible density.

Our implementation uses the latter scheme. Per cell, one thread is devoted to one placeholder. Empty places are filled with virtual particles. Each thread i of a cell c_0 loads the data of placeholder i of cell c_n from global memory and stores it into shared memory. It synchronises with the other threads of the same cell c_0 to ensure loading has finished. Now it computes the interactions of particle i with all particles in shared memory. Note that this is done for virtual particles, too. These steps are performed for the centre cell, $c_n = c_0$, and all neighbour cells, $n = 1 \ldots 26$.

But the force computation is not the only task. As particles move, the cell lists have to be updated. While this is straight forward on a single CPU, the parallel

version comes with some difficulties. If a cell realises that one of its particles is about to move to a neighbour cell, it cannot move the particle there without the risk of memory–write conflicts and data inconsistency.

To safely update a cell in a parallel environment, we first remove all particles from the list which left the cell. Then all particles from neighbouring cells are checked to see if they moved in and need to be added to this list. Double–buffering ensures that all old lists stay intact until all cells have updated their list. Both for removing the particles from a cell that have left the cell and adding the particles that have moved in, we have to first test where a particle belongs, and then update the corresponding particle list by either deleting or adding particles.

Testing particles can be done in parallel. Each thread of a cell computes the cell id for one particle and stores it in shared memory. Now one thread sequentially loops over these particles and adds those with a correct cell id to the list. To prevent memory–write conflicts, this task has to be performed by a single thread per cell, leaving all other threads idle.

Updating a cell list requires all particles from this cell and its neighbour cells to be loaded from memory. In order not to load the same data twice, we perform this task during force calculation, not directly after the integration of positions. As a drawback the cell lists are not precisely up–to–date, but one time-step behind. In order not to neglect any interactions, the cells have to have a side length larger than the maximum interaction range plus a so-called skin of thickness λ, where λ is the maximum particle displacement per time step. For MD simulations it is common practise to use an even larger skin and therefore update the cell lists only every couple of time steps.

3.2.2 Results

As for the N-squared MD algorithm in Section 3.1, we compared the GPU implementation with its CPU counterpart. Because the algorithmic complexity exceeded the capabilities of Cg, only CUDA could be used.

The system size lower limit is given by the requirement to have at least 3 cells per dimension at a density of $\rho = 1.0$. The upper limit for the density was given by the array size associated with every cell, which was $n = 32$ in the data presented here. For a minimum cell size of $r_c = 2.5\sigma$ and a density of $\rho = 1.0$, on average $\rho\, r_c^3 \approx 16$ places per cell are occupied. This implies that most of the time half of the threads are calculating interactions of virtual particles which do not contribute. This deteriorates at lower densities.

The run time is density dependent: the more particles per cell, the more interactions have to be computed, and the computation time rises. This is true for the CPU version. However, our GPU version behaves differently. In contrast to the CPU

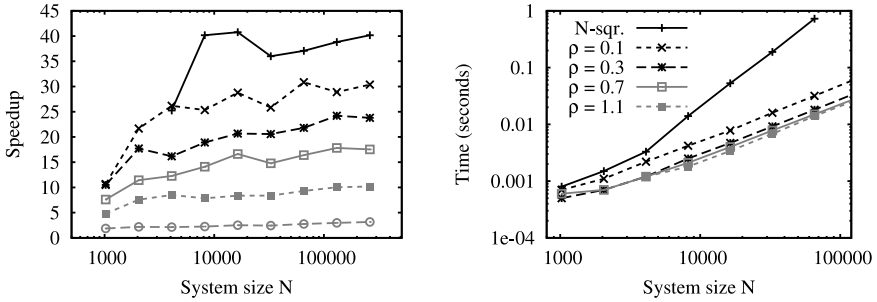


Figure 3.4: Left: Speedup of the GPU version at various densities (from top to bottom: $\rho = [1.1, 0.9, 0.7, 0.5, 0.3, 0.1]$). The kinks and bumps are reproducible. For details, see main text. Right: absolute times (in seconds) for a single MD step. The density-independent N-squared MD data is presented as a reference. The intersection points with this reference data would indicate the break-even points for the cell-lists algorithm. But for all system sizes shown here, the cell-lists MD is faster than its N-squared counterpart.

version, the GPU version's run time is dominated by the total number of cells, not by the number of interactions per cell. This is because interactions are always calculated for all placeholders; at low densities, most of them are however empty. At constant number of particles, the number of cells decreases with the density, and therefore the run time decreases. This effect saturates once all placeholders of a cell are used.

The left graph of Figure 3.4 shows the speedup factor for our GPU implementation. At the lowest density of $\rho = 0.1$, the GPU version is twice as fast as the CPU version. At higher densities, the GPU outperforms the CPU by up to a factor 40. The errors for these speedup factor are smaller than the symbol sizes and the kinks and bumps reproducible. They relate to the cell size, which fluctuates in order to get an integer number of cells per dimension. Assume a box length of $L_x = 11\sigma$ and a minimum cell size of $r_c = 2.5\sigma$; then the number of cells for this dimension is $n_x = \text{int}[L_x/r_c] = \text{int}[4.4] = 4$ and the actual cell size is $r_c' = L_x/n_x = 2.75\sigma$. This increase of 10% results in 33% more particles per cell, leading to 77% more interactions, decreasing the CPU performance. But for the GPU version a few more threads compute real particle interactions instead of virtual ones, resulting in no penalty.

The absolute computation times required per MD time step are depicted in the right graph of Figure 3.4. For comparison, the (density independent) N-squared MD data is shown as well. The cell-list data feature a different slope than the N-squared data, indicating linear and quadratic scaling, respectively. Intersection points with the N-squared curve would indicate the break even points, where both algorithms perform equally well. However, for all system sizes and densities depicted in Figure 3.4, the cell-lists version performed better than its N-squared counterpart.

3.3 Random number generation

For many applications in computer simulations, e. g. Monte Carlo simulations or Molecular Dynamics simulations with a stochastic thermostat, a large quantity of (pseudo–)random numbers is required. Typically, simple linear congruential generators such as the `lrand48` are used [91]. Given a number x_n, the following number in its series is generated as follows:

$$x_{n+1} \equiv a x_n + c \mod 2^{48} \tag{3.5}$$

where a and c are some integer constants. The pseudo–random number x_{n+1} is then converted to the pseudo–random number Y_{n+1} of the required data type. For Y_{n+1}, one uses the d most–significant bits of x_{n+1}, where d is the bit size of the required data type (e. g. $d = 31$ for nonnegative 32–bit integers).

3.3.1 Implementation details

To parallelise generation rule (3.5), we note that

$$x_{n+m} \equiv A x_n + C \mod 2^{48} \tag{3.6}$$

where

$$A \equiv a^m \mod 2^{48}, \quad \text{and} \quad C \equiv \sum_{i=0}^{m} a^i c \mod 2^{48}. \tag{3.7}$$

The random number generator is then implemented as follows. We choose a number S of random numbers to generate in parallel. To start the random number generator, we choose a seed x_0, and generate the first x_i, $i = 1 \ldots S$ according to the serial rule (3.5). The next set of S pseudo–random numbers is then generated from this set according to rule (3.6). Since the calculation of x_{i+S} only requires knowledge about x_i, all x_{i+S}, $i = 1 \ldots S$ can be calculated in parallel. If some multiple RS of S random numbers is required, each set x_i, x_{i+S}, …, $x_{i+(R-1)S}$ can be calculated independently by an isolated processor.

For the implementation on current GPUs, this is very convenient: Using S independent threads, each thread first loads its current state x_i. Then, it generates the following pseudo–random numbers x_{i+S}, calculates the output value Y_{i+S} and stores it. This step is repeated for all x_{i+nS}, until in total R random numbers have been generated and stored. Finally, the thread saves the current state $x_{i+(R-1)S}$. For our test, we used $S = 6144$ independent threads, grouped into 32 blocks of 192 threads each.

Note that all arithmetics is done modulo 2^{48}. However, GPUs (as well as standard CPUs) do not offer 48–bit data types. In principle, a 48–bit number can be represented by three 16–bit numbers, but for performance reasons it is better to represent

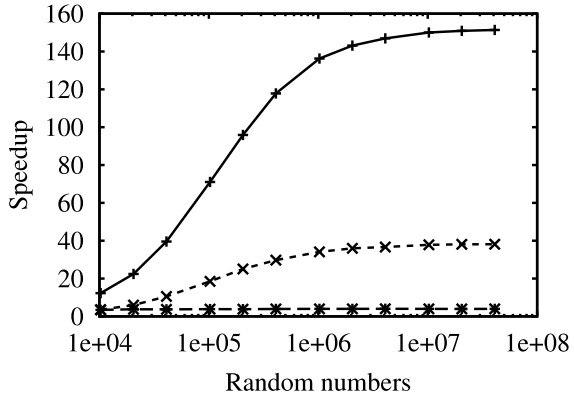


Figure 3.5: Speedup of a self-written CPU and a GPU version of the `rand48` random number generator versus the standard `glibc` implementation. From top to bottom: GPU vs. CPU (glibc), GPU vs. CPU (optimized), CPU (optimized) vs. CPU (glibc).

the 48–bit number as one 64–bit integer or two 32–bit integers. We have chosen to represent the x_n by two 32–bit integers, which contain the 24 most–significant and 24 lowest–significant bits.

3.3.2 Results

We compare our implementation of the `lrand48` random number generator on the GPU both to the standard GNU–libc `lrand48()` function as well as a self–written CPU version using 64–bit arithmetics. For each implementation, we measured the time necessary to generate the first N random numbers of the `lrand48` series for $10,240 \leq N \leq 40,960,000$. The resulting speedup factors of our implementations relative to the GNU–libc implementations are shown in Figure 3.5.

The optimised CPU version is consistently faster than the system implementation by a factor of almost four. This simply demonstrates the high 64–bit performance of current PC processors. However, the GPU achieves a much higher performance. For generating more than a million random numbers, the GPU is faster than the standard–libc `lrand48` by a factor of 150. Compared to our optimised CPU–version, the speedup is still almost 40.

Although the speedup factor for this pure integer arithmetics problem is therefore not as high as for typical floating–point problems, the GPU is still competitive for the generation of random numbers. Moreover, our implementation stores the output random numbers in the relatively slow main memory of the graphics card. Depending on the problem at hand, it is however often possible to generate random numbers on the fly, which will increase the speedup factor.

3.4 Summary and outlook

The computational power of recent graphics cards is fifty times as large as the power of a conventional processor. It has been shown previously that this speed can be harvested for many problems, e. g. matrix multiplication or the calculation of electrostatic interactions. In this work, we have demonstrated that it is possible to run a conventional MD simulation entirely on a graphics card. The simulations run 25-80 times faster than on a single conventional processor, at comparable prices. Similar results are reported by Anderson et al. [93] using a different technical approach. This shows that it is also possible to harvest this computational power for MD simulations.

Although our code features only the simple Lennard–Jones potential, it is trivial to replace this potential by other pair-potentials, including the coulombic interaction. By this, our code can in fact be used for many systems of interest. Moreover, the GPUs are indeed general purpose processors by now, and therefore it should be possible to implement many other techniques equally efficiently, such as Ewald summation methods or SHAKE for constrained dynamics. Currently the GPUs are limited to single precision floating point operations. For long non–thermalized simulations, in which energy conservation is crucial, this precision might not be sufficient. However, double–precision GPUs are expected for the end of this year.

While MD simulation techniques can be easily ported onto the GPU architecture, this does not hold for the equally wide–spread family of Monte–Carlo methods. Tomov et al. [87] have implemented a MC scheme for the 2D Ising model showing that lattice-based probabilistic simulations can be ported to the GPUs SIMD architecture. However, off-lattice many particle MC simulations are difficult to parallelise, both on conventional parallel architectures and on SIMD hardware. Reasons for this are the random acceptance moves causing unpredictable branching, and the permanent access to global information to obey detailed balance.

The difference in computational power between conventional processors and GPUs is expected to increase further. At the end of this year, NVIDIA GPUs are expected to reach Teraflops performance on a single card, and will feature double precision floating point operations, at a rate of 250 Gigaflops. Even with current off–the–shelf PC mainboards it is possible to build systems equipped with four graphics cards. A single PC can therefore obtain Teraflops performance, and a small cluster of such PCs provides a computational power of 10 Teraflops for a price of less than $100,000.

Chapter 4

Vapor-crystal nucleation close below the triple point

with A. Page and R. Sear

We present the results of Monte Carlo simulations of crystal nucleation from the vapor phase. We studied the Lennard-Jones system at conditions close to, but below, the triple point. This system is expected to show surface melting. The nucleation pathway that we observe consists of two distinct steps. In the first step, a liquid droplet nucleates from the vapor. Its nucleation rate depends strongly on the vapor supersaturation. In the second step, the final crystal phase nucleates in the liquid droplet, provided that this liquid droplet exceeds a minimum size. Our simulations show that within a liquid droplet the crystal nucleation rate does not depend on the vapor supersaturation.

In a recent independent study Chen et al. [J. Phys. Chem. B, **112***, 4069 (2008)] investigated the same phenomenon using umbrella sampling to compute free-energy barriers and hence nucleation rates. We use a different numerical approach where we focus on computing the nucleation rates directly using forward-flux sampling. Our results agree with the findings of Chen et al. and both methods observe two step nucleation. This finding indicates that this nucleation process can be described with a quasi-equilibrium theory. Due to different cutoffs for the interaction potential the results cannot be compared quantitatively.*

A first order phase transition starts with nucleation, if a free-energy barrier separates the parent phase from the phase that is thermodynamically more stable. Nucleation from this metastable initial state will take place when a rare fluctuation allows the system to surmount the free-energy barrier. Although the phase that forms during nucleation must be more stable than the parent phase, it is not necessarily the phase that is thermodynamically most stable [7]. A metastable phase may nucleate if the transition to that phase is kinetically favored. A phase transition to a thermodynamically stable phase may therefore proceed via a two step process. In the first nucleation event a metastable intermediate phase is formed; this then goes on to form the final phase via a second nucleation event.

There is experimental evidence that such a two-step nucleation scenario may be relevant to protein crystallization. For instance, lysozyme is sometimes found to separate into liquid-like aggregates before crystallization [8, 9]. This is at conditions at which lysozyme is only thermodynamically stable in the dilute fluid phase or in the crystal phase; the transition between a dilute and a concentrated liquid solution is located in a metastable region of the phase diagram where the solid phase is thermodynamically stable. Interestingly, crystal nucleation of these proteins appears to be near the metastable liquid-liquid coexistence curve [10, 94]. Simulations of enhanced crystal nucleation close to the metastable liquid-liquid critical point show that the system first forms dense, liquid-like clusters [10]. Crystallites then form inside these liquid clusters. Several theoretical models provide a rationale for the behavior observed in the simulations [11–14].

In the present paper we study the vapor of the Lennard-Jones system at temperatures below its triple point at chemical potentials where the crystal phase is the stable phase. Under these conditions, this system also acts as a simple model for "nano-colloids" in dilute solution. We find that, starting from a vapor, the phase transition does not occur directly to the crystal phase but via a liquid intermediate phase. The surface tension of the liquid-vapor interface is less than the surface free-energy density of the crystal-vapor interface, making the free-energy barrier to liquid droplet formation lower than to crystal formation. We expect that any phase transition, where an intermediate with a faster kinetic transition exists, will occur via a similar two-step process.

In a recent article, Chen et al. have used umbrella sampling to study two-step nucleation in the Lennard-Jones system [95], where they applied classical nucleation theory (CNT) to estimate nucleation rates and critical nucleus sizes. In the present work, we present an independent study of the same phenomenon using forward-flux sampling (FFS) [60, 61]. This technique primarily yields nucleation rates rather than free-energy barriers, and allows to estimate the critical nucleus size directly from the simulation data. In addition, FFS includes the effect of kinetics which can seriously alter the observed nucleation phenomenon [65]. Where a comparison can be made,

our results are consistent with those of Ref. [95] (see Section 4.4).

The remainder of the present paper is organized as follows. Section 4.1 deals with the simulation details and introduces the techniques that we have used. In Section 4.2 we present our results, and in Section 4.3 we explain our result in the framework of classical nucleation theory (CNT).

4.1 Simulation details

Throughout this work we report Metropolis Monte Carlo simulations on a model system consisting of particles interacting through a truncated and shifted Lennard-Jones (LJ) pair potential.

$$U(r) = \begin{cases} U_{LJ}(r) - U_{LJ}(r_c) & ; r \leq r_c \\ 0 & ; r > r_c \end{cases} \tag{4.1}$$

where the full (i.e. not truncated) LJ interaction is given by

$$U_{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{4.2}$$

Here, ε is the unit of energy, σ is the unit of length and r_c the interaction cutoff distance. Note that the choice of interaction cutoff has a significant effect on the free energy of the system and on properties such as coexistence lines and surface tensions [96].

In what follows, we use reduced units. We define the reduced distance as $r^* = r/\sigma$ and the reduced potential energy as $u^* = U\varepsilon^{-1}$. All other reduced quantities (e.g. the pressure $P^* = P\sigma^3\varepsilon^{-1}$, the density $\rho^* = \rho\sigma^3$, and the temperature $T^* = k_B T\varepsilon^{-1}$) follow. All quantities reported in this work are stated in these reduced units. Therefore, we omit the superscript star (*) from here on.

We performed all vapor-liquid simulations in the grand-canonical ensemble, where temperature, volume and the chemical potential are kept constant, and the particle number is allowed to fluctuate. This ensures a constant vapor pressure and minimizes finite size effects. The resulting nucleation rates are presented per unit volume of the vapor phase.

Our liquid-solid simulations from a liquid droplet embedded in the vapor were performed both in the canonical and grand-canonical ensemble. The initial liquid droplet was taken both from grand-canonical vapor-liquid simulations, as well as from melting a fcc crystal at high temperatures, quenching it to $T = 0.45$, and relaxing it for a long time. The results were unaffected by the preparation method.

Periodic boundary conditions were applied to all sides of the cubic simulation boxes. In the canonical ensemble simulations the box sizes were chosen to be $l^3 =$

20^3 for all but the largest droplet size, in which case the simulation was performed in a box of $l^3 = 30^3$. For all grand canonical simulations the simulation box size was $l^3 = 44^3$. The trial move size is chosen to be $\Delta x = 0.11$ in the canonical ensemble and $\Delta x = 0.2$ in the grand-canonical ensemble. In the grand canonical ensemble each particle was moved on average 20 times before a insertion/removal move was performed.

To capture nucleation events, we applied the forward-flux sampling (FFS) scheme, a rare-event technique developed by Allen et al. [60, 61]. FFS primarily yields nucleation rates. For the critical nucleus size an estimate is given by the order parameter value of the FFS interface at which the probability to reach the final state exceeds 0.5. The error is then given by the FFS interface spacing. All nucleation rates and estimates for the critical cluster size were obtained by averaging over five independent FFS runs with at least 50 successful paths per interface. As the unit of time τ in our FFS simulations, we use a Monte Carlo sweep, which corresponds to one trial displacement move per particle. For a given Monte Carlo step size, τ can be related to the time in a molecular dynamics simulation by comparing the self-diffusion coefficient computed both from Monte Carlo and Molecular Dynamics simulations. All nucleation rates are presented as the number of nuclei that form per unit volume and unit time, that is $[k] = [\tau^{-1}\sigma^{-3}]$. In addition, we also present the liquid-to-solid nucleation rates per droplet, $[kV_D] = [\tau^{-1}]$, where V_D is the droplet volume. In the grand-canonical simulation, in which the droplet keeps growing, the initial droplet volume is used. To properly compare the nucleation rates between different ensembles, care has to be taken that the rates are expressed in the same unit of time. Furthermore, in the grand-canonical ensemble simulation the rate of particle insertion and deletion affects the density profile in the vicinity of a droplet surface, effectively changing the diffusion coefficient. However, small changes in our particle swap rates did not noticeably affect the nucleation rates. Therefore we neglected this effect throughout this work.

Order parameters

For the vapor-to-liquid nucleation we used an order parameter based on Stillinger's overlapping sphere criterion [66]. It defines a particle to be in a high-density phase if it has at least one neighbor within a distance 1.5 corresponding roughly to the first minimum of the pair correlation function of a bulk liquid. Then a cluster analysis is performed on all high-density particles and the number of particles in the largest cluster is taken as the order parameter. Note that this order parameter does not distinguish between an ordered or disordered high-density phase, and therefore does not favor one phase (liquid or crystal) over the other. For a detailed description of this order parameter we refer the reader to Ref. [67, 97].

As order parameter for the liquid-to-solid nucleation we applied the local bond-order parameter [57, 98]. This order parameter assigns each particle a 13-dimensional vector capturing its local environment,

$$\Big(q_6(i)\Big)_m = \frac{1}{N_b(i)} \sum_j Y_{6m}(\hat{\mathbf{r}}_{ij}), \tag{4.3}$$

where Y_{6m} denotes a 6-th order spherical harmonic with components m ranging from $-6 \leq m \leq 6$. $N_b(i)$ is the number of nearest neighbors of particle i, and $\hat{\mathbf{r}}_{ij}$ a unit vector connecting the centers of mass of particles i and j. The sum is over all neighboring particles j within a cutoff distance $r = 1.5$. In a second step the order parameter computes the dot product $\mathbf{q}_6(i) \cdot \mathbf{q}_6(j)$ between each particle i and all its neighbors j, effectively comparing the particles' neighborhoods. If $\mathbf{q}_6(i) \cdot \mathbf{q}_6(j)$ exceeds a threshold of 0.65, particles i and j are considered to form a 'link'. Only if a particle's total number of links $n(i)$ exceeds 5 links is it considered to be solid-like. In a final step Stillinger's criterion is applied to all solid-like particles to identify the number of particles in the largest solid-like cluster, which is used as order parameter. All parameters for this order parameter can be obtained by minimizing the overlap between distributions of an equilibrated bulk solid and a metastable bulk liquid at working conditions. Note that slight modification of these values affected neither the nucleation rate nor the nucleation pathway of our simulations. But it does affect the size of the clusters identified, and therefore the critical cluster size, too.

This local bond-order parameter was designed to detect nucleation in the bulk. In order to verify that nucleation at the droplet surface is not neglected, we also performed simulations with the modified local bond-order parameter used by Mendez-Villuendas and Bowles [73]. This modification ensures that surface particles, too, can be identified as solid-like particles and are taken into account properly. However, compared to the unmodified order parameter no difference in the nucleation behavior was found.

Simulation conditions

In this study, the interaction cutoff distance was set to $r_c = 2.5$. As the choice of cutoff distance affects the free energies of the system, we computed the free energies by using thermodynamic integration for the liquid [92] and the Einstein crystal method for the fcc solid [99]. Below the triple point the vapor behaves effectively like an ideal gas and its free energy can be computed analytically. We located the triple point to be at $T^*_{TP} \approx 0.65$ and $P^*_{TP} \approx 0.00271$.

Throughout this work the temperature was fixed at $T = 0.45$ ($0.692\,T_{TP}$). The vapor pressure was varied between $P_v = 1 \times 10^{-4}$ and $P_v = 5 \times 10^{-4}$. The bulk densities, obtained from NPT simulations with $N = 2028$ particles, are $\rho_L = 0.905$ for the

P_v [10^{-4}]	1.0	2.0	3.0	4.0	5.0	6.0
S_s	4.39	8.77	13.16	17.54	21.93	26.28
S_l	2.33	4.67	7.00	9.34	11.67	14.01
$\Delta\mu_s$	-0.67	-0.98	-1.16	-1.29	-1.39	-1.47
$\Delta\mu_l$	-0.38	-0.69	-0.88	-1.01	-1.11	-1.19

Table 4.1: For both the liquid (subscript l) and the fcc solid (subscript s) phase, the vapor supersaturation S and the difference in chemical potential $\Delta\mu$ with respect to the vapor phase are presented as a function of the vapor pressure. The coexistence pressure for vapor-liquid is $P_{vl}^{coex} = 4.28 \times 10^{-5}$ and for vapor-solid it is $P_{vs}^{coex} = 2.28 \times 10^{-5}$. The difference in chemical potential between fcc solid and liquid is $\Delta\mu_{sl} = \mu_s - \mu_l = -0.29$.

$P_V[10^{-4}]$	$\ln k$	n^*
2.0	-117.3 ± 0.6	155 ± 35
3.0	-75.4 ± 1.2	72 ± 8
4.0	-56.6 ± 0.8	50 ± 4
5.0	-47.6 ± 0.9	38 ± 4
6.0	-39.4 ± 0.7	34 ± 4

Table 4.2: Vapor-liquid nucleation rates k and the critical cluster size n^* as a function of the vapor pressure P_V. All nucleation rates are obtained with FFS by using 50 paths for each interface and are averaged over 5 independent runs. The rates shown here are number of nuclei per unit volume and per Monte Carlo cycle.

liquid and $\rho_S = 0.989$ for the fcc solid. For both the liquid and solid phases, Table 4.1 lists the associated vapor supersaturation and the difference in chemical potential to the vapor phase. It is worth mentioning that at such low pressures the free energy of the high-density phases does not noticeably change with pressure. Therefore, once the coexistence pressure is known, the difference in chemical potential can be computed directly from the vapor supersaturation.

4.2 Results

The aim of this work is to investigate the pathway for vapor-to-solid nucleation close to but below the triple point. Due to the existence of a large nucleation barrier this phenomenon is not accessible in a direct brute-force simulation. Even with forward-flux sampling, a rare-event technique to overcome such barriers, we could not observe direct vapor-to-solid nucleation. Instead, we find that a solid can be formed in a two-step process, where first a liquid droplet is nucleated from the vapor, and in a second

nucleation event the crystal forms within the liquid droplet. We will discuss each nucleation step separately.

4.2.1 Droplet nucleation from the vapor

From the Lennard-Jones phase diagram it is known that, at our conditions (see Section 4.1), the crystal phase is the thermodynamically stable phase [100]. In addition, our chemical potential calculations (see Table 4.1) show that the liquid phase, also, has a lower free energy than the vapor phase. As a consequence, both a liquid and a solid can form. In order not to bias which high-density phase nucleates from the vapor, we use the local-density based order parameter discussed in Section 4.1. This order parameter does not distinguish between ordered and disordered phases. In a canonical simulation the vapor is depleted of particles as the liquid droplet grows. For this reason we perform our simulations in the grand-canonical ensemble, which mimics conditions of constant vapor pressure.

Our simulations show only the nucleation of a liquid droplet. Analysis of the liquid droplets with a local bond order parameter confirmed that the droplets did not contain any crystallites. The observed liquid clusters are on average compact spherical objects, which confirms previous results by ten Wolde et al. [97] and Wedekind et al. [67]. The critical cluster size is approximated to be the FFS interface nearest to where the probability to grow to the final phase exceeds $P = 0.5$. This will have a error equal to the gap between interfaces. A snapshot of a critical cluster is shown in Figure 4.2a. The compact spherical shape is clearly visible. The droplets continue growing, and for all clusters up to $N = 3000$ particles, we did not observe any spontaneous crystallization.

The nucleation rates for the liquid are presented in Table 4.2 and in Figure 4.1. The rate depends strongly on the vapor pressure. This is due to the fact that the free energy of the vapor is strongly affected by a slight change in pressure, whereas the free energy of the liquid is not.

This nucleation of a liquid phase, and the knowledge that the crystal phase has a lower free energy than the liquid phase, led us to investigate crystal nucleation within a liquid droplet.

4.2.2 Crystal nucleation within liquid droplets

The nucleation of the crystal phase inside a liquid droplet was simulated in the canonical ensemble. We performed simulations to investigate how the size of the liquid droplet affects the nucleation of the crystal phase. The nucleation rates were obtained with the forward-flux sampling [60] technique, using the local bond-order parameter [57, 98] discussed in Section 4.1. Simulations were performed on systems of size

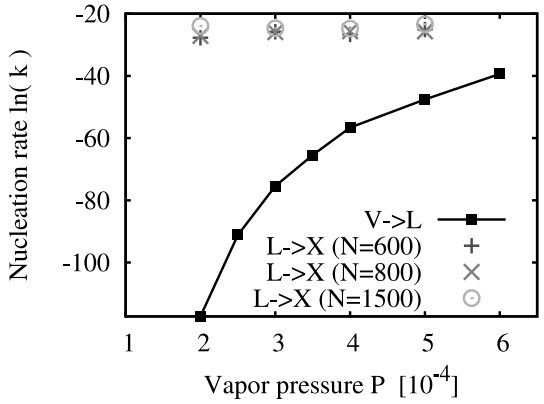


Figure 4.1: The logarithmic rate is plotted against the vapor pressure for the vapor-to-liquid (filled squares) and liquid-to-crystal (unfilled symbols) nucleation processes. The rates are expressed per unit volume of the vapor and liquid phase, respectively. Results for initial droplet sizes of 600 (triangles), 800 (circles) and 1500 (squares) are shown. The line through the filled symbols is a guide for the eye and errors are within the size of the symbols.

N = 448, 612, 700, 850 and 2028. The vast majority of particles formed part of the liquid droplet with, on average, less than 7 particles forming the coexisting vapor. The vapor density, and hence the vapor pressure does not noticeably change as the crystal nucleates and grows inside the liquid droplet.

For all but the smallest system size with $N = 448$, a stable or metastable crystallite was nucleated. Nucleation occurs in the core of the droplet and not at the surface like the freezing of gold nano-clusters [73]. In the Lennard-Jones system close below its triple point the crystal-vapor interface has a higher free-energy cost than the liquid-vapor interface and surface melting is expected [101]. This phenomenon can be seen in Figure 4.2c, a snapshot of a crystallized cluster. Notice that a disordered layer of approximately one particle thickness separates the crystal and vapor phases. When we continue the simulations after crystal nucleation is complete, we find that the particles in this disordered layer diffuse around the surface of the crystal cluster. Thus the mono-layer is liquid-like.

In small droplets of 448 particles we find that all crystallites that form are unstable and quickly melt. The absence of a stable or metastable crystallites suggests that the crystal phase for these small droplets is unstable due to the high free-energy cost of the liquid-crystal interface. For the droplets that crystallized, the natural logarithms of the nucleation rates are shown in Table 4.3. The nucleation rate per droplet, directly provided by our FFS simulations, is an extensive property; we therefore expressed

No. of particles N_D	$\ln(kV_D)$	$\ln(k)$	n^*
448	-	-	-
612	-23.7 ± 1.5	-30.2 ± 1.5	180
700	-22.1 ± 1.0	-28.8 ± 1.0	190
850	-20.7 ± 1.0	-27.6 ± 2.0	180
2028	-21.1 ± 1.6	-28.8 ± 1.6	190

Table 4.3: For a liquid droplet embedded in a vapor, this table shows the logarithmic rate for liquid-to-crystal nucleation and the critical cluster size as a function of the number of particles N_D in the liquid droplet. All simulations were performed in the canonical ensemble. The rates are expressed per droplet, $\ln(kV_D)$, and per unit volume, $\ln(k)$, where $V_D \approx N_D/\rho_L$ is the droplet volume. The unit of time is a Monte Carlo cycle. The critical cluster results have an error of ± 10.

the rates per unit volume of the liquid, too. The droplet volume is estimated by $V_D \approx N_D/\rho_L$, where N_D are the number of particles in the droplet, and ρ_L is the liquid density at coexistence.

Further analysis of our FFS data shows that the critical cluster occurs within a range of cluster sizes from 180 to 200 particles. This range is the same for all system sizes consisting of 700 to 2028 particles. A snapshot of a critical cluster is shown in Figure 4.2b.

4.2.3 Crystallization in the grand-canonical ensemble

To test out calculations of the crystal nucleation rates in the canonical ensemble, we also performed simulations in the grand-canonical ensemble. Apart from the ensemble both simulation methods were identical. As starting configurations we used post-critical liquid droplets with $N = 600$, $N = 800$, and $N = 1500$ particles.

Figure 4.1 shows logarithmic nucleation rates for the vapor-liquid and liquid-crystal against vapor pressure. Nucleation inside the liquid droplet is not affected significantly by the vapor pressure. The pressure inside the droplet is approximated to be

$$P_{droplet} = P_{vapor} + P_{Laplace} \tag{4.4}$$

where

$$P_{Laplace} = \frac{2\gamma_{lv}}{R}. \tag{4.5}$$

Using the virial pressure tensor we computed that $\gamma_{lv} = 1.07$. Taking the droplet in Figure 4.2b as an example, the radius is approximately $R \approx 5$, giving a Laplace pressure $P_{Laplace} \approx 0.4$. So although an increase in the vapor pressure does increase

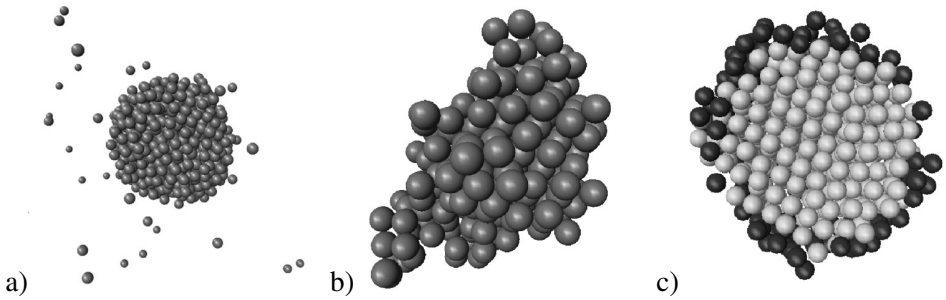


Figure 4.2: a) Snapshot of a liquid cluster from a grand-canonical simulation clearly showing its compact spherical shape. b) A critical crystal cluster from a canonical simulation containing 2028 particles. Only the crystal particles are shown in this snapshot. This crystal cluster contains 192 crystal particles. Snapshots of the critical cluster from simulations containing 700 and 850 particles are indistinguishable from the cluster shown here. c) Cross section of a crystal cluster in a system size of 850 particles. Light particles are crystalline, dark are liquid. The crystal is coated in a liquid-like layer. This layer is approximately 1 particle in thickness. In snapshots b and c the crystal particles were characterized by bond-order parameters as detailed in Section 4.1.

the pressure inside the droplet, we see this increase is small compared to the Laplace pressure, $10^{-4} \ll 0.4$. The liquid and crystal phases are dense hence any increase in pressure of this order (i.e. 10^{-4}) does not notably change the chemical potential. This is in contrast to the vapor phase where small changes in the vapor pressure cause large changes in the chemical potential and hence in the free-energy barrier.

In the grand-canonical ensemble, the liquid droplets keep growing during the FFS simulations. At a pressure of 3×10^{-4} the droplet initially containing 1500 particles grew to 6581 ± 3349 particles whereas the $N = 600$ droplet only grew to 1875 ± 323 particles. The larger increase in particle number of the $N = 1500$ droplet is due to the larger surface area than the smaller droplets. We expect this increase in the droplet size to be the cause of the systematically higher rates for the $N = 1500$ droplet. We point out that due to the droplet growth the droplet volume is not a well-defined quantity. For the nucleation rate per unit volume we therefore use the initial droplet volume for the normalization.

However, the observed nucleation rates (see Table 4.4) are comparable to the rates obtained from canonical simulations (see Table 4.3) and both methods yield the same value for the critical nucleus. Nucleation of the liquid phase from the vapor is the rate-limiting step. At a vapor pressure of $P_V \approx 10^{-3}$ the nucleation rates of the liquid and crystal phases are of the same order of magnitude. We expect that, as the

P_V [10^{-4}]	2.0	3.0	4.0	5.0
$\ln(k_{600}V_{600})$	-21.4 ± 2.0	-19.4 ± 0.9	-20.3 ± 1.5	-18.7 ± 1.4
$\ln(k_{800}V_{800})$	-20.5 ± 1.8	-18.3 ± 1.0	-19.0 ± 0.7	-19.0 ± 2.5
$\ln(k_{1500}V_{1500})$	-16.6 ± 1.5	-17.4 ± 2.3	-17.5 ± 2.1	-16.2 ± 2.5
$\ln(k_{600})$	-27.9 ± 2.0	-25.9 ± 0.9	-26.8 ± 1.5	-25.2 ± 1.4
$\ln(k_{800})$	-27.3 ± 1.8	-25.1 ± 1.0	-25.8 ± 0.7	-25.8 ± 2.5
$\ln(k_{1500})$	-24.0 ± 1.5	-24.8 ± 2.3	-24.9 ± 2.1	-23.6 ± 2.5

P_V [10^{-4}]	2.0	3.0	4.0	5.0
$n^*_{(600)}$	180	180	180	190
$n^*_{(800)}$	180	180	170	170
$n^*_{(1500)}$	170	160	180	170

Table 4.4: For a liquid droplet embedded in a vapor the logarithmic rate for liquid-to-crystal nucleation and the critical cluster size is shown as a function of vapor pressure. Results for different initial droplet sizes $N_D = 600, 800, 1500$ are presented both as rates per droplet, $\ln(k_N V_N)$, and as rates per unit volume, $\ln(k_N)$, where $V_N \approx N_D/\rho_L$ is the droplet volume. All simulations were performed in the grand canonical ensemble, therefore the particle numbers N_D and the associated droplet volume V_D are those at the start of the simulation. The error for the critical cluster size is ± 20.

pressure increases, the nucleation of the crystal phase inside the liquid droplet will become the rate limiting step.

Note that for a proper comparison of canonical and grand-canonical nucleation rates both need to be expressed in the same unit of time τ (see Section 4.1 for details). Therefore, the comparison remains qualitative.

4.2.4 Direct crystal nucleation from the vapor

Between the vapor-solid and metastable vapor-liquid coexistence curves ($P_{vl} = 2.28 \times 10^{-5} < P < P_{vs} = 4.28 \times 10^{-5}$), the liquid phase has a higher free energy than the vapor phase. Here we would expect direct nucleation to the crystal phase instead of the two step nucleation detailed in this paper. At these conditions we could not get our system to crystallize at all. No direct vapor-crystal nucleation was observed in any of our simulations. This illustrates the dramatic effect of the intermediate metastable liquid on the crystal nucleation rate.

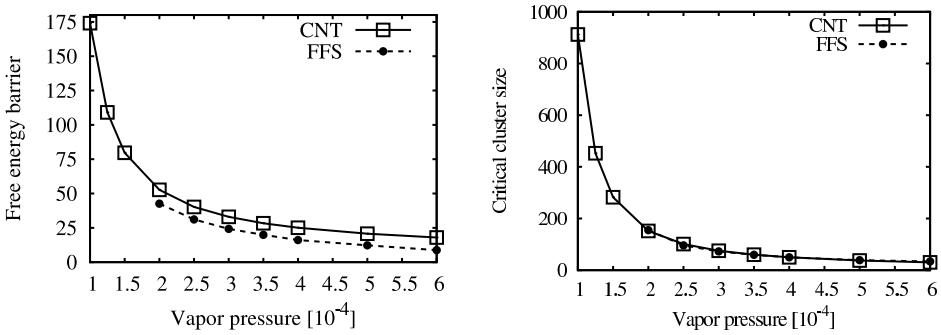


Figure 4.3: Comparison of CNT predictions versus the simulation results for the vapor-to-liquid nucleation. Both the nucleation barrier height ΔG^* (left) and the critical cluster size n^* (right) are plotted as a function of vapor pressure. The constant offset in barrier height is attributed to the accumulated error in the CNT description for very small clusters [26].

4.3 Classical Nucleation Theory

Since the direct nucleation of the crystal phase from the vapor can not be seen in simulation we decided to estimate the nucleation rate via CNT [102]. This theory assumes that the crystal nuclei are perfectly spherical and incompressible. The resulting free-energy barrier is then

$$\Delta G^* = \frac{16\pi}{3} \frac{\gamma_{vs}^3}{\rho_s^2 |\Delta\mu_s|^2}, \tag{4.6}$$

where $\Delta\mu_s$ is the difference in chemical potential between the vapor and the solid from Table 4.1. The vapor-crystal surface tension γ_{vs} is unknown and must be estimated. As lower bound we use the planar vapor-liquid surface tension γ_{vl}, which can be computed from the virial pressure tensor obtained from simulation [92]. We find $\gamma_{vl} = 1.07$. Note that the surface tension depends strongly on the interaction cutoff distance, and for $r_c = 2.5$ it significantly deviates from the untruncated Lennard-Jones system. Using $\gamma_{vl} = 1.07$, $\Delta\mu_s = -0.29$ and $\rho_s = 0.989$ we obtain $\Delta G^* = 250$ for a lower bound for the free energy barrier. The resulting nucleation rate is essentially zero, which explains why not direct crystal nucleation was observed.

Using the same value for the surface tension and the data from Table 4.1 the simulation data can be compared to CNT. In literature it is common to compare nucleation free-energy barriers rather than nucleation rates. As FFS yields nucleation rates directly, we approximate the barrier height using the CNT rate expression to obtain

$$\Delta G^*_{FFS} = -k_B T \ln\left(k_{FFS}/J_0\right), \tag{4.7}$$

where J_0 is the kinetic prefactor from the CNT rate expression. Note that, in contrast to the work of Chen et al. , errors in the CNT kinetic prefactor will influence the estimate of the barrier height computed by FFS. But since these errors have logarithmic dependence we considered them negligible. The kinetic prefactor J_0 can be approximated by [103],

$$J_0 \approx Z\rho_v^2 D R^*, \tag{4.8}$$

where D is the self-diffusion coefficient and

$$Z = \sqrt{\frac{|\Delta\mu|}{6\pi T N^*}} \tag{4.9}$$

is the Zeldovich factor [104]. Note that the above expression for J_0 assumes that the dynamics of droplet nucleation and growth is diffusive. Such a description is compatible with the use of Monte Carlo simulations to model the dynamics of a diffusive nano-colloidal system. For a dilute vapor the self-diffusion coefficient can be estimated from $D = \Delta x^2/\tau = 0.04$, the MC step size squared per MC cycle. The CNT critical droplet radius R^* is given by

$$R^* = \frac{2\gamma_{vl}}{\rho_l |\Delta\mu_l|}. \tag{4.10}$$

For the conditions that apply to the present simulations, we find that J_0 varies between $J_0 \approx 2.2 \times 10^{-11}$ and 1.6×10^{-9} for pressures ranging between $P_v = 1.0 \times 10^{-4}$ and 5.0×10^{-4}. The results are shown in Figure 4.3. The functional form of the CNT barrier is in good agreement with our simulation results, but a constant offset is observed. This is consistent with earlier work on vapor-liquid nucleation [26, 95, 97, 105]. But even though many offset values are reported in literature, the applied interaction cutoffs, temperatures and supersaturations vary strongly making a direct comparison difficult. Since the precise composition of the barrier height offset to CNT is still debated [26, 105–108] we only compare the total offset $D(T)$. From our simulations we obtain $D(T) = 7.5$, which is in good agreement with the results from ten Wolde and Frenkel, $D(T) = 5.3$, that were obtained with the same truncation and shift, but at a slightly higher temperature of $T = 0.741$. These values are lower than those reported by McGraw et al. ($D(T = 0.8) \approx 19 - 20$), Merikanto et al. [1] ($D(T = 0.7) \approx 18 - 19$), and Chen et al. [2] ($D(T = 0.6) \approx 22$ to $D(T = 0.45) \approx 27$), who used the full Lennard-Jones potential. This deviation is not unexpected and can be attributed to the effect of the interaction potential's long-range tail on the surface tension γ and, even more, on its curvature correction $B(T)$.

[1] Only $B(T)$ and γ were provided in Ref. [26]. For the estimation of $D(T) = B(T) + A_1\gamma - |\Delta\mu|$ we assumed $\rho_L \approx 0.84$ and $\Delta\mu = -1.0$.

[2] Data extracted from the graphical representation in Ref. [95].

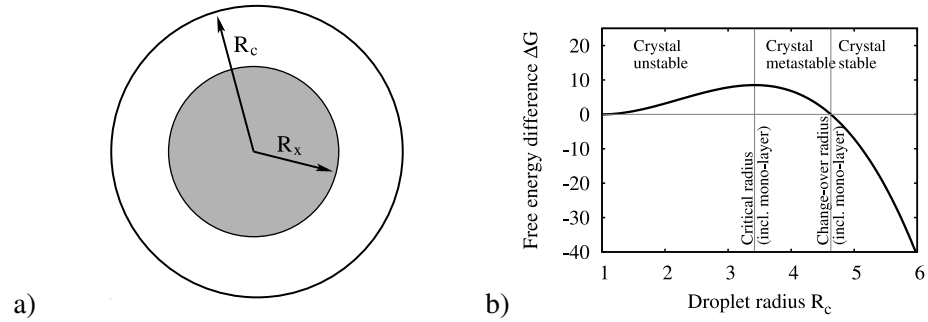


Figure 4.4: a) Diagram of the assumed geometry of a crystal cluster growing in a liquid droplet. The size of the crystal cluster, R_x, is always at least one less than the size of the liquid droplet it grows within, R_c, to account for the liquid-like mono-layer. b) Free-energy difference between a purely liquid droplet and the crystal state, a droplet containing a crystal of radius $R_x = R_c - 1$, the largest crystal the droplet can support. Depending on the droplet radius R_c, the crystal state of the droplet will either be unstable, metastable or stable with respect to the purely liquid droplet. Note the droplet has a fixed radius and we do not refer to the *bulk* phase stability.

In order to use CNT to analyze the nucleation of crystallites inside liquid droplets, we use a simplified model: first, we assume that the crystallite and liquid droplet are prefect spheres. Second, the crystal is assumed to grow from the center of the liquid droplet (i.e. we assume bulk rather than surface nucleation). Furthermore, as suggested by our simulations, we assume that the droplet is always covered with at least a mono-layer of liquid. Finally, we postulate that the surface free energies are independent of each other no matter how close the crystal-liquid and liquid-vapor interfaces are. A diagram of this crude but effective model is shown in Figure 4.4a. Using this model, the free-energy barrier for crystal nucleation can be approximated by Eq. (4.6) where $\Delta\mu_s$ is exchanged with $\Delta\mu_{lx}$, the difference between the bulk liquid and crystal chemical potentials (from Table 4.1), and γ_{vs} is exchanged with γ_{lx}, the liquid-crystal surface tension. As an estimate we take $\gamma_{lx} = 0.347$. This value is the planar surface tension for the (111) crystal plane calculated by Davidchack and Laird [109] at the triple point. They used a truncated and force shifted Lennard-Jones potential at the same cutoff as our simulations but as we are working under the triple point we expect the actual value to be lower than this. This simple model gives an estimated free-energy barrier of $\Delta G^*_{lx} = -8.5 (\Delta G^*_{lx}/kT = -18.9)$.

Further, assuming that the surface tension does not depend on the radius of curvature of the droplet (a rather drastic assumption [25]), all crystallites have the same free-energy curve. Note that, unlike for bulk systems, this free-energy curve is ter-

minated at $R_x = R_c - 1$, the droplet radius minus the liquid mono-layer. The droplet therefore has two distinct states: the liquid state referring to a purely liquid droplet, and the crystalline state referring to the droplet containing a crystal of maximum size ($R_x^{(max)} = R_c - 1$).

The free-energy difference between these two states is shown in Figure 4.4b as a function of the droplet radius, R_c. We now introduce the change-over radius R_x^{co}, which defines the point where both states, the liquid state and the crystalline state, have the same free energy. Note that this is not exactly the radius where there is an equal probability to find droplet on the liquid or crystalline side of the nucleation barrier as the ratio of these probabilities also depends on the ratio of the "volume" in the free-energy landscape on either side of the barrier. However, if these (logarithmic) phase-space volume corrections are small, R_x^{co} provides a good approximation of the point where crystal and liquid are equally likely.

We can now identify three different classes for the droplet radius where the crystalline state is either a) unstable, b) metastable or c) stable with respect to the droplet in the liquid state (see Figure 4.4b).

a) Unstable: $R_c \leq R_x^* + 1$, where R^* is the critical radius of a crystal nucleus in a bulk liquid. If the droplet is smaller than the crystal critical radius any crystal clusters formed are unstable and will only be observed as rare fluctuations.

b) Metastable: $R_x^* + 1 < R_c \leq R_x^{co} + 1$. Above the critical radius there is a free-energy barrier associated with the crystalline state returning to the liquid state. The droplet has a small probability of being in the crystalline state but, as remelting is an activated process, the lifetime of these metastable crystalline states may be long. Since the crystalline state has a higher free-energy the droplet spends the majority of its time in the liquid state. This "metastable" regime ends for $R_c \geq R_x^{co} + 1$ where the crystalline state becomes more stable than the liquid).

c) Stable: $R_c > R_x^{co} + 1$. The change-over radius determines the minimum size for the crystalline state to be more stable than the liquid. In this class the liquid state is now metastable.

The crystallites inside the droplets follow the same free-energy curve and therefore the free-energy barrier is independent of the size of the droplet. We observe this feature in our simulation results where the liquid-crystal nucleation rates are to a good approximation independent of droplet size. A similar scenario is found in the simulation of explosive melting of polymer crystallites [110].

CNT predicts the critical radius and the change-over radius to be $R_x^*[CNT] = 2.4$ and $R_x^{co}[CNT] = 3.6$, respectively. A comparison with the simulation results remains qualitative, as the characterization of the crystal phase, and therefore the critical cluster size, depends on the definition of order parameter used for analysis (see Section 4.1). Analyzing our FFS data yields a critical cluster size of $n_{sim}^* \approx 190$ particles which corresponds to a radius of $R_x^*[SIM] \approx 3.6$ (without liquid mono-layer).

This critical radius obtained from FFS is close to the droplet size $R_c = 4.8$ (see Figure 4.4a) found for a $N = 448$ system size. At this proximity to the critical radius it is not surprising we find the droplet to be stable in the liquid state. For droplets of size $N \approx 600$, the crystalline state is expected to be metastable, and for $N \approx 700$ both states should be observable. Droplets exceeding $N \approx 800$ are expected to be stable in the crystal phase.

4.4 Discussion

This chapter details the nucleation pathway from a metastable vapor just below the triple-point temperature to the crystal phase. This process is relevant for the fabrication of nano-crystals, and it can also act as a model for crystallization of "nano-colloids" from a dilute solution.

At these conditions, we find that the vapor-to-crystal nucleation occurs in two stages. First, an intermediate liquid phase is nucleated rather than the thermodynamically most stable crystal phase. The final crystal phase then nucleates in the liquid phase in a second independent step. This two-stage process suggests that both the vapor-liquid and liquid-crystal free-energy barriers are lower than that of a direct nucleation of the crystal from the vapor [7, 111]. This is supported by both an analysis of surface free energies and a comparison with CNT. Direct nucleation of the crystal is not observed and a CNT estimate of this nucleation rate is essentially zero.

A recent publication by Chen et al. came to conclusions similar to ours, on the basis of a different numerical approach [95]. Chen et al. obtained free-energy curves from aggregation-volume-bias Monte Carlo simulations with umbrella sampling. As the reaction coordinate for the crystal phase they used the global bond order parameter, which detects the overall crystallinity of the liquid cluster [56]. Both nucleation rates and critical cluster size were estimated using CNT. In contrast, we focus on computing nucleation rates directly rather than free-energy barriers by using the forward-flux sampling technique, which can be used both for equilibrium and non-equilibrium processes [61]. As order parameter the size of the largest crystal cluster in the liquid droplet was used. The critical nuclei were then obtained directly from analysis of our FFS data.

To compare with the work of Ref. [95] we must translate nucleation rates in free-energy barriers. The fact that we obtain reasonable agreement with Ref. [95] suggests that CNT works rather well for this system, and that this process can be described with a quasi-equilibrium theory. In addition, note that both umbrella sampling and forward-flux sampling essentially yields the same results for the Lennard-Jones system. As reported by Sanz et al. this need not to be true for every system [65].

A quantitative comparison between the present simulations and those of Ref. [95]

is unfortunately not possible as the studies used different cutoffs for the Lennard-Jones potential. The change in cutoff affects the chemical potentials for both the liquid and crystal phases, and it also changes the surface tensions [96]. As both properties have a large impact on the resulting nucleation free energy barrier, the comparison remains qualitative.

To conclude, this work details the process of crystal nucleation from the vapor phase. The two-stage nucleation found is potentially a very common phenomenon that could occur in fields as diverse as protein crystallization and ice formation.

Acknowledgements

The authors thank P. Charbonneau for a careful reading of the manuscript.

Chapter 5

Liquid wetting can enhance vapor-crystal nucleation in a pore

In this chapter we present a numerical study of the effect of a hemispherical pore on the nucleation of Lennard-Jones crystals from the vapor phase. At conditions below the triple point, bulk crystal nucleation in this system is known to proceed via a two-step process (vapor → liquid → crystal) [see Chapter 4].

The presence of a pore strongly affects the nucleation scenario. For a narrow range of pore radii, we find that vapor-liquid nucleation itself becomes a two-step process (1. pore filling, 2. break-out into the bulk). With increasing pore radius the rate for pore-filling decreases while the rate for growing out of the pore increases. As a consequence, the overall nucleation rate shows a non-monotonic behavior. This observation confirms predictions by Page and Sear [18] for pore nucleation based on a numerical study of pore nucleation in an Ising lattice model.

As is to be expected, both the maximum nucleation rate and the associated optimal pore size depend on the adsorption strength of the pore. We find that the maximum nucleation rate exceeds the rate for nucleation at a planar wall with the same attraction.

Finally, we study the rate of crystal nucleation inside the liquid that has nucleated inside pores. We find that the pore-induced vapor-liquid nucleation is the rate-limiting step for crystal nucleation. This implies that crystal nucleation can be enhanced by a judicious choice of the wetting properties of a micro-porous nucleating agent.

Freezing in three dimensional systems is a first order phase transition. As a consequence, it is possible to supercool or super-compress the liquid parent phase until crystallization happens spontaneously on the timescale of the experiment. Often the crystallization process starts with a nucleation step during which a crystalline nucleus is formed that can then grow to macroscopic size, and the overall rate of crystallization is determined frequently by the nucleation step. Moreover, the structure of the crystal that grows to macroscopic size is often determined by the structure of the crystal nucleus that forms during the rate limiting step. For this reason it is of both fundamental and practical interest to gain insight into the mechanisms that can be used to control the rate and the pathway of crystal nucleation.

Nucleation can proceed either homogeneously, that is in the bulk of the parent phase, far away from walls or other external influences, or heterogeneously at the surface of a "seed". In most cases of practical interest, heterogeneous nucleation dominates the rate of crystal formation [16].

For this reason, heterogeneous crystal nucleation has been studied extensively ever since the early work of Ostwald [7]. During the past decade, there has been something of a revival of studies of heterogeneous nucleation both from a theoretical/numerical perspective and from experiments. This renewed interest was stimulated on the one hand by the availability of suitable colloidal model systems that make it possible to follow the dynamics of individual particles during nucleation in real space and real time [112, 113] and, on the other, by the development of novel computational techniques that made it possible to study the early stages of crystal nucleation in great detail [56, 57].

Heterogeneous nucleation of liquids and crystals has been studied both theoretically and in simulation for a variety of geometries such as nucleation on planar walls [27, 114] and in slit pores [115–117], on cylinders [118–120], patches [121, 122], patterned templates [123, 124], spheres [125, 126] and in pores [18, 103]. Experimental studies include colloidal crystallization on patterned templates [113, 127, 128] and spherical impurities [129, 130]. However, in many cases a detailed description of the nucleation pathway or a quantitative analysis of the nucleation rates is still lacking. For a detailed overview we refer the reader to the reviews by Gelb et al. [131] and Sear [103].

Experiments by Chayen and co-workers [132, 133] on the crystallization behavior of a number of different proteins demonstrated that the rate of protein crystallization is strongly enhanced in the presence of a micro-porous medium. This intriguing observation stimulated subsequent theoretical work by Page and Sear [18] who studied the nucleation behavior of a (two-dimensional) Ising model in a rectangular indentation in a wall. This lattice model did indeed show that pores can enhance nucleation and that this effect depends strongly on the pore radius. However, a lattice model such as the one used in Ref. [18] cannot capture one of the most

interesting features of protein crystallization, namely that it is often most effective in a temperature-concentration range where the protein solution is close to liquid-liquid demixing [10, 134, 134]: the Ising model does not distinguish between the liquid and the crystalline state. It is therefore of interest to study pore nucleation in an off-lattice model for which the liquid and the crystalline states are distinct. With such a model study, we can look in more detail at the ways in which a porous medium affects the pathway for crystallization. In particular, we can address questions such as: Does a crystal form immediately in the pore or does crystallization proceed via an intermediate liquid phase? Another advantage of an off-lattice model is the fact that the diameter of the pore need not be commensurable with the lattice spacing of the crystal (in a normal lattice model, this commensurability is inevitable). If a crystal does not "fit" optimally in a pore, one might expect that this affects the free-energy barrier for crystal nucleation. As we will show below, this is indeed the case.

In the present chapter we first investigate the effect of a spherical pore on the nucleation pathway for vapor-liquid nucleation. In Section 5.2.1 we explore nucleation at a planar (weakly) attractive wall. Next, we consider nucleation on a planar circular attractive patch, and finally, we consider a pore. Section 5.2.2 focuses on liquid-crystal nucleation. We find that, as in the case of homogeneous nucleation of Chapter 4, this is an independent nucleation event that can be treated separately. Finally, Section 5.3 integrates the resulting picture for the overall nucleation pathway in systems with surface heterogeneities.

5.1 Simulation details

The present study is based on Metropolis Monte Carlo simulations of a model system consisting of particles interacting through a truncated and shifted Lennard-Jones (LJ) pair potential [92].

$$U(r) = \begin{cases} U_{LJ}(r) - U_{LJ}(r_c) & ; \quad r \leq r_c \\ 0 & ; \quad r > r_c \end{cases}$$

where the full (i.e. not truncated) LJ interaction is given by

$$U_{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right].$$

Here, ε is the unit of energy, σ is the unit of length and r_c the interaction cutoff distance. Note that the choice of interaction cutoff has a significant effect on the free energy of the system and on properties such as coexistence lines and surface tensions [62, 96]. However, the qualitative features of the phase diagram remain unaffected. Hence, the model that we use is adequate for the present study that focuses on generic effects, rather than on the properties of a specific experimental system.

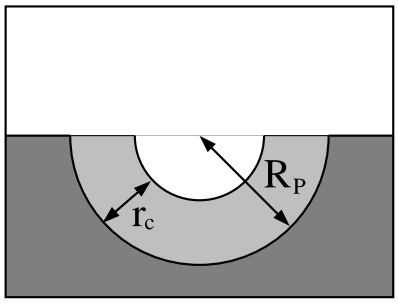


Figure 5.1: Schematic side-view of a pore. The dark-gray area indicates the the space that is inaccessible to LJ particles due to the presence of a hard wall. The range of the weak attraction is depicted in light-gray. Molecules in the white area are outside the range of interaction of the wall. The radius of the pore is denoted by R_p and the range of attraction of the pore walls is denoted by r_c.

Three different geometries were used to study heterogeneous nucleation: a planar wall, a planar circular patch, and a spherical pore. Figure 5.1 depicts a schematic side view of such a pore, which consists of an open half-sphere embedded in a hard wall. It shows the relevant parameters: the pore radius R and the pore's interaction range, r_c. All surface geometries are weakly adsorbing due to an effective potential given by the Lennard-Jones 9-3 potential,

$$U_w(r) = \left\{ \begin{array}{ll} \varepsilon_w \left[(\sigma/r)^9 - (\sigma/r)^3 \right] + U_w(r_c) & ; r < r_c \\ 0 & ; r \geq r_c \end{array} \right\},$$

which results from integration of the pair-interactions over the half-space assuming a constant particle density. Here, r is defined as the distance perpendicular to the surface. The potential $U_w(r)$ was cut and shifted at a distance r_c, and we choose r_c and σ to be the same for both $U_w(r)$ and $U(r)$.

In what follows, we use reduced units. We define the reduced distance as $r^* = r/\sigma$ and the reduced potential energy as $u^* = U/\varepsilon$. All other reduced quantities (e.g. the pressure $P^* = P\sigma^3\varepsilon^{-1}$, the density $\rho^* = \rho\sigma^3$, and the temperature $T^* = k_B T\varepsilon^{-1}$) follow. All quantities reported in this work are expressed in these reduced units. Therefore, we omit the superscript asterisk (*) from here on.

We performed all vapor-liquid simulations in the grand-canonical ensemble, where temperature, volume and the chemical potential are kept constant, and the particle number is allowed to fluctuate. This ensures a constant vapor pressure and minimizes finite size effects. In the simulations of liquid-solid nucleation, we consider a liquid droplet in contact with both the vapor phase and the surface. These simulations were performed in the canonical ensemble to keep the size of the liquid droplet constant.

The initial liquid droplet was taken from grand-canonical vapor-liquid simulations, then equilibrated at a high temperature to remove all crystallinity, whilst constraining the system (using umbrella sampling) to prevent evaporation. The droplet was then quenched to $T = 0.45$ and equilibrated using umbrella sampling to keep the droplet in the meta-stable liquid phase. This procedure allowed us to produce meta-stable liquid droplets of the required size.

Periodic boundary conditions were applied in the y- and z-directions of the cubic simulation boxes, and a planar hard wall prevented particles from escaping in the x-direction. For all simulations we chose the box size to be $l^3 = 50^3$ and the particle displacement trial move size to be $\Delta x = 0.2$. In the grand canonical ensemble each particle was moved on average 20 times before an insertion/removal move was performed.

Rare event sampling

In order to study the properties of the system on the path towards nucleation, we applied an umbrella sampling [52] (US) scheme, where a biasing potential is used to constrain the size of the nucleus to be close to a certain size n_0. By computing the bias-corrected cluster size probability distribution $P_0(n)$ for many different n_0, a series of overlapping windows is obtained. By fitting these windows together, the free energy as function of cluster size can then computed from the overall probability histogram, $G(n) = -k_B T \ln P(n) + c$, where c is a constant offset. To be more precise, we define $P(n)$ as the ratio of $M(n)$, the number of clusters of size n, divided by the (average) total number of particles N. For large cluster sizes, $P(n) \ll 1$ and the probability of observing more than one cluster of this size (or of a larger size) in a simulation is negligible. In that case the probability to find a cluster of size n in the system can be approximated by the probability to find that the *largest* cluster in the system contains n particles. This facilitates the umbrella sampling because we can then apply a bias that controls only the size of the largest cluster in the system [57]. However, when computing $P(n)$ for small clusters, we cannot use this approach as there will be many clusters with sizes comparable to, or larger than n. In that case, we sample $P(n)$ from an unbiased run taking *all* cluster sizes into account. This is important because if we would constrain the size of the largest cluster for small n (thereby cutting off the cluster distribution for all sizes larger than n), we would be truncating the natural cluster-size distribution in the system. As a consequence, the free energy of the system would then appear to increase as $n \to 1$ and as a result there would be a spurious (and system-size dependent) minimum in the free energy $G(n)$ at small n.

For the computation of the vapor-liquid nucleation free-energy barriers we started from the pure vapor phase. For the liquid-crystal nucleation, we first equilibrated a

P_v	S_s	S_l	$\Delta\mu_s$	$\Delta\mu_l$
1.0×10^{-4}	4.39	2.33	-0.67	-0.38
3.0×10^{-4}	13.16	7.00	-1.16	-0.88

Table 5.1: For the two vapor pressures used in this work the vapor supersaturation $S = P/P_{coex}$ and the difference in chemical potential $\Delta\mu$ with respect to the vapor phase are presented for both the liquid (subscript l) and the fcc solid (subscript s) phases. The vapor-liquid coexistence pressure is $P_{vl}^{coex} = 4.28\times10^{-5}$ and the vapor-solid coexistence pressure is $P_{vs}^{coex} = 2.28\times10^{-5}$. The difference in chemical potential between fcc solid and liquid is $\Delta\mu_{sl} = \mu_s - \mu_l = -0.29$.

liquid droplet of $n_l \approx 700$ particles, biasing the system to prevent any crystallization. Then we performed US simulations in the NVT ensemble to compute the nucleation free-energy barrier.

Order parameters

For the vapor-liquid nucleation we used an order parameter based on Stillinger's overlapping-spheres criterion [66]. It defines a particle to be in a high-density phase if it has at least one neighbor within a distance 1.5 corresponding roughly to the first minimum of the pair correlation function of a bulk liquid. Then a cluster analysis is performed on all high-density particles and the number of particles in the largest cluster is taken as the order parameter. Note that this order parameter does not distinguish between an ordered or disordered high-density phase, and therefore does not favor one phase (liquid or crystal) over the other. For a detailed description of this order parameter we refer the reader to Refs. [67, 97].

As order parameter for the liquid-solid nucleation we applied the local bond-order parameter [57, 98]. This order parameter assigns each particle a 13-dimensional vector capturing its local environment,

$$\Big(q_6(i)\Big)_m = \frac{1}{N_b(i)}\sum_j Y_{6m}(\hat{\mathbf{r}}_{ij}), \tag{5.1}$$

where Y_{6m} denotes a 6-th order spherical harmonic with components m ranging from $-6 \le m \le 6$. $N_b(i)$ is the number of nearest neighbors of particle i, and $\hat{\mathbf{r}}_{ij}$ a unit vector connecting the centers of mass of particles i and j. The sum is over all neighboring particles j within a cutoff distance $r = 1.5$. In a second step the order parameter computes the dot product $\mathbf{q}_6(i)\cdot\mathbf{q}_6(j)$ between each particle i and all its neighbors j, effectively comparing the particles' neighborhoods. If $\mathbf{q}_6(i)\cdot\mathbf{q}_6(j)$ exceeds a threshold of 0.65, particles i and j are considered to form a 'link'. Only if a particle's total number of links $n(i)$ exceeds 5 links is it considered to be solid-like. In a final

step Stillinger's criterion is applied to all solid-like particles to identify the number of particles in the largest solid-like cluster, which is used as order parameter. All parameters for this order parameter can be obtained by minimizing the overlap between distributions of an equilibrated bulk solid and a metastable bulk liquid at working conditions. Note that neither the nucleation rate nor the nucleation pathway of our simulations were sensitive to (moderate) variations in the definition of the order parameters. However, such changes do affect the apparent size of the clusters that we detect, including the critical cluster size.

The local bond-order parameter that we used was originally designed to detect nucleation in the bulk. In order to verify that nucleation at the droplet surface is still properly detected, we also performed simulations with the modified local bond-order parameter used by Mendez-Villuendas and Bowles [73]. This modification ensures that surface particles, too, can be identified as solid-like particles and are taken into account properly. However, no significant differences were observed.

For all simulations involving a heterogeneity, the order parameter is adapted to identify only clusters in contact with the heterogeneity. This means that in addition to the procedure mentioned above, only those clusters are considered that contain at least one particle that is within the effective interaction range of the surface heterogeneity. We refer to this modification as a *restricted order parameter* (rOP). rOP's were applied to detect heterogeneous vapor-liquid and liquid-crystal nucleation.

Simulation conditions

As in Chapter 4, we truncated and shifted the Lennard-Jones potential at $r_c = 2.5$. Chapter 4 also reports the coexistence data and vapor-liquid surface tension for this model system in the relevant temperature range. Here, we fix temperature at $T = 0.45$, in which case the vapor-liquid and vapor-crystal coexistence pressures are $P_{vl}^{coex} = 4.28 \times 10^{-5}$ and $P_{vx}^{coex} = 2.28 \times 10^{-5}$, respectively, and the vapor-liquid surface tension is $\gamma_{vl} = 1.07$.

We present results for two vapor pressures, $P_v = 1 \times 10^{-4}$ and $P_v = 3 \times 10^{-4}$, with the corresponding bulk densities $\rho_L = 0.905$ for the liquid and $\rho_S = 0.989$ for the fcc solid [135]. For both the liquid and solid phases, Table 5.1 lists the associated vapor supersaturation and the difference in chemical potential to the vapor phase. At such low pressures the free energy of the high-density phases does not change noticeably with pressure. Therefore, once the coexistence pressure is known, the difference in chemical potential can be computed directly from the vapor supersaturation.

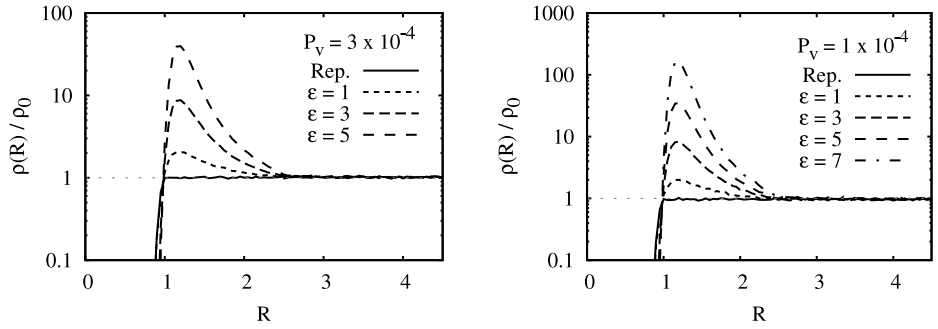


Figure 5.2: Vapor density perpendicular to a weakly attractive planar wall for various wall-particle interaction strengths ε at a vapor pressure $P_v = 3 \times 10^{-4}$ (left) and $P_v = 1 \times 10^{-4}$ (right). For reference a purely repulsive wall (Rep.) is shown. The repulsive potential has the same functional form as the attractive potential, but is truncated and shifted at the potential minimum to remove the attractive part.

5.2 Results

As was shown in Chapter 4 and Ref. [95], the bulk vapor-crystal nucleation for the present model system proceeds via two independent nucleation events. First, a metastable liquid droplet nucleates directly from the vapor phase. Subsequently, a crystal nucleus emerges within the liquid droplet. The two events happen on very different timescales and are therefore effectively decoupled except that the droplet needs to exceed a minimum size in order to sustain a stable crystal nucleus. As we show below, the pathway for heterogeneous nucleation follows essentially the same route as its homogeneous analog, which means that the nucleation of the liquid and of the crystal are decoupled and can be treated independently.

5.2.1 Vapor-liquid nucleation

As a first step, we compute the vapor density as function of the distance perpendicular to a planar wall for various wall-particle interaction strengths ε. This computation is performed with a supersaturated vapor under the same conditions used in the subsequent nucleation studies. From this preliminary calculation we can estimate the range of adsorption strengths for which heterogeneous liquid-vapor nucleation (and, *a fortiori*, liquid-crystal nucleation) is negligible on the timescale of our simulations. The results are presented in Figure 5.2 for a vapor pressure $P_v = 3 \times 10^{-4}$ (left) and $P_v = 1 \times 10^{-4}$ (right). The vapor density is rescaled by the bulk vapor density $\rho_0(P)$. As is to be expected for a dilute vapor, we find the vapor density to scale with the wall

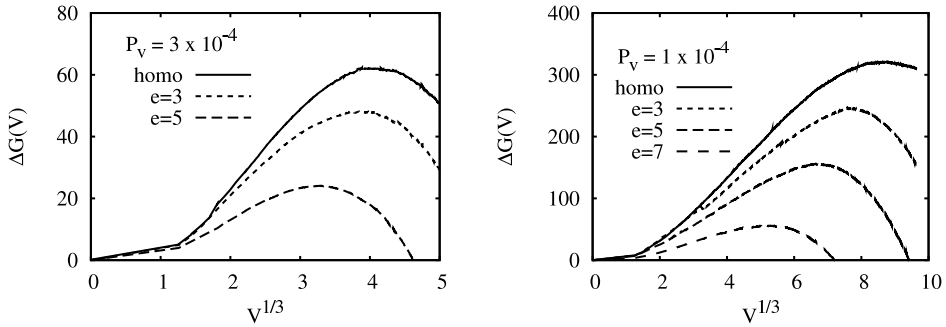


Figure 5.3: Vapor-liquid free-energy barriers as function of effective droplet radius $V^{1/3}$ for nucleation on a weakly attractive planar wall with varying wall-strength ε at a vapor pressure $P_v = 3 \times 10^{-4}$ (left) and $P_v = 1 \times 10^{-4}$ (right). The barrier for homogeneous nucleation is shown for reference.

potential $U_w(R)$ according to the Boltzmann factor, $\rho(R) = \rho_0(P) \exp\{-\beta U_w(R)\}$ (data not shown). For stronger adsorption, where the vapor density is strongly increased, this expression requires corrections to capture to the increasing inter-particle interactions. For wall-particle interaction strengths $\varepsilon > 5$ ($P_v = 3 \times 10^{-4}$, left) and $\varepsilon > 7$ ($P_v = 1 \times 10^{-4}$, right) spontaneous vapor-liquid nucleation was observed.

Next, we compute the free-energy barriers for heterogeneous nucleation on a planar weakly attractive wall. It is important to note that the definition of a free-energy barrier for heterogeneous nucleation requires some care as, unlike the nucleation rate, it is not an experimental observable. Rather, the free-energy barrier provides us with a measure of the probability to observe a nucleus of size n in a given system. This probability is made intensive by dividing the expected number of nuclei of size n by N, the number of particles in the system where nucleation takes place. If nucleation takes place in the bulk, N is simply the total number of particles in the system. However, if nucleation occurs on a wall, a result independent of system volume can be obtained by dividing the number of nuclei of size n that are formed on the wall by the total number of particles in contact with the wall. When comparing the rates of homogeneous and heterogeneous nucleation, one should then take into account how many particles in the system are in contact with the wall. In this way one accounts for the fact that the rate of heterogeneous nucleation is proportional to the total "active" surface area in the system (in other words, that it is proportional to the density of seeds). In what follows, we follow the above approach and define the barrier for heterogeneous nucleation as $\Delta G_{het}(n) = -k_B T \ln P_{het}(n)$, where we define $P_{het}(n)$ as the ratio of $M_{het}(n)$, the number of clusters with size n in contact with the heterogeneity, divided by the (average) total number of particles in contact with the heterogeneity,

N_{het}. We identify a particle or clusters to be in contact with the heterogeneity using the restricted order parameter as described in Section 5.1.

The resulting free-energy barriers are presented in Figure 5.3 as function of droplet volume (to be more precise, we show the barrier as a function of the cube root of this volume. In what follows, we denote this quantity as the "effective droplet radius"). Again, the vapor pressures are $P_v = 3 \times 10^{-4}$ and $P_v = 1 \times 10^{-4}$ for the left and right panel, respectively. When the attraction due to the wall is very weak, the computed barriers converge to the homogeneous case, as is to be expected. The barrier for homogeneous nucleation is plotted for the sake of comparison in Figure 5.3. With increasing attraction the liquid-vapor nucleation barrier decreases until spontaneous nucleation occurs (not shown - as, in this case, it becomes impossible to measure a free-energy barrier). For the rest of this work we choose our interaction strengths to be $\varepsilon = 5$ for $P_v = 3 \times 10^{-4}$ and $\varepsilon = 7$ for $P_v = 1 \times 10^{-4}$. Under these conditions the size of the critical nucleus is still sufficiently small to be studied in US simulations for the system sizes that we used.

To obtain a rough estimate of the wall-liquid contact angle, we fit the expression from classical nucleation theory (CNT) [4] with our simulation data (see Appendix 5.A for details on the CNT expression), which yields $\Theta = 0.33\pi$ ($P_v = 1 \times 10^{-4}$, $\varepsilon = 7$) and $\Theta = 0.43\pi$ ($P_v = 3 \times 10^{-4}$, $\varepsilon = 5$). The CNT estimates for the heterogeneous nucleation barrier discussed below are based on these estimates of the contact angle. As our estimates of the contact angles are not very accurate, the comparison with our simulation results is mainly intended to gain a qualitative understanding of the observed trends. A quantitative comparison would require more precise knowledge of the contact angles and the solid-liquid-vapor line tensions.

Having established the limiting cases, i.e. homogeneous nucleation and heterogeneous nucleation on a planar wall, we focus on the effect of a planar circular weakly-adsorbing patch on the nucleation rate. In particular, we are interested in the dependence of the nucleation rate on the radius of the patch. Figure 5.4 shows the nucleation barriers as function of the effective droplet radius $V^{1/3}$. The theoretical CNT prediction is shown in the upper panel and our simulation results in the lower panel, both for a vapor pressure $P_V = 1 \times 10^{-4}$. Valencia and Lipowsky [122] predicted (on the basis of CNT) that nucleation could be a two-step process under certain conditions. A similar conclusion was reached by Smorodin [121] on the basis of similar theoretical arguments. Indeed, under the conditions studied in the present simulations, CNT predicts a two-step barrier for the vapor density $P_v = 1 \times 10^{-4}$ and a patch radius around $R = 6.0$, as depicted by the inset in upper panel of Figure 5.4. However, the effect is very weak compared to the overall barrier height, and our simulations show no evidence for a two-step nucleation process.

The dependence of the height of the nucleation barrier on the patch radius is presented in Figure 5.5. The nucleation barrier decreases monotonically from the

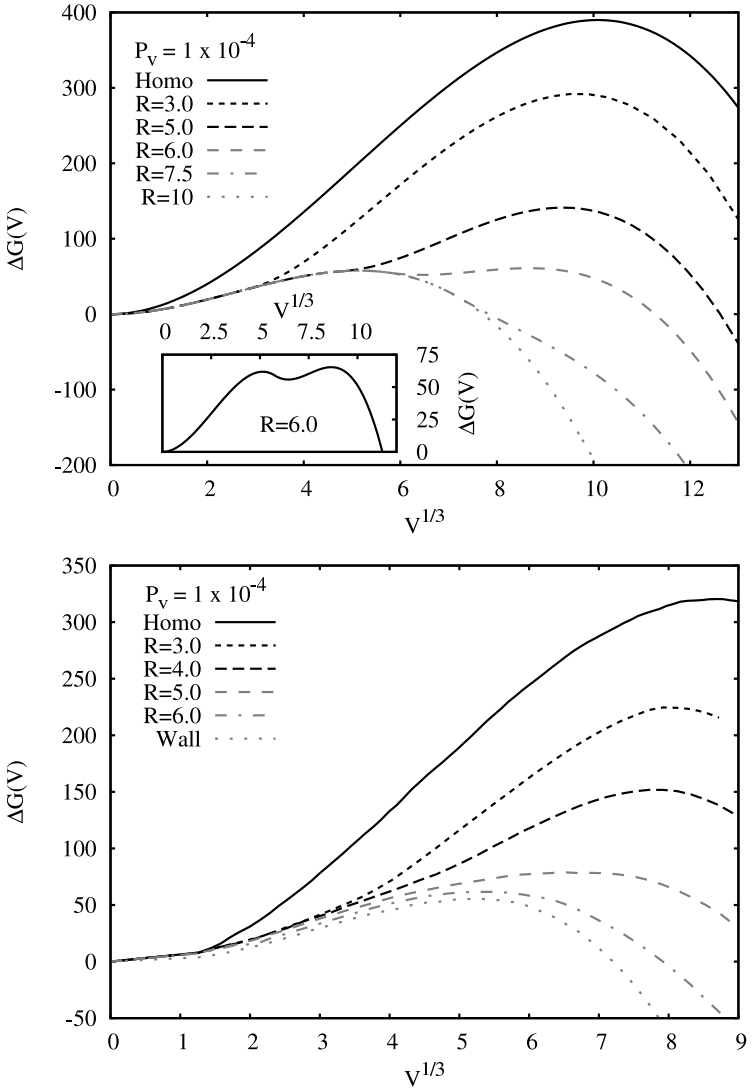


Figure 5.4: Vapor-liquid free-energy barriers as function of effective droplet radius $V^{1/3}$ for nucleation on a weakly attractive circular patch at vapor pressure $P_v = 1 \times 10^{-4}$. Top: CNT prediction for various patch radii. The inset shows a close-up on a weakly double-peaked barrier predicted for $R = 6$. Bottom: The simulation results, although qualitatively similar to the CNT predictions, show no evidence for a two-step process.

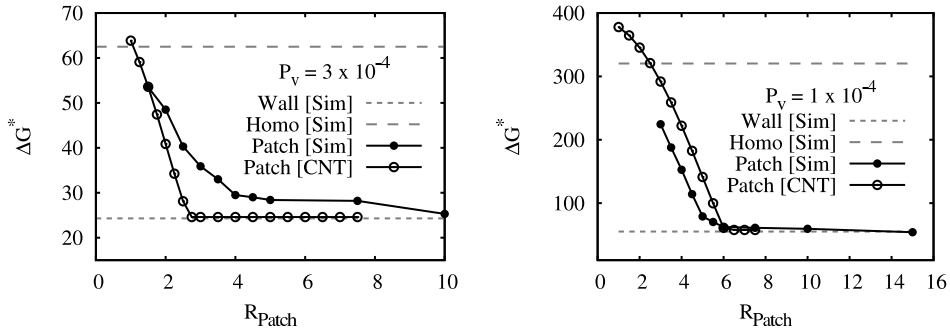


Figure 5.5: Vapor-liquid free-energy barrier heights for nucleation on a weakly attractive circular patch at vapor pressures $P_v = 3 \times 10^{-4}$ (left) and $P_v = 1 \times 10^{-4}$ (right) as function of patch radius R_{Patch} (filled circles). The limiting cases of homogeneous nucleation and nucleation on a planar weakly-attractive wall are presented for reference (horizontal lines), as well as the theoretical predictions using CNT (open circles).

homogeneous limit for a very small patch radius to the heterogeneous planar wall limit for large patch radii. The planar wall limit is reached at a patch radius $R_{max} \approx R^*_{homo} \sin\Theta$, which corresponds to the contact circle radius of a critical nucleus on a planar wall. The CNT predictions are shown for reference. Qualitatively, the CNT predictions agree well with our simulations results.

Finally, we consider a pore geometry, which is modeled as an open, weakly-attractive hemi-spherical cavity embedded in an otherwise purely repulsive (i.e. hard) planar wall. For a schematic side-view see Fig. 5.1.

Figure 5.6 presents the nucleation free-energy barriers as function of the effective droplet diameter for a vapor pressure $P_V = 1 \times 10^{-4}$. Again, the upper panel contains the CNT prediction, and the lower panel the simulation results. As expected, the nucleation barriers approach the corresponding limiting cases for very large and very small pores. Interestingly, for intermediate pore sizes, CNT predicts a double-barrier corresponding to a two-step process, and this feature is indeed reproduced by our simulations. From a visual inspection of simulation snapshots of configurations at the local and global maximum of the free-energy barrier, we identify the first step to be a pore-filling event and the second step to be the nucleation out of the pore into the bulk. Figure 5.8 presents simulation snapshots along such a pathway. It shows that first a critical nucleus is required to nucleate a liquid inside the pore (left), the system then spends a certain time in the metastable filled pore state (center), and finally a second nucleation event is required to break out of the pore (right). This confirms the predictions of Page and Sear for nucleation in a pore for an Ising model system [18].

The heights of the nucleation free-energy barrier are shown in Figure 5.7 as func-

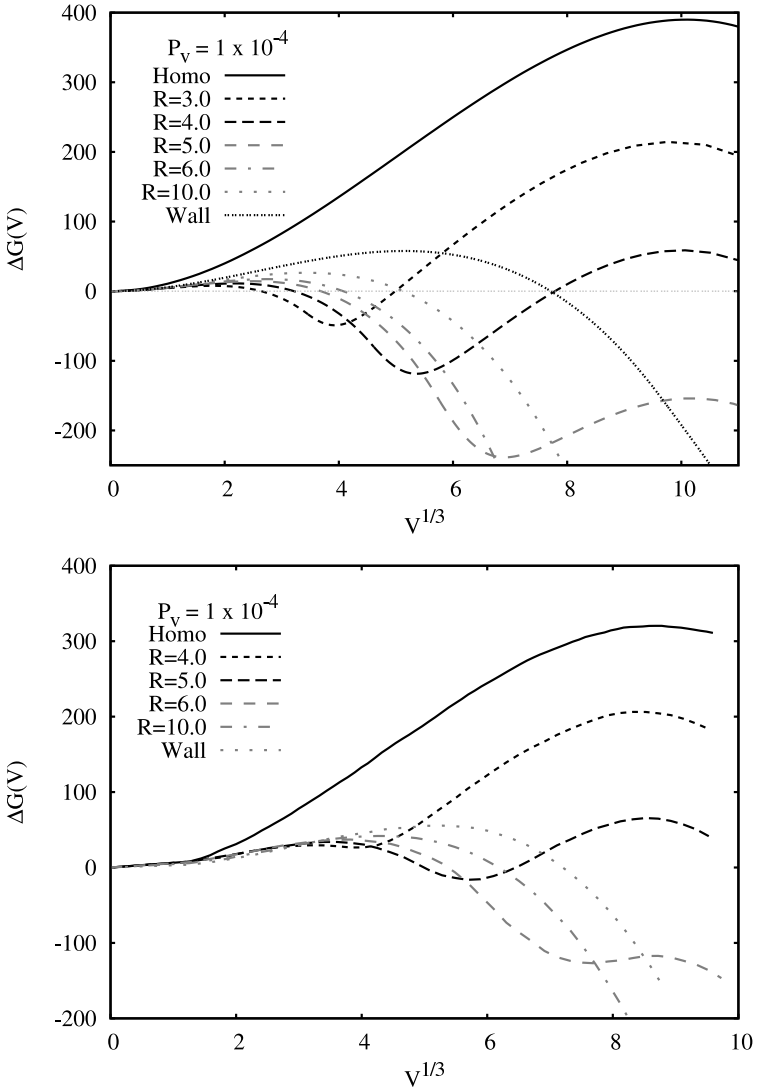


Figure 5.6: Vapor-liquid free-energy barriers as function of effective droplet radius $V^{1/3}$ for nucleation in a weakly attractive spherical pore at vapor pressure $P_v = 1 \times 10^{-4}$. Top: CNT prediction. Bottom: Simulation results.

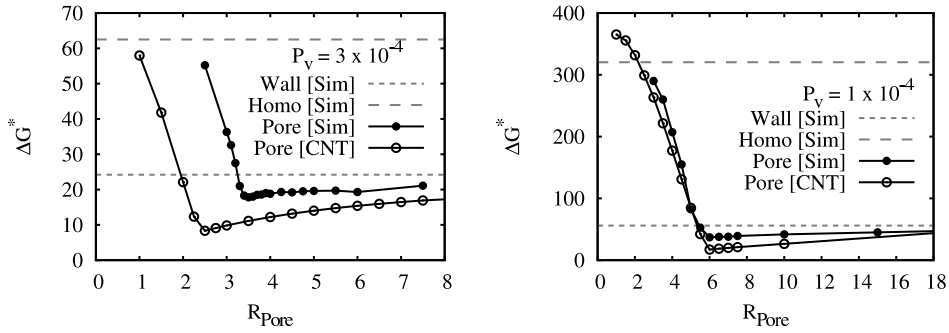


Figure 5.7: Vapor-liquid free-energy barrier heights for nucleation in a weakly attractive spherical pore at vapor pressures $P_v = 3 \times 10^{-4}$ (left) and $P_v = 1 \times 10^{-4}$ (right) as function of pore radius R_{Pore} (filled circles). The limiting cases of homogeneous nucleation and nucleation on a planar weakly-attractive wall are presented for reference (horizontal lines), as well as the theoretical predictions using CNT (open circles).

tion of pore size, for both $P_v = 3 \times 10^{-4}$ (left) and $P_v = 1 \times 10^{-4}$ (right). The bottleneck for the overall nucleation rate is the crossing of the higher free-energy barrier shown in this plot. Again, good qualitative agreement with CNT is found. For nucleation in a pore, the dependence on R is clearly non-monotonic, featuring a global minimum of the free-energy barrier. This finding supports the suggestion by Chayen and co-workers [133] that in a porous medium with a broad distribution of pore sizes only a small number of pores (namely those that correspond to the minimum in the overall nucleation barrier) dominate the overall nucleation process.

5.2.2 Liquid-crystal nucleation

We now focus on the liquid-crystal nucleation. In Chapter 4 we showed that, in the case of homogeneous nucleation, the nucleation of a crystal in a previously nucleated droplet is relatively fast. In other words: the liquid-crystal nucleation event is not the rate-limiting step. Further, we found that a droplet of a minimum size is required to host a crystal nucleus, and that this is approximately the size of a supercritical crystal nucleus plus a liquid mono-layer.

Following the same procedure as in the case of vapor-liquid nucleation, we first assess the limiting cases: homogeneous nucleation and heterogeneous nucleation on a planar wall. The crystallization rate within a liquid droplet was reported in Chapter 4 and found to be independent of the droplet size and vapor pressure. For the sake of completeness we here compute the crystallization free-energy barrier for homogeneous nucleation in a bulk liquid in the NpT ensemble. When we compare the

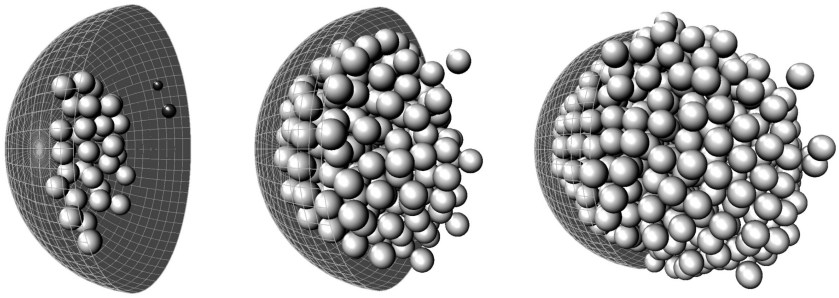


Figure 5.8: Snapshots from the two-step nucleation process in a pore ($P_v = 1 \times 10^{-4}$, $R = 5.5$). From left to right: critical nucleus for pore filling ($n_1^* \approx 50$), metastable filled pore ($n \approx 270$), and the critical nucleus for growing out of the pore ($n_2^* \approx 650$).

present results with those reported in Chapter 4 we find good agreement.

For the crystal nucleation on a planar wall we find that our system crystallizes spontaneously on the time scale of our simulations for both attraction strengths, $\varepsilon = 5$ and $\varepsilon = 7$. Similar behavior was observed in a study by Page and Sear on the pre-freezing transition of the same model system [136]. This high rate of spontaneous crystal nucleation implies that crystallization may take place during US simulations that probe the free-energy barrier for vapor-liquid nucleation. To find out whether this is indeed the case, we analyzed whether the configurations generated during the US simulations of the pathway for vapor-liquid nucleation contained any crystalline particles. In the left panel of Figure 5.9 we plot the number of crystalline particles n_x, identified using bond-order parameters [57], versus the number of liquid particles n_l, identified using Stillinger's overlapping spheres criterion [66]. Two observations can be made. First of all, all larger droplets ($n_l > 200$), contained a crystalline cluster. If we consider that a crystal is always covered by a thin layer of liquid, as was found in Ref. [135], we can fit the results to an empirical formula assuming that the crystal takes the same shape as the liquid droplet,

$$n_x = \left[(n_l)^{1/3} - a\right]^3 + c,$$

where c is a constant offset and the difference in density between the liquid and the crystal is ignored. The thickness Δ of the liquid layer can then be extracted by dividing a by the geometrical prefactor linking the droplet volume to the radius of a spherical cap with liquid-wall contact angle Θ, $\Delta = a\left[\frac{\pi}{3}(1-\cos\Theta)^2(2+\cos\Theta)\right]^{-1/3}$. Our fit yields $a = 0.748 \pm 0.1$. Assuming a wall-liquid contact angle of $\Theta = 0.33\pi$, obtained from matching the simulation barrier height for vapor-liquid nucleation on a

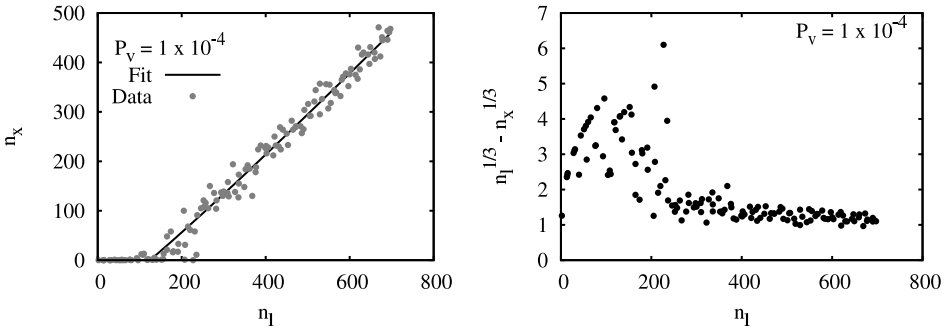


Figure 5.9: Analysis of the US data for nucleation on a planar wall. Left: largest cluster size n_x using bond-order parameters versus the largest cluster size n_l using a liquid order parameter. A minimum droplet size of $n_l \approx 200$ is required to host a stable crystal. Large droplets consist of a crystal cluster covered by a liquid mono-layer. The solid line is a fit to an empirical formula for a crystal with a liquid mono-layer (see main text for details). Right: The difference in the cube root of the cluster size between liquid and solid clusters, $n_l^{1/3} - n_x^{1/3}$. This difference converges to a constant value for large clusters indicating that the crystal surface is covered by a liquid layer of constant thickness.

planar wall to the CNT prediction, we arrive at a thickness of $\Delta \approx 0.88$, which agrees well with the assumption of a liquid mono-layer surrounding the crystallite. The difference $a = n_l^{1/3} - n_x^{1/3}$ is also plotted in the right panel of Figure 5.9 and appears to converge to a constant value for large clusters.

Second, we observe that small droplets, $n < 200$, contain very little crystalline order. This means that for these small droplet sizes, crystallization will not affect the computed vapor-liquid free-energy curve. As in the case of homogeneous nucleation, the liquid droplet has to exceed a critical size before crystals can form easily within it. In view of the high (spontaneous) crystal nucleation rate in large droplets, this might seem to be a rather surprising result, because the crystal critical nucleus is expected to be very small. However, due to the small wall-liquid contact angle, a droplet with $n \approx 200$ is only a few layers thick. Crystal nucleation in such a pancake-shaped droplet is not determined by the total number of particles in the droplet, but by its thickness. Inside thin droplets, liquid-crystal nucleation is expected to be strongly suppressed, and that is indeed what we observe.

Next, we assess crystal nucleation in a finite-sized circular patch. Again, we observe spontaneous crystal nucleation for all but the smallest patch radii. Using US we estimated the crystal nucleation barrier to be very small ($\Delta G^* < 10 k_B T$). Figure 5.10 shows the results from analysis of the crystallinity inside the liquid clusters.

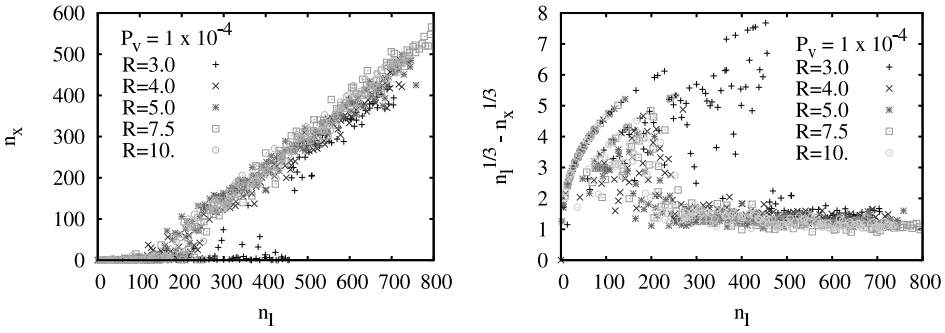


Figure 5.10: Analysis of the US data for nucleation on a planar patch. Different symbols correspond to different patch radii. Left: largest cluster size n_x using bond-order parameters versus the largest cluster size n_l using a liquid order parameter. A minimum droplet size of $n_l \approx 200$ is required to host a stable crystal. Large droplets consist of a crystal cluster covered by a liquid mono-layer. Right: The difference in the cube root of the cluster size between liquid and solid clusters, $n_l^{1/3} - n_x^{1/3}$. This difference converges to a constant value for large clusters indicating that the crystal surface is covered by a liquid layer of constant thickness.

From the plot we conclude that, as in the case of nucleation on a planar wall, a minimum droplet size is required to support a stable crystal. As before, we find that the resulting crystal fills the droplet except for a liquid mono-layer.

For crystallization in a pore the curvature can cause stress and frustration for a regular arrangement of particles, and thus might inhibit crystal nucleation. Therefore, we computed the free-energy barrier for crystal nucleation for those systems within a pore filled with a meta-stable droplet (data not shown). For small pores no stable crystal grows in these droplets. For the largest pore that can contain meta-stable liquid droplets ($R = 6$), we observe a small crystal-nucleation barrier of $\Delta G^* \approx 7.5 k_B T$. This barrie height is slightly less than the nucleation barrier for growing out of the pore ($\Delta G^* \approx 10 k_B T$). Hence in these pores crystallization is expected to precede pore break-out. The resulting crystallization results in the formation of a stable crystal of $n \approx 140$ particles.

For pore radii $R > 6$ a fully-filled pore is not a metastable state anymore, and the liquid continuously grows into the bulk. In this situation, the pore's attraction and curvature can still influence the rate for crystal nucleation, but this step is never rate-limiting.

5.3 Conclusion

The focus of the present chapter is on heterogeneous nucleation of liquids and crystals. We studied nucleation on a planar wall, on a planar circular patch, and in a spherical pore. For all surface geometries we find the nucleation pathway to follow the homogeneous two-step route, with the vapor-liquid nucleation as rate-limiting step. Comparing the simulation results to the corresponding CNT predictions we find good qualitative agreement. However, a quantitative comparison was not possible due to the fact that the wall-liquid and wall-crystal surface tensions are not known with sufficient accuracy, and the line tensions of the nuclei on the surface are unknown. Yet line tensions are known to have a pronounced effect on the nucleation barrier [27, 137, 138].

For a planar circular patch the free-energy barrier for vapor-liquid nucleation shows a monotonic decrease with increasing patch radius, interpolating between the barrier for homogeneous nucleation and the one for nucleation on a planar wall. For the parameter range studied in our simulations, we did not observe two-step vapor-liquid nucleation. Moreover, CNT calculations using as input the parameters that characterize the systems that we studied in the simulations also failed to exhibit two-step nucleation. These findings indicate that the two-step nucleation scenario on a planar circular patch that was predicted by Refs. [121, 122] occurs for conditions that are outside the domain that we could study in our simulations. Furthermore, our simulations show that liquid-crystal nucleation happens spontaneously for large patches and at a planar wall, provided that the liquid droplet exceeds a minimum size of $n \approx 200$ particles.

For a hemi-spherical pore we find a richer behavior. Very small and very large pores exhibit a vapor-liquid nucleation free-energy barrier that approaches the ones for homogeneous nucleation and heterogeneous nucleation on a planar wall, respectively. For intermediately-sized pores, the liquid nucleation itself is a two-step process. In the first nucleation event a droplet nucleates inside the pore and grows to fill the entire pore. From this metastable state, a second nucleation event is required to grow out of the pore. The overall nucleation rate shows a non-monotonic dependence on the pore radius with an optimum pore size for which the barrier is lowest. This confirms a previous study on nucleation in and out of pores for an Ising model system [18]. Depending on the strength of attraction the minimum barrier can be even lower than for nucleation on a planar wall. The liquid-crystal nucleation typically follows once the droplet is large enough. For a small pore the droplet extends into the bulk and the crystal nucleation is homogeneous. For large pores the crystal forms in contact with the pore surface at a very high rate - in practice, there is no nucleation barrier. Only for the largest pore within which a metastable droplet could be maintained, did we find the filled pore to be large enough to host a stable crystallite. In

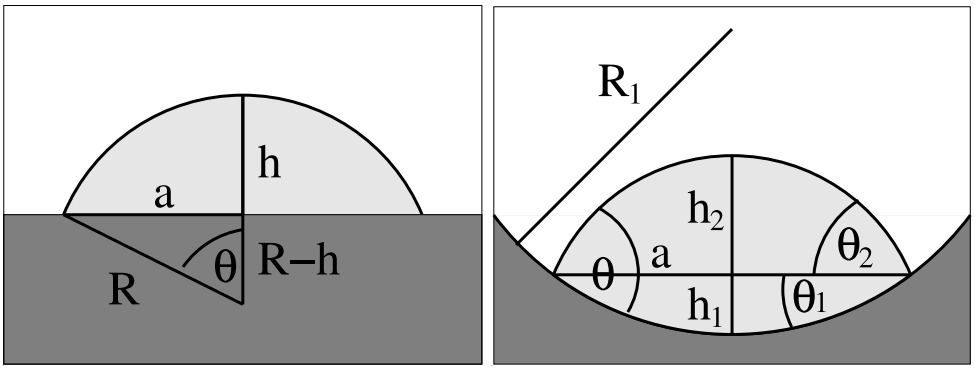


Figure 5.11: Left: liquid droplet on a planar surface, surrounded by vapor. The droplet's shape corresponds to a spherical cap where the geometrical parameters are the droplet radius R, the contact angle Θ, the height of the droplet h, and the radius a of the contact circle between droplet and surface. Right: liquid droplet in a spherical pore, surrounded by vapor. The droplet is composed of two spherical caps with the corresponding parameters R_i, a_i, h_i and Θ_i, with $i = 1$ for the lower cap and $i = 2$ for the upper cap. The contact angle is denoted Θ (without subscript). For the relation between these parameters see main text.

that case the liquid crystallizes before it grows out of the pore.

In conclusion, our simulations show that a weakly attractive circular patch or hemi-spherical pore on an otherwise non-adsorbing substrate can reduce significantly the free-energy barrier for nucleation. Both geometries are most efficient if their radius is close to or slightly larger than the critical radius for homogeneous nucleation, at which point they strongly accelerate the formation of exactly one nucleus. A widely-spaced array of adsorbing patches or pores with a broad distribution of radii crafted on an otherwise non-adsorbing surface should therefore make an efficient nucleation agent for the formation of high-quality crystals. We point out that the present work is limited to smooth surfaces. Further research is required on (crystal) nucleation at surfaces that are rough on the scale of a particle diameter. In such cases, roughness is expected to have a pronounced effect on the free-energy barrier for crystal nucleation.

Acknowledgements

The authors thank S. Abeln, A. Vila Verde, and K. Shundyak for fruitful discussions.

5.A CNT derivation

Assume a droplet in contact with an attractive planar wall like depicted in the left panel of Figure 5.11. We denote the surface tension between the wall-vapor, wall-liquid and vapor-liquid interface by σ_{wg}, σ_{wl} and σ_{lg}, respectively, and the contact angle by Θ. The CNT expression for the free-energy of this droplet is given by [16]

$$\Delta G = \sigma_{lg} A_{lg} - |\Delta\mu| \rho V_l + (\sigma_{wl} - \sigma_{wg}) A_{wl},$$

where A_{lg} and A_{wl} refer to the vapor-liquid and wall-liquid surface area, respectively, V_l to the droplet volume, ρ to the bulk liquid density, and $\Delta\mu$ to the difference in chemical potential between the bulk vapor and bulk liquid. Using the Young-Dupré relation for the contact angle, $\sigma_{wg} = \sigma_{wl} + \sigma_{lg}\cos\Theta$, we obtain

$$\Delta G = \sigma_{lg} A_{lg} - |\Delta\mu| \rho V_l - \sigma_{lg} A_{wl} \cos\Theta \tag{A5.1}$$

The volume and the surface area of this droplet is given by the expression for a spherical cap,

$$\begin{aligned} V_{cap}(R,\Theta) &= \frac{1}{4}(1-\cos\Theta)^2(2+\cos\Theta)\frac{4\pi}{3}R^3 \\ &= \Phi_V(\Theta) V_s(R) \\ A_{cap}(R,\Theta) &= \frac{1}{2}(1-\cos\Theta) 4\pi R^2 \\ &= \Phi_A(\Theta) A_s(R), \end{aligned}$$

where $V_s(R) = 4\pi/3R^3$ and $A_s(R) = 4\pi R^2$ is the volume and surface area, respectively, of a full sphere with radius R, and $\Phi_A(\Theta) = 0.25(1-\cos\Theta)^2(2+\cos\Theta)$ and $\Phi_V(\Theta) = 0.5(1-\cos\Theta)$ are contact-angle dependent surface and volume modifiers, respectively. Inserting the volume and area of our spherical cap in the CNT expression yields the barrier height and critical radius for nucleation on a planar attractive wall [16],

$$\Delta G^* = \Delta G^*_{homo} \Phi_V(\Theta), \tag{A5.2}$$

with $\Delta G^*_{homo} = 16\pi/3\sigma_{lg}^3 \rho^{-2} |\Delta\mu|^{-2}$ the critical CNT barrier height for homogeneous nucleation. The critical nucleus radius is the same for both homogeneous nucleation and nucleation on a planar wall, $R^* = 2\gamma\rho^{-1}|\Delta\mu|^{-1}$, and the radius of the wall-liquid contact circle is $R = R^* \sin\Theta$.

Circular patch

For an attractive circular patch with radius R_P the free energy has to be slightly modified. We can identify three distinct regions [121, 122, 139]: in regime *I* ($R <$

$R_P/\sin\Theta_P$) the droplet is small enough to fit entirely on the circular patch with the natural patch-liquid contact angle Θ_P. Regime *II* applies if the droplet covers the entire patch with a contact angle $\Theta_P < \Theta < \Theta_W$, where Θ_W is the wall-liquid contact angle. In the final regime *III* ($R > R_P/\sin\Theta_W$) the droplet extends beyond the circular patch, and the contact angle is that with the wall. The CNT expression for the free energy difference is hence

$$\begin{aligned}\Delta G_I(R) &= \sigma_{lg}A_S(R)\Phi_A(\Theta_P)\\ &\quad -\sigma_{lg}\pi R^2\sin^2\Theta_P\cos\Theta_P\\ &\quad -|\Delta\mu|\rho V_S(R)\Phi_V(\Theta_P)\end{aligned}$$

$$\begin{aligned}\Delta G_{II}(\Theta) &= \sigma_{lg}A_S(R_P/\sin\Theta)\Phi_A(\Theta)\\ &\quad -\sigma_{lg}\pi R_P^2\cos\Theta_P\\ &\quad -|\Delta\mu|\rho V_S(R_P/\sin\Theta_P)\Phi_V(\Theta)\end{aligned}$$

$$\begin{aligned}\Delta G_{III}(R) &= \sigma_{lg}A_S(R)\Phi_A(\Theta_W)\\ &\quad -\sigma_{lg}\pi R_P^2\cos\Theta_P\\ &\quad -\sigma_{lg}\pi(R\sin\Theta_W - R_P)^2\cos\Theta_W\\ &\quad -|\Delta\mu|\rho V_S(R)\Phi_V(\Theta_W)\end{aligned}$$

Spherical pore

Next, consider a droplet in a spherical pore, as sketched in the righthand panel of Figure 5.11. The droplet volume consists of two spherical caps, each with its own radius R_i and contact angle Θ_i. From the figure it is obvious that $\Theta_P = \Theta_1 + \Theta_2$, where Θ_P is the liquid-pore contact angle. The lower cap has the curvature of the spherical pore, so $R_1 = R_P$. Both caps share the radius of the circle where they meet, so $a_1 = a_2 \equiv a$. The radius R_i of a cap is related to a by $a = R_i\sin\Theta_i$, so the radii are related by $R_2 = R_P\sin\Theta_1/\sin\Theta_2 = R_P\sin\Theta_1/\sin\Theta_P - \Theta_1$. Analogous to the droplet on a circular patch we identify three regimes: in regime *I* the droplet does not fill the pore, i.e. $\Theta_1 < \pi/2$. Note that this is independent of the actual contact angle Θ_P. In the second regime, the droplet fills the pore, $\Theta_1 = \pi/2$, and grows until the upper cap angle Θ_2 reaches the contact angle with the surrounding wall, $\Theta_P - \pi/2 \leq \Theta_2 \leq \Theta_W$. Here, we point out that for intermediately to strongly wetting pores, $\Theta_P < \pi/2$, the upper cap angle $\Theta_2 = \Theta_P - \pi/2$ starts out negative. In regime *III* the upper cap grows beyond the pore onto the surrounding wall, taking a mushroom-like shape, $R > R_P$.

Computing the free energy for the three different regimes then yields

$$\begin{aligned}
\Delta G_I(\Theta_1) &= \sigma_{lg} A_S(R_P) \frac{\Phi_A(\Theta_P - \Theta_1)\sin^2\Theta_1}{\sin^2\Theta_P - \Theta_1} \\
&\quad -\sigma_{lg} A_S(R_P)\Phi_A(\Theta_1)\cos\Theta_P \\
&\quad -|\Delta\mu|\rho V_S(R_P)\frac{\Phi_V(\Theta_P - \Theta_1)\sin^3\Theta_1}{\sin^3\Theta_P - \Theta_1} \\
&\quad -|\Delta\mu|\rho V_S(R_P)\Phi_V(\Theta_1) \\
\Delta G_{II}(\Theta_2) &= \sigma_{lg} A_S(R_P)\Phi_A(\Theta_2)\sin^{-2}\Theta_2 \\
&\quad -\sigma_{lg} A_S(R_P)\Phi_A(\Theta_1)\cos\Theta_P \\
&\quad -|\Delta\mu|\rho V_S(R_P)\Phi_V(\Theta_2)\sin^{-3}\Theta_2 \\
&\quad -|\Delta\mu|\rho V_S(R_P)\Phi_V(\Theta_1) \\
\Delta G_{III}(R) &= \sigma_{lg} A_S(R)\Phi_A(\Theta_W) \\
&\quad -\sigma_{lg}\pi(R^2 - R_P^2)\cos\Theta_W \\
&\quad -\sigma_{lg}\frac{1}{2}A_S(R_P)\cos\Theta_P \\
&\quad -(\frac{1}{2}V_S(R_P) + V_S(R)\Phi_V(\Theta_W))\rho|\Delta\mu|.
\end{aligned}$$

Chapter 6

Disordered substrates favor heterogeneous crystal nucleation of model globular proteins

In this chapter we investigate by means of numerical simulations the pathways of heterogeneous crystal nucleation in a simple model system for globular proteins. We consider smooth, periodically structured, and disordered surfaces of planar walls and cylindrical cavities, modeling adsorption to container walls and micro-porous media.

Under conditions similar to those where protein crystallization is observed experimentally, we find that the presence of a planar wall with suitable wetting conditions induces spontaneous crystallization. Our results on periodically structured surfaces agree well with the behavior reported for colloidal hard-sphere systems [123], with the exception that for the model studied here a strong lattice mismatch between substrate and adsorbed crystal leads to the formation of a thin, dense layer of a disordered phase, which lowers the free-energy barrier to subsequent crystal nucleation.

Our most striking results relate to the properties of disordered substrates. As in the case of a strong mismatch of the periodically structured wall, a disordered wall adsorbs a thin liquid film. For a cylindrical cavity of suitable size we find that disorder leads to a liquid filling the entire pore, and we show that such an environment facilitates fast crystal nucleation without any of the mismatch-induced strain effects that occur in the case of nucleation on an ordered but non-commensurate substrate.

First-order phase transitions such as crystallization typically proceed via nucleation and subsequent growth [16]. In the first stage of this process, spontaneous thermal fluctuations cause the formation of a microscopic nucleus of the new phase. If such a nucleus is small it has a high chance to dissolve again, but if the fluctuation is large enough, exceeding the so-called critical nucleus size, the new phase is likely to grow to a macroscopic size. The tipping point corresponds approximately to the top of a free-energy barrier separating the meta-stable parent phase (e.g. the liquid) from the thermodynamically stable phase (e.g. the crystal). Provided that this barrier is much larger than the thermal energy k_BT, nucleation becomes a rare event limiting the rate of formation of the new phase.

An effective way to enhance the rate of spontaneous nucleation is to lower the free-energy barrier, e.g. by increasing the supersaturation. However, increasing the supersaturation also affects the growth speed, and in the case of crystallization often results in the formation of a defect-rich crystal or even of a poly-crystalline material. For the fabrication of high-quality single crystals, such as those required for protein structure analysis, such defects should be avoided.

Heterogeneous nucleation offers an alternative route to lower the free-energy barrier. In contrast to homogeneous nucleation of a pure bulk substance, the term *heterogeneous* refers to the presence of foreign particles or container walls that can act as seeds facilitating the formation of nuclei. For crystallization this means that heterogeneous nucleation can occur at conditions where the rate of homogeneous nucleation is negligible and where the rate of the crystal growth is still slow. It would therefore be advantageous to design substrates that can act as heterogeneous nucleation agents. But this is complicated by the fact that the process of heterogeneous nucleation is still poorly understood on a molecular scale.

Over the past two decades much progress has been reported in the study of crystal nucleation. On the one hand, the advent of colloids as experimental model-system allows for a direct comparison of experimental observations with numerical many-particle simulations [112, 113]. On the other hand, advances in the field of rare-event algorithms make it easier to study the kinetics of nucleation-dominated phase transformations in simulations [23, 44, 49, 57]. Research is further stimulated by the demand to control nucleation for the assembly of nano-colloids into photonic materials or for the fabrication of high-quality protein crystals for structural analysis.

Both experimentally and theoretically, much attention has been paid to the study of structured surfaces to control crystal nucleation and growth. For instance, experiments on colloidal systems showed that a crystal template can strongly enhance or suppress epitaxal crystal growth depending on the ratio of the lattice spacing of the template and the epitaxially grown crystal [127]. These experiments inspired subsequent theoretical [140, 141] and numerical [123] studies. Surprisingly, a highly ordered substrate is not necessarily the best nucleation agent: experiments by Chayen et

al. [17, 132, 133] on heterogeneous protein crystallization suggested that highly disordered micro-porous seeds can act as universal nucleation agents for the formation of high-quality single protein crystals. The molecular mechanism of nucleation in a pore has subsequently been studied theoretically using an Ising lattice model [18] and by means of numerical simulations in a Lennard-Jones model accounting for effects from a high-density meta-stable disordered phase [142]. Another experimental study of protein crystallization revealed that surface roughness has a pronounced effect on the nucleation rate [143], but to our knowledge, systematic theoretical or numerical studies on this topic are still lacking.

In this chapter we investigate by means of numerical simulations the pathways of heterogeneous crystal nucleation in a simple model system for globular proteins. We consider smooth, periodically structured, and disordered surfaces of planar walls and cylindrical cavities, modeling adsorption to container walls and micro-porous media. Under conditions similar to those where protein crystallization is observed experimentally, we find that the presence of a planar wall with suitable wetting conditions induces spontaneous crystallization. Our results on periodically structured surfaces agree well with the behavior reported for colloidal hard-sphere systems [123], with the exception that for the model studied here a strong lattice mismatch between substrate and adsorbed crystal leads to the formation of a thin, dense layer of a disordered phase, which lowers the free-energy barrier to subsequent crystal nucleation. Our most striking results relate to the properties of disordered substrates. As in the case of a strong mismatch of the periodically structured wall, a disordered wall adsorbs a thin liquid film. We find that for a roughness with an amplitude comparable to the size of protein particles both the amount and the isotropy of the adsorbed liquid are maximal, which helps to explain some of the experimental observations [143]. For a cylindrical cavity of suitable size we find that disorder leads to a liquid filling the entire pore, and we show that such an environment facilitates fast crystal nucleation without any of the mismatch-induced strain effects that occur in the case of nucleation on an ordered but non-commensurate substrate.

6.1 Simulations details

Our aim is to study heterogeneous crystal nucleation in a many-particle system with a phase diagram similar to that of many globular-protein solutions. To this end, we must pay special attention to the choice of the effective pair interaction between our model proteins. Such pair potentials combine excluded volume interactions at distances less than the effective diameter of the molecules with a short-range attraction at slightly larger separations. Shortening the range of attraction has a pronounced effect on the topology of the phase diagram: for long-range attractions, the phase behav-

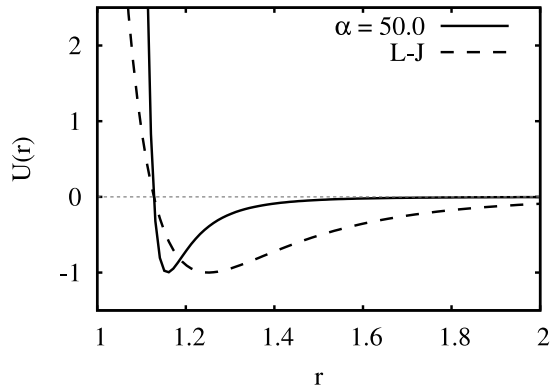


Figure 6.1: Conventional Lennard-Jones potential (dashed line) and the modified Lennard-Jones potential (solid line) according to Eqn. 6.1 with $\alpha = 50$. In order to facilitate direct comparison of the two model potentials, the distance-scale for the Lennard-Jones potential has been shifted to match the zero-crossing of the $\alpha = 50$ potential. For both potentials we set $\varepsilon = \sigma = 1$.

ior is similar to that of Argon: the critical temperature is well above the triple-point temperature and between those two temperatures it is possible to observe a thermodynamically stable liquid phase. As the range of attraction is shortened, the vapor-liquid coexistence curve drops below the vapor-solid coexistence and, as a consequence the critical point is in the meta-stable region below the freezing curve. For even shorter-ranged attractions an iso-structural transition in the solid phase is expected to emerge. Potentials that exhibit this sequence of scenarios include a hard-core potential with an attractive square well [144, 145], a hard-core attractive Yukawa [146–149], the so-called $2n - n$ potentials [150], and the modified Lennard-Jones [10]. We discuss the latter model in more detail in the next paragraph.

On the basis of the analysis of a large number of data on real protein solutions, Rosenbaum and co-workers [134, 151] argued that all potentials with a 'similar' ratio of the range of attraction to the range of (hard-core) repulsion should exhibit similar phase behavior. A more quantitative formulation of this generalized corresponding states principle has been proposed by Noro and Frenkel [152].

Inter-particle potential

For the present study we use the modified Lennard-Jones potential with $\alpha = 50$ that was proposed by ten Wolde and Frenkel as a simple model for protein-protein inter-

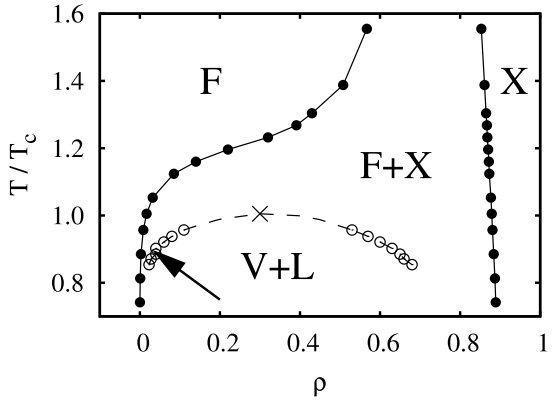


Figure 6.2: Phase diagram for the modified Lennard-Jones potential (Eqn. 6.1) for $\alpha = 50$. The stable phases are the fluid (F) and the crystal (X). Their coexistence region (solid line) envelops the metastable vapor-liquid (L+V) coexistence curve (dashed line) including its critical point indicated by the large cross. The arrow indicates the working conditions for this chapter. Data from Ref. [10].

actions [10]:

$$U(r) = \begin{cases} \infty & r \leq \sigma \\ \frac{4\varepsilon}{\alpha^2}\left[\left(\left(\frac{r}{\sigma}\right)^2 - 1\right)^{-6} - \alpha\left(\left(\frac{r}{\sigma}\right)^2 - 1\right)^{-3}\right] & r \geq \sigma \end{cases} \tag{6.1}$$

The hard-core repulsive part prevents penetration allowing for both Monte Carlo (MC) and Molecular Dynamics (MD) simulations. For $\alpha = 50$ the stickyness parameter τ [152] equals that of the hard-core attractive Yukawa system for $\kappa = 7$ known to resemble protein phase behavior [134]. More work on the phase diagram has been reported by Pagan et al. [153], who computed more accurately the metastable fluid-fluid coexistence in the vicinity of the critical point, and by Lutsko and Nicolis [154] who studied the entire phase diagram. Crystallization dynamics was explored by Costa et al. [155], and nucleation free-energy barriers were presented by ten Wolde and Frenkel [10] and by Lutsko and Nicolis [14].

Numerical approach

In what follows, we use reduced units. We define the reduced distance as $r^* = r/\sigma$ and the reduced potential energy as $u^* = U/\varepsilon$. All other reduced quantities (e.g. the pressure $P^* = P\sigma^3\varepsilon^{-1}$, the density $\rho^* = \rho\sigma^3$, and the temperature $T^* = k_B T\varepsilon^{-1}$) follow. All quantities reported in this work are expressed in these reduced units. Therefore, we omit the superscript asterisk (*) from here on.

T/T_c	T_m/T_c	T	P	$\Delta\mu$	ρ_f	ρ_c	μ_f	P_{ig}
0.89	1.04	0.374	0.0104	-0.748	0.035	0.877	-1.400	0.00885
1.00	1.14	0.418	0.0406	-0.610	0.300	0.872	-1.165	0.0262
1.09	1.22	0.456	0.1200	-0.455	0.452	0.869	-0.971	0.0543
2.23	2.39	0.931	4.1490	-0.208	0.652	0.814	*	*

Table 6.1: Nucleation conditions from Ref. [156]. Note that the difference in chemical potential is approximated using the equation $\Delta\mu \approx \Delta h(T_m - T)/T_m$ as stated in Ref. [10]. Values for ρ_f, μ_f and P_{ig} are obtained from NPT simulations and from Widom insertion in the NVT ensemble. P_{ig} is the pressure of the particle reservoir in the grand-canonical ensemble.

For the present work we set $\alpha = 50$, and truncate and shift the interaction at a distance $r_c = 2.0\sigma$. The resulting potential is plotted in Figure 6.1 along with a conventional Lennard-Jones potential for reference. The corresponding phase diagram is depicted in Figure 6.2 using the data of Ref. [10]. Table 6.1 lists the working conditions for homogeneous nucleation studied in Ref. [10]. We focus only on a temperature slightly below the critical temperature, $T = 0.89T_c = 0.374$, at a fluid density in the vicinity of the meta-stable vapor-liquid coexistence line (see Table 6.1), where surface pre-wetting might occur. At this point in the phase diagram, which is indicated by an arrow in Fig. 6.2, the crystal (X) is the thermodynamically stable phase, the fluid (F) is meta-stable, and the liquid (L) is unstable.

We performed Metropolis Monte Carlo simulations in the grand-canonical ensemble, where temperature, volume, and the chemical potential are kept constant, and the particle number is allowed to fluctuate. This ensures a constant fluid pressure and minimizes finite size effects. The maximum particle displacement is set to $\Delta = 0.2\sigma$, and each particle is moved on average 20 times before an attempted particle insertion/removal move. The simulation box size is chosen such as to keep the bulk volume constant at $V_{bulk} = 12,000$, and periodic boundary conditions were applied in both y and z directions. A planar hard wall opposite to the heterogeneous surface prevents particles from escaping the box in the x direction. For the surface-particle interaction we apply the same functional form as for inter-particle interactions, but vary the interaction strength ε. For the computation of free-energy barriers we use umbrella sampling with a bias on the number of particles in the largest crystalline cluster as outlined in Ref. [57]. Crystalline particles are identified with bond-order parameters using a neighbor distance cutoff $r_c = 1.35$, a correlator threshold $\mathbf{q}_6(i) \cdot \mathbf{q}_6^*(j) > 0.55$ to form a link, and a number of links threshold of $n_l > 7$ to form a solid-like particle. This definition was used for all calculations on heterogeneous nucleation. In the case of homogeneous nucleation, where fluid particles have on average less than 7 neighbors, we applied a modification to the last step of the order

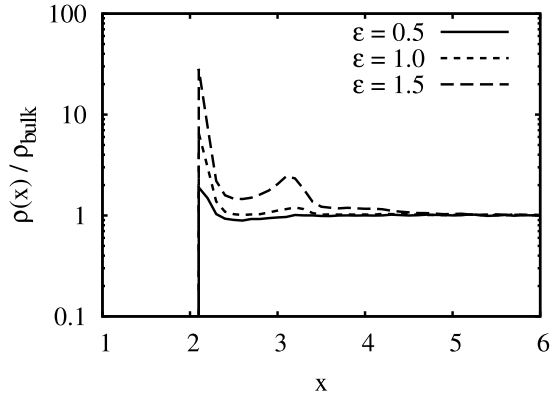


Figure 6.3: Relative fluid density $\rho(x)/\rho_{bulk}$ as function of distance x perpendicular to a weakly-attractive planar wall for various interaction strengths ε.

parameter, that changes the threshold for the number of links to a threshold for the *fraction* of links, i.e. the number of links divided by the number of neighbors [73]. Note that we checked whether the choice of order parameter affects the nucleation free-energy barrier height for heterogeneous nucleation, but no significant difference was observed.

6.2 Results

Before considering heterogeneous nucleation, we first verify that our simulations for homogeneous nucleation yield results consistent with Ref. [10], in particular since our numerical approach differs in two aspects. First, in contrast to the Molecular Dynamics simulations in the isobaric-isoenthalpic (NPH) ensemble of Ref. [10] we perform Monte Carlo simulations in the grand-canonical (μVT) ensemble. Second, instead of using two order parameters we have only a single order parameter, which takes into account that particles at the surface of a crystal nucleus have less nearest neighbors compared to bulk crystal particles. We compute the nucleation free-energy barrier at three temperatures, $T = 0.374$, $T = 0.42$, and $T = 0.456$, and find good agreement with the values reported in Ref. [10].

6.2.1 Heterogeneous nucleation on a smooth surface

First, we consider the simplest geometry for heterogeneous nucleation: a smooth planar adsorbing wall. For the interaction between the particles and the wall we apply the same potential as for the inter-particle interaction, but we vary the attraction strength by changing the prefactor ε. Direct simulations identify whether particles adsorb to

the wall. From these simulations we obtain the vapor-density profile perpendicular to the wall. We find spontaneous crystallization for all simulations within 10^6 MC cycles given that the wall-particle interaction strength is sufficient high, namely $\varepsilon \geq 2$. For those simulations without spontaneous crystallization Figure 6.3 shows the vapor density profile $\rho(x)$ as function of distance x perpendicular to the wall, normalized by the bulk fluid density ρ_{bulk}. Surprisingly, we never observe a disordered high-density layer adsorbed on the surface, although we simulate at conditions very close to the meta-stable vapor-liquid coexistence line. We will come back to this point later.

Next, we reduce the attractive area of the wall to a circle of radius R, and modify the remainder of the wall to interact purely repulsively. This geometry is of interest for its potential to induce the growth of high-quality single crystals: it can lower the nucleation free-energy barrier while supporting only one crystal nucleus at a time. The key question is how the free-energy barrier changes with radius. More specifically, we ask whether there exits a range of patch radii such that the patch supports only one nucleus at a time, yet lowers the nucleation free-energy barrier almost to that of a planar wall? Our simulations show spontaneous nucleation given the wall strength exceeds a certain value. Patches with $R < 3$ do not show any spontaneous crystallization, but once the patch radius exceeds $R \geq 3$, crystal growth is observed for all but the weakest adsorptions with $\varepsilon = 2$. From $R = 5$ on all samples crystallized, which indicates that the free-energy barrier is virtually non-existent and the planar wall limit is reached. At this size the patch is still small enough to only support a single nucleus, and should qualify for a good nucleation agent. We come back to this point in the discussion.

Another geometry of interest is a hemi-spherical cavity as a model for a shallow pore. In a recent study on vapor-crystal nucleation of a conventional Lennard-Jones system we found that the presence of a hemi-spherical pore strongly enhances the nucleation rate [142]. This effect can be understood by recalling that, just below the triple point, vapor-crystal nucleation is expected to proceed via a two-step process [14, 95, 135]. In a first event a meta-stable liquid droplet nucleates. The crystal then forms inside this droplet in a second independent step. In the case of heterogeneous nucleation, the hemi-spherical pore does not affect crystal nucleation, but facilitates the formation of the liquid phase, and because the latter is rate-limiting it enhances the overall vapor-crystal nucleation rate [142].

However, there is one important difference between the model system studied here and the conventional Lennard-Jones system. At conditions studied here the liquid phase is not even meta-stable but *unstable*. This means that, in contrast to the situation discussed in Ref. [142], it is not possible to form a meta-stable liquid droplet in the bulk vapor. If a pore fills with a liquid it is stabilized purely by the impurity and will not grow out of the pore. On the other hand, if no liquid forms, the crystal could still benefit by attaching directly to the surface. It then depends on the curva-

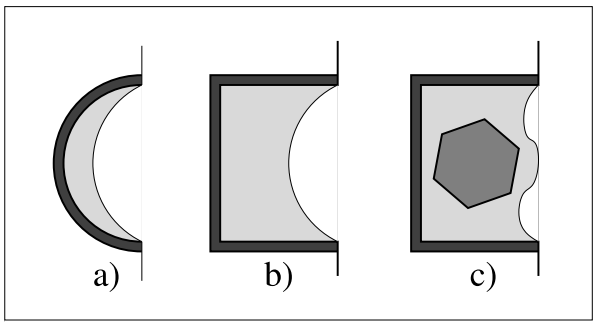


Figure 6.4: Schematic side view of a hemi-spherical (a) and cylindrical (b,c) pore wetted by a liquid. The dark gray area indicates the attractive part of the pore, and the light-gray area the adsorbed liquid. Panel c) also shows a crystal nucleus fully immersed in the liquid.

ture of the sphere whether or not it is possible to nucleate a stress-free crystal. Our simulations show no sign of spontaneous crystallization, regardless of the sphere's radius and attraction strength. This contrasts with the behavior at a planar wall or on a circular patch. Instead, we find that a monolayer adsorbs on the surface, and for strong attraction short-ranged ordering appears that optimizes the packing at the curved surface. For very small pores this packing extends to fill the entire pore, but it is incompatible with the crystal lattice and hence does not induce crystallization.

In a final step, we therefore consider a cylindrical pore embedded in an otherwise hard wall. This geometry combines two possible mechanisms to enhance nucleation: on the one hand its planar bottom resembles the planar circular patch and facilitates ordering. On the other hand, a cylindrical cavity, like the hemi-spherical pore, adsorbs potentially a high-density liquid phase. And deep cylinders with a depth larger than its radius, $D > R$, might sustain a liquid where a shallow hemi-spherical pore might have none. This difference is depicted schematically in Figure 6.4 showing both a hemi-spherical (a) and cylindrical pore (b). The cylinder can be filled, but its side walls can also introduce frustration for a growing crystal if the cylinder's diameter is incommensurable with the lattice spacing.

To properly simulate this geometry care has to be taken which distance is used for the potential, as artifacts can arise where the bottom meets the side walls. To minimize these effects we choose to take the distance to the closest surface $r = \min(r_{side}, r_{bottom})$, which can be either to the side wall r_{side} or to the cylinder bottom, r_{bottom}. In our simulations we find that pores with a small radius $R \leq 4$ never induce spontaneous nucleation within 10^6 MC cycles. In contrast, all pores with $R \geq 7$ crystallize, provided the interaction strength is sufficiently high ($\varepsilon \geq 2$ as in the case of the planar wall). A qualitatively different behavior is observed for a narrow region

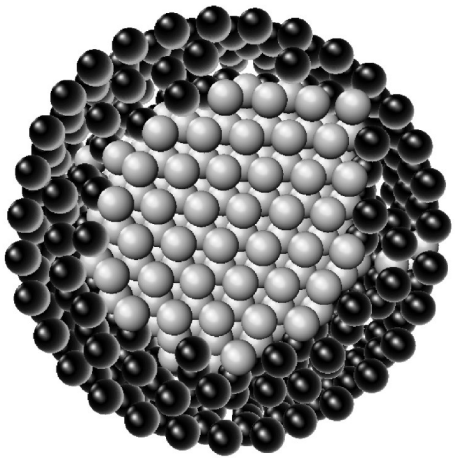


Figure 6.5: Snapshot of a cut through a smooth cylindrical pore with radius $R = 7$, depth $D = 10$, and attraction strength $\varepsilon = 5$, after $T = 2.1 \times 10^5$ MC time steps. The light-colored particles indicate the emerging crystal, dark-colored ones the layering along the cylinder's side walls.

of radii, $4.5 \leq R \leq 5.5$, in particular for deep pores. Here, we find that only a weak attraction leads to proper crystal growth, and visual inspection reveals that stronger attraction leads to a competition between crystal growth and attachment of particles on the cylinder's side walls. Particles at the cylinder's bottom eventually start to crystallize, while particles at the cylinder's side walls arrange in radially symmetric layers. These two packings are incompatible and the resulting frustration prevents the crystal to grow to the open end of the cylinder. Figure 6.5 depicts a cut through such a cylinder ($R = 7$, $D = 10$, after 2.1×10^5 MC cycles) showing the emerging crystal (light-colored particles) and the quasi-liquid layering at the cylinder's side walls (dark-colored particles). This behavior is typical for strongly adsorbing cylinders of all radii, even those showing crystallization. In the latter case, a structural rearrangement takes place which allows the crystal to eventually grow out of the cylinder into the bulk. Although this might also happen for smaller cylinders, we do not observe this in the course of our simulations.

6.2.2 Heterogeneous nucleation on a structured surface

On an atomistic scale all surfaces are rough. Hence, when we speak about "smooth" surfaces we refer to a situation where the surface roughness is small compared to the particle size. For colloids, it is possible to construct templates that are effectively smooth, but on the nanometer scale of proteins some atomistic roughness is

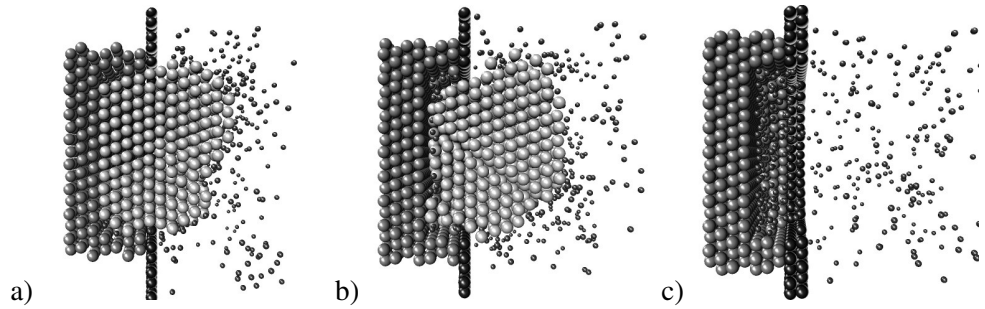


Figure 6.6: Cut through simulation-snapshots of a fluid in contact with a structured wall. The cylinders with radius $R = 7$ and depth $D = 3$ are cut from an fcc crystal with its 111 plane normal parallel to the x-axis, at densities $\rho = 0.8$ (a), $\rho = 1.0$ (b), and $\rho = 1.2$ (c). For both panel a) and b) the fcc density is close to the equilibrium density of the "protein" crystal, $\rho = 0.88$. In c) the lattice spacing of the substrate is significantly decreased ($\approx 10\%$), and the mismatch prevents spontaneous crystallization. Color coding: black: hard-sphere wall particles, dark-gray: attractive cylinder particles, light-gray: crystalline particles, small black: liquid and fluid particles.

inevitable. Therefore, it is more natural to consider structured templates for protein crystal nucleation. The simplest structured template is one that consists of particles that are immobile, but otherwise identical to those of the protein solution. In this work we consider both periodic and disordered arrangements.

Periodic structures

Earlier experimental and theoretical studies on colloidal and hard-sphere crystallization in the presence of structured surfaces report that templates are highly effective provided the crystal and substrate lattice spacing mismatch is not too large [123, 127, 141]. In what follows, we first consider a planar wall represented by the surface of a 111 fcc crystal phase at densities $\rho = [0.8, 0.88, 1.0, 1.2]$. The equilibrium density of the target protein crystal is $\rho = 0.88$. Direct simulations show that for all substrate densities but $\rho = 1.2$ a crystal forms spontaneously at the wall. At a density of $\rho = 1.2$ we observe a thin layer of a disordered high-density liquid phase. This contrasts with our observations for a smooth planar wall, where we find either a crystal or only a slight adsorption of the fluid phase. However, the formation of a thin liquid film that forms on the $\rho = 1.2$ substrate does not lead to subsequent spontaneous crystallization.

Next, we consider a structured circular patch. To construct such a patch, we start from a planar wall, but impose the condition that only those particles within a radial distance to the center of the patch, $(y^2 + z^2) \leq R_P^2$, interact via the modified

Lennard-Jones potential; all other particles experience only a hard-core interaction (i.e. excluded volume). As expected all samples with a patch radius $R \geq 5$ spontaneously crystallized, except for the patches with a fcc density of $\rho = 1.2$, for which a thin liquid film is observed. For small patches $R < 5$ a small crystal forms, but it is too small to spontaneously grow into the bulk.

Next, we created a structured cylindrical pore in a flat fcc substrate block by removing all particles located inside the cylinder. On top of the cylinder we added a monolayer of particles with hard-core interaction, such that 'protein' particles are attracted only to the cylinder walls and not to the planar outside surface. Varying the cylinder radius, depth, and the underlying fcc density had little effect on the overall behavior of the system. For all cylinders with $R \geq 5$ a crystal formed inside and grew out of the pore into the bulk. Only for a fcc density of $\rho = 1.2$ we did not observe spontaneous crystallization. Instead, a thin liquid layer wetted the cylinder's surface, and for cylinders up to $R = 5$ the liquid filled the pore. Figure 6.6 shows simulation snapshots for a cylinder with radius $R = 7$ and depth $D = 3$ for a fcc density of $\rho = 0.8$ (a), $\rho = 1.0$ (b), and $\rho = 1.2$ (c). For a slight mismatch between the lattice spacing λ of the surface and the equilibrium crystal a crystal spontaneously forms. But the snapshots indicate that the sign of mismatch, $\Delta\lambda = \lambda_{surface} - \lambda_{protein}$, influences the growth of the crystal. For a positive mismatch as depicted in Fig. 6.6a the crystal seems to grow several layers with the imposed density, after which the internal stress is released in defects causing the formation of several independent islands, a phenomenon known as Stranski-Krastanov growth [157]. For a negative mismatch, the stress inside the crystal is too large to be contained, and a defect plane emerges already in the first crystal layer. For even larger mismatches spontaneous crystallization is suppressed, and only a thin layer of liquid particles is adsorbed.

Disordered structures

The large effect of mismatch on the adsorption and crystallization behavior led us to investigate surfaces that are not only micro-porous but also disordered on the scale of the constituent particles. Such structures were obtained by preparing a hard-sphere fluid configuration at a density just below the freezing density. The density of the configuration was then rescaled to a density of $\rho = 1.2$ to match the density of the previous fcc surface. Note that after rescaling some of the particles do overlap, but since these particles are immobile and no interactions are computed between them, overlap has no effect on the protein particles.

In studying the effect of the disordered substrate, we again start with a planar wall cut from the bulk disordered phase at density $\rho = 1.2$. After 10^6 MC cycles of a grand-canonical Monte Carlo simulation we find no evidence for crystallization but, as in the case of the fcc surface, we observe the formation of a thin liquid layer.

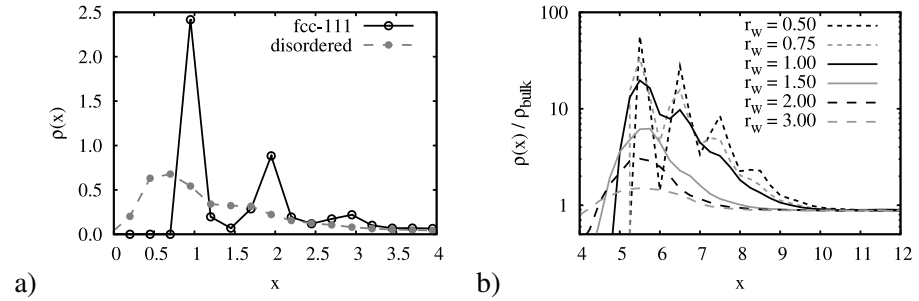


Figure 6.7: Density profile of the adsorbed liquid as function of distance x perpendicular to a wall. a) Density profile at a 111 fcc surface at density $\rho = 1.2$ (≈ 10 % lattice mismatch) and at a disordered wall with $r_w = r_p$. b) Density profile at a disordered wall with varying particle size r_w.

Figure 6.7a shows the computed density profile as function of distance perpendicular to the wall, both for the disordered and the fcc surface. The profile reveals that the liquid film on both surfaces is structurally very different. Whereas the disordered surface sustains a liquid with little short-ranged order, the periodic fcc lattice induces strong layering as revealed by the oscillating density profile.

To explore the effect of surface roughness on the amount and the local structure of the adsorbed liquid, we computed the density profile for various sizes of the wall particles (r_w). To achieve this size change we modified the interaction potential between surface particles and "protein" particles by shifting the r-axis, which ensures that the range of attraction is not affected. The resulting potential is then $U(r;r_w) = U[r-(r_w-r_p)]$, where r_p denotes the radius of the "protein" particles and $U(r)$ is given by Eqn. 6.1. For equally sized particles, $r_p = r_w$, the shift is zero. Note that we keep the *packing fraction* $\eta = (4\pi/3)r_w^3\rho$ of the wall constant (i.e. not the number density ρ) to correct for the volume increase of the wall particles. The resulting density profiles are presented in Figure 6.7b. We find that decreasing the particle size leads to an increase in density of the adsorbed fluid. Moreover, for wall particles smaller than the "protein" particles a layering in the liquid is induced, as is revealed by the oscillating density profile. For optimal crystal nucleation conditions the liquid should not impose any stress on the crystal, and hence layering should be avoided. At the same time, the liquid density should be as high as possible to reduce the free-energy barrier for crystal nucleation. Therefore, the 'optimal' roughness adsorbs a high-density liquid without a local structure (such as layering), and from Figure 6.7b this corresponds to the case where the wall particles are of the same size as the "protein" particles.

For homogeneous protein crystal nucleation there is evidence that large density fluctuations due to the presence of a meta-stable critical point lower significantly

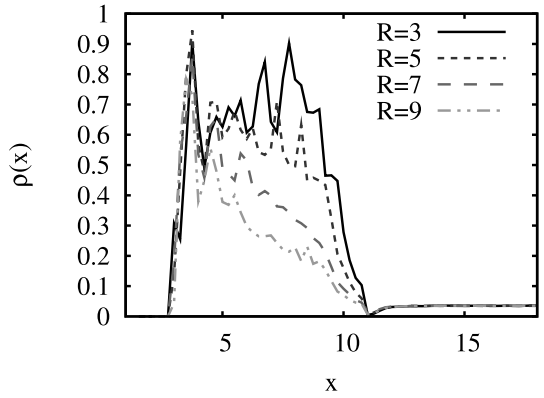


Figure 6.8: Density profile as function of x-position in a disordered cylindrical pore with radius $3 \leq R \leq 9$ and depth $D = 7$. The pore's bottom is at $x_{min} \approx 3$, and its top at $x_{max} \approx 10$. For small radii, $R \leq 5$, the pore is filled with a high-density phase, which shows some layering indicated by the oscillating density profile. For $R = 7$ the density is still partially filled, but the layering has disappeared. Further increasing the pore radius leads to a decrease in density indicating that the pore's surface is wet, but the pore is not filled anymore. The density is normalized by the liquid bulk density ($\rho_l = 0.66$).

the free-energy barrier for nucleation [10, 11]. It is therefore plausible that also a liquid adsorbed on a surface facilitates crystallization. To test this hypothesis we perform umbrella sampling simulations to compute the crystal nucleation free-energy barrier as described in Section 6.1. Indeed, we find that the presence of the liquid lowers significantly the free-energy barrier to nucleation, from $\Delta G^*_{homo} \approx 80 k_B T$ for homogeneous nucleation to $\Delta G^*_{wall} \approx 50 k_B T$ at the disordered surface.

Since we observe a liquid wetting the disordered wall, we check whether this is also the case for a cylindrical cavity. The disordered pore is constructed analogously to the fcc pore. Scanning a broad range of parameters we find an interesting behavior. The results are presented in Figure 6.8, where we plot the average liquid density along the cylinder axis for pores of different radius. Thin cylinders with $R < 7$ exhibit the highest liquid density, but the oscillating profile indicates strong layering, which is likely to intervene with the crystal formation. For radii around $R = 7$ the density is still very high, but the oscillations have disappeared, indicating that the structure of the liquid is not strongly influenced by the walls. For large cylinders $R > 7$ the density decreases, and visual inspection of our simulation snapshots confirms that only the surface of the pores are wetted, but the pore does not fill with a dense liquid phase.

From the density analysis we conclude that a deep cylinder with radius $R = 7$, filled with a high-density isotropic liquid, should provide a good crystal nucleation

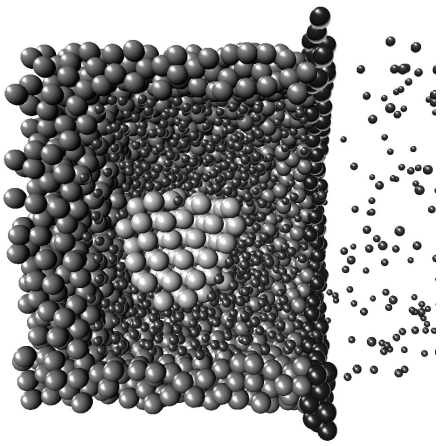


Figure 6.9: Cut through a snapshot of a direct simulation of a fluid in contact with a structured wall. The cylinder with radius $R = 7$ and depth $D = 7$ is cut from an disordered phase (see main text for details) at a density $\rho = 1.2$. A crystal forms inside the high-density liquid which partially fills the pore. Color coding: black: hard-sphere wall particles, dark-gray: attractive cylinder particles, light-gray: crystalline particles, small black: liquid and fluid particles.

site. Indeed, we find that the cylindrical cavity lowers the free-energy barrier height further from $\Delta G^*_{wall} \approx 50 k_B T$ for a disordered surface to $\Delta G^*_{pore} \approx 25 k_B T$ for a cylindrical pore. But interestingly, we observe two different initial pathways in our umbrella simulations. In both pathways a crystal nucleus forms inside the cylinder, but in one path the crystal forms at the wall and in the other away from it, fully immersed in the liquid phase. In our simulations using umbrella sampling both routes can be followed initially, but once the cluster exceeds 50 particles, further growth of the cluster at the wall turns out to be extremely difficult. It seems that strong internal stresses prevent further growth, and crystallization is inhibited. In contrast, the route that starts with a fully immersed crystal leads to a low barrier and to well-ordered crystals. The latter pathway is schematically depicted in Fig. 6.4c, and a corresponding simulation snapshot is presented in Fig. 6.9 showing an emerging crystal that is fully immersed in the liquid, i.e. away from the cylinder walls.

6.3 Discussion

For the growth of high-quality protein crystals that are suitable for protein structure analysis it is essential to work at a low supersaturation. Although a higher supersaturation enhances the crystal nucleation rate, it also affects the crystal growth speed and therefore leads to the formation of defect-rich crystals or even a poly-crystalline ma-

terial. However, an enhancement of the rate of crystal nucleation is necessary to obtain crystals within experimental time-scales, and heterogeneous nucleation promises an alternative route: a nucleation agent can induce crystallization at a supersaturation where homogeneous crystal nucleation in the bulk is negligible and the crystal growth speed is slow. From a practical perspective, such a nucleation agent should be "universal" in the sense that it should be effective for a broad range of proteins. Moreover, to prevent the formation of a poly-crystalline material, it also should facilitate the formation of only a few well-separated crystal nuclei. But, although the requirements of a good nucleation agent are well-defined, for its implementation a detailed understanding of the molecular mechanism is still lacking. The aim of the simulations that we presented in this chapter was to arrive at a set of criteria for the design of such a nucleation agent.

We started with the simulation of crystal nucleation on a planar surface, both smooth and periodically structured. We find that a smooth surface induces spontaneous crystallization for all but the weakest adsorption. However, such an aspecific behavior is less likely to lead to high-quality single crystals than to a poly-crystalline material, since many crystal nuclei form simultaneously at all over the surface. Reducing the adsorbing area to a planar circular patch might solve this problem. However, we are not aware of an experimental study on such nucleation agents. The limited success of planar adsorbing surfaces indicates that this mechanism depends strongly on how a protein adsorbs to a surface, and hence on the specific interaction between a protein and the surface. For a periodically structured surface we observe spontaneous crystallization provided the lattice mismatch between the template and the crystal is not too large. The size of such a template could be tailored to sustain exactly one nucleus preventing poly-crystallinity. However, such a design is not very universal, because the optimal lattice spacing varies strongly between different proteins. Nucleation agents that do seem effective in experiment for a broad range of proteins are micro-porous materials. And although both theory and simulation confirmed the effectiveness of a pore as nucleation site, our simulation results show that a smooth adsorbing cavity is not enough to facilitate crystal nucleation for our model protein. This observation motivated us to re-think the nucleation pathway.

In all simulations that showed spontaneous crystal nucleation the crystal formed directly in contact with the surface. Such a route depends on the details of how proteins adsorb to a surface, which are not captured by our model. But, in order to qualify for a "universal" a nucleation agent, we argued that the underlying mechanism should be generic, i.e. protein-independent. Based on our observation that a periodically structured surface with a strong lattice mismatch adsorbs a thin film of a high-density disordered protein phase (a "liquid"), we consider an alternative pathway for heterogeneous protein crystal nucleation, one that proceeds via an intermediate liquid phase. Such a pathway was already proposed for homogeneous

protein nucleation in the vicinity of a meta-stable critical point [10, 11], where the strong density fluctuations lower the free-energy barrier for nucleation. The heterogeneous nucleation analog would then require a nucleation agent to adsorb a liquid rather than a crystal. Following this line of thought, we investigated crystal nucleation on rough disordered surfaces. We find that the surface that best adsorbs a liquid, i.e. the highest amount of liquid with a maximum structural isotropy, has a surface roughness on the scale of the protein particles. This observation agrees well with experimental results on the effect of surface roughness on the protein crystallization behavior [143]. Furthermore, our simulations show that a deep cylindrical pore with a rough disordered surface fills with such a liquid provided it is of suitable size. In such a pore the crystal nucleus can grow fully immersed in the liquid, i.e. away from the surface and stress-free, which lowers significantly the free-energy barrier for crystal nucleation. Therefore, our results provide an rationalization for the experimental results on micro-porous media of Chayen and co-workers [132, 133] that goes beyond the effect of the pore size on the nucleation free-energy barrier [18, 103, 142], as it stresses the pronounced effect that surface roughness has on the behavior of protein crystallization.

In conclusion, our numerical simulations provide a molecular mechanism for protein crystal nucleation in micro-porous media that relies on the combined effect of liquid adsorption and surface roughness. The proposed nucleation pathway can be tested experimentally with colloidal systems, which allow good control of the interparticle interactions. Disordered substrates can be fabricated by sedimentation of poly-disperse colloids with a roughness controlled by the particle size distribution, and it should be possible to create cylindrical cavities of various sizes. Moreover, the particles that form the pore's inside could be functionalized differently from the particles at the outer surface, such that adsorption takes place only inside the cylindrical cavities. Finally, the pore size distribution should be adjusted such that the optimal pore size, which supports the formation of a liquid, should be located in the tail of the distribution ensuring that only a few pores contribute actively as nucleation sites. However, the fabrication of nucleation agents for protein crystallization requires further experiments, since the surface properties need to be controlled on the nanometer scale.

Acknowledgements

We thank S. Abeln, A. Vila Verde, and K. Shundyak for fruitful discussions, and D. Zwicker for a careful reading of the manuscript.

Chapter 7

Geometrical Frustration: A Study of $4d$ Hard Spheres

with B. Charbonneau and P. Charbonneau

The smallest maximum kissing-number Voronoi polyhedron of 3d spheres is the icosahedron and the tetrahedron is the smallest volume that can show up in Delaunay tessellation. No periodic lattice is consistent with either and hence these dense packings are geometrically frustrated. Because icosahedra can be assembled from almost perfect tetrahedra, the terms "icosahedral" and "polytetrahedral" packing are often used interchangeably, which leaves the true origin of geometric frustration unclear. Here we report a computational study of freezing of 4d hard spheres, where the densest Voronoi cluster is compatible with the symmetry of the densest crystal, while polytetrahedral order is not. We observe that, under otherwise comparable conditions, crystal nucleation in 4d is less facile than in 3d. This suggest that it is the geometrical frustration of polytetrahedral structures that inhibits crystallization.

Most glasses form under conditions where the thermodynamically stable state of the system is crystalline. Good glass formers should therefore be poor crystallizers. Geometrical frustration is one of the factors that may prevent the formation of the ordered phase and therefore help physical glass formation [19]. There is also evidence that such frustration increases the height of the crystallization-nucleation barrier of liquid metals [158]. Isotropic simple liquids are often considered frustrated because the five-fold symmetry of the liquid icosahedron cannot pack as a regular lattice. This scenario contrasts with what happens in a fluid of 2d disks, where hexagonal order is both locally and globally preferred and where crystallization is particularly easy.

Several physical mechanisms have been proposed to support the formation of icosahedra. On the one hand, Frank, considering the optimal way for kissing spheres to cluster around a central one, found the icosahedron to be more stable than the cubic lattice unit cells for the Lennard-Jones model [20]. Though the original argument relies on the energetics of spurious surface effects [159], mean-field studies correcting for solvation leave the result unchanged [160, 161]. The icosahedron, the smallest maximum kissing-number Voronoi polyhedron, is optimally packed. It offers the most free volume to surface spheres, so it is also preferred entropically. On the other hand, the polytetrahedral scenario ascribes the presence of icosahedra to their facile assembly from quasi-perfect tetrahedra, themselves the smallest Delaunay decomposition of space [21, 22]. But is it the packing of Voronoi polyhedra or the packing of Delaunay hyper-triangles that counts? Experiments [162–164] and simulations [165, 166] only manage to identify icosahedral order in limited quantities, even in deeply supercooled systems. Recent studies indicate that liquid polytetrahedral order is a lot more varied [167, 168] than the icosahedral picture suggests. Yet, because of the geometrical ambiguity, the equation of the icosahedron with frustration is difficult to asses.

Looking at crystallization in a system where polytetrahedral frustration does not correspond to a symmetric closed-shell structure like the icosahedron would help. Precisely such an example is provided by the freezing of 4d spheres that we study in this Chapter. It is, of course, somewhat unsatisfactory to perform a numerical study of a system that cannot be probed experimentally. However, there are other examples (e.g. renormalization-group theory) where higher-dimensional model systems serve as a very useful reference state for the theoretical description of our 3d world. The objective of the numerical study that we report here is therefore not to present quantitative estimates of crystal nucleation barriers in 4d (even though we obtain these numbers too), but to shed more light on the nature and role of geometrical frustration and the ease of crystallization.

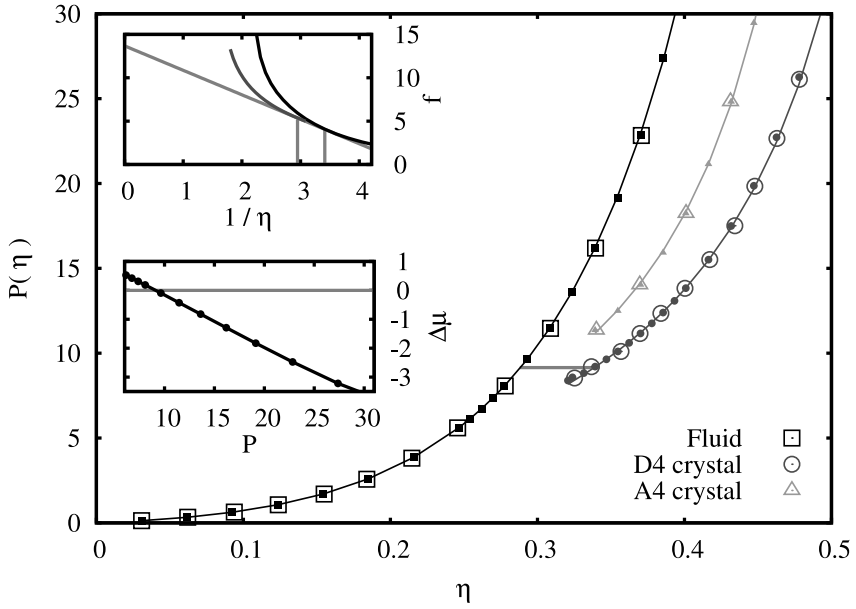


Figure 7.1: Equations of state of 4d hard spheres at constant V (closed) and P (open) extend earlier molecular dynamics results [169]. Padé approximant for the fluid [170] and Speedy fits for the crystals [171] are given for reference (lines). Bottom inset: at coexistence chemical potentials are equal, thus $P_{\text{coex}} = 9.15$ and $\mu_{\text{coex}} = 13.68$. Top inset: the common tangent to the free energy curves pinpoints the phase transition boundaries: $\eta_{\text{freeze}} = 0.288$ and $\eta_{\text{melt}} = 0.337$.

7.1 Phase diagram of 4d hard spheres

The D_4 crystal phase is formed by stacking, without voids, 24-cell Platonic polytopes [172, 173]. In general, D_d lattices are obtained by inserting an additional sphere in each void of a d-dimensional cubic lattice. In 3d the spacing between the spheres on the original cubic lattice increases to form a body-centered-cubic crystal; in 4d the additional sphere fits perfectly in the hole and leads to a unique, high symmetry crystal with maximal volume fraction $\eta = \pi^2/16 \approx 0.617$. There exist other dense 4d lattices, such as A_4 and A_4^*, but D_4 packs over 10% more densely and offers more nearest-neighbor contacts. D_4's unit cell, the 24-cell, is made of 24 octahedral cells and is a Platonic polytope that has no analog in other dimensions [173]. Placing 24 kissing spheres around a central one in the 24-cell arrangement is the densest closed-shell cluster of 4d spheres [174] and is postulated to be unique [175]. Even accounting for solvation effects, clusters with the 24-cell geometry are locally preferred. Unlike in 3d, for an equal number of particles 4d polytetrahedral clusters

do not form more interparticle contacts than the 24-cell, and their slightly larger radius offers less, not more stabilization [160]. The symmetry match between the 24-cell and the D_4 lattice therefore guarantees that no frustration arises from maximally kissing clusters. But neither the 24-cell nor any other unit cell can be assembled from (nearly) regular $4d$ tetrahedra. Four-dimensional spheres are thus an ideal system to clarify the origin of geometrical frustration. An earlier compaction study of $4d$ spheres indirectly hinted that spontaneous crystallization might be slow [176], but this work could not disentangle the different contributing factors, because such an analysis requires knowledge of the equilibrium phase diagram, of the dynamical properties of the fluid phase, and of the crystal nucleation barriers. Our computational study addresses these questions. To this end, we first locate the $4d$ freezing transition, quantify the fluid order, and then compute the free energy barrier to nucleation at different supersaturations.

Interestingly, although the equations of state of both the fluid and the crystal phases of $4d$ hard spheres were computed in the early 80's [169], we are not aware of any numerical determinations of the solid-fluid coexistence point. Using a quasi-Maxwell construction [177] at the crystal stability limit [176], we can use these results to approximate the coexistence range $\eta_{\text{coex}} = 0.29 - 0.35$, but this is insufficiently accurate. To the best of our knowledge, density functional theory has only been applied to the fluid-A_4 coexistence [178]. In order to precisely locate the freezing point, we thus performed standard NVT-Monte Carlo (MC) simulations to compute the equation of state of hard spheres, outside the range studied in Ref. [169]. As a test, we also performed constant NPT simulations and verified that the two techniques yielded consistent results. In what follows, we use the particle diameter σ as our unit of length and the thermal energy k_BT as our unit of energy. The equation of state of $4d$ spheres is related to the value of the pair-distribution function $g(r)$ at contact $Pv_0/\eta = 1 + 8\eta g(1^+)$, where v_0 is the volume of a $4d$ sphere and $g(1^+)$ is the value of the radial distribution function at contact [170]. The results for the fluid and two crystal phases are presented in Fig. 7.1 for systems containing 2048 (D_4) and 4096 (fluid and A_4) particles. The equation of state could not be calculated for A_4^*, because it is mechanically unstable, which makes it unlikely to contribute to the crystallization process. We won't consider it further. To locate the fluid-solid coexistence regime, we need to determine the absolute free energy of the solid at least at one point [92]. The absolute Helmholtz free energy per particle f of the D_4 and A_4 crystals at $\eta = 0.37$ is obtained by the Einstein-crystal method [99]. The free energy at other densities can then be obtained by thermodynamic integration. We find D_4 to be the thermodynamically stable crystal phase. The fluid-D_4 coexistence pressure P_{coex} (Fig. 7.1 lower inset), allows to read off the melting and freezing densities by common tangent construction (Fig. 7.1 higher inset). The resulting two-phase region $\eta_{\text{coex}} = 0.288 - 0.337$ is compatible with the rough estimate above. The thermody-

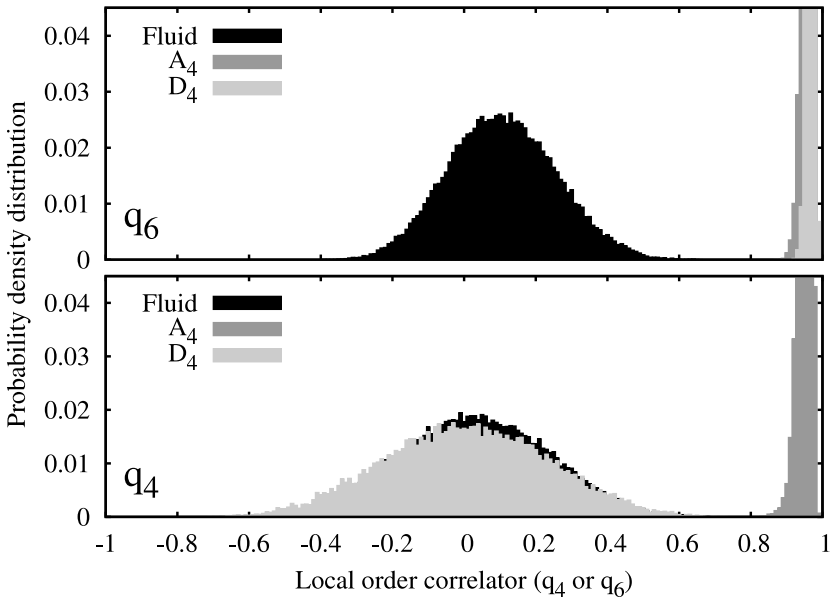


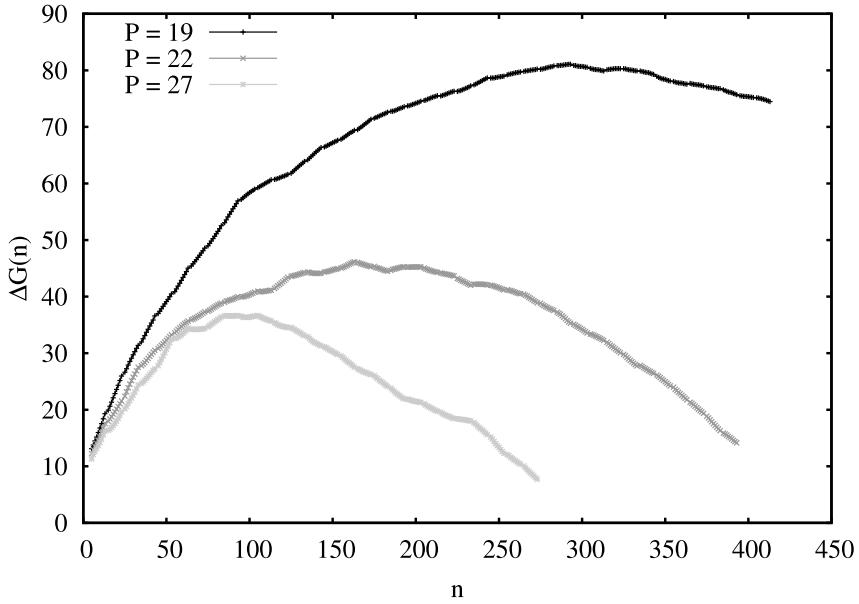
Figure 7.2: Distribution of the local order correlator with $l = 4$ (top) and $l = 6$ (bottom) at $P = 19$.

namic driving force for crystallization in the supersaturated fluid at constant pressure is the difference in chemical potential $\Delta\mu = \mu_{D_4} - \mu_{\text{fluid}}$ between the two phases displayed in the lower inset of Fig. 7.1.

7.2 Crystallization free-energy barrier in $4d$

To characterize the structure of the fluid and identify the formation of crystallites, we need a local criterion that distinguishes crystal from fluid. Studies in $2d$ and $3d$ suggest that order parameters derived from invariant combinations of spherical harmonics Y_l of degree l might suffice [98, 179]. In high dimensions, it is more convenient to rewrite the second-order invariant in terms of Gegenbauer polynomials G_l^n, where $n = d/2-1$, using the sum rule [175]. The $(l+1)^2$ $4d$ spherical harmonics give

$$G_l^1(\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2) = \frac{2\pi^2}{(l+1)^2} \sum_{m=1}^{(l+1)^2} Y_l^m(\hat{\mathbf{r}}_1)\overline{Y_l^m(\hat{\mathbf{r}}_2)}, \tag{7.1}$$



P	$\Delta\mu$	ΔG^*	γ_{CNT}	n^*	n^*_{CNT}
19	-1.8	81	1.80	157	133
22	-2.3	42	1.94	75	60
27	-3.2	37	2.4	40	35

Figure 7.3: Free-energy barriers for 4096 particles at various supersaturations, along with the corresponding CNT and simulation parameters. The critical cluster size n^* is obtained using a stricter order parameter (see text).

where $\hat{\mathbf{r}}_i$ are unit vectors. The local order correlator is

$$q_l^{i,j} = \mathbf{q}_l(i) \cdot \mathbf{q}_l(j) = \frac{1}{N(i)N(j)} \sum_{\alpha=1}^{N(i)} \sum_{\beta=1}^{N(j)} G_l^1(\hat{\mathbf{r}}_{\alpha i} \cdot \hat{\mathbf{r}}_{\beta j}), \tag{7.2}$$

where the indices α and β run over the number of neighbors contained within a distance equal to the first minimum of $g(r)$. The local order correlation distinguishes between different geometrical environments: q_6 set apart fluidlike particles from those within a D_4 or an A_4 lattice, while q_4 discriminates between the two crystals (Fig. 7.2).

As freezing in $4d$ is a first-order phase transition, we expect crystallization to proceed via nucleation and growth. A Landau free energy analysis predicts that crystals with reciprocal lattice vectors forming equilateral triangles should initiate the nucleation [180]. Though this argument has met only limited success in $3d$ [98], in $4d$

it supports the preferential nucleation of D_4, in line with the thermodynamic drive. To estimate the ease of crystallization, we compute the free energy barrier for crystal nucleation ΔG^*. Classical nucleation theory (CNT) [4] derives from the thermodynamic drive $\Delta\mu$ and the interfacial free energy γ of a spherical crystallite a free energy functional that depends on the size n of the crystallite

$$\Delta G(n) = S_d(n/\rho_x)^{(d-1)/d}\gamma - n\Delta\mu, \tag{7.3}$$

where ρ_x is the crystal density at a given pressure and the shape-dependent prefactor is $S_4 = (128\pi^2)^{1/4}$ for $4d$ spheres. The resulting maximal barrier height is then

$$\Delta G^*(n^*) = \frac{27\pi^2\gamma^4}{2\rho_{D_4}^3\Delta\mu^3} \tag{7.4}$$

at the critical cluster size n^*. The rate of nucleation per unit volume I is given by $I = \kappa\exp(-\Delta G^*)$, where κ is a kinetic prefactor that is proportional to the diffusion coefficient in the fluid phase [98]. Though schematic this level of theory is sufficient to clarify the contribution of geometrical frustration through an analysis of γ. Within the CNT framework the geometrical mismatch in $3d$ between icosahedral and crystal order should lead to a relatively large γ, while in $4d$ one might expect γ to be small if the locally preferred cluster scenario is valid, but not for polytetrahedral frustration.

Results for $3d$ crystallization are available [98], so only a few $4d$ barriers are needed to complete the picture. Crystallization being a rare event in this regime, we perform constant-pressure MC runs with umbrella sampling to bias the growth of a crystal cluster from the fluid [92]. A standard algorithm is employed to identify the crystallites [98, 179]. We *link* pairs of nearest neighbors with $q_6 > 0.4$. If a particle has more than five links it is deemed crystalline. The number of particles in the largest crystallite is then the order parameter. The resulting free energy profiles are presented in Fig. 7.3. Though q_6 does not discriminate between A_4 and D_4 crystals, further checks with q_4 show that only the latter nucleates. In $4d$ a low q_6 cutoff value is required, because of the minimal overlap between the crystalline and fluid regions (Fig. 7.2), and consequently, non-compact clusters are initially observed. Though the clusters irreversibly compactify, the process can be very slow. To reduce the computational burden, the system is first equilibrated by growing the total number of *links* in the largest crystallite. The low q_6 cutoff also artificially inflates the measured critical cluster size. A fit to the CNT functional form (Eq. 7.3) is thus of little use in extracting γ. However, because the barrier height is unaffected by this biasing choice and $\Delta\mu$ is known, we can obtain γ from Eq. 7.4 directly. To validate the implied size of the CNT critical cluster n^*_{CNT} we compare it to the cluster size obtained by imposing a purely crystalline linking criterion $q_6 > 0.65$ to the configurations at the top of the barrier. The difference between the two (Fig. 7.3) is no more than 25%, which is remarkably good in this context.

The results of Fig. 7.3 allow us to conclude that the very slow crystallization of $4d$ spheres observed in the study of Ref. [176] is due to the presence of a considerably higher $4d$ nucleation barrier than at the same supersaturation in $3d$. Slow nucleation could also be due to a low value of the kinetic prefactor κ, which would require that the diffusion of particles in the dense fluid be anomalously slow. But simulations with the code of Ref. [176] show no evidence for slow diffusion, not even at the highest pressures studied. The slow crystallization is thus consistent with a high degree of geometrical frustration in $4d$ fluids. Based on the similarity between the number of neighbors within the first peak of $g(r)$ and the maximal kissing number Skoge *et al.* speculated that high dimensional fluids contain a number of deformed crystalline unit cells, rather than polytetrahedral structures [176]. However, the clear difference between the fluid and the 24-cell shown by the local order correlator (Fig 7.2) suggests this not to be the case. The similarity between the kissing number and the number of first neighbors can instead be explained by a wide first peak of $g(r)$ (not shown) that accommodates non-kissing neighbors in polytetrahedral clusters. Because the "locally preferred" 24-cell has little to do with geometrical frustration, our results support the generic polytetrahedral structures as the source of frustration. By dimensional analogy, we infer that the "locally preferred" icosahedron is not singular, but instead one of the many possible geometrically frustrating structures, and explains its limited presence in fluids. The dimensionless surface free-energy density $\gamma\sigma^{d-1}/(k_BT)$ is at least two to three times larger in $4d$ than in $3d$, which indicates that geometrical frustration is surprisingly rather weak in $3d$. It is this weakness that helps make hard sphere crystallization so prevalent. The interesting puzzle is therefore not to identify the origin of $3d$ frustration, but the source of its mildness. One possibility is that the tetrahedra that are part of the face-centered cubic (fcc) structure (none are found in D_4) relax the geometrical frustration and therefore reduce the interfacial tension. Another possibility is that the "planetary perturbations" that allow to exchange the positions of spheres at the surface of an icosahedron by sliding, go through a cuboctahedron configuration, which is the fcc unit cell [173]. If common, this phenomenon would imply that not all polytetrahedral structures are equally frustrating and that icosahedra might in fact be early nucleation sites.

The large values for the height of the nucleation barrier of $4d$ crystals, as well as the evidence (Fig. 7.2) that the local structures in the fluid and the D_4 crystal are rather different, indicate that it is the Delaunay packing that matters. This finding underlines that one should be rather careful in caricaturing the nature of frustration as icosahedral in $3d$ liquids. Icosahedra are but one of the many possible polytetrahedral arrangements and little indicates that it plays a more prominent role in geometrical frustration than others. Note that the difficulty to crystallize $4d$ spheres makes them, as well as their higher dimensional equivalents, promising testing grounds for theories of packing and glass-forming liquids.

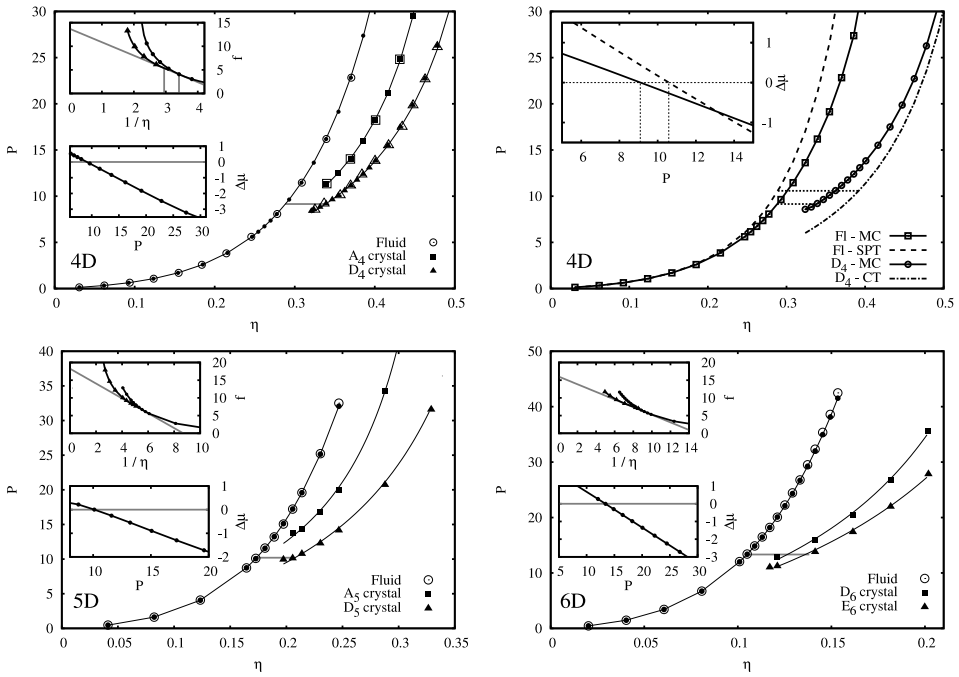


Figure 7.4: Monte Carlo EoS of the fluid and the two densest known ordered phases in 4D [181], 5D, and 6D computed in the canonical ensemble (points) with Padé approximants of the virial expansion for the fluid [170] and the Speedy fits to the crystal phase results [171] (solid lines). Insets give $\Delta\mu$ and the common tangent construction for determining coexistence between the fluid and the densest crystal phase. The top-right panel contrasts 4D MC and SPT/CT EoS as well as chemical potential (inset).

7.A Higher-dimensional phase diagrams and jamming

In the next two sections of this appendix we follow the line of thought from the previous work on four-dimensional hard spheres to extend the discussion to 5D and 6D hard-sphere systems, where we put some theoretical predictions to the test. The fluid equation of state (EoS) reported in Figure 7.4 agrees with earlier 4D, 5D [169, 170, 183, 184], and 6D [185] simulation results as well as 5-4 Padé approximants of the virial expansion [170, 186]. Small deviations are only observed at the highest densities [170, 185]. Crystal phase EoS for D_4 and D_5 were first obtained from simulation in the early 80's, but without reference free energies [169, 176], and we are not aware of any 6D simulation results. As expected from free volume arguments, the densest known lattice is the phase with the lowest pressure, at densities where it is mechanically stable, which makes it the most free energetically favorable of

d	P^{coex}	μ^{coex}	$\eta_{\text{f}} - \eta_{\text{x}}$	$\eta_{\text{f}} - \eta_{\text{x}}$ mod. Ref. [176]	$\eta_{\text{f}} - \eta_{\text{x}}$ SPT/CT [182]	η_{cp}
3 [92]	11.564	17.071	$0.494 - 0.545$	-	-	0.741
4 [181]	9.15	13.7	$0.288 - 0.337$	$0.29 - 0.35$	$0.29 - 0.39$	0.617
5	10.2	14.6	$0.174 - 0.206$	$0.18 - 0.22$	$0.27 - 0.35$	0.465
6	13.3	16.0	$0.105 - 0.138$	-	$0.20 - 0.25$	0.373

Table 7.1: Coexistence parameters from Monte Carlo simulations compared with previous simulation estimates (see text) and SPT/CT results from Ref. [182] (4D is a new calculation). The volume fraction of the densest known lattice η_{cp} is also included for reference [173, Chap 1. § 1.5]

the ordered phases. Crystallization to lattice geometries other than the most stable structure would thus only be possible at high pressures. In 4D a Landau free energy treatment kinetically favors the D_4 ground state [180]. But even in 5D and 6D we do not expect the difference in fluid-crystal interfacial free energy to affect the nucleation barrier sufficiently for other phases to kinetically interfere with the crystallization of the thermodynamic crystal form [187].

The coexistence conditions obtained from the MC EoS and free energy reference point for 5D and 6D hard spheres along with the 3D [92] and 4D [181] data can be seen in Table 7.1. Skoge et al. offered upper bounds to the 4D and 5D coexistence regimes by using the pressure of the fluid at the density at which the simulated D_d crystal becomes mechanically unstable as an estimate of coexistence pressure P^{coex} [176]. A more accurate estimate of P^{coex} can be obtained from the same data by using instead a quasi-Maxwell construction around the limit of mechanical stability [177]. We include the results of this last analysis and the coupled fluid scaled-particle theory (SPT) and crystal cell theory (CT) coexistence determination of Finken et al. [182] in Table 7.1 as well. To the best of our knowledge, density functional theory (DFT) coexistence information have only been reported for the non-equilibrium 4D fluid-A_4 coexistence [178], which does not lend itself to a meaningful comparison. Finken et al. refer to their DFT coexistence calculations, but do not report them [182].

Our MC results are at least an order or magnitude more precise than the estimates from Ref. [176], which permits a clearer assessment of the SPT/CT predictions. As can be seen in Table 7.1, the coupled theory overshoots the coexistence regime, and does so more in higher dimensions. Yet SPT reportedly overestimates the third-order virial coefficient of the fluid, which should instead result in a depressed coexistence regime and become less relevant with increasing dimensionality [182]. This discrepancy suggests that CT significantly underestimates the crystal pressure near coexistence, as can be checked in Fig. 7.4 (b) for 4D. The high compressibility of hard

sphere crystals near the limit of mechanical stability is a collective effect that is not captured by the mean-field nature of CT. SPT/CT thus leads to a higher coexistence pressure and crystal coexistence density. The fluid coexistence density barely differs from the MC results in 4D, but in higher dimensions the crystal branch gets longer than the fluid branch, which introduces a larger error in P^{coex}. Consequently, the coexistence fluid density shifts up and the theoretical treatment grows progressively worse with dimension. A cancelation of errors is therefore responsible for the SPT/CT predictions to remain reasonably on target for hard spheres in low dimensions. SPT/CT nonetheless predicts certain dimensional trends correctly. For instance, the relative width of the coexistence regime $\Delta\eta_{\text{coex}}/\eta_{\text{x}}$, which is thought to go to unity for large dimensions, does increase appreciably from below 10% in 3D to over 20% in 6D. Also, the crystal volume fraction at coexistence η_{x} decreases much faster than the close-packed volume fraction η_{cp}, which leaves the phase diagram dominated by the ordered phase. In dimensions much higher than those we can access here, it has however been suggested that the fluid packs more efficiently than any crystals [188]. If it is indeed the case, this last trend should eventually reverse.

It is interesting to note that P^{coex} does not change monotonically with density, but goes through a minimum in 4D. The nonmonotonic behavior of P^{coex} might be due to D_4's particularly well-suited nature to fill 4D Euclidian space. A D_4 lattice can be generated by placing a sphere in each of the voids of a 4D simple cubic lattice. These new spheres are equidistant to the ones on the simple cubic frame, so the resulting lattice is twice denser. The corresponding construction D_3^*, or body centered-cubic lattice, in 3D packs much less efficiently, because the simple cubic frame needs to be extended to insert the new spheres. Though D_4 does not appear as a singularity in the dimensional trend of dense packings [173, Chap. 1, §1.5], its specificity might simply be overshadowed by other dimensional trends to which P^{coex} is less sensitive. We therefore conjecture that the non-monotonic coexistence pressure is a symmetry signature that should also be present in 8D, 12D, 16D, and 24D, where other singularly dense lattices are known to exist.

Another noteworthy observation concerns the high pressure fluid limit. The results presented in this work are for equilibrium or metastable equilibrium systems only. But extrapolating to infinite pressures from the densest supersaturated fluid using the free-volume functional form suggested in Ref. [189], allows us to extract the density of the corresponding jammed system. It is not the maximally-dense random jammed (MRJ) density, but it should asymptotically approach it, as suggested in Ref. [190]. We obtain $\eta_{\text{MRJ}} \approx 0.47$, 0.31, and 0.21, in 4D, 5D, and 6D respectively, which is in very good agreement with the results from the direct non-equilibrium compression [176] and a similar extrapolation in Ref. [190].

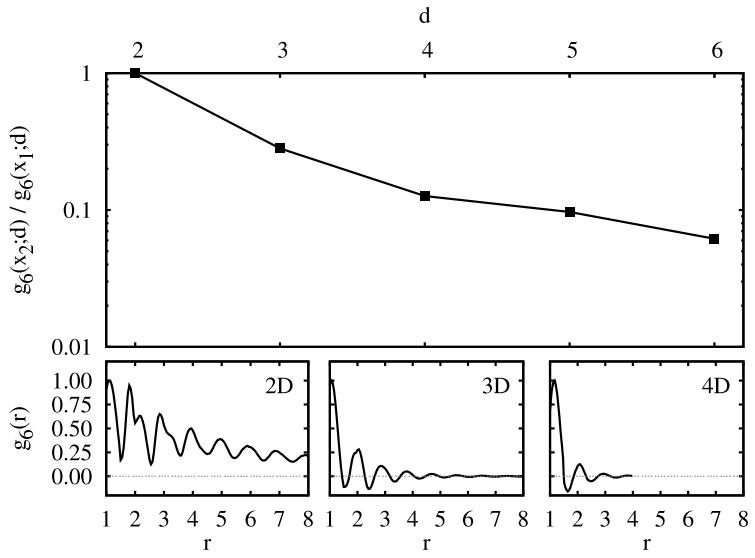


Figure 7.5: Impact of dimensionality on the decay or orientational order. Top: Second to first peak ratio of the radial decay of the orientational order parameter $g_6(r)$ at fluid coexistence. The line is a guide to the eye. Bottom: Decay of $g_6(r)$ in the fluid near the hexatic phase in 2D and at coexistence in 3D and 4D.

7.B Higher-dimensional bond-order correlators

Skoge et al., considering the radial pair distribution function $g(r)$, found that higher-order unconstrained spatial correlations vanish with increasing dimensionality [176]. In accordance to the "decorrelation principle" [188], we also expect orientational correlations of order l $g_l(r)$ to decay more rapidly with increasing dimensionality. As seen in Figure 7.5, the 2D hexatic signature gives rise to long-ranged orientational correlations that diverges on approaching coexistence [191, 192]; in 3D the orientational order stretches over a couple of particle radii, but stays finite even in the supersaturated regime; and in higher dimensions the correlations are very small and decay with dimension, as can be assessed from the second to first peak ratio of $g_6(r)$ in Fig. 7.5.

The authors of Ref. [176] also remarked that the number of particles counted in the first peak of $g(r)$ for supersaturated fluids matches the number of kissing neighbors in the densest known lattice phase for a given dimension. They hypothesized that disordered packings in higher dimension might thus be built of deformed crystal unit cells, in contrast to the three-dimensional case where "icosahedral" order dominates the packing. The distributions of local bond-order correlators, which shows how the relative crystal and fluid local orders evolve with dimensionality paint a different pic-

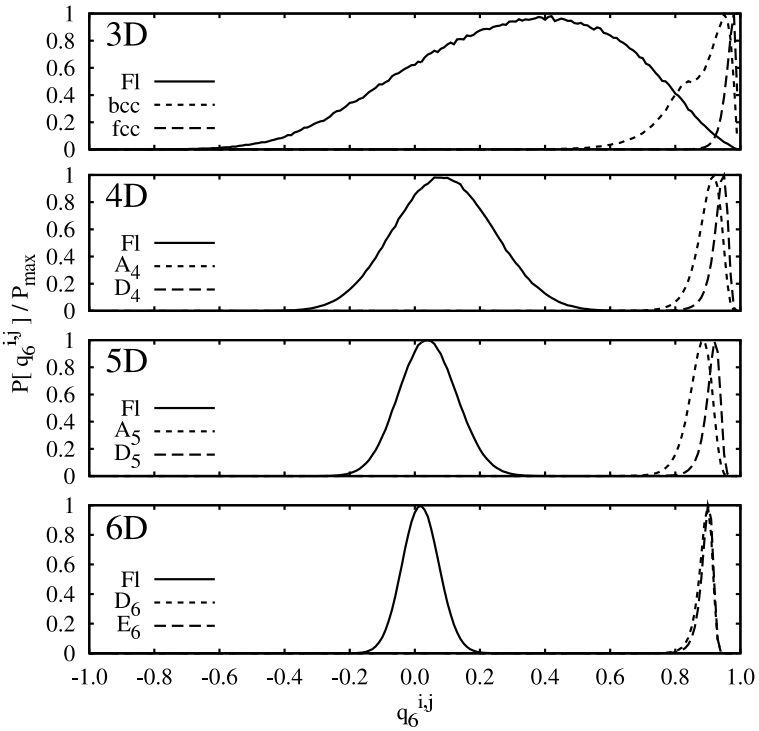


Figure 7.6: Second-order $l = 6$ invariants for dense crystal phases and the fluid. With increasing dimension fluid and crystalline local bond-order parameters become increasingly distinct.

ture (see Fig. 7.6 and Fig. 7.7). Both the second- and third-order invariants in 4D to 6D, capture no significant overlap between the liquid and crystal local bond-order parameters, which contrasts to 2D and 3D where the bond-order distributions overlap significantly [193]. Moreover, the distinction between the fluid and crystal phases increases with dimension, which suggest that the fluid and crystal local orders are just as or more distinct with increasing dimensionality, not less [1]. In 4D, where the maximally-kissing cluster 24-cell [174] is also the unit cell of the crystal, but is not a simplex-based structure, no hint of the presence of 24-cell in the fluid is captured by the bond-order distribution (Fig. 7.6 and Fig. 7.7) [181]. For a simplex-based cluster to have as many nearest neighbors around a central sphere as in the crystal, the first neighbor spheres cannot all be kissing the central sphere at the same time. They have to fluctuate in and out of the surface of the central sphere. This variety of possible

[1] In the trivial cases where the l-order invariants are incompatible with the crystal symmetry, the bond-order distribution is centered around zero and overlaps with the fluid distribution [57]. The difference between the fluid and crystal local orders should then be assessed by other l-order invariants.

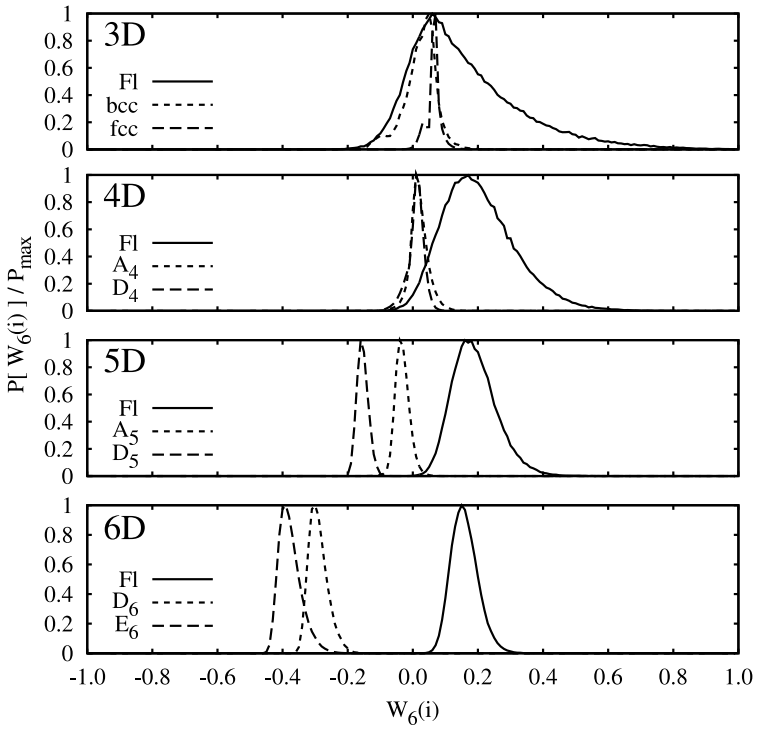


Figure 7.7: Third-order $l = 6$ invariants for dense crystal phases and the fluid. With increasing dimension fluid and crystalline local bond-order parameters become increasingly distinct.

configurations is what broadens the first peak of $g(r)$ and by ricochet its second peak as well [176]. Though they are harder to illustrate geometrically, similar phenomena are also expected in higher dimensions. The bond-order distribution is therefore fully consistent with a fluid structure dominated by simplex-based order, but not with the presence of deformed crystal unit cells.

7.C Methodology

For this study the particle diameter σ sets the unit of length and the thermal energy k_BT sets the unit of energy. For hard interactions this choice can be done without loss of generality, because entropy is the sole contributor to the free energy. Spheres become less efficient at filling space with increasing dimension. Though with our choice of units the fluid densities of interest ρ increase, the corresponding volume

fraction η

$$\eta \equiv \rho V_d 2^{-d} = \rho \frac{\pi^{d/2}}{\Gamma(1+d/2)2^d}, \tag{A7.1}$$

steadily decreases, because the volume of a d-dimensional hard sphere of radius $1/2$ in $\mathbb{R}^d$, $V_d 2^{-d}$, decreases faster. We report most quantities as volume fractions, but we also use ρ when it simplifies the notation.

7.C.1 Phase diagram

In order to precisely locate the freezing point, we perform standard NVT-Monte Carlo (MC) simulations to compute the fluid and crystal hard sphere equations of state (EoS) [92]. We consider the crystal phase postulated to be the densest in a given dimension, in addition to a less dense crystal for reference. The generating lattice vectors are selected in order to keep the size of the simulation box commensurate with the crystal dimensions to a minimum, as described in Appendix 7.C.4. The densest known close-packed structures in 5D and 6D are degenerate through layering, the same way that hexagonal close-packed and face-centered cubic packings are degenerate layerings in 3D [194]. For convenience we choose the most symmetric of these for the crystal lattice [173]. As in 3D, this decision should have but a minimal impact on the phase diagram [195]. Also, with increasing dimensionality layered structures show a growing similarity in their local two- and three-particle distributions, because only one of the spatial dimensions is affected by layering. Our choice of one of the layered phases should thus only have a small impact on the analysis.

The hard sphere equation of state of is related to the value of the radial pair distribution function $g(r)$

$$P/\rho = 1 + B_2 \rho g(1^+), \tag{A7.2}$$

where $B_2 = V_d/2$ is the second virial coefficient and $g(1^+)$ is extrapolated at contact [170]. The results are obtained for systems containing $N = 2048$ (D_4), 4096 (4D fluids and A_4), 3888 (5D fluids and D_5), 14400 (A_5), 2048 (D_6), 10000 (6D fluids), and 17496 (E_6) particles. To locate the fluid-crystal coexistence regime, we determine the absolute Helmholtz free energy per particle f of the crystal for a reference point [92], using the Einstein-crystal method [99]. The reference point is obtained for D_4 and A_4 crystals at $\eta = 0.37$, for D_5 and A_5 at $\eta = 0.21$ and for D_6 and E_6 at $\eta = 0.12$. The excess free energy at other crystal densities is then obtained by thermodynamic integration of the EoS. For the fluid the EoS is integrated from the ideal gas limit. The chemical potential

$$\mu(\rho) = f + P/\rho \tag{A7.3}$$

allows to identify the fluid-crystal coexistence pressure P^{coex} and μ^{coex} by finding where the chemical potential difference between the two phases $\Delta\mu(P) = 0$. The density of the coexisting phases is then obtained from a constant NPT MC computation. This approach is formally equivalent to the common tangent construction, but we find it to be numerically more efficient.

7.C.2 Order parameters

To characterize the structure of the fluid and crystal phases we need a criterion to quantify local ordering. Studies in 2D and 3D suggest that order parameters derived from rotationally-invariant combinations of the m different spherical harmonics Y_l^m of degree l might suffice [57, 98, 196]. Here, we consider second- and third-order invariants. Though a 4D canonical spherical harmonics basis [197, Sec. 9.6] and both its second- and third-order invariants [198] are known, in higher dimensions it rapidly becomes analytically intractable to identify a basis composed of weight vectors for the representation of SO(d) [199]. It is therefore more convenient to rewrite the invariants as polynomials of the vector inner products. For the second-order invariants, one simply uses the Gegenbauer polynomials $G_l^{d/2-1}$ obtained from the sum rule [200, Thm. 9.6.3]. For instance, the sum over the $(l+1)^2$ 4D spherical harmonics for unit vectors $\hat{\mathbf{r}}_i$ can be rewritten as

$$G_l^1(\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{r}}_2) = \frac{2\pi^2}{(l+1)^2} \sum_{m=1}^{(l+1)^2} Y_l^m(\hat{\mathbf{r}}_1)\overline{Y_l^m(\hat{\mathbf{r}}_2)}. \tag{A7.4}$$

The second-order local bond-order correlator $q_l(i,j)$ is obtained by summing over the $N(i)$ and $N(j)$ neighbors of particles i and j within a distance equal to the first minimum of $g(r)$. By letting the indices α and β run over these neighbors we find

$$q_l(i,j) = \mathbf{q}_l(i) \cdot \mathbf{q}_l(j) = \frac{\sum_{\alpha=1}^{N(i)} \sum_{\beta=1}^{N(j)} G_l^{d/2-1}(\hat{\mathbf{r}}_{\alpha i} \cdot \hat{\mathbf{r}}_{\beta j})}{N(i)N(j)}. \tag{A7.5}$$

Third-order rotationally invariant polynomials $\tilde{w}_l$ analogous to the Gegenbauer polynomials can also be obtained [187]. A classical theorem due to Weyl allows us to rewrite the third-order local bond-order correlator up to a dimension-dependent multiplicative constant c_l^d in terms of inner products,

$$W_l(i) = c_l^d \frac{\sum_{\alpha,\beta,\delta}^{N(i)} \tilde{w}_l^d(\hat{\mathbf{r}}_{i\alpha} \cdot \hat{\mathbf{r}}_{i\beta}, \hat{\mathbf{r}}_{i\alpha} \cdot \hat{\mathbf{r}}_{i\delta}, \hat{\mathbf{r}}_{i\beta} \cdot \hat{\mathbf{r}}_{i\delta})}{[q_l(i,i)]^{3/2}}. \tag{A7.6}$$

In 3D and 4D, the constant c_l^d can be set by comparing with the expression available in the literature. In higher dimensions, we choose the normalization for which

the polynomial equals unity when evaluated on three orthogonal unit vectors i.e., $c_l^d \tilde{w}_l^d(0,0,0) = 1$. Note that because of the rotational symmetry, triplets with permutated indices can be summed only once by correcting for the multiplicity. This simplification offers an important computational advantage. Though the use of rotationally-invariant polynomials for the computation of the bond-order parameters is mainly used for analytical convenience, it is also worth noting that for large l and at low densities, it is computationally more efficient than the standard spherical harmonics decomposition.

7.C.3 Generalized Classical Nucleation Theory

Classical nucleation theory (CNT) [4] uses the difference in chemical potential between the bulk phases and the fluid-crystal interfacial free energy $\gamma_{\mathrm{f-x}}$ of a spherical crystallite to obtain a free energy functional

$$\Delta G(n) = A_d (n/\rho_{\mathrm{x}})^{(d-1)/d} \gamma_{\mathrm{f-x}} - n\Delta\mu, \tag{A7.7}$$

of the number of particles n in the crystallite. The functional further depends on the crystal density ρ_{x} at the supersaturated fluid pressure and a geometrical prefactor A_d. For hard spheres $A_d = S_{d-1} V_d^{-(d-1)/d}$, where $S_{d-1} = dV_d$ is the surface area of a d-dimensional unit sphere. The resulting barrier height at the critical cluster size n^* is then

$$\Delta G^*(n^*) = \frac{(d-1)^{d-1}\pi^{d/2}}{\Gamma(d/2+1)2^d} \frac{\gamma_{\mathrm{f-x}}^d}{\rho_{\mathrm{x}}^{d-1}\Delta\mu^{d-1}}. \tag{A7.8}$$

In the high dimensional limit the barrier asymptotically approaches

$$\Delta G^*(n^*) \sim (2\pi e d)^{d/2} \frac{\gamma_{\mathrm{f-x}}^d}{\rho_{\mathrm{x}}^{d-1}\Delta\mu^{d-1}}. \tag{A7.9}$$

The rate of nucleation per unit volume k is then $k = \kappa \exp(-\Delta G^*)$, where κ is a kinetic prefactor proportional to the diffusion coefficient in the fluid phase [98]. The kinetic prefactor has weak dimensionality dependence, which we won't need to consider here. Though schematic, this level of theory is sufficient to clarify the contribution of geometrical frustration through an analysis of $\gamma_{\mathrm{f-x}}$. Within the CNT framework geometrical frustration between ordered and disordered phases should lead to a relatively large $\gamma_{\mathrm{f-x}}$, and thus to a high crystallization free energy barrier. On the contrary, geometrically similar phases should have small $\gamma_{\mathrm{f-x}}$ and $\Delta G^*(n^*)$.

7.C.4 Lattice generating matrices

The canonical lattice representations are often given in a form that simplifies the notation [173, Chap. 4], but does not necessarily allow for a convenient physical

construction for simulation purposes. D_d packings are checkerboard lattices, which are algorithmically simple to generate. A_4 and A_5 are dense packings with generating matrices

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 1/2 & \sqrt{2}/2 & 1/2 \\ 0 & 0 & 0 & 1 \\ \sqrt{10}/4 & 0 & 1/4 & 1/2 \end{pmatrix} \tag{A7.10}$$

and

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & -1/2 & \sqrt{3}/2 & 0 & 0 \\ 0 & 0 & -\sqrt{3}/3 & \sqrt{6}/3 & 0 \\ \sqrt{10}/4 & 0 & 0 & \sqrt{6}/4 & 0 \\ \sqrt{10}/5 & 0 & 0 & 0 & \sqrt{15}/5 \end{pmatrix} \tag{A7.11}$$

respectively. E_6 is a cut through the E_8 generalization of the diamond lattice, for which we use the generating matrix

$$\begin{pmatrix} 0 & 0 & 0 & \sqrt{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{3} & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1/2 & -1/2 & 0 & -\sqrt{3}/2 & -\sqrt{3}/2 \\ -1/2 & 1 & -1/2 & -\sqrt{3}/2 & 0 & -\sqrt{3}/2 \\ 1/2 & 1/2 & 1/2 & \sqrt{3}/2 & \sqrt{3}/2 & \sqrt{3}/2 \end{pmatrix}. \tag{A7.12}$$

From these generating matrices we can construct a unit cell commensurable with our hyper-rectangular simulation box. The sides of the unit cell u_i are obtained by linear combinations of the matrix' row vectors such that only one non-zero element remains. The lattice sites located within the unit cell borders then correspond to the particle positions within the unit cell. Following this recipe, our A_4 unit cell has relative side dimensions $l = (\sqrt{10}, 1, \sqrt{2}, 1)$ with $n_u = 8$ particles in the unit cell, A_5 yields $l = (\sqrt{10}, 1, \sqrt{3}, \sqrt{6}, \sqrt{15})$ with $n_u = 120$, and E_6 yields $l = (3, 3, 3, \sqrt{3}, \sqrt{3}, \sqrt{3})$ with $n_u = 24$.

Appendix A

A simple parameter-free nearest-neighbor algorithm

In this appendix we propose a parameter-free algorithm for the identification of nearest neighbors in many-particle systems. Compared to a Voronoi construction (or its dual, the Delaunay triangulation), which is the most commonly used parameter-free algorithm, the main advantages of the proposed algorithm are its low computational cost and its robustness against thermal fluctuations in the particle positions, in particular for particles at interfaces. We compare the present algorithm to a fixed-distance cutoff and a Delaunay construction analyzing Lennard-Jones and carbon bulk phases and two-phase simulation data.

For many-particle systems the identification of a particle's nearest neighbors is a recurrent task, both in simulation and structural analysis [92]. Algorithms solving this task can be classified as either tunable or parameter-free, with the most prominent representatives the fixed-distance cutoff and the Voronoi construction, respectively.

A fixed-distance cutoff [92] is generally applied in simulations with particles interacting through a short-range potential, where each neighbor is an interaction partner and the cutoff distance corresponds to the interaction range. For structural analysis the cutoff is less well-defined, but arguably the first minimum of the pair correlation function $g(r)$ is a natural choice, as it relates to the neighbors in the first coordination shell. However, the precise location of this minimum depends both on the system's details and on external conditions, and must therefore be determined every time either one of them is changed.

In contrast, the Voronoi construction [201] is based on a purely geometric construction and requires no tuning of parameters. In addition to the actual identification of nearest neighbors this method also provides geometric properties such as the edges, faces, and polyhedra shared with these neighbors - information that can be used itself for structural analysis and classification. But these advantages come at a significant computational cost, which is why this method is generally applied only in post-analysis, not on-the-fly during simulations. Another disadvantage of the Voronoi construction is its sensitivity to thermal fluctuations and the counter-intuitive results that it yields at interfaces. Several extensions to resolve these issues have been proposed, but they often introduce parameters and further increase the computational cost.

In this appendix we propose a simple parameter-free algorithm for the identification of nearest neighbors. Like the Voronoi construction it requires no tuning of parameters, but its computational cost is only slightly higher than that of the fixed-distance cutoff, making it suitable for the use during simulations. For a comparison of our algorithm with the fixed-cutoff distance and Voronoi construction, we apply it to Lennard-Jones liquid and fcc crystal bulk phases, to the 3-coordinated liquid carbon and graphite phases, and to the 4-coordinated liquid carbon and diamond phases. To test its behavior at an interface, where both the Voronoi construction and the fixed-distance cutoff are reported to yield unsatisfying results [202], we also analyze a liquid-crystal and a vapor-crystal two-phase system. We compute the distributions of the number of nearest neighbors, show how they relate to the pair correlation function, and use the neighbor data as input for structural analysis using bond-order parameters. We find that our algorithm yields results similar to those of the fixed-distance cutoff, is more robust than the Voronoi construction against fluctuations, and performs well at interfaces.

The remainder of this appendix is organized as follows: in the next section we explain the basic algorithm, and show how it can be extended to obtain the next-

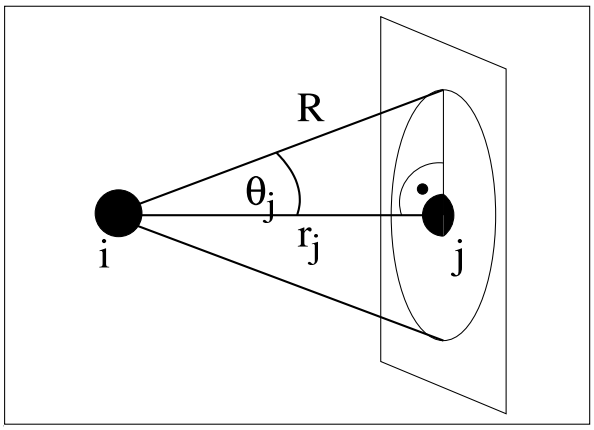


Figure A.1: Definition of the solid angle θ_j associated with a neighbor j of a particle i. Here, r_j is the distance between both particles, and R is the neighbor shell radius.

nearest neighbors or to include an optional tuning parameter. In Section A.2 we then present the results of our analysis and discuss the algorithm's advantages and disadvantages. In the last section we summarize our findings and discuss possible applications.

A.1 Method

Consider a particle i surrounded by its neighbor particles $\{j\}$. If we assume that the neighboring particles are organized in a shell-like structure, we can assign to each neighbor j a solid angle Θ_j based on the neighbor's distance $|\mathbf{r}_j|$ and a yet undetermined shell radius R. This definition is schematically depicted for a single neighbor in Figure A.1. We now define the neighborhood of a particle i to consist of all particles $\{j\}$ such that the sum of all solid angles equals 4π.

$$4\pi = \sum_{j=1}^{m} 2\pi[1 - cos(\theta_j)] = \sum_{j=1}^{m} 2\pi(1 - r_j/R). \tag{A.1}$$

Here, r_j refers to the distance of neighbor j and m denotes the number of neighbors for particle i. Neither the number m nor the shell radius R are known *a priori*, but one fixes the other.

For a system at any density, it is reasonable to assume that the *nearest* neighbors are the ones that define the local environment of a particle but, in general, we know neither how far away these neighbors may be, nor how many there are. The aim of a neighbor-searching algorithm is to identify the relevant nearest neighbors within the

cut-off distance R, but exclude the closest next-nearest neighbor:

$$r_m \leq R < r_{m+1}. \tag{A.2}$$

Combining Eqn. A.1 and Eqn. A.2 leads to a condition for the determination of the neighbor shell radius,

$$R^{(m)} = \frac{\sum_{j=1}^{m} r_j}{m-2} < r_{m+1}, \tag{A.3}$$

where $R^{(m)}$ refers to the shell radius containing m particles. To solve this inequality, we start with the smallest number of neighbors capable of satisfying Eqn. A.1, $m = 3$, and increase m iteratively. The smallest m that satisfies the condition A.3 yields the number of neighbors with $R^{(m)}$ the corresponding neighbor shell radius. It is straightforward to show that the algorithm convergences, because the neighbor distance increases monotonically due to the sorting, $r_{m+1} \geq r_m$, and the cutoff radius $R^{(m)}$ decreases monotonically ($R^{(m+1)} \leq R^{(m)}$) as long as $r_{m+1} \leq R^{(m)}$.

1.1.1 Scheme

Following the procedure outlined in the previous paragraph we propose the following implementation:

1. Sort all possible neighbors $\{j\}$ by their distance r_j in increasing order.
2. Start with $m = 3$ (minimum number of neighbors).
3. Compute $R^{(m)} = \sum_{j=1}^{m} r_j/(m-2)$.
4. If $(R^{(m)} > r_{(m+1)})$, then increase m and goto 3).
5. Else, m is the number of neighbors for particle i, and $R^{(m)}$ the associated neighbor shell radius.

1.1.2 Extension to a tunable algorithm

The algorithm assumes that the sum of all solid angles subtended by neighboring particles should add up to 4π. But, as the solid angles overlap anyway, this condition is somewhat arbitrary. One could therefore generalize the algorithm defining the target solid angle to be $4\pi\alpha$, where α is a positive constant.

$$4\pi\alpha = \sum_{j=1}^{N_b(i)} 2\pi[1 - cos(\theta_j)] = \sum_{j=1}^{N_b(i)} 2\pi(1 - r_j/R). \tag{A.4}$$

The condition to stop at the current iteration then becomes

$$R^{(m)} = \frac{\sum_{j=1}^{m} r_j}{m - 2\alpha} < r_{m+1}. \tag{A.5}$$

Note, however, that this version of the algorithm would not be parameter-free. In what follows, we shall limit ourselves to the case $\alpha = 1$.

1.1.3 Extension to obtain the next-nearest neighbors

The algorithm can be extended to yield the next-nearest neighbors. For this task the algorithm is performed twice as follows: in the first run, the nearest neighbors are computed without any modifications. Then all nearest neighbors are discarded from the list of possible neighbors, and the algorithm is run a second time. Because the algorithm is scale-free, no modification of the algorithm is required, and the second neighbor shell is obtained. Note that just increasing the solid angle to 8π does not work, as the solid angle contribution of the nearest neighbors would dominate due to the large cutoff distance $R^{(m)}$.

1.1.4 Sample preparation and implementation details

We tested the nearest-neighbor algorithm on a number of different systems that had been studied earlier in the context of crystal or liquid nucleation. Specifically, we prepared configurations of the following model systems:

Lennard Jones liquid and crystal

To prepare configurations representative of a Lennard-Jones fcc crystal and liquid phases we performed Monte Carlo simulations in the isothermal-isobaric ensemble for particles interacting via a truncated and shifted Lennard-Jones pair potential [92] with a cutoff distance of 2.5σ. For both phases, a system of $N = 4000$ particles was prepared at the reduced temperature $T = 0.92$ and pressure $P = 5.68$. At these conditions, which correspond to 20% undercooling with respect to coexistence, the liquid phase is metastable with respect to the fcc crystal phase, but a nucleation free-energy barrier of $\Delta G^* \approx 20 k_B T$ prevents spontaneous crystallization on simulation time scales [179].

Carbon liquid and crystal

Configurations representative of carbon in different phases were prepared by C. Valeriani using the LCBOPI$^+$ potential that is described in Appendix A of Ref. [203]. The structures were generated under the same conditions that were used in the study on

diamond nucleation [204], namely $P = 30\,GPa$ and $T = 3750\,K$ for the 3-coordinated liquid and graphite, and $P = 85\,GPa$ and $T = 5000\,K$ for the 4-coordinated liquid and diamond phases. Both conditions correspond to 25% undercooling with a nucleation free-energy barrier equal to or larger than $\Delta G^* = 25 k_B T$ preventing spontaneous crystallization of the metastable liquid phase. All systems contain $N = 1000$ particles, except for the graphite crystal with $N = 960$ particles. For details on the simulation methods and the semi-empirical interaction potential we refer to Ref. [204].

Liquid-crystal two-phase system

To prepare configurations representative of a two-phase liquid-crystal system, we used the truncated and shifted Lennard-Jones potential model that was discussed above at the same conditions, but with a system size of $N = 8000$ particles. To stabilize the solid-liquid interface, the equilibration was biased with a quadratic potential on the number of solid-like particles to prevent further growth of the crystal phase. Solid-like particles were identified with the bond-order parameters according to Ref. [57].

Crystal-vapor two-phase system

To prepare configurations representative of a two-phase vapor-crystal system we applied the modified Lennard-Jones potential introduced by ten Wolde and Frenkel [10]. These authors showed that for a proper choice of parameters the interaction is sufficiently short-ranged to result in a phase diagram similar to that of global proteins, for which the vapor-liquid coexistence line including its critical point moves below the fluid-crystal coexistence and thereby becomes entirely metastable. Our samples are prepared in the grand-canonical ensemble at conditions close to the metastable critical point, $T_c \approx 0.42$ and $\mu \approx -1.17$. Again, a bias on the number of crystalline particles was applied to prevent the crystal from growing. To prevent wetting of the crystal phase with a high-density liquid, we identified the crystal particles using a high-density order parameter, which requires a particle to have more than 9 neighbors within a cutoff distance of $r_c = 1.5\sigma$ [66].

Instead of the full Voronoi construction, we computed its dual, the Delaunay triangulation. It yields essentially the same information, but does not construct the polygons covering the Wigner-Seitz cell, and is therefore computationally less expensive. For the Delaunay triangulation we used the 3D triangulation kernel from the Open Source Computational Geometry Algorithms Library (CGAL [205]). Because CGAL does not support periodic boundary conditions yet, we included up to 8 periodic copies of particles located in the vicinity of the simulation box borders. To speed up the proposed algorithm we made use of Verlet lists to determine the set of pos-

sible neighbors for each particle. Although this method involves a cutoff parameter, we could have chosen a parameter-free algorithm like a binary space partitioning tree (octree). Any domain-decomposition method suffices as long as it provides enough particles for the algorithm to converge.

A.2 Results

Before we analyze the neighborhoods obtained by our algorithm and compare those with the results from a fixed-distance cutoff (CUTOFF) and Delaunay construction (DELAUNAY), we first discuss several inherent properties of our solid-angle based nearest neighbor algorithm (SANN).

The first property concerns symmetry. For both a fixed-distance cutoff and the Voronoi construction, the neighbors are symmetric in the sense that if particle i is neighbor of j, j is also a neighbor of particle i. Although our algorithm is scale-free, this symmetry is not ensured, because every particle has its own neighbor shell radius. Hence the distance between both particles can be smaller than the shell radius of particle i and larger than that of particle j at the same time. However, our analysis shows that for the Lennard-Jones system the fraction of asymmetric neighbors is below 2%, and in the carbon system below 5%.

For the second property we assign to each particle a local volume. The Voronoi algorithm yields the Wiegner-Seitz cell as obvious choice for the local volume, and by construction the sum of all local volumes adds up to the total system volume. For both the SANN and CUTOFF algorithms one can think of many different definitions of a local volume, e.g. related to a neighbor's distance or solid angle, but there exists no inherent definition.

Next, we quantify our earlier statement concerning the relative computational cost. To assess the performance, we compare the average compute time required to calculate the neighborhood for all particles in a system. Our test machine is a Linux workstation with the GCC 3.4.4 compiler and a 32-bit Pentium IV processor running at 3.2 GHz. We find that our algorithm takes on average twice as long as the fixed-distance cutoff [1], but compared to the Delaunay triangulations it runs more than 7 times faster for the Lennard-Jones phases and more than 12 times faster for the carbon phases [2].

For the sake of convenience we rewrite Eqn. A.3 to express the shell radius as an

[1] For both the CUTOFF and the SANN algorithms we use cell lists. However, the SANN algorithm used a cutoff which is 2σ larger than the cutoff for the fixed-distance algorithm.

[2] Data averaged over 20 different samples.

average distance to the nearest neighbors times a correction term,

$$R^{(m)} = \frac{\sum_{j=1}^{m} r_j}{m} \frac{m}{m-2} = \left\langle r \right\rangle_m \frac{m}{m-2}, \tag{A.6}$$

where $\left\langle r \right\rangle_m$ denotes the arithmetic average distance of all m neighbors. The correction factor $m/(m-2)$ can be interpreted as a measure of how well m neighbors approximate a spherical arrangement.

Eqn. A.6 leads directly to the next property: for a closed-packed arrangement of kissing spheres, the algorithm always identifies the correct number of nearest neighbors. Since all neighbors are kissing, their distances to the central particle are all equal, as is their average. Second neighbors are ignored if the correction term is smaller than the ratio of the second-nearest-neighbor distance over the nearest-neighbor distance.

The last property concerns the application of the algorithm to a higher-dimensional space, e.g. to hyper-spherical particles. Although designed for our three-dimensional space, we point out that the algorithm is valid without modification for any space-dimension $D \geq 2$. However, results for 4D to 6D hard hyper-spheres showed a decrease in quality with increasing dimension.

In what follows we apply the proposed algorithm to several simulation samples and compare it against both the fixed-distance cutoff criterion and the Delauany triangulation.

1.2.1 Nearest neighbors

First, we compute the number of nearest neighbors for all test configurations using all three neighborhood algorithms. For the fixed-distance cutoff criterion we set the cutoff equal to the minimum of the pair correlation function, which yields $r_c = 1.5$ and $r_c = 1.35$ for the Lennard-Jones liquid and fcc crystal phases, respectively, and $r_c = 2.0$ for all carbon phases. The number of nearest neighbor distributions are presented in the panels a) of Fig. A.2 (Lennard-Jones), Fig. A.3 (3-coordinated carbon), and Fig. A.4 (4-coordinated carbon).

For the Lennard-Jones phases the distributions of the number of nearest neighbors are fairly similar, but show some systematic differences (see Fig. A.2a). The Delaunay triangulation slightly overestimates the number of nearest neighbors and peaks around 14 neighbors for both the liquid and the fcc phase. This behavior originates from fluctuations which cause next-nearest neighbors to share occasionally a small face with the center particle [206]. There exist extensions to the Delaunay construction that aim to increase the robustness against fluctuations [70, 207, 208]. But many of these extensions introduce non-inherent parameters and are as such not parameter-free. Therefore, we will not consider them here. The fixed-distance cutoff

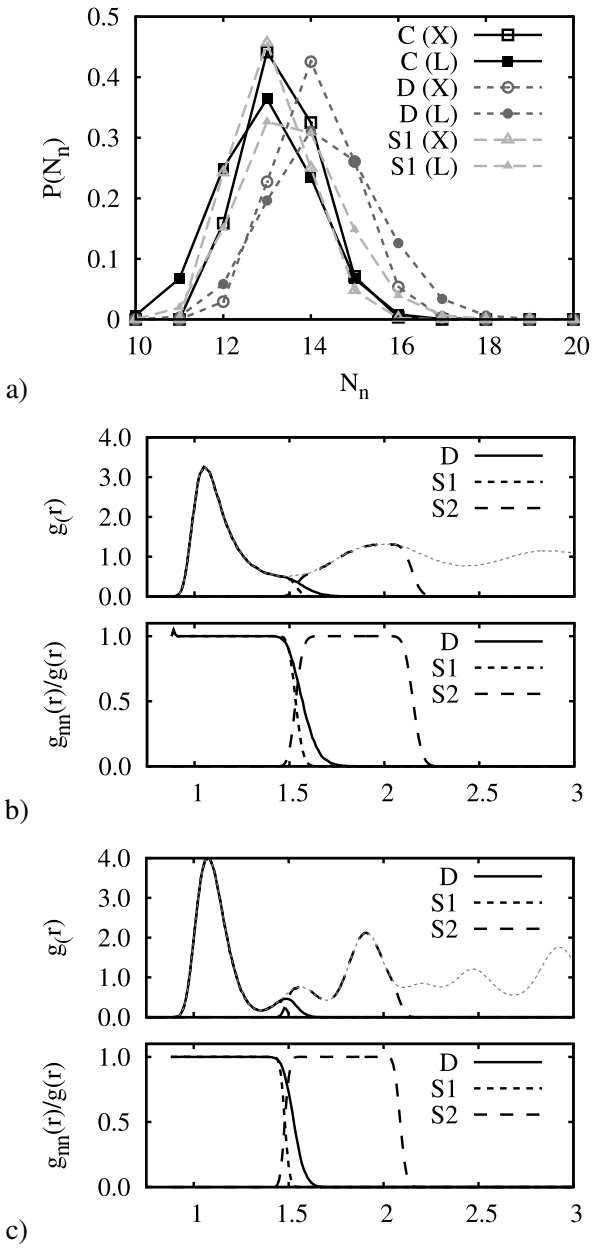


Figure A.2: a) Number of nearest neighbors distribution for both a Lennard-Jones liquid (L) and fcc crystal (X) obtained by all three methods. Panels b) and c) plot the pair correlation functions $g(r)$, considering all particles (dotted line), and $g_{nn}(r)$, considering only nearest neighbors obtained by Delaunay (D) and SANN's first (S1) and second (S2) neighbor shell, for both a liquid (b) and an fcc crystal (c). In addition, the fraction $g_{nn}(r)/g(r)$ is shown.

exhibits a distribution which peaks at the expected 12 neighbors for the fcc crystal, and at 13 for the liquid. The distribution obtained using the SANN algorithm overlaps almost entirely with the cutoff distribution for the fcc crystal phase. For the liquid the distribution broadens a little and the peak slightly shift to a lower value, which contrasts the other methods.

For the 3- and 4-coordinated carbon phases, Figure A.3a and A.4a, the distributions are quite different. The fixed-distance cutoff, using the first minimum of the pair correlation function, yields a distribution peaked sharply around 3 and 4 particles for a 3- and 4-coordinated carboin, respectively. The SANN algorithm peaks around 10 and 12 particles, and the Delaunay triangulation peaks at 18 and 20 (diamond). This means that the latter two algorithms include particles from the second and third neighbor shells. For the comparison in Fig. A.3 and A.4 we therefore set the cutoff for the fixed-distance cutoff algorithm to the *second* minimum of the pair correlation function, hence including the second neighbor shell. One might argue that including next-nearest neighbors is incorrect. However, as we will discuss in Section 1.2.2, with this extended neighborhood a 6^{th} order local bond-order parameter can distinguish better between disordered and ordered carbon phases than the 3^{rd} order harmonics based on the 1^{st} neighbor shell as used in Ref. [204]. This is of interest in the study of crystal nucleation, where the system is biased to form a crystal nucleus with an order parameter that should not favor any particular crystal lattice.

In order to get a better understanding about which particles are identified as nearest neighbors, we compute the pair correlation function using either all particles (for reference) or only nearest neighbors using the different neighborhood algorithms. We label them $g(r)$ and $g_{nn}(r)$, respectively. The fixed-distance cutoff method is not tested here as it yields, by definition, $g(r)$ exactly up to the cutoff radius, from which on it is zero. The results are depicted in panels b (liquid) and c (crystal) of Figures A.2 to A.4. The upper graphs each show the pair correlation functions $g(r)$ and $g_{nn}(r)$, and the lower graph the fraction $g_{nn}(r)/g(r)$. At a given distance r the latter fraction is one if all particles of this distance are identified as nearest neighbors, and reduces to zero if none of these particles are considered neighbors. The steeper the ratio drops, the less fluctuations there are in the selection of the neighbors. In addition to the nearest neighbors the graphs also show results for the next-nearest neighbors obtained from the SANN method as described in Section 1.1.3.

For the Lennard-Jones system Figure A.2b and A.2c show that the pair correlation function for the Delaunay triangulation is identical with the reference $g(r)$ up to the first minimum, and in the fcc crystal even slightly beyond that. Although it decreases quickly afterwards it still includes particles from the second neighbor shell (see Fig. A.2c upper panel). In contrast, the $g_{nn}(r)$ using the SANN algorithm drops to zero at the first minimum for both the liquid and crystal phases, and therefore barely includes next-nearest neighbors. The second neighbor shell does not yield

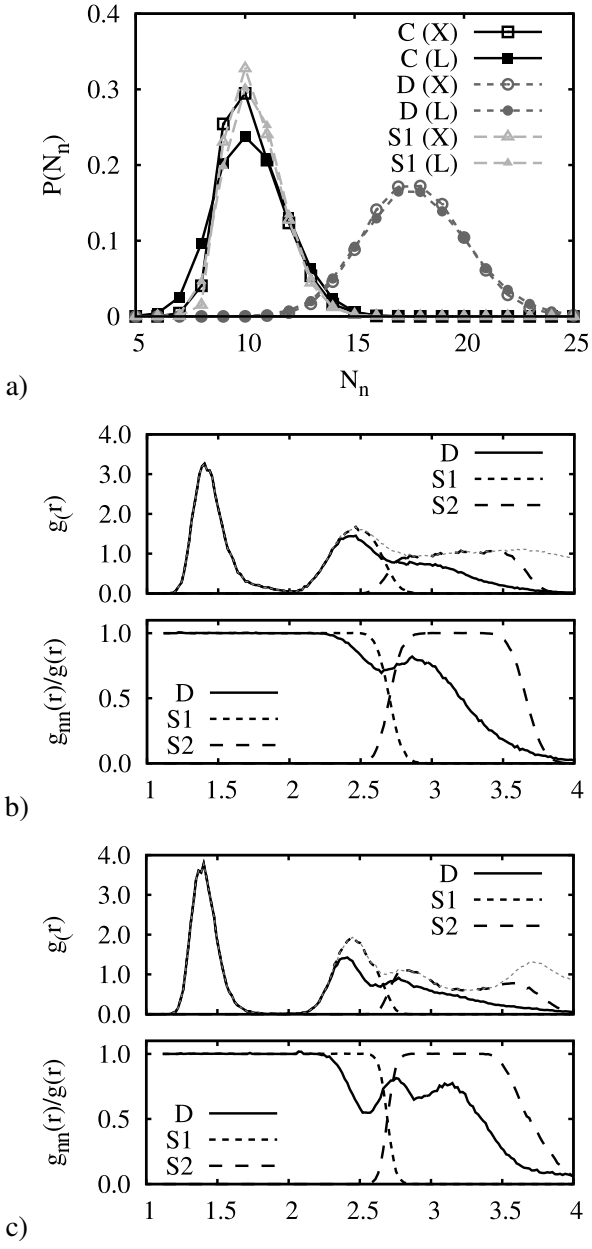


Figure A.3: a) Number of nearest neighbors distribution for both a 3-coordinated carbon liquid (L) and graphite crystal (X) obtained by all three methods. Panels b) and c) plot the pair correlation functions $g(r)$, considering all particles (dotted line), and $g_{nn}(r)$, considering only nearest neighbors obtained by Delaunay (D) and SANN's first (S1) and second (S2) neighbor shell, for both a liquid (b) and an fcc crystal (c). In addition, the fraction $g_{nn}(r)/g(r)$ is shown.

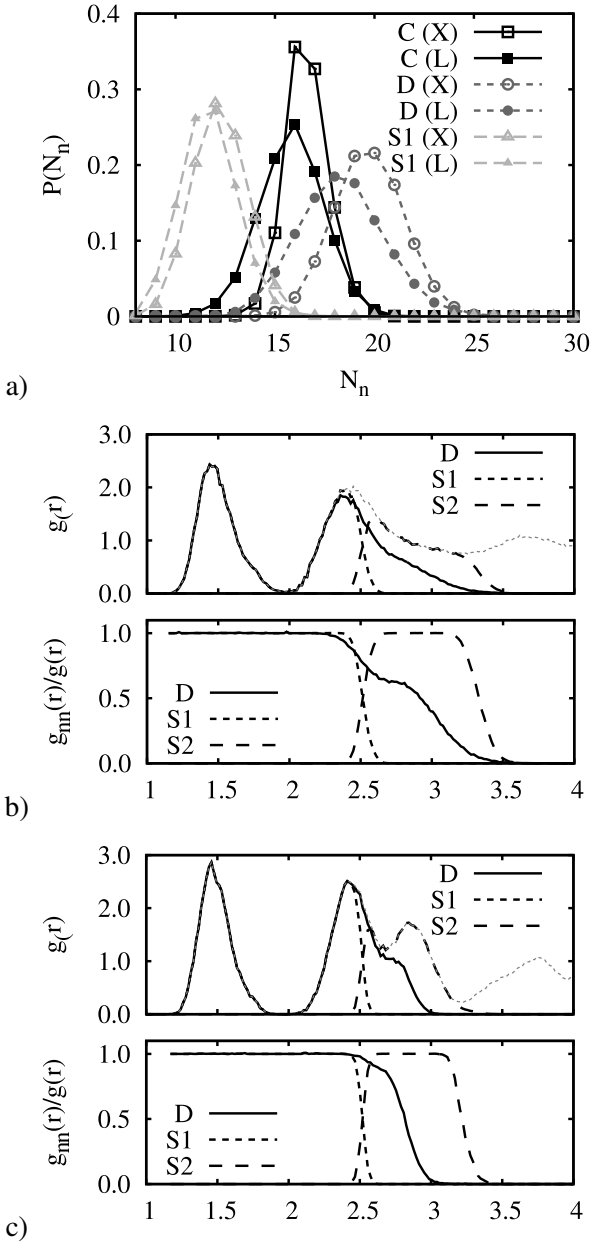


Figure A.4: a) Number of nearest neighbors distribution for both a 4-coordinated carbon liquid (L) and diamond crystal (X) obtained by all three methods. Panels b) and c) plot the pair correlation functions $g(r)$, considering all particles (dotted line), and $g_{nn}(r)$, considering only nearest neighbors obtained by Delaunay (D) and SANN's first (S1) and second (S2) neighbor shell, for both a liquid (b) and an fcc crystal (c). In addition, the fraction $g_{nn}(r)/g(r)$ is shown.

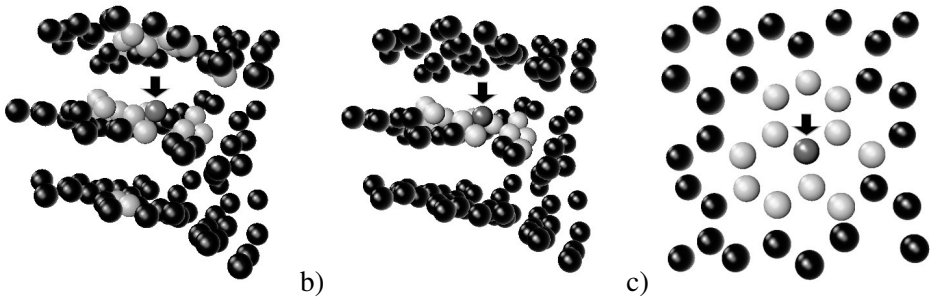


Figure A.5: Simulation snapshot showing the neighbors (light gray) of a center particle (dark-gray) indicated by the arrow. Surrounding particles that are not part of the neighborhood are shown in black. In contrast to the SANN algorithm (panel b) the Delauany construction (panel a) includes particles from different layers. Panel (c) presents a top-view on the center layer of panel (b), which shows that the neighborhood includes the complete first neighbor shell and all but one particle from the second.

such accurate results, but it still gives a reasonable approximation for the next-nearest neighbors.

For the 3-coordinated carbon phases, the 2$^{\text{nd}}$ and 3$^{\text{rd}}$ neighbor shells are so close together that their peaks in the pair correlation function merge. The number of neighbors including the second and third neighbor shell are then 9 and 12, respectively. The number of nearest neighbors obtained from SANN are distributed around a peak at 10, and the corresponding $g_{nn}(r)$ drops to zero at approximately the second minimum. A particularly interesting behavior is shown by the Delaunay $g_{nn}(r)$ (see Fig. A.3b and A.3c). It extends well beyond the second minimum, does not include all particles anymore, and the fraction $g_{nn}(r)/g(r)$ does not decrease monotonically to zero. To explain this behavior, we point out that 3-coordinated carbon forms layers, which are several particle diameters apart. Whereas the fixed-distance cutoff and the SANN algorithms both consider only close in-plane neighbors, the Delaunay triangulation also includes particles from neighboring layers, which are much further apart. Although geometrically correct, these additional neighbors are part of a completely different neighborhood, and may distort results for local quantities. Figure A.5a shows such a neighborhood as obtained from the Delaunay construction, which clearly includes neighbors from different layers. In contrast, the SANN algorithm selects only in-plane neighbors (Fig. A.5b) from the first and second neighbor shell (Fig. A.5c, top view).

In contrast to the 3-coordinated carbon phases, the 4-coordinated phases are not organized in sheets. As in the previous case, both SANN and Delaunay include next-nearest neighbors (see Fig. A.4b and A.4c). But unlike before, SANN does

not include the full 2nd neighbor shell anymore, and its $g_{nn}(r)$ starts to decrease a little too early. For the liquid, the Delaunay algorithm also ignores some of the next-nearest neighbors. But in contrast to SANN its fraction $g_{nn}(r)/g(r)$ decreases more slowly. For the diamond crystal phase, the Delaunay triangulation includes not only the complete 2nd neighbor shell, but also parts of the 3rd .

1.2.2 Application to bond-order parameters

Next, we apply investigate the effect of a nearest-neighbor algorithm on the local bond-order analysis for crystal nucleation. The original order parameters proposed by Steinhardt et al. [72] were based on the idea to expand the neighborhood of particles using spherical harmonics to quantify a system's local symmetry. The algorithm was later refined by ten Wolde et al. [57] for the study of nucleation, and has proven a useful tool even for higher-dimensional systems [181, 187]. To compute the bond-order parameter each particle i is assigned a $(2l+1)$-dimensional complex vector $\mathbf{q}_l(i)$ capturing its local neighborhood,

$$q_l^m(i) = \frac{1}{N_b(i)} \sum_j Y_{lm}(\hat{\mathbf{r}}_{ij}), \tag{A.7}$$

where $N_b(i)$ denotes the number of nearest neighbors, $Y_{lm}(\hat{\mathbf{r}}_{ij})$ is the spherical harmonics of order l with components $-l \leq m \leq l$, $\hat{\mathbf{r}}_{ij}$ is the unit vector pointing from the center of i to its neighbor j, and the sum runs over all neighbors $\{j\}$ of particle i. From this we can construct a measure for the neighborhood similarity of two particles,

$$q_l(i,j) = \frac{\mathbf{q}_l(i) \cdot \mathbf{q}_l^*(j)}{|\mathbf{q}_l(i)|\,|\mathbf{q}_l(j)|}, \tag{A.8}$$

where the superscript star denotes the complex conjugate. We call the $q_l(i,j)$ the local bond-order correlator, which is one if both particles are in an identical ordered environment. To distinguish reliably between solid-like and liquid-like particles, particularly in an undercooled liquid, additional steps are required to increase the contrast. But since the neighborhood algorithm enters at this stage we will not follow the procedure to the end, but instead compare the local bond-order correlators.

For the Lennard-Jones phases all neighborhood algorithms performs equally well, as is to be expected from the similarity in both the number of nearest neighbors and neighbor-based radial distribution function $g_{nn}(r)$ of Fig. A.2. However, for the carbon phases the situation is different. Ref. [204] reported that a 3rd order harmonic can capture the symmetry of all carbon crystal phases by including only particles from the 1st neighbor shell. However, since both the Delaunay construction and the SANN algorithm include particles from the 2nd neighbor shell, we find that $l = 6$, rather than

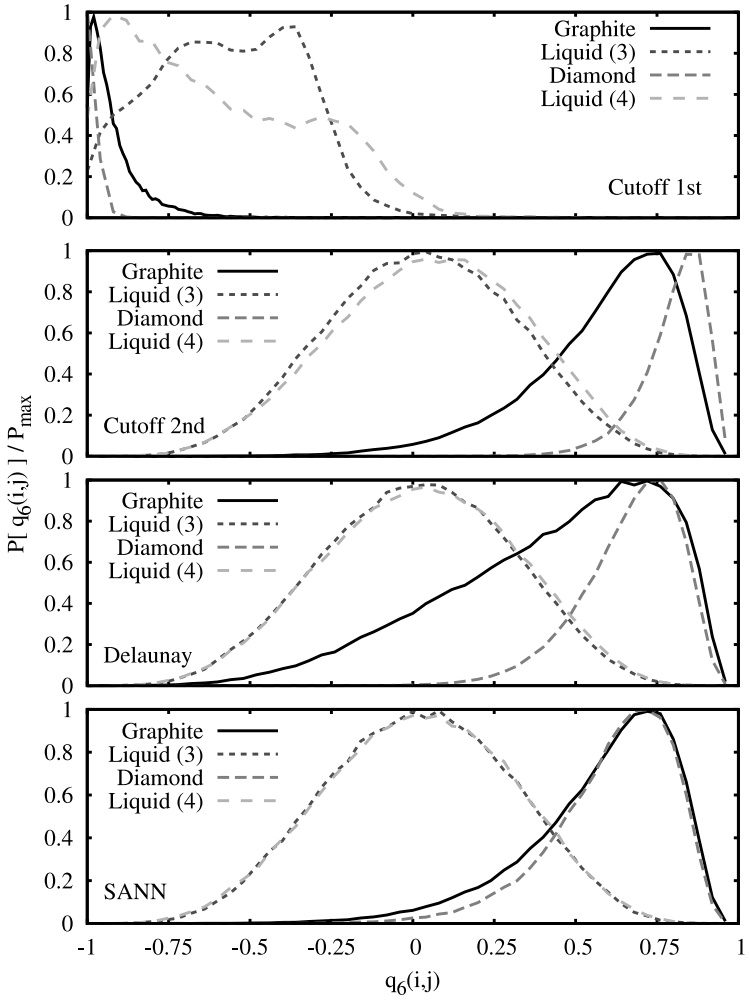


Figure A.6: Distribution of the bond-order correlator for the carbon phases. The top panel uses a 3rd order harmonics and a cutoff including only particles from the 1st neighbor shell, as used in Ref. [204]. The resulting liquid distributions are not well-separated from the solid. The other three panels show a 6th order harmonics with the extended neighborhood using a fixed-cutoff criterion, the Delaunay construction, and the SANN algorithm (top to bottom). Here, both the liquid and crystal distributions are nicely peaked and reasonably well separated.

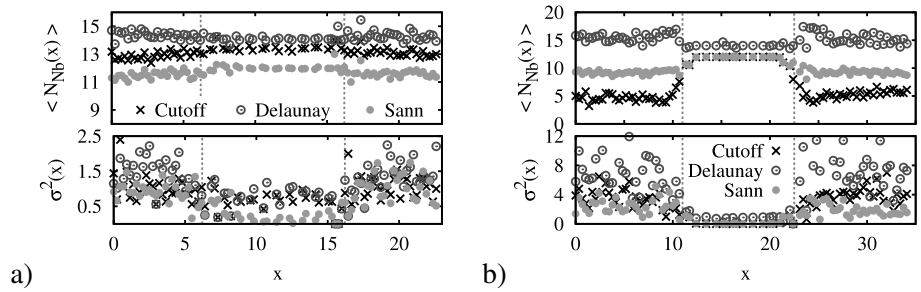


Figure A.7: Results for two-phase samples in a slab-geometry, with the two interfaces oriented normal to the x-direction. The two phases are liquid-crystal (a), and fluid-crystal (b). For both samples the upper panel shows, as function of x-position, the average number of nearest neighbors, $\langle N_{Nb}(x)\rangle$, and the lower panel the corresponding standard deviation, $\sigma^2 = \langle N_{Nb}^2(x)\rangle - \langle N_{Nb}(x)\rangle^2$.

$l = 3$, is commensurable with the symmetry of this extended environment. To make a proper comparison we set the fixed-distance cutoff to 2.7 to include next-nearest neighbors. Figure A.6 presents the bond-order correlator distributions for the carbon phases, each curve normalized by its maximum P_{max}. The top panel shows results for the 3$^{\text{rd}}$ order harmonics with a fixed-distance cutoff including only nearest neighbors ($r_c = 2.0$), as used in Ref. [204]. The crystal structures are sharply peaked next to -1, but the liquid distributions are broad and overlap strongly with the crystal distributions. Panels b) to d) show 6$^{\text{th}}$ order harmonics with the extended neighborhood using the fixed-distance cutoff criterion, the Delaunay construction, and the SANN algorithm, respectively. Here, the liquid and crystal distributions are quite well distinguished, and SANN performs slightly better than the other two methods.

1.2.3 Application to interfaces

In a final step, we apply all algorithms to two two-phase systems with planar interfaces (slab geometry): a Lennard-Jones liquid-crystal system and *modified* Lennard-Jones fluid-crystal system. The latter interaction potential was introduced by ten Wolde and Frenkel [10] and includes a tuning parameter α that controls the interaction range. For a value $\alpha = 50$ it reproduces the phase diagram of globular proteins featuring only the fluid and the crystal as thermodynamically stable phases, and a meta-stable vapor-liquid coexistence line hidden in the fluid-crystal two-phase region. Our sample is prepared at conditions close the the meta-stable critical point ($T = 0.42$, $P = 0.01$, $\rho_f = 0.3$). The two phases are arranged in a slab-geometry such that the interfaces are normal to the x-direction. We now compute both the number of nearest neighbors as function of x-position, $\langle N_{Nb}(x)\rangle$, and the corresponding standard

deviation, $\sigma^2 = \langle N_{Nb}^2(x)\rangle - \langle N_{Nb}(x)\rangle^2$. For the fixed-distance algorithm the cutoff is set to $r_c = 1.5\sigma$ resembling the first minimum of the radial distribution function of the liquid phase.

The results are presented in Figure A.7. In the case of liquid-crystal coexistence, Fig. A.7a, all the algorithms give fairly similar results. But compared to both SANN and CUTOFF the Delaunay tessellation shows the least distinct change in the number of nearest neighbors when crossing the interface. This effect becomes even more pronounced at interfaces between a high-density and a low-density phase, such as the vapor-crystal coexistence (Fig. A.7b). For the low-density phase, where the CUTOFF method identifies on average around 5 neighbors, the Delaunay triangulation determines a strongly fluctuating average of 15 neighbors, a number that even exceeds the average for the crystal phase. Here, the SANN algorithm provides a more reasonable average of 10 neighbors, and shows considerably less fluctuations. Also, for particles at the interface, the SANN algorithm chooses its neighbors mostly from the interface, and, unlike the Delaunay construction, does not include particles that are located far-off in the vapor.

A.3 Conclusion

In this appendix we proposed a simple parameter-free algorithm for the detection of a particle's nearest neighbors. In this algorithm the m nearest neighbors are each assigned a solid angle depending on their distance to the central particle and the neighbor shell radius $R^{(m)}$, which follows from the requirement that the total sum of all solid angles adds up to 4π. In an iterative procedure m is increased to include more neighboring particles until the monotonically decreasing $R^{(m)}$ becomes smaller than the distance to the $(m+1)^{\text{th}}$ neighbor. Using a common method for domain decomposition, like cell lists, Verlet lists, or a binary space-partitioning tree (octree), the computational cost for this algorithm are roughly an order of magnitude smaller than the cost for a Voronoi construction (or its dual, the Delaunay triangulation). In addition, our scale-free algorithm is easily adapted to include a tuning parameter, or to yield only particles from the 2^{nd} neighbor shell.

We discussed the advantages and limitations of our algorithm, in particular its computational cost and the broken neighbor symmetry. The algorithm was then applied to analyze supercooled liquid and crystal phases of Lennard-Jones, 3-coordinated and 4-coordinated carbon, and compared to both a fixed-distance cutoff criterion and a Delaunay triangulation. For the Lennard-Jones phases, where the Delaunay triangulation is known to be strongly affected by fluctuations, our algorithm reproduces very well the number of nearest-neighbor distribution of a well-tuned fixed-distance cutoff. For the carbon phases, our algorithm includes the second neighbor shell, like

the Delaunay triangulation, but avoids spurious neighbors in neighboring graphite layers.

Further, we deployed the neighbor information of all algorithms as input for bond order analysis, which is used in crystal nucleation studies for the identification of solid-like particles in a supercooled metastable liquid. Comparing the bond-order correlator distributions, we find little difference between the algorithms, indicating that the proposed algorithm is suitable for structure analysis.

In a final step we focused on particles at the interface of a two-phase system. We find that for two high-density phases all algorithms give reliable results. But at the interface between a high-density crystal and a low-density fluid, our algorithm is shown to be more robust against thermal fluctuations than the Voronoi construction.

Given the low computational cost of our algorithm, we argue that it is well suited not only for post-analysis, but also during simulations. It reliably identifies the nearest neighbors, and its behavior for graphite and at a two-phase interface suggests its application to situations where the Voronoi construction suffers from distorted polyhedra, as is the case in structural analysis of protein folding trajectories.

Acknowledgements

The authors thank K. Shundyak for fruitful discussions and L. Bossen for a careful reading of the manuscript.

Bibliography

[1] Wikipedia. Crystallization (engineering aspects)— Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Crystallizer (2009). [Online; accessed 6th July 2009].

[2] D. Turnbull. *J. Chem. Phys.* **18**, 198 (1950).

[3] D. Turnbull and R. E. Cech. *J. App. Phys.* **21**, 804 (1950).

[4] M. Volmer and A. Weber. *Z. Phys. Chem.* **119**, 277 (1926).

[5] R. Becker and W. Döring. *Annalen der Physik* **416**, 719 (1935).

[6] J. W. Gibbs. *The Scientific Papers of J. Willard Gibbs*. Dover, New York (1961).

[7] W. Ostwald. *Z. Phys. Chem.* **22**, 289 (1897).

[8] M. Muschol and F. Rosenberger. *J. Chem. Phys.* **107**, 1953 (1997).

[9] Y. G. Kuznetsov, A. J. Malkin, and A. McPherson. *J. Cryst. Growth* **232**, 30 (2001).

[10] P. R. ten Wolde and D. Frenkel. *Science* **277**, 1975 (1997).

[11] V. Talanquer and D. W. Oxtoby. *J. Chem. Phys.* **109**, 223 (1998).

[12] R. P. Sear. *J. Chem. Phys.* **114**, 3170 (2001).

[13] G. Nicolis and C. Nicolis. *Physica A* **323**, 139 (2003).

[14] J. F. Lutsko and G. Nicolis. *Phys. Rev. Lett.* **96**, 046102 (2006).

[15] J. R. Savage and A. D. Dinsmore. *Phys. Rev. Lett.* **102**, 198302 (2009).

[16] P. G. Debenedetti. *Metastable Liquids*. Princeton University Press (1996).

[17] E. Saridakis and N. E. Chayen. *Trends in Biotech.* **27**, 99 (2009).

[18] A. J. Page and R. P. Sear. *Phys. Rev. Lett.* **97**, 065701 (2006).

[19] G. Tarjus, S. A. Kivelson, Z. Nussinov, and P. Viot. *J. Phys.: Condens. Matter* **17**, R1143 (2005).

[20] F. C. Frank. *Proc. R. Soc. London A* **215**, 43 (1952).

[21] D. R. Nelson. *Defects and geometry in condensed matter physics*. Cambridge University Press, New York (2002).

[22] J.-F. Sadoc and R. Mosseri. *Geometrical Frustration*. Cambridge University Press, Cambridge (1999).

[23] C. Valeriani, E. Sanz, and D. Frenkel. *The Journal of Chemical Physics* **122**, 194501 (2005).

[24] R. C. Tolman. *J. Chem. Phys.* **17**, 333 (1949).

[25] T. Zykova-Timan, C. Valeriani, E. Sanz, D. Frenkel, and E. Tosatti. *Phys. Rev. Lett.* **100**, 036103 (2008).

[26] J. Merikanto, E. Zapadinsky, A. Lauri, and H. Vehkamaki. *Phys. Rev. Lett.* **98**, 145702 (2007).

[27] S. Auer and D. Frenkel. *Phys. Rev. Lett.* **91**, 015703 (2003).

[28] C. J. Geyer. In *Computing Science and Statistics: Proceedings on the 23rd Symposium on the Interface*, pp. 156–163 (1991).

[29] C. J. Geyer. *Stat. Sci.* **7**, 473 (1992).

[30] C. J. Geyer and E. A. Thompson. *J. Amer. Stat. Assoc.* **90**, 431 (1995).

[31] Q. Yan and J. J. de Pablo. *J. Chem. Phys.* **111**, 9509 (1999).

[32] M. A. Miller and D. Frenkel. *Phys. Rev. Lett.* **90**, 135702 (2003).

[33] S. Tănase-Nicola and J. Kurchan. *Phys. Rev. Lett.* **91**, 188302 (2003).

[34] J. Tailleur and J. Kurchan. *Nature Physics* **3**, 203 (2007).

[35] F. Wang and D. P. Landau. *Phys. Rev. Lett.* **86**, 2050 (2001).

[36] F. Wang and D. P. Landau. *Phys. Rev. E* **64**, 056101 (2001).

[37] A. Laio and M. Parrinello. *Proc. Nat. Acad. Sci.* **99**, 12562 (2002).

[38] A. Barducci, G. Bussi, and M. Parrinello. *Phys. Rev. Lett.* **100**, 020603 (2008).

[39] M. Bonomi, A. Barducci, and M. Parrinello. *J. Comp. Chem.* **30**, 1615 (2009).

[40] E. Darve and A. Pohorille. *J. Chem. Phys.* **115**, 9169 (2001).

[41] L. Rosso, P. Mináry, Z. Zhu, and M. E. Tuckerman. *J. Chem. Phys.* **116**, 4389 (2002).

[42] L. Maragliano and E. Vanden-Eijnden. *Chem. Phys. Lett.* **426**, 168 (2006).

[43] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler. *J. Chem. Phys.* **108**, 1964 (1998).

[44] C. Dellago, P. G. Bolhuis, and D. Chandler. *J. Chem. Phys.* **108**, 9236 (1998).

[45] W. E, W. Ren, and E. Vanden-Eijnden. *Phys. Rev. B* **66**, 052301 (2002).

[46] W. E, W. Ren, and E. Vanden-Eijnden. *J. Phys. Chem. B* **109**, 6688 (2005).

[47] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti. *J. Chem. Phys.* **125**, 024106 (2006).

[48] L. Maragliano and E. Vanden-Eijnden. *J. Chem. Phys.* **128**, 184110 (2008).

[49] M. Venturoli, E. Vanden-Eijnden, and G. Ciccotti. *J. Math. Chem.* **45**, 188 (2009).

[50] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral. *Chem. Phys. Lett.* **156**, 472 (1989).

[51] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden. *Chem. Phys. Chem.* **6**, 1809 (2005).

[52] G. M. Torrie and J. P. Valleau. *J. Comp. Phys.* **23**, 187 (1977).

[53] T. S. van Erp, D. Moroni, and P. G. Bolhuis. *J. Chem. Phys.* **118**, 7762 (2003).

[54] D. Moroni, T. S. van Erp, and P. G. Bolhuis. *Phys. Rev. E* **71**, 056709 (2005).

[55] P. G. Bolhuis. *J. Chem. Phys.* **129**, 114108 (2008).

[56] J. S. van Duijneveldt and D. Frenkel. *J. Chem. Phys.* **96**, 4655 (1992).

[57] P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel. *Faraday Discuss.* **104**, 93 (1996).

[58] A. Warmflash, P. Bhimalapuram, and A. R. Dinner. *J. Chem. Phys.* **127**, 154112 (2007).

[59] A. Dickson, A. Warmflash, and A. R. Dinner. *J. Chem. Phys.* **130**, 074104 (2009).

[60] R. J. Allen, P. B. Warren, and P. R. ten Wolde. *Phys. Rev. Lett.* **94**, 018104 (2005).

[61] R. J. Allen, D. Frenkel, and P. R. ten Wolde. *J. Chem. Phys.* **124**, 024102 (2006).

[62] C. Valeriani, R. J. Allen, M. J. Morelli, D. Frenkel, and P. R. ten Wolde. *J. Chem. Phys.* **127**, 114109 (2007).

[63] P. Hänggi, P. Talkner, and M. Borkovec. *Rev. Mod. Phys.* **62**, 251 (1990).

[64] J. Wedekind, R. Strey, and D. Reguera. *J. Chem. Phys.* **126**, 134103 (2007).

[65] E. Sanz, C. Valeriani, D. Frenkel, and M. Dijkstra. *Phys. Rev. Lett.* **99**, 055501 (2007).

[66] F. H. Stillinger. *J. Chem. Phys.* **38**, 1486 (1963).

[67] J. Wedekind and D. Reguera. *J. Chem. Phys.* **127**, 154516 (2007).

[68] M. Tanemura, Y. Hiwatari, H. Matsuda, T. Ogawa, N. Ogita, and A. Ueda. *Prog. Theo. Phys.* **58**, 1079 (1977).

[69] Y. Hiwatari, T. Saito, and A. Ueda. *J. Chem. Phys.* **81**, 6044 (1984).

[70] W. Brostow, M. Chybicki, R. Laskowski, and J. Rybicki. *Phys. Rev. B* **57**, 13448 (1998).

[71] J. D. Honeycutt and H. C. Andersen. *J. Phys. Chem.* **91**, 4950 (1987).

[72] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti. *Phys. Rev. B* **28**, 784 (1983).

[73] E. Mendez-Villuendas and R. K. Bowles. *Phys. Rev. Lett.* **98**, 185503 (2007).

[74] H.-J. Limbach, A. Arnold, B. A. Mann, and C. Holm. *Comp. Phys. Comm.* **174**, 704 (2006).

[75] LAMMPS. Benchmarks of the LAMMPS Molecular Dynamics Simulator. http://lammps.sandia.gov/bench.html (2005).

[76] V. Sunderam. *Concur.: Pract. and Exp.* **2**, 315 (1990).

[77] E. Gabriel, G. E. Fagg, G. Bosilca, T. Angskun, J. J. Dongarra, J. M. Squyres, V. Sahay, P. Kambadur, B. Barrett, A. Lumsdaine, R. H. Castain, D. J. Daniel, R. L. Graham, and T. S. Woodall. In *Proceedings, 11th European PVM/MPI Users' Group Meeting*, pp. 97–104. Budapest, Hungary (2004).

[78] L. Dagum and R. Menon. *IEEE Computational Science & Engeneering* **5**, 46 (1998).

[79] J. Makino and M. Taiji. *Scientific simulations with special purpose computers: The GRAPE systems*. John Wiley & Sons (1998).

[80] R. Susukita, T. Ebisuzaki, B. G. Elmegreen, H. Furusawa, K. Kato, A. Kawai, Y. Kobayashi, T. Koishi, G. D. McNiven, T. Narumi, and K. Yasuoka. *Comp. Phys. Comm.* **155**, 115 (2003).

[81] R. J. Rost. *OpenGL Shading Language*. Pearson Education, 1st edition (2004).

[82] R. Fernando and M. J. Kilgard. *The Cg Tutorial: The Definitive Guide to Programmable Real-Time Graphics*. Addison Wesley Professional (2003).

[83] NVIDIA. *CUDA Programming Guide Version 1.0* (2006).

[84] K. Moreland and E. Angel. *Graphics Hardware,* pp. 112–119 (2003).

[85] E. S. Larsen and D. McAllister. *Proceedings of the 2001 ACM/IEEE conference on Supercomputing,* p. 55 (2001).

[86] W. Li, X. Wei, and A. Kaufman. *The Visual Computer* **19**, 444 (2003).

[87] S. Tomov, M. McGuigan, R. Bennett, G. Smith, and J. Spiletic. *Comp. Graph.* **29**, 71 (2005).

[88] S. F. Portegies Zwart, R. G. Belleman, and P. M. Geldof. *New Astronomy* **12**, 641 (2007).

[89] R. G. Belleman, J. Bédorf, and S. F. Portegies Zwart. *New Astronomy* **12**, 103 (2008).

[90] J. Yang, Y. Wang, and Y. Chen. *J. Comp. Phys.* **221**, 799 (2007).

[91] T. O. Group. *The Single UNIX Specification, Version 3* (2002). Http://www.unix.org/version3.

[92] D. Frenkel and B. Smit. *Understanding Molecular Simulation*. Academic Press (2002).

[93] J. A. Anderson, C. D. Lorenz, and A. Travesset. *J. Comp. Phys.* **227**, 5342 (2008).

[94] O. Galkin and P. G. Vekilov. *Proc. Nat. Acad. Sci.* **97**, 6277 (2000).

[95] B. Chen, H. Kim, S. J. Keasler, and R. B. Nellas. *J. Phys. Chem. B* **112**, 4067 (2008).

[96] V. G. Baidakov, G. G. Chernykh, and S. P. Protsenko. *Chem. Phys. Lett.* **321**, 315 (2000).

[97] P. R. ten Wolde and D. Frenkel. *J. Chem. Phys.* **109**, 9901 (1998).

[98] S. Auer and D. Frenkel. *Nature* **409**, 1020 (2001).

[99] D. Frenkel and A. J. C. Ladd. *J. Chem. Phys.* **81**, 3188 (1984).

[100] M. A. van der Hoef. *J. Chem. Phys.* **113**, 8142 (2000).

[101] R. Ohnesorge, H. Löwen, and H. Wagner. *Phys. Rev. E* **50**, 4801 (1994).

[102] D. Turnbull and J. C. Fisher. *J. Chem. Phys.* **17**, 71 (1949).

[103] R. P. Sear. *J. Phys.: Condens. Mat.* **19**, 033101 (2007).

[104] J. Zeldovich. *J. Exp. Theor. Phys.* **12**, 525 (1942).

[105] R. McGraw and A. Laaksonen. *J. Chem. Phys.* **106**, 5284 (1997).

[106] S. L. Girshick and C.-P. Chiu. *J. Chem. Phys.* **93**, 1273 (1990).

[107] A. Laaksonen, I. J. Ford, and M. Kulmala. *Phys. Rev. E* **49**, 5517 (1994).

[108] M. Horsch, J. Vrabec, and H. Hasse. *Phys. Rev. E* **78**, 011603 (2008).

[109] R. L. Davidchack and B. B. Laird. *J. Chem. Phys.* **118**, 7651 (2003).

[110] W. Hu, D. Frenkel, and V. B. F. Mathot. *J. Chem. Phys.* **118**, 3455 (2003).

[111] I. N. Stranski and D. Totomanov. *Z. Phys. Chem.* **163**, 399 (1933).

[112] U. Gasser, E. R. Weeks, A. Schofield, P. N. Pusey, and D. A. Weitz. *Science* **292**, 258 (2001).

[113] A. van Blaaderen, J. P. Hoogenboom, D. L. J. Vossen, A. Yethiraj, A. van der Horst, K. Visscher, and M. Dogterom. *Faraday Discuss.* **123**, 107 (2003).

[114] S. Toxvaerd. *J. Chem. Phys.* **117**, 10303 (2002).

[115] M. Miyahara and K. E. Gubbins. *J. Chem. Phys.* **106**, 2865 (1997).

[116] V. Talanquer and D. W. Oxtoby. *J. Chem. Phys.* **114**, 2793 (2001).

[117] C. Qiu, T. Qian, and W. Ren. *J. Chem. Phys.* **129**, 154711 (2008).

[118] M. W. Maddox and K. E. Gubbins. *J. Chem. Phys.* **107**, 9659 (1997).

[119] B. Husowitz and V. Talanquer. *J. Chem. Phys.* **121**, 8021 (2004).

[120] B. Husowitz and V. Talanquer. *J. Chem. Phys.* **122**, 194710 (2005).

[121] V. E. Smorodin. *Langmuir* **10**, 2250 (1994).

[122] A. Valencia and R. Lipowsky. *Langmuir* **20**, 1986 (2004).

[123] A. Cacciuto and D. Frenkel. *Phys. Rev. E* **72**, 041604 (2005).

[124] H. Wang, H. Gould, and W. Klein. *Phys. Rev. E* **76**, 031604 (2007).

[125] A. Cacciuto, S. Auer, and D. Frenkel. *Nature* **428**, 404 (2004).

[126] R. P. Sear. *EPL* **83**, 66002 (2008).

[127] A. van Blaaderen, R. Ruel, and P. Wiltzius. *Nature* **385**, 321 (1997).

[128] I. B. Ramsteiner, K. E. Jensen, D. A. Weitz, and F. Spaepen. *Phys. Rev. E* **79**, 011403 (2009).

[129] V. W. A. de Villeneuve, R. P. A. Dullens, D. G. A. L. Aarts, E. Groeneveld, J. H. Scherff, W. K. Kegel, and H. N. W. Lekkerkerker. *Science* **309**, 1231 (2005).

[130] V. W. A. de Villeneuve, D. Verboekend, R. P. A. Dullens, D. G. A. L. Aarts, W. K. Kegel, and H. N. W. Lekkerkerker. *J. Phys.: Condens. Mat.* **17**, S3371 (2005).

[131] L. D. Gelb, K. E. Gubbins, R. Radhakrishnan, and M. Sliwinska-Bartkowiak. *Rep. Progr. Phys.* **62**, 1573 (1999).

[132] N. E. Chayen, E. Saridakis, R. El-Bahar, and Y. Nemirovsky. *J. Molec. Bio.* **312**, 591 (2001).

[133] N. E. Chayen, E. Saridakis, and R. P. Sear. *Proc. Nat. Acad. Sci. (USA)* **103**, 597 (2006).

[134] D. F. Rosenbaum, P. C. Zamora, and C. F. Zukoski. *Phys. Rev. Lett.* **76**, 150 (1996).

[135] J. A. van Meel, A. J. Page, R. P. Sear, and D. Frenkel. *J. Chem. Phys.* **129**, 204505 (2008). See Chapter 4 of this Thesis.

[136] A. Page and R. Sear. *in prep.* (2009).

[137] A. I. Hienola, P. M. Winkler, P. E. Wagner, H. Vehkamaki, A. Lauri, I. Napari, and M. Kulmala. *J. Chem. Phys.* **126**, 094705 (2007).

[138] P. M. Winkler, A. Hienola, G. Steiner, G. Hill, A. Vrtala, G. P. Reischl, M. Kulmala, and P. E. Wagner. *Atmos. Research* **90**, 187 (2008).

[139] P. Lenz and R. Lipowsky. *Phys. Rev. Lett.* **80**, 1920 (1998).

[140] M. Heni and H. Löwen. *Phys. Rev. E* **60**, 7057 (1999).

[141] M. Heni and H. Löwen. *Phys. Rev. Lett.* **85**, 3668 (2000).

[142] J. A. van Meel and D. Frenkel. *in prep.* (2009). See Chapter 5 of this Thesis.

[143] H. Liu, S. K. Kumar, and F. Sciortino. *J. Chem. Phys.* **127**, 084902 (2007).

[144] L. Vega, E. de Miguel, L. F. Rull, G. Jackson, and I. A. McLure. *J. Chem. Phys.* **96**, 2296 (1992).

[145] J. R. Elliott and L. Hu. *J. Chem. Phys.* **110**, 3043 (1999).

[146] M. H. J. Hagen and D. Frenkel. *J. Chem. Phys.* **101**, 4093 (1994).

[147] M. Dijkstra. *Phys. Rev. E* **66**, 021402 (2002).

[148] L. Mederos and G. Navascues. *J. Chem. Phys.* **101**, 9841 (1994).

[149] D. Fu, Y. Li, and J. Wu. *Phys. Rev. E* **68**, 011403 (2003).

[150] G. A. Vliegenthart, J. F. M. Lodge, and H. N. W. Lekkerkerker. *Physica A* **263**, 378 (1999).

[151] D. F. Rosenbaum and C. Zukoski. *J. Cryst. Growth* **169**, 752 (1996).

[152] M. G. Noro and D. Frenkel. *J. Chem. Phys.* **113**, 2941 (2000).

[153] D. L. Pagan, M. E. Gracheva, and J. D. Gunton. *J. Chem. Phys.* **120**, 8292 (2004).

[154] J. F. Lutsko and G. Nicolis. *J. Chem. Phys.* **122**, 244907 (2005).

[155] D. Costa, P. Ballone, and C. Caccamo. *J. Chem. Phys.* **116**, 3327 (2002).

[156] D. Frenkel and P. ten Wolde. In M. Lal, R. A. Mashelkar, B. D. Kulkarni, and V. M. Naik (editors), *Structure and Dynamics of Materials in the Mesoscopic Domain*, pp. 139–149. Imperial College Press - The Royal Society (1999).

[157] I. N. Stranski and L. V. Krastanow. *Abh. Math.-Nat. Klasse* **146**, 797 (1939).

[158] T. Schenk, D. Holland-Moritz, V. Simonet, R. Bellissent, and D. M. Herlach. *Phys. Rev. Lett.* **89**, 075507 (2002).

[159] J. P. K. Doye and D. J. Wales. *J. Phys. B: At., Mol. Opt. Phys.* **29**, 4859 (1996).

[160] S. Mossa and G. Tarjus. *J. Chem. Phys.* **119**, 8069 (2003).

[161] S. Mossa and G. Tarjus. *J. Non-Cryst. Sol.* **352**, 4847 (2006).

[162] A. Di Cicco, A. Trapananti, S. Faggioni, and A. Filipponi. *Phys. Rev. Lett.* **91**, 135505 (2003).

[163] T. Aste, M. Saadatfar, and T. J. Senden. *Phys. Rev. E* **71**, 061302 (2005).

[164] A. Di Cicco and A. Trapananti. *J. of Non-Cryst. Sol.* **353**, 3671 (2007).

[165] T. Kondo and K. Tsumuraya. *J. Chem. Phys.* **94**, 8220 (1991).

[166] N. Jakse and A. Pasturel. *Phys. Rev. Lett.* **91**, 195501 (2003).

[167] A. V. Anikeenko and N. N. Medvedev. *Phys. Rev. Lett.* **98**, 235504 (2007).

[168] A. V. Anikeenko, N. N. Medvedev, and T. Aste. *Phys. Rev. E* **77**, 031101 (2008).

[169] J. P. J. Michels and N. J. Trappeniers. *Phys. Lett. A* **104**, 425 (1984).

[170] M. Bishop and P. A. Whitlock. *J. Chem. Phys.* **123**, 014507 (2005).

[171] R. J. Speedy. *J. Phys.: Condens. Mat.* **10**, 4387 (1998).

[172] H. S. M. Coxeter. *Regular Polytopes*. Dover, New York (1973).

[173] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*. Springer-Verlag, New York (1988).

[174] O. R. Musin. *Ann. Math.* **168**, 1 (2008).

[175] F. Pfender and G. M. Ziegler. *Notices of AMS* **51**, 873 (2004).

[176] M. Skoge, A. Donev, F. H. Stillinger, and S. Torquato. *Phys. Rev. E* **74**, 041127 (2006).

[177] W. B. Streett, H. J. Raveche, and R. D. Mountain. *J. Chem. Phys.* **61**, 1960 (1974).

[178] J.-L. Colot and M. Baus. *Phys. Lett. A* **119**, 135 (1986).

[179] P. R. ten Wolde, M. J. Ruiz-Montero, and D. Frenkel. *J. Chem. Phys.* **104**, 9932 (1996).

[180] S. Alexander and J. McTague. *Phys. Rev. Lett.* **41**, 702 (1978).

[181] J. A. van Meel, D. Frenkel, and P. Charbonneau. *Phys. Rev. E* **79**, 030201 (R) (2009). See Chapter 7 of this Thesis.

[182] R. Finken, M. Schmidt, and H. Löwen. *Phys. Rev. E* **65**, 016108 (2001).

[183] M. Luban and J. P. J. Michels. *Phys. Rev. A* **41**, 6796 (1990).

[184] L. Lue. *J. Chem. Phys.* **122**, 044513 (2005).

[185] L. Lue and M. Bishop. *Phys. Rev. E* **74**, 021201 (2006).

[186] M. Bishop and P. A. Whitlock. *J. Stat. Phys.* **126**, 299 (2007).

[187] J. A. van Meel, B. Charbonneau, A. Fortini, and P. Charbonneau. *in prep.* (2009).

[188] S. Torquato and F. H. Stillinger. *Phys. Rev. E* **73**, 031106 (2006).

[189] R. D. Kamien and A. J. Liu. *Phys. Rev. Lett.* **99**, 155501 (2007).

[190] G. Parisi and F. Zamponi. *arXiv* p. 0802.2180 (2008).

[191] B. I. Halperin and D. R. Nelson. *Phys. Rev. Lett.* **41**, 121 (1978).

[192] D. Frenkel and J. P. McTague. *Phys. Rev. Lett.* **42**, 1632 (1979).

[193] S. Auer and D. Frenkel. *J. Chem. Phys.* **120**, 3015 (2004).

[194] J. H. Conway and N. J. A. Sloane. *Discr. Comput. Geom.* **13**, 383 (1995).

[195] P. G. Bolhuis, D. Frenkel, S. C. Mau, D. A. Huse, and L. V. Woodcock. *Nature* **388**, 235 (1997). 0028-0836.

[196] D. R. Nelson and J. Toner. *Phys. Rev. B* **24**, 363 (1981).

[197] M. Hamermesh. *Group Theory and its Application to Physical Problems*. Dover Publications, New York (1962).

[198] D. R. Nelson and M. Widom. *Nucl. Phys. B* **240**, 113 (1984).

[199] D. J. Rowe, P. S. Turner, and J. Repka. *J. Math. Phys.* **45**, 2761 (2004).

[200] G. E. Andrews, R. Askey, and R. Roy. *Special functions*, volume 71 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge (1999).

[201] G. F. Voronoï. *J. Reine Ang. Math.* **133**, 97 (1907).

[202] J. P. J. M. van der Eerden, J. Makkinje, and T. J. H. Vlugt. *J. Cryst. Growth* **275**, 83 (2005). Proceedings of the 14th International Conference on Crystal Growth and the 12th International Conference on Vapor Growth and Epitaxy.

[203] J. H. Los, L. M. Ghiringhelli, E. J. Meijer, and A. Fasolino. *Phys. Rev. B* **72**, 214102 (2005).

[204] L. M. Ghiringhelli, C. Valeriani, E. J. Meijer, and D. Frenkel. *Phys. Rev. Lett.* **99**, 055702 (2007).

[205] S. Pion and M. Teillaud. In C. E. Board (editor), *CGAL User and Reference Manual*. 3.4 edition (2008).

[206] J. P. Troadec, A. Gervois, and L. Oger. *EPL* **42**, 167 (1998).

[207] D. Bandyopadhyay and J. Snoeyink. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 410–419. Society for Industrial and Applied Mathematics, New Orleans, Louisiana (2004).

[208] D.-Q. Yu, M. Chen, and X.-J. Han. *Phys. Rev. E* **72**, 051202 (2005).

Summary

First-order phase transitions often start with nucleation, which refers to the spontaneous formation of a microscopic amount of the new phase due to thermal fluctuations. If the size of such a fluctuation exceeds a certain threshold, the so-called critical nucleus size, the new phase has a high probability to grow to macroscopic dimensions. Although this process is well understood from a phenomenological perspective, the design of nucleation agents to facilitate the growth of high-quality crystals requires insight into the physical mechanism on the microscopic scale.

The research presented in this Thesis aims to assess the physical mechanism of nucleation on the microscopic scale using many-particle simulations. The main emphasis of this work is on homogeneous and heterogeneous crystal nucleation of nano-colloids and proteins from dilute solution, but also other aspects of this field of research are addressed, such as alternative compute hardware, nearest neighbor algorithms, and the origin of geometrical frustration in liquids.

In Chapter 3 we discuss a topic common to all numerical simulations: the compute hardware. Simulations are typically performed on a computer's *Central Processing Unit* (CPU), such as a Pentium® or an Athlon® processor. The CPU's architecture is highly optimized for the execution of complex applications, dedicating significantly more transistors to program flow control and data caches than to so-called arithmetic logic units (ALUs), which perform integer and floating point operations. However, numerical simulations are often limited by the preformance of these ALUs. Recent fully programmable *Graphic Processing Units* (GPUs) offer an alternative to CPUs as compute hardware. Driven by the demands of 3D video games they are designed to process large amounts of image data in parallel. In Chapter 3 we demonstrate that a conventional Molecular Dynamics simulation can be rewritten to run entirely on a GPU reducing its run-time by up to a factor 80 for an n-squared algorithm. Further, we show how cell lists domain decomposition can be parallelized efficiently resulting in a maximum 35-fold speed-up. We conclude this chapter with a discussion on the advantages and limitations of graphics hardware for the use in scientific simulations.

The next chapter concerns homogeneous crystal nucleation of nano-colloids from

dilute solution. The model system consists of spherical particles interacting via a Lennard-Jones pair potential under conditions below the triple point. Here, the vapor is the meta-stable parent phase, the crystal the stable final phase, and the liquid is an intermediate meta-stable phase with a free energy inbetween the other two phases. Using forward-flux sampling paired with a local bond-order parameter we find no evidence for a direct vapor-crystal phase transition. Instead, using a density-based order parameter, we find the transition to proceed via a two-step process: first a meta-stable liquid droplet forms and subsequently a crystal emerges within this droplet. Both nucleation events are independent and can be treated separately. Our findings confirm earlier simulations based on quasi-equilibrium umbrella sampling. Our simulations reveal that a minimum droplet size is required to host a stable crystal cluster, and that the crystal is always covered by a liquid mono-layer. The overall nucleation rate is limited by the vapor-liquid nucleation step, and hence depends sensitively on the vapor pressure. We compare our simulation results to predictions from classical nucleation theory (CNT) and literature results using umbrella sampling. We find good qualitative agreement, which suggests that this phase transition can be treated by an equilibrium theory.

In Chapter 5 we extend the study on vapor-crystal nucleation of Lennard-Jones particles to the case of heterogeneous nucleation on smooth weakly-adsorbing surfaces. We investigate the microscopic mechanism of nucleation in the presence of an attractive planar wall, planar circular patch, and hemi-spherical pore. Nucleation in a pore is of particular interest due to experimental evidence that a porous medium may function as "universal" nucleation agent for protein crystallization. As in the case of homogeneous nucleation we observe the vapor-crystal nucleation pathway to proceed via a intermediate liquid phase, and that both nucleation events can be treated independently. For liquid nucleation, we find good qualitative agreement with CNT. Only for the circular patch there exists a narrow range of parameters for which CNT predicts a weak double-peak in the free-energy barrier that is not reproduced by our simulations. We find crystal nucleation to be spontaneous, that is without any noticable barrier, for both the planar wall and the circular patch, provided the liquid droplet exceeds a minimum size. For nucleation in a hemi-spherical pore CNT predicts a double-peaked free-energy barrier for a wider range of parameters, and here the behavior is nicely reproduced by our simulation. We find that the first peak corresponds to filling the (microscopic) pore, and the second barrier separates the liquid from growing into the bulk. On increasing the pore radius the computed overall nucleation rate goes through a minimum before it approaches, in the limit of an infinitely large pore, the rate for nucleation on a planar wall. In contrast to the planar wall and circular patch, crystal nucleation is not induced by the surface of the pore, but proceeds homogeneously in the bulk. Although it is possible for a crystal to form within the pore, our simulations suggest that it is much more likely that crystal nucleation

occurs after the liquid has grown out of the pore. In conclusion, our simulations provide further evidence that porous media with a broad distribution of pore sizes can significantly enhance crystal nucleation from dilute solution, and that the process is dominated by a few pores with optimal size.

Chapter 6 builds upon the experiences from the last two chapters and investigates heterogeneous nucleation of a model protein at smooth and structured surfaces. Due to the very short-range nature of the "protein" particles' interaction the phase diagram features only the fluid and the crystal as thermodynamically stable phases, and the meta-stable vapor-liquid binodal is hidden within the fluid-crystal coexistence region. Under conditions where protein crystallization is observed experimentally, the liquid is unstable with respect to both the fluid and the crystal, which contrasts the situation in the previous two chapters. As a consequence, at a smooth planar wall and patch we find direct vapor-crystal nucleation without the liquid as intermediate phase. Moreover, it appears that a smooth hemi-spherical pore does not facilitate crystallization at all. Instead, particles adsorb to the pore's surface in a packing that is incommensurable with any periodic structure, effectively inhibiting further crystal growth. At a periodically structured surface, represented by a 111-fcc surface of particles that are immobile but otherwise identical to the "protein" particles, crystallization is almost instantly induced provided the lattice mismatch between template and protein crystal is not too large. Interestingly, a large lattice mismatch leads to the adsorption of a thin layer of liquid. This effect is even more pronounced at a disordered surface. For cylindrical cavities with a disordered surface and suitable radius the liquid fills the pore. In such a geometry the free-energy barrier to crystallization is significantly reduced as compared to homogeneous nucleation, and a crystal nucleus can form fully immersed in the liquid allowing for stress-free growth. These results suggest that the molecular mechanism for crystal nucleation induced by micro-porous media relies on the combined effect of liquid adsorption and pore filling, and hence depends on both pore size and surface roughness.

The final chapter of this Thesis discusses the origin of geometrical frustration in liquids. Geometrical frustration is conjectured to prevent crystallization and therefore help glass formation. It arises from a competition between the local and global packing. Locally, particles can arrange to achieve a density higher than that of a crystal. But such a local order cannot be used to fill space. Once the system density exceeds a certain threshold a global packing becomes favorable. During the crystallization process particles have to leave their local arrangement in order to adapt to the global lattice. For hard-sphere systems this leads to the entropic free-energy barrier to crystal nucleation. This contrasts the behavior of two-dimensional hard disks, for which hexagonal packing is both locally and globally prefered, and crystallization is particularly easy. In three-dimensional (3D) Euclidean space, there exist two local structures that are considered to cause geometrical frustration. The first is the icosa-

hedron as the smallest maximum kissing-number Voronoi polyhedron, and the second is the tetrahedron as the smallest volume that can show up in a Delaunay tessellation. Because the icosahedron can be constructed from almost-perfect tetrahedra, both are often used interchangeably, which leaves the true origin of frustration unclear. In Chapter 7 we leave the familiarity of our 3D world and investigate crystallization of four-dimensional (4D) hard (hyper-)spheres, because in 4D the smallest maximum kissing-number polyhedron *is* commensurable with the densest crystal lattice. This allows us to assess whether it is the icosahedron or the tetrahedron that causes the frustration. Our simulations reveal a free-energy barrier that is significantly larger in 4D than in 3D at comparable conditions, clearly identifying tetrahedral order as the origin of frustration. Moreover, the high free-energy barrier to crystallization makes 4D hard spheres a suitable model system to test theories of glass transition and jamming. In the appendix of this chapter we extend the discussion to 5D and 6D hard spheres presenting equations of state and quantifying the growing dissimilarity between higher-dimensional fluid and crystal phases using second- and third-order bond-order invariants.

This Thesis is concluded with an Appendix on nearest neighbor algorithms, where we propose a simple parameter-free algorithm based on the sum of solid angles. The determination of the nearest neighbors of a given particle is a recurrent task in both numerical simulations and structural analysis. Algorithms can be classified as either tunable or parameter-free, with their most prominent representatives the fixed-distance cutoff (CUTOFF) and the Voronoi construction, respectively. The former one considers two particles to be neighbors if their distance is smaller than a manually-tuned cutoff value. Using a common domain decomposition scheme, such as cell lists or Verlet lists, the CUTOFF algorithm is very efficient and therefore often used during simulation. However, each time the model system or working conditions have changed the cutoff distance has to be tuned again. In contrast, the Voronoi construction is based on a purely geometrical construction and requires no manual input. Its disadvantage is its significantly increased computational cost and its sensitivity to positional perturbations caused by thermal fluctations. With the solid-angle based nearest neighbor (SANN) algorithm we propose a method that is parameter-free, but computationally less expensive than the Voronoi construction. Its disadvantages are that it does not preserve neighbor symmetry, which means that if A is a neighbor of B, B does not have to be a neighbor of A. We apply the three methods to the liquid and crystal phases of a Lennard-Jones and a carbon system, and also investigate their behavior at a two-phase interface. We find that SANN works reliably on all systems tested. Moreover, compared to the Voronoi construction SANN is more robust against thermal fluctuations and, in the case of graphite, does not include spurious neighbors from neighboring layers. With these properties we conclude that SANN is suitable for the use during simulation and for the analysis of protein structures.

Samenvatting (Dutch summary)

Een faseovergang, zoals het bevriezen van water, begint met nucleatie. Dat betekent dat een microscopische hoeveelheid van de nieuwe fase, een zogenoemde kiem, spontaan ontstaat door middel van thermische fluctuaties. Als zo een fluctuatie groot genoeg is, dat wil zeggen uitkomt boven de *kritieke kiemgrote*, dan is de kans groot dat de nieuwe fase doorgroeit tot een macroscopische hoeveelheid. Ondanks dat nucleatie goed is begrepen vanuit een fenomenologisch oogpunt ontbreekt er gedetaileerd inzicht in het microscopisch gedrag. Maar juist het microscopisch mechanisme is belangrijk voor bijvoorbeeld de voorspelling van juiste omstandigheden voor het kweken van hoogwaardige kristallen.

Het doel van het onderzoek in dit proefschrift is dan ook nader inzicht te verkrijgen in nucleatie op de schaal van enkele deeltjes met behulp van computersimulaties. Het oogmerk ligt vooral op homogene en heterogene nucleatie van kristallisatie vanuit de gas fase in modelsystemen voor nano-colloïden en eiwitten. Maar ook andere aspecten van dit onderzoeksgebied worden verder uitgewerkt, zoals nieuwe mogelijkheden in computer hardware voor simulaties, een algoritme ter bepaling van de dichst bijzijnde buren, en de oorsprong van geometrische frustratie in vloeistoffen. Na een korte introductie geeft hoofdstuk 2 een overzicht over de theorieën van nucleatie en zeldzame gebeurtenissen, en de numerieke technieken die in dit proefschrift zijn gebruikt.

Hoofdstuk 3 bevat de eerste onderzoeksgegevens en gaat over nieuwe ontwikkelingen op het gebied van computer hardware. Grafische kaarten (eng.: graphic processing unit, GPU) zijn oorspronkelijk ontwikkelt voor het bewerken van 3D (video-)beelden op de computer. Gedreven door de enorme interesse in computerspelletjes heeft de GPU in de afgelopen 20 jaren een enorme ontwikkeling ondergaan, van een simpele chip tot een krachte rekenmachine. Hedendaagse GPUs bezitten een vermogen dat ver uitsteekt boven de gebruikelijke centrale processor (eng.: central processing unit, CPU), en kunnen sinds kort ook voor toepassingen buiten de grafische wereld gebruikt worden. Hun kracht ontlenen zij aan parallellisatie, wat alleen voordelig is als een programma dat ook kan benutten. Zoals wij in hoofdstuk 3 laten zien is de moleculaire dynamica (MD) simulatie uitermate geschikt voor een

GPU. Vergeleken met de traditionele CPU-versie is een GPU-versie tot en met 80 keer sneller, waardoor berekeningen van een jaar in enkele weken voltooid kunnen worden.

In hoofdstuk 4 wordt de homogene kristallisatie van nano-colloïde in verdunde oplossing uitgewerkt. Als model dient een systeem van bolvormige deeltjes die via een Lennard-Jones potentiaal wisselwerken, dat heet met zachte aantrekking op korte afstand en een sterke afstoting op hele korte afstand. Bij een temperatuur en druk waaronder het kristal de thermodynamisch stabiele fase is, beginnen wij een gas te simuleren. Met behulp van "forward-flux sampling", een techniek om zeldzame gebeurtenissen te simuleren, kunnen wij de overgang van gas naar kristal forceren en gericht naar het gedrag van de deeltjes kijken. Hierbij vinden wij dat de gasdeeltjes niet direct een kristal vormen. In plaats daarvan ontstaat er eerst een vloeistof druppel. Deze groeit dan door tot dat een bepaald volume bereikt is. Pas dan is kristallisatie mogelijk. En zelfs als de druppel volledig is gekristalliseerd is het kristal nog bedekt met een dun laagje vloeistof.

Voor homogene kristallisatie ontdekten wij ook dat de snelheid van de gas→kristal overgang gelimiteerd wordt door de gas→vloeistof overgang. Als deze vereenvoudigd wordt kan de gehele overgang versneld worden. Dit is dan ook het doel van het onderzoek in hoofdstuk 5. Hier kijken wij naar heterogene nucleatie, dus met behulp van verontreinigingen zoals een stofdeeltje of een muur. In hetzelfde modelsysteem als in hoofdstuk 4 plaatsen wij een muur met een licht absorberend oppervlak, zodat de deeltjes bij voorkeur in de buurt van de muur zitten. Onze simulaties tonen aan dat het bevriezingsproces zich hetzelfde gedraagt als bij homogene nucleatie, maar dan wel vele malen sneller. De gas→vloeistof overgang hangt sterk af van de aantrekkingskracht van de muur, en het kristal vormt bijna onmiddellijk aan de muur zodra voldoende vloeistof ter beschikking staat. Wij zijn bijzonder geïnteresseerd in het gedrag bij een absorberende porie, omdat experimenten hebben aangetoond dat een poreus materiaal uitstekend eiwitkristallisatie bevordert. In ons model vinden wij dat een porie met de juiste dimensies de kristalvorming niet direct vereenvoudigt, maar wel de drempel voor vloeistofnucleatie sterk verlaagt. Wij zien dan ook dat het gehele gas→kristal proces in een optimale porie zelfs sneller kan verlopen dan bij een gladde muur. Deze resultaten steunen recente theorieën over de werkwijze van poreus materiaal.

In onze simulaties kunnen wij de experimenten op het gebied van eiwitkristallisatie nog iets verder tegemoet komen. Door de afstand voor effectieve attractie van onze deeltjes nog iets te verkorten krijgen wij een gedrag dat typisch is voor een groot aantal bolvormige eiwitten. In tegenstelling tot nano-colloïden hebben eiwitten geen thermodynamisch stabiele vloeistof fase, waardoor de overgang gas→vloeistof→kristal niet waarschijnlijk lijkt. Er is dan wel bekend dat sterke dichtheidsfluctuaties, zoals in de buurt van het kritieke punt, de kristallisatie bevorderen, maar een echte twee-

stappen overgang is niet bekend. Onze simulaties tonen dan ook dat de overgang bij een absorberende muur direct van gas naar kristal gaat. Het nadeel van dit soort gedrag is dat het sterk afhangt van de wisselwerking tussen eiwit en muur, en dat verschilt per eiwit. Deze hypothese wordt gesteund door experimenten die laten zien dat een muur bijna nooit effect heeft. Verrassend is dat kristallisatie in een porie helemaal niet bevorderd wordt. In plaats van een kristal ontstaat een wanordelijke structuur die incompatibel is met het kristalrooster.

Zover hebben wij alleen naar gladde oppervlakken gekeken. Maar op de schaal van een eiwit is het veel waarschijnlijker dat een oppervlak ruw of gestructureerd is. Een interessant resultaat blijkt uit onze berekening met een ruw oppervlak: in plaats van een kristal absorbeert een dun vloeistoflaagje. Deze wisselwerking is minder specifiek, en de vloeistoflaag kan ook kristalgroei bevorderen. Onze simulaties laten zelfs zien dat zich in een porie met een ruw oppervlak genoeg vloeistof kan verzamelen om een kristal groter dan de kritieke kiem te laten groeien. Dit verklaart waarom poreus materiaal ook bij eiwitten efficiënt kristalnucleatie induceert.

In het laatste hoofdstuk van dit proefschrift verlaten wij onze bekende driedimensionale (3D) wereld en bekijken kristalnucleatie van harde bollen in vier dimensies (4D). Dit bijzondere systeem geeft de mogelijkheid een antwoord te krijgen op een oude vraag over de oorsprong van “geometrische frustratie” in 3D vloeistoffen. Geometrische frustratie benoemt het conflict tussen locale en globaal orde van deeltjes. Lokaal kunnen enkele deeltjes zich heel dicht ordenen, maar deze groepjes kunnen niet als bouwstenen gebruikt worden om de ruimte geheel te vullen, waardoor er gaten onstaan. De dichtheid van het gehele systeem kan verhoogd worden wanneer alle deeltjes op een globaal rooster zitten, en dus een kristal vormen. Tijdens de overgang van vloeistof naar kristal moeten deeltjes uit hun lokale orde overstappen op het globale rooster. Dit proces verlaagt tijdelijk de entropie, wat leidt tot de vrije energiebarrière bij kristallisatie in harde bollen. In twee dimensies met harde schijven bestaat dit conflict niet, want het hexagoon is de dichtste lokale en globale pakking, en kristallisatie gaat heel makkelijk. Over de oorzaak van geometrische frustratie in 3D concurreren twee theorieën. De oudere theorie beschouwt de meest efficiënte pakking rond een centraal deeltje, het icosaëder, als het belangrijkste lokale arrangement. De nieuwere theorie beweert dat het tetraëder, het kleinste volume-element in drie dimensies, verantwoordelijk is voor de frustratie. Het bijzondere van 4D harde bollen is dat de meest efficiënte pakking rond een centraal deeltjes juist het kristalrooster vormt, en alleen het (hyper-)tetraëder als potentiële oorzaak voor frustratie overblijft. 4D harde bollen zijn dus een ideaal testsysteem om deze vraag uit te zoeken. Uit onze simulaties blijkt dat kristallisatie in 4D nog veel moeilijker is dan in 3D bij vergelijkbare condities. Daardoor is aangetoond dat het tetraëder de sleutel is tot frustratie. Door deze hoge barrière voor kristallisatie blijken 4D harde bollen ook nog eens een goed testsysteem te zijn voor theorieën over de glasovergang en “jam-

ming". In de appendix van dit hoofdstuk breiden wij de discussie uit tot 5D en 6D systemen om enkele theorische voorspellingen voor hoger-dimensionale vloeistoffen te testen.

Dit proefschrift sluit af met een appendix over algoritmen ter bepaling van de dichst bijzijnde buren. Informatie over de buren van een deeltje is van belang tijdens een simulatie ter berekening van wisselwerkingen, maar ook nadien voor structurele analyse. Er wordt onderscheid gemaakt tussen afstembare en parameter-vrije algoritmen. De bekendste vertegenwoordiger van het eerste type is de afstembare maximale afstand drempel (eng.: fixed-distance cutoff), een snel algoritme dat veel wordt gebruikt tijdens simulaties. Het nadeel is dat telkens wanneer het systeem of de thermodynamische condities veranderd wordt de afstand opnieuw afgestemd moet worden. Dit nadeel is niet van toepassing voor de Voronoi constructie, een parameter-vrije algoritme gebaseerd op een puur geometrische constructie. Het nadeel hiervan is dat de berekening duur is en het gevoelig is voor deeltjes fluctuaties. Als alternatief van beiden stellen wij in deze appendix een nieuw parameter-vrij algoritme voor. Deze is gebaseerd op een simpelere constructie, waardoor de berekening duidelijk sneller wordt, maar niet alle symmetrieën bewaard blijven. Wij vergelijken de drie algoritmen door ze toe te passen op vloeistoffen en kristallen uit Lennard-Jones en koolstof deeltjes. Wij laten zien dat ons algoritme betrouwbaar functioneert en, vergeleken met de Voronoi constructie, goedkoper in rekentijd en minder gevoelig voor fluctuaties is. De eigenschappen van ons algoritme wijzen erop dat het uitermate geschikt is voor gebruik in simulaties, maar ook voor structurele analyse van eiwitten.

Acknowledgements

Over the last four years there were many people who contributed directly or indirectly to this work. Some provided scientific input, some spiritual support, and others made Amolf such an enjoyable working place for me. And although I know I can never thank everyone, in what follows I try to pay some tribute to these people.

First of all my gratitude goes to my supervisor Daan. He managed the keep this fragile balance between guidance and freedom. This allowed me to make mistakes, discover many aspects of science, and to explore a variety of different topics. But at the same time, whenever I got lost in the dark jungle of options, he provided the light that got me back on track. He posesses the invaluable ability to provide - within a flick of an eye - excellent feedback on a scientific problem, and re-establish both confidence and motivation.

During my interview at Amolf I was foretold that much scientific guidance comes from discussions with other "group" members, a statement I am happy to confirm. In this regard I like to mention Ana, Axel, Bela, Bianca, Chantal, Kostya, Live, Patrick, Pieter Rein, Sanne, and Simon. They were always willing to explain a C++ algorithm, a simulation technique, a mathematical derivation, to discuss any problem at hand, or to listen to my wildest theories. But I also like to include my collaborators from outside Amolf: Amanda, Benoit, Richard, Robert, and Simon. They broadened my perspective in our plenty discussions by challenging my point of view, be it on the nucleation rate units, the implementation of higher-dimensional third-order invariants, or the advantages of CUDA over CG. And when it came to finally writing down all parts for this booklet, I gratefully acknowledge Ana, David, Dorit, Filipe, Laurens, Nienke, Patrick, Pieter Rein, and Wiet for a careful reading, Rutger for his thesis template, and Kostya for his Latex support.

But scientific input is not the only ingredient. It is the warm, open, cheerful atmosphere that made working here very special to me, achieved by the fun I had with all the "overloop" people. Whether there was cake during coffee, a discussion at lunch, an "occasional" afternoon espresso, a chat in the sun on the grass, one of the countless parties, or a "PV uitje", I always felt happy and surrounded by friends. And a special thanks goes to Nienke, my room mate over the past four years. Her

joyful spirit never failed to make me see things in a positive way, and I shed quite some tears with her rolling on the floor laughing.

And still more is needed to successfully accomplish a thesis: it lies with the people in the support facilities. Here, I like to mention Jan and Lars, who helped me in my countless attemps to use the clusters, and Wouter, how made all the legal issues of moving to Amsterdam look like a piece of cake. But there are many others that deserve my thanks, like the reception, the storage, the accounting, the catering, and many more, who make everything at Amolf run smoothly.

So far I only mentioned people I met during my time at Amolf. But work means nothing without a chance to relax from its strains, both mentally and physically. Nothing tastes better than a beer after an exhausting hour of squash with Chris, Evert-Jan, Fedor, Gabriel, Jeroen, and Tom, or during the victorious WoW sessions with Christof, Florian, and Matthias. Or in a bar in Vienna with Ahmed, Alex, Artur, Boris, Cornelia, Florian, Iris, Karin, Lukas, May, Philip, Patricia, Sonja, and Teresa, who rearranged their schedules on short notice to make time for "a beer" with me when I was around. And last but not least I want to thank my family, who offered their support and encouragement whenever needed, and of course Dorit, who always stood by my side as my pillar of strength.

In the end, I apologize to all those people I did not name here explicitely. If you feel you should have been mentioned, consider yourself placed on these pages, and accept my sincere gratitude.

About the author

Born in Amsterdam, The Netherlands, Jacobus van Meel grew up in Biberach a. d. Riss, Germany. He obtained his secondary education at the Wieland Gymnasium Biberach, where he followed initially the language track (Latin, English, French), but in 1999 he graduated with a major in physics and maths. From 1997 to 1998 he was also a participant of the MINT course for talented maths and physics students at the University of Ulm. For his studies he moved to Vienna, Austria, where he obtained his MSc in physics with distinction under supervision of Prof. dr. H. Posch. During his studies he worked from 2001 to 2003 part time as a Software Architect for BOC Information Systems, an international company specialized in IT-based management consulting. Since 2005 he works as a PhD student in the computational physics group of Prof. dr. D. Frenkel at the FOM institute for atomic and molecular research (AMOLF) in Amsterdam. In his spare time, he enjoys playing Squash or planning the next wilderness canoeing tour.