Clustering and nucleation in metastable fluids of hard polyhedra

by

Samanthule Nola

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Macromolecular Science and Engineering)
in The University of Michigan
2016

Doctoral Committee:

Professor Sharon C. Glotzer, Chair
Professor Ronald G. Larson
Assistant Professor Xiaoming Mao
Professor Michael J. Solomon

For Ruth and Orion.

ACKNOWLEDGEMENTS

Many people have contributed to this work. First my advisor, Sharon Glotzer, for making this possible, for suggesting new approaches, providing a foundation and framework for thinking about entropy, and for putting together an amazing group where I have been able to learn the computational side of research. Thanks to Karen Coulter for making it all possible. Thanks to those who have funded this work; the NSF and IGERT. Thanks to my group mates for numerous discussions and resources. Particularly Richmond and Julia. Also Wenbo, Eric, Jens, Greg, Caroline, Josh, and others, for helping shape my understanding of entropic self assembly, simulations, software design, and more. Thanks also to colleagues in other groups: Matt, Steve, Shaurjo, Yansha, Mark, Katie, Mike, Gun-Ho, Ying, and too many more to mention here.

Thanks to Rick Laine for getting me interested in graduate research, and Max Shtein for opportunities to conduct experimental research and help learning to present research effectively to different audiences. Thanks to the professors in the UofM MSE department, particularly Michael Falk and Tresa Pollock for encouraging my interest in materials science and teaching the fundamentals I have built on, as well as professor Halloran for research experience and professor Thornton for a first introduction to computational research. To Mark Batell for showing me the elegance of describing the world mathematically, and Jason Davis who (inadvertently) led me to pursue mathematics beyond a few required classes. Thanks to Dave for sharing enthusiasm for physics and engineering, and making university seem like a possible goal.

Thanks to Jimmy, who has worked with me toward shared goals, and been a part of my career and educational goals for many years. Thanks to Ally who has given me back the sense of belonging I lost for awhile. To my family, particularly my father who never made *too* much of the fact that he defended his dissertation on the day I was born, and Ivan, who likes to talk about math to anyone who will listen. Thanks to the Morris community.

Thanks to Ruth, for sharing her growing understanding of the world with me and reminding me to relax. And to Orion, for sharing his enthusiasm for everything with me, and reminding me to play.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF APPENDICES

Appendix

LIST OF ABBREVIATIONS

γ interfacial energy

ϕ packing fraction

P_i $P(\lambda_{i+1}|\lambda_i)$

ASTS Adaptive Segmented Transition Sampling

CN Coordination Number

CNT Classical Nucleation Theory

DDT Disdyakis Triacontahedron

FCC Face Centered Cubic

FFS Forward Flux Sampling

HCP Hexagonal Close Packed

IQ Isoperimetric Quotient

MC Monte Carlo

PPTIS Partial Path Transition Interface Sampling

RHD Rhombic Dodecahedron

RHT Rhombic Tricontahedron

TGR Trigyrate Rhombicosidodecahedron

PHC Pentagonal Hexacontahedron

TIC Truncated Icosahedron

TIS Transition Interface Sampling

TPS Transition Path Sampling

TRD Trapizo-Rhombic Dodecahedron

TST Transition State Theory

ABSTRACT

Clustering and nucleation in metastable fluids of hard polyhedra

by

Samanthule Nola

Chair: Sharon C. Glotzer

Crystallization on the colloidal scale is influenced by numerous factors. I have taken advantage of recent advances in computational techniques and resources to focus on systems of hard polyhedra and investigate the differences in crystallization processes, nucleation, clustering, and local order in metastable fluids due to particle shape.

The crystallization behavior of hard particle systems is influenced by the strength of local orientational preferences, and the affinity of particles for aligning face-to-face, leading to bias toward specific local packing arrangements such as the locally dense icosahedral or hexagonally close-packed configurations. We report the driving force needed for crystallization of a selection of polyhedra known to assemble into the face centered cubic crystal: Disdyakis Triacontahedron (DDT), Trigyrate Rhombicosidodecahedron (TGR), Truncated Icosahedron (TIC), Rhombic Tricontahedron (RHT), and Rhombic Dodecahedron (RHD). Shapes that promote local environments (arrangements of neighbors in the fluid) that are incompatible with the local environment in the crystal require a similar driving force to spheres to observe crystallization directly. For these shapes the crystallization probability increase relatively slowly with the increasing driving force, so the fluid remains metastable to a

higher driving force. Polyhedra that promote local environments in the fluid consistent with the local environment of a particle in the crystal, or that had low energetic barriers to rearrangement of neighbors relative to particle orientation, are observed to crystallize more easily than spheres. Local order and emergent entropic preference for face-to-face contact play roles in clustering and crystallization of metastable hard polyhedra fluids.

In order to better understand the transition from fluid to solid, I have extended non-biasing forward flux-type rare event sampling methods, and generated ensembles of non-equilibrium states intermediate between fluid and solid.

CHAPTER I

Background

1.1   Computation

Working with simulations, we have the privilege of building for ourselves an empty world to play with. A universe in which there is nothing that we did not place and define explicitly. Every object and interaction is described . Once a system is described we are able to use computational tools to allow the system to evolve, and observe what changes take place and what configurations result. This gives us a wealth of detailed information on exactly how our explicitly defined simulated system behaves.

Knowing what happens in exact detail for the simulated system is interesting in itself, but gains additional value through comparison to physical experimental systems. When the results of a simulated system agree with what can be observed by physical experiments, we can see that the components included in our description of the system include all necessary aspects of the real system needed to cause the behavior observed. We can then use the detailed information produced by the simulation to understand underlying mechanisms in the physical system.

Another interesting consideration in computational research is how to retrieve useful results and conclusions from the wealth of raw data generated by the simulation. The methods and challenges to resolving useful information from raw simulation data

depend on the type of simulation. Here we are using atomistic simulations, considering discrete objects with well-defined positions rather than property fields as in the case of discrete element methods. Starting from knowing the positions of objects over time, we are interested in finding patterns. Patterns can emerge at a variety of length scales in both position and time.

Having every aspect of the little simulated world explicitly defined to represent an aspect of a physical system does not necessarily mean that we are aware of every possible non-physical contribution to the computational results. Computational mechanisms such as random number generators meant to add a random component from a defined distribution, discretization of time, the order in which physically simultaneous components are calculated, and the limit of precision in values, such as distances, stored by a computer, can all potentially introduce artifacts into simulations. Computational materials is a maturing field at this point, and many possible sources of non-physicality are well understood, and can be accounted for reliably. [39, 56].

Once all errors and artifacts due to simulation methods are accounted for, we need to establish the degree to which the interactions and objects we simulate will reflect real world systems; the real world is messy. Every experimentalist knows that you cannot know and control every variable involved in an experiment. The goal is to control every relevant variable, and that is where the power of simulations comes into play.

The results of physical experiments are physically true for the system in use, but all details of the system cannot be known completely. Simulations complement experiments by showing what results from an explicit model. The power is when experiment and simulation results agree, showing that the simulation model includes all components of the real system that were needed to produce the behavior of interest. The detailed understanding of the system that comes from knowing explicitly, with-

out needing indirect observation, exact system evolution, is then useful in deducing important physical mechanisms.

In this work we are using simulations to understand the effects of certain shape parameters on a non-equilibrium thermodynamic process. Rather than trying to simulate all relevant aspects of a physical system, we consider systems with only a simplified set of properties, in order to isolate the emergent system behaviors associated with those properties. Rather than use simulations to understand a specific behavior in a specific system, we are investigating the contribution of a specific property to a specific behavior: the contribution of shape to self-assembly. Understanding the contribution of shape will accord us access to a parameter that can be independent of other system parameters such as material, and can in some cases alter the overall system in a meaningful way.

1.2   Hard spheres

Hard spheres have been thoroughly studied as a model system for nucleation and growth processes. Early computational work, and even significant ongoing work focuses on hard spheres as a first approximation to real systems. We will review some of the work done on hard spheres, how it was understood and what relationships and methods were used and gained in those studies. From there we will consider adaptations needed to understand how local anisotropy can lead to additional behaviors in systems that are otherwise similar to spheres.

When we describe hard spheres we are taking advantage of our idealized empty computational universe to understand how a system will behave with only some of the properties of a physical system. Hard spheres act as ideal van der Waals gases, with no attractions but hard cores, offering a way to observe and understand theorized behavior proposed in statistical mechanics. Hard spheres also approximate colloids such as those made by polymerizing micelles, and any other system where we believe

the behavior of the system is not significantly influenced by attraction, repulsion, friction, or viscous effects.

The behavior of hard spheres is fully governed by the statistics of most the probable states, depending entirely on entropic rather than enthalpic interactions. These fluids have no attraction, so they have no distinction between gas and liquid. Throughout this work we will discuss fluids that are dense enough for all particles to be interacting with multiple neighbors, and where diffusion is significant and there is no crystalline order. Hard sphere systems are well studied in the literature, and we use them along with other hard particle systems to investigate metastable equilibrium states, metastability, and transitions that take a system from a metastable local minimum, through non-equilibrium states, to a lower free energy state.

Monte Carlo (MC) simulations sample phase space. For ergodic equilibrium systems they provide understanding of available states and their relative probabilities. Using a Metropolis criterion for move acceptance, the system samples phase space by picking a particle in the system, generating either a translation vector or a rotation quaternion, and either accepting or rejecting that trial move based on how the energy of the system would change if the move was accepted. MC is particularly well suited to hard particle systems because the free energy, ΔU, due to a trial move is always 0 (if no overlap occurs) or ∞ (if an overlap occurs). This method will, for ergodic systems, sample states proportionally to their free energy, Giving us equilibrium states as the most probable. For systems with non-ergodic behavior, such as metastable fluids barred from accessing the low energy crystalline state, MC samples metastable states as a local equilibria until the barrier to crystallization is surmounted. Here we have three types of moves that are proposed for the system: translation moves that translate a single particle, rotation moves that rotate a single particle, and (for isobaric (NPT) simulations where the box size changes to maintain a constant pressure) box size moves that shrink or expand one axis of the box.

1.2.1 Entropy

Entropy can be broadly defined as $S = k_B \ ln(W)$, where k_B is the Boltzmann constant and W is the number of possible microstates. The number of microstates can be explicitly described in tractable systems, such as a handful of dice, and can be reliably described statistically in relatively simple systems such as an ideal gas, and even moderately complex systems such as mixed gasses in equilibrium with a solid. The entropy of a system depends on what microstates are accessible, and unique, which depends on many factors such as the distinguishability of constituent particles.

Entropy is an integral component in the energy balance leading to hard sphere crystallization, and is critical to the ordering behaviors observed in systems of polyhedra. The tendency of hard spheres to crystallize can be understood by considering the free energy, including the contribution of entropy. Free energy is minimized, and entropy is maximized, when the spheres have the additional 'wiggle room' allowed by the close packed arrangement [8, 7].

Highly spherical polyhedra have non-trivial rotational entropy due to differences in their interaction with their local environment at different orientations, while spheres have no orientational contribution. Orientational entropy contributes to the system's free energy, and can increase the drive to order in cases where particles in a rotator crystal are more free to rotate than they are in the fluid. The contribution this makes to the system depends on how the particle's orientation relates to its local environment. A more direct question is, how do properties of a shape affect entropy of a monodisperse system of particles of that shape? Introductory statistical mechanics tells us that a cone will have two more degrees of freedom in its position than a sphere, and a cone with a ridge along the long axis will have a third rotational degree of freedom.

But what happens when that cylinder has identical ridges 180° apart? According to [26], it still has meaningful orientations in three dimensions, but now half of those

are duplicates, cutting the number of options in half, adding a $-ln(2)$ term to the free energy. For systems with N_{sym}-fold symmetry, it would add $-ln(N_{sym})$ to the free energy. Proper treatment for multiple symmetries in three dimension must be analogous, but the details of how to count distinguishable states remain subject to situation-specific adaptation [48, 21].

1.3 Crystal nucleation

The formation of a solid phase from a metastable fluid can happen by several mechanisms. For monodisperse systems such as studied here, homogeneous crystallization is commonly a nucleation and growth type process.

Nucleation occurs when a system goes from a metastable local energy minimum state by passing through a higher energy intermediate state, before reaching a lower energy state. In different systems this takes a wide variety of forms. In real systems there are impurities and edges that provide sites for heterogeneous nucleation, but in most ideal boundary-less systems heterogeneous nucleation is not available, so nucleation can only be caused by homogeneous mechanisms of density fluctuations in the fluid. An idealized description of nucleation and growth, Classical Nucleation Theory (CNT) is a well established model for understanding how nucleation happens [7].

1.3.1 Classical Nucleation Theory

In a homogeneous nucleation and growth process, random fluctuations in the system create regions, nuclei, with lower free energy due to their arrangement. If one of these regions becomes large enough that its free energy benefit ($\propto$ volume) overcomes the interface energy penalty ($\propto$ surface area), it becomes more probable that the nucleus will grow than dissipate. Once this critical nucleus is achieved, growth proceeds by addition of particles to the nucleus surface at an attachment rate

determined by system properties such as diffusion rate and the local energy change associated with attachment.

Classical Nucleation Theory (CNT) describes this process for uniform spheres forming a spherical nucleus out of a uniform system, and finds the free energy nucleation barrier (ΔG) as

$$\Delta G = 4/3\pi R^3 \rho_S \Delta\mu + 4\pi R^2 \gamma \tag{1.1}$$

in terms of the number density in the solid nucleus (ρ_S), the difference in chemical potential between a particle in the solid and in the liquid ($\Delta\mu$), the radius of the nucleus (R), and the solid-liquid interfacial energy density (γ).

Three fundamental underlying assumptions of CNT are that the parent phase is uniform outside of the nucleus, the nucleus is uniform throughout its volume, and the interface between them is well-defined, uniform, and of negligible volume. Proper CNT assumes a spherical nucleus; modified versions allow for nuclei with known shapes and facets, and different interfacial energy penalties, γ, for different growth facets [46]. CNT is a useful conceptual model, and a useful practical tool in many large-scale systems and in systems where the deviations from the three fundamental assumptions can be quantified and are minimal. However, for colloidal systems without strong directors to narrow the transition path, CNT is best reserved as a framework.

In this model, the critical nucleus can be found at the maximum free energy by setting the derivative of Eq. 1.1 to 0: $\Delta G^* = 16\pi/3 * \gamma^3/(\rho_S|\Delta\mu|)^2$, is $R^* = -2\gamma/(\rho_S\Delta\mu)$, assuming $\Delta\mu < 0$.

The nucleation rate can then be found as the product of the probability of overcoming the nucleation barrier ($\propto exp(-\Delta G^*/(k_B T))$) and a kinetic prefactor Γ

$$Nucleation\ Rate = \Gamma * exp\left[-\frac{16\pi}{3k_B T}\frac{\gamma^3}{(\rho_S|\Delta\mu|)^2}\right]. \tag{1.2}$$

The interfacial free energy density γ is an important property in CNT, and is understood from first principles in only a few well studied systems. Empirical values for γ are widely cataloged and used in simulations and analytical studies. The kinetic prefactor Γ is related to the number density in the liquid, the diffusion rate in the parent phase, the shape of the energy barrier, and a diffusion related attachment term [10]. CNT typically takes Γ as a system constant, but recent reviews of crystallization such as [65] describe attachment as an activated process and Γ as an Arrhenius function.

In some types of systems substantial modifications are made to CNT to reflect, for example, anisotropy of the critical nucleus. Although it does not completely capture the physical process, CNT is a useful and nearly ubiquitous approximation. Crystal growth has been described for decades by CNT models, with an ever evolving variety of modifications to both the nucleation and growth portions to describe specific systems [9, 13, 15, 37, 52, 36].

1.3.2   Nucleation and growth of colloidal and hard sphere crystals

By studying systems of monodisperse hard spheres, where spheres interact only with elastic collisions and by excluding others from occupying the same local volume, Auer, Frenkel, and others have made a great deal of progress understanding the nucleation and growth process [30, 10, 19, 35, 47].

Spheres will fill space with a maximum packing fraction (ϕ) of 0.74, when assembled into a HCP or FCC structure [29]. It is intuitive that spheres under pressure or confined to a volume approaching $\phi = 0.74$ will assume a close packed structure. What is more interesting is that non-interacting spheres will assemble into a close-packed crystal, typically FCC, at ϕ values barely greater than 0.5.

Examination of process dynamics have recently proposed a solution to the long standing question of why face centered cubic (FCC) is preferred over hexagonal close

packed (HCP) in many systems, despite having the same density and extremely similar free energies [11]. Heitkam et al. [30] have shown that the forces acting during the assembly of a third layer in HCP stacking tend to destabilize the second layer more than if the particles approach the third layer as FCC. The HCP and FCC structures are both comprised of hexagonally close packed layers, having either an ABAB stacking in the case of HCP, or ABCABC stacking in the case of FCC. I do not differentiate explicitly between FCC and HCP local environments as HCP stacking faults are inconsequential for my purposes, but do note that all systems studied here are primarily FCC.

While hard spheres follow a nucleation and growth type process, they have been found to match poorly with CNT. Solutions of CNT that either fit both the interfacial tension and the kinetic prefactor allow significant variation, and those that use analytical approximations of one of those values have been found to give unphysical values for the other [7].

1.4 Entropic assembly of polyhedra

Recent experimental work on shaped nanoparticles has been dependent on the ability to fabricate faceted nano particles, for example, by inhibiting growth selectively on specific crystal planes of single crystals growing out of solution [46, 15, 31]. There is an additional challenge in fabricating small particles that interact with a hard potential, since surfaces and edges of crystals tend to have unsatisfied bonds, giving them energetic interactions. This can be avoided in some systems by using surfactants to stabilize facets and edges [46], but this becomes increasingly difficult as the particle size decreases, approaching the size scale of the surfactants [22]. The hard particle limit can be realized, however, for micron-sized colloids.

In simulation we can isolate the contribution of purely entropic "forces" by specifying that these systems have only hard interactions, neither attracting nor repelling at

a distance, and strictly excluding other particles from penetrating their own volume. The study of entropic assembly due to shape is a growing field [24], including examinations of single shapes such as tetrahedra [26, 16] and different types of ordering, such as by Keys and Glotzer [37]
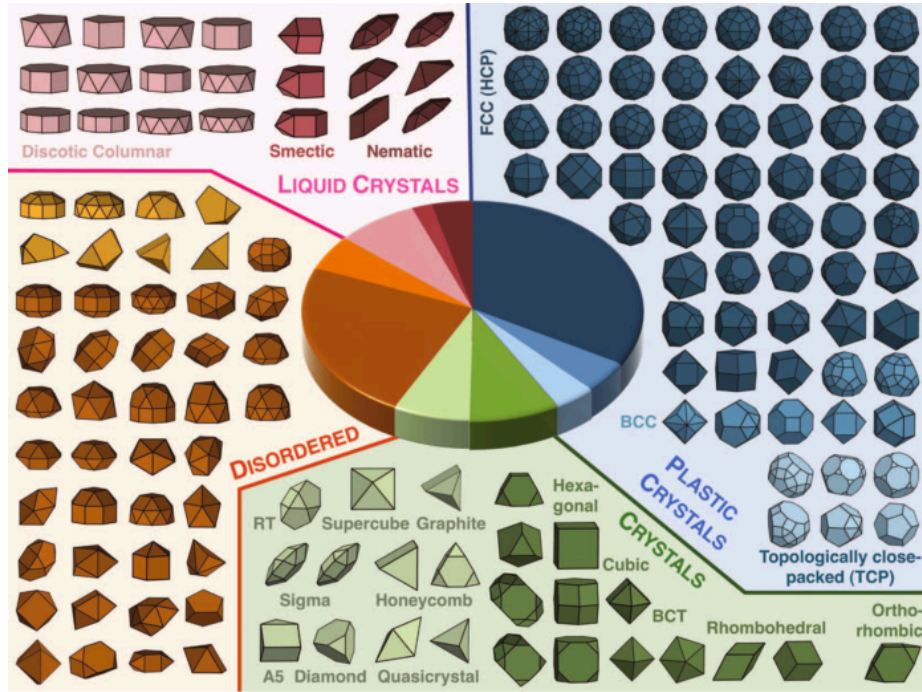
Figure 1.1: Structures formed by the 145 polyhedra studied by Damasceno *et.al.*, reproduced from [17].

Investigations of 145 hard convex polyhedra were recently conducted by Damasceno et al. [17], using Monte Carlo (MC) computer simulations. They showed many polyhedra organizing into ordered structures ranging from Face Centered Cubic (FCC) to the complex γ-brass structure with a 52-particle unit cell. In the 53 systems that formed FCC crystals, which, for convenience, were not distinguished from HCP, the particles were found to be positioned in a crystal lattice without correlated particle orientations. This phenomenon is known as a rotator crystal, or plastic crystal.

Particle orientation is significant in understanding the entropy of the system, since rotational freedom of the anisotropic particles contributes significantly to the system's

entropy.

Elongated, rod-like, polyhedra tended to form nematic phases with their long axes parallel, and disk-like polyhedra tended to form nematic or columnar phases. This is consistent with studies of prolate and oblate spheroids [44].

Damasceno *et al.* found two primary properties of polyhedra that largely predict their final structure: the isoperimetric quotient (IQ) and the coordination number (CN). The coordination number is the average number of neighbors a particle will have, and thus is a property of a system of particles. CN describes the local environment in the fluid, rather than directly describing the shape of the polyhedra.

Sphericity is measured by the isoperimetric quotient (IQ) where $0 < IQ \leq 1$ and $IQ = 1$ for a perfect sphere. The IQ is a measure of the curvature of an object, and in the case of polyhedra (flat faces with discontinuous edges) is reduced to the squared ratio of the volume V of a polyhedron to the volume V_{sph} of the sphere with the same surface area S: $IQ = (\frac{V}{V_{sph}})^2 = \frac{36V^2}{\pi S^3}$. In this work we consider a selection of FCC formers chosen from those identified by Damasceno et al. [17].

CHAPTER II

Methods

2.1 Monte Carlo simulation

We use HPMC, the Monte Carlo (MC) module of the open software HOOMD-blue [5], to simulate the self-assembly behavior of hard polyhedral particles [4, 23, 3]. We fix the maximum translational trial move sizes to the same value for all systems, and fix maximum rotational trial move sizes individually for each shape to allow for significant differences in rotational diffusion and symmetries between shapes. Pressures are given in units of k_bT/V_p, with the particle volume V_p chosen to be unity. All fluid systems contain 4096 identical unit volume particles; this size is sufficient to exclude box size effects in similar systems [36], especially for the small unit cells we have investigated. Independent NPT fluid simulations were initialized as dilute cubic lattices and equilibrated by running for a minimum of 80,000 MC sweeps, or until the system had equilibrated to a stable density, *i.e.* a stable box volume. The metastable fluids are then sampled for up to 40 million MC sweeps. Systems for calculating the equation of state in the solid phase contain 4000 particles, and are initialized in a perfect lattice at the approximate packing fraction (ϕ), and then allowed to relax for one million MC sweeps before taking volume measurements. Approximate ϕ for each pressure were estimated by comparison to known states, and then refined by manual simulations allowing non-physical changes to the system density to find what

density was maintained by systems at fixed pressure. Example data can be found in Appendix A.1

2.2 Metastable equilibration

We sample the fluid at pressures up to the point where one third of systems randomly initialized at an appropriate ϕ crystallize within one million steps. This point is set as the practical upper limit of metastability for each system, where the nucleation barrier is $\Delta G^* \sim k_B T$, slightly below the true limit where $\Delta G^* = 0$. We refer to this point as the metastable fluid limit (MFL). An exact MFL point can be found from free energy curves as the point where the free energy of the stable solid intersects close enough to the minimal free energy of the fluid that the fluid is a shoulder rather than a second well. In this case we are considering the MFL as the point where the barrier between fluid and crystal energy wells is on the order of $k_B T$, as indicated by crystallization occurring from the fluid too readily to allow thorough sampling. Supporting data is available in Appendix A.1.

The distribution of cluster sizes is determined from sampling the equilibrated fluid after the system density has stabilized. We analyze between 2,000 and 20,000 independent snapshots of the metastable fluid at each pressure. From observations of the fluid we can calculate the probability $P_i(X)$ for a particle i to find itself in a cluster of size X as $P_i(X) = \frac{Obs_X * X}{N * Obs_{tot}}$. Here Obs_X is the count of observed clusters of size X, Obs_{tot} is the count of observations of the system, and N is the number of particles in the system.

2.3 Order parameter

We find clusters using ten Wolde's method, which uses the Steinhardt $l = 6$ spherical harmonic-based method to measure 6-fold motifs in the positions of neigh-

bors around each particle [49, 53]. This clustering method uses three cutoff values: $r_{cut} = 1.6$ is the cutoff distance for neighboring particles, $d_{cut} = 0.7$ is for how well-ordered the local environment must be for a particle to be considered solid-like, and $S_{cut} = 8$ is the minimum number of solid-like neighbors a particle must have to be considered part of a cluster. These values are consistent with prior work such as [19]. For the unit volume shapes, $r_{cut} = 1.6 \sim 1.3\times$ unit volume sphere diameter is consistent with common values used in literature. Values for d_{cut} and S_{cut} were chosen by computing the largest cluster size in each frame of trajectories with crystallization events and selecting values that captured an increase in cluster size prior to crystallization of the system.

We first calculate the complex vector $q_{6m}(i)$ for each particle i, quantifying how well the local environment matches the 6-fold (l=6) spherical harmonic Y_{6m}

$$ q_{6m}(i) = \frac{1}{N_N(i)} \sum_{j=1}^{N_N(i)} Y_{6m}(r_i - r_j). \tag{2.1} $$

Each particle's $q_{6m}(i)$ is then compared to that of each neighboring particle as

$$ d_6(i, j) = \frac{\sum_{m=-6}^{6} q_{6m}(i) q_{6m}(j)^*}{\sqrt{\sum_{m=-6}^{6} |q_{6m}(i)|^2} \sqrt{\sum_{m=-6}^{6} |q_{6m}(j)|^2}} \tag{2.2} $$

If the average $d_6(i, j)$ for neighbors j of i is greater than d_{cut} , i is said to be in a solid-like environment. A particle with a solid-like environment that has a sufficient number of neighbors also in solid-like environments is considered to be part of a cluster. The size of a cluster is calculated by counting adjacent clustered particles.

2.4   Coexistence

We used interfacial pinning, developed by Pedersen [45], Thapar and Escobedo [55] and implemented here by R. Newman, to determine coexistence pressures that

were in turn used to calculate the chemical potential *via* thermodynamic integration, in the manner of Thapar and Escobedo [54]. Interfacial pinning uses a simulation box containing both solid and fluid phases, with an interfacial area that is held constant. We applied a harmonic biasing potential to stabilize the phase coexistence and maintain equal amounts of the two phases. Systems are run at various pressures; coexistence occurs where no bias is required to maintain a stable volume of each phase.

2.5   Chemical potential

In order to compare between hard particle NPT systems, we calculate the driving force for crystallization at each pressure. In the NPT the Gibbs free energy reduces to $dG = -SdT + VdP + \mu dN = 0 + 0 + VdP$. The driving force for crystallization at coexistence is by definition 0, so the driving force for crystallization, the free energy change (work released) per particle due to crystallization at a pressure P^x is the chemical potential, calculated by integrating the difference between fluid and solid equations of state from the coexistence pressures (determined from interfacial pinning) up to the pressures of interest. $\Delta\mu = \int_{P_{coex}}^{P^x} VdP$.

2.6   PMFT

We calculated the potential of mean force and torque (PMFT) to understand the local environments in the fluids. The PMFT is related to the logarithm of the probability of a particle having a neighbor at a given location, in energy units [60]. Local particle neighborhoods are described with respect to a central particle's position and orientation. The relative likelihood for a neighbor to be at a specific distance and angular location in the surrounding volume is found by binning neighboring particle centers in the volume around each particle in many uncorrelated simulation frames.

Because spheres are isotropic, they have no distinguishable orientation attached to them, and consequently PMFTs of systems of spheres will always be isotropic.

Investigation of crystallization process for spheres vs. rhombic dodecahedra

3.1   Abstract

In this chapter we present a comprehensive study comparing how the same entropically stabilized crystal structure forms in two systems of monodisperse hard particle systems: spheres and rhombic dodecahedra. We investigate these systems by reporting cluster size distributions and crystallization probabilities across the range of pressures where nucleation is observed in Monte Carlo simulations, as well as free energy barriers for nucleation. In addition we quantify the local arrangements around the particles in both fluid and solid phases and discuss the effects on nucleation. We find that rhombic dodecahedra promote motifs in the fluid that resemble the eventual crystal, leading to fluid that is more similar to the crystal. This similarity of fluid corresponds to lower energy barriers to nucleation, and lower driving forces needed for crystallization.

3.2   Introduction

The contents of this chapter reflect work done in collaboration with co-authors Richmond Newman, Julia Dshemuchadse, and Sharon C. Glotzer. The contents are

in preparation for publication, and are modified slightly here to highlight individuals' contributions.

As a model case to investigate the nucleation behavior of anisotropic particles, we choose the rhombic dodecahedron. The rhombic dodecahedron is the shape of the Voronoi cell of the cubic close-packed (*ccp*) sphere arrangement, which is the face-centered cubic Bravais crystal lattice. We will refer to this as FCC throughout this work. In a study of 145 different polyhedra, 51 different shapes were found to self-assemble FCC [17], among which the rhombic dodecahedron was the least spherical – as ranked by the isoperimetric quotient – ($IQ = 0.74$), whereas the sphere is by definition the most spherical ($IQ = 1$). Another recent study of hard polyhedra described emergence of heterogeneity in orientational order preceding crystallization in monodisperse systems of three shapes with low sphericity[54]. They considered crystallization into differing structures, with differing local motifs, and found that the polyhedra caused local orientational order in the fluid that promoted crystallization by lowering the free energy barrier to nucleation. We find qualitative agreement with their overall findings, and focus our attention on differences in the crystal formation processes between shapes that form the same final crystal structure.

These hard-particle systems crystallize *via* a nucleation and growth type process [7, 10, 43, 54]: they need to overcome an energetic barrier to the spontaneous formation of a "critical" nucleus of sufficient size. The critical nucleus is defined as a cluster that has a 50% probability of growing to span the entire system. The size of the critical nucleus corresponds to the point where the free energy gained in the cluster volume exceeds the energetic penalty that stems from the unfavorable interfacial free energy arising from frustrated particle geometries occurring at the crystal-fluid interface. Here we investigate in detail the conditions under which the formation of crystal nuclei is favored in fluids of spheres *vs.* rhombic dodecahedra, highlighting the contribution of particle anisotropy to the process of nucleus formation.

Free energy barriers to nucleation were computed by R. S. Newman using umbrella sampling, measuring the bias needed to observe clusters of a given size in a metastable fluid. In the equilibrated stable and metastable fluid regime we sampled the cluster size distribution directly, however, to gather adequate statistics of larger clusters within the metastable regime we used umbrella sampling [57, 6, 7, 10]. This bias sampling method involves running several simulations with harmonic potentials that bias the system towards a certain cluster size. The statistics of the unbiased ensemble near the target cluster size can be calculated from the collected data *via* the recorded amount of bias energy, and free energy curves were combined from simulations of different target values into a continuous curve using the weighted histogram analysis method [25]. Coexistence points between the fluid and solid were found by interfacial pinning, performed by R. Newman.

3.3   Results

The equations of state (EOS) for both rhombic dodecahedra and spheres are depicted in Fig. 3.1(a). The fluid of rhombic dodecahedra has a significantly lower packing fraction than the fluid of spheres at any given pressure. This arises from shape anisotropy: rhombic dodecahedra have a greater circumradius than spheres of equal volume, allowing for collisions at greater inter-particle distances, which leads to increased pressure of the system at similar densities. a FCC crystal of rhombic dodecahedra also has a lower packing fraction than the FCC crystal of spheres over this pressure range, despite the fact that rhombic dodecahedra are space-tessellating in the limit of infinite pressure, while the spheres have a maximum packing fraction of approximately 0.74 in FCC [28, 29]. For both shapes the packing fraction difference between the fluid and crystal increases with pressure. The slope of the EOS is slightly steeper for rhombic dodecahedra at high pressures, *i.e.*, rhombic dodecahedra have a higher compressibility. This means that rhombic dodecahedra require a smaller

increase in pressure for the same increase in packing fraction than spheres do at the same pressure.

In Fig. 3.1(b) the chemical potential difference of the fluid and solid states is plotted *vs.* the pressure for systems of spheres and rhombic dodecahedra. The chemical potential difference ($\Delta\mu$) between the fluid and solid is calculated by integrating the volume per particle between those phases from the coexistence pressure. When we compare $\Delta\mu$ of spheres to that of RHD we can see that the curves look similar because the difference in density between fluid and solid phases is similar for these two shapes. The difference for rhombic dodecahedra is initially slightly smaller but increases slightly with pressure, shown in the equation of states above. While there is some difference in the energy from thermodynamic integration, the difference in $\Delta\mu$ between spheres and RHD is mostly due to the difference in their coexistence points.

The behavior of solid-like clusters in the fluid and during nucleation of the crystalline phase in both systems is illustrated in Fig. 3.2, where we plot the probability to find clusters of different sizes at pressures approaching crystallization, as well as the crystallization probability, the free energy barriers *vs.* cluster sizes, and the free energy maxima *vs.* the chemical potential.

Probabilities $P_i(X)$ for a particle i to be in a cluster of size X are shown in Fig. 3.2(a) for both spheres and rhombic dodecahedra. At low $|\Delta\mu|$, small clusters are more common in fluids of spheres than in fluids of rhombic dodecahedra, but trends change with increasing $\Delta\mu$. The apparent lack of small clusters in the fluid of rhombic dodecahedra is likely a consequence of the aforementioned lower packing fraction at small values of $\Delta\mu$. There is a crossover point where larger clusters become more prevalent for rhombic dodecahedra than for spheres. At $|\Delta\mu| > 0.25$. A sharp increase in the formation of larger clusters occurs for the fluid of rhombic dodecahedra near $|\Delta\mu| > 0.4$, and slightly further from coexistence rhombic dodecahedra start crystallizing at a measurable rate as seen in Fig. 3.2(e). This leads us to consider

(a)

(b)
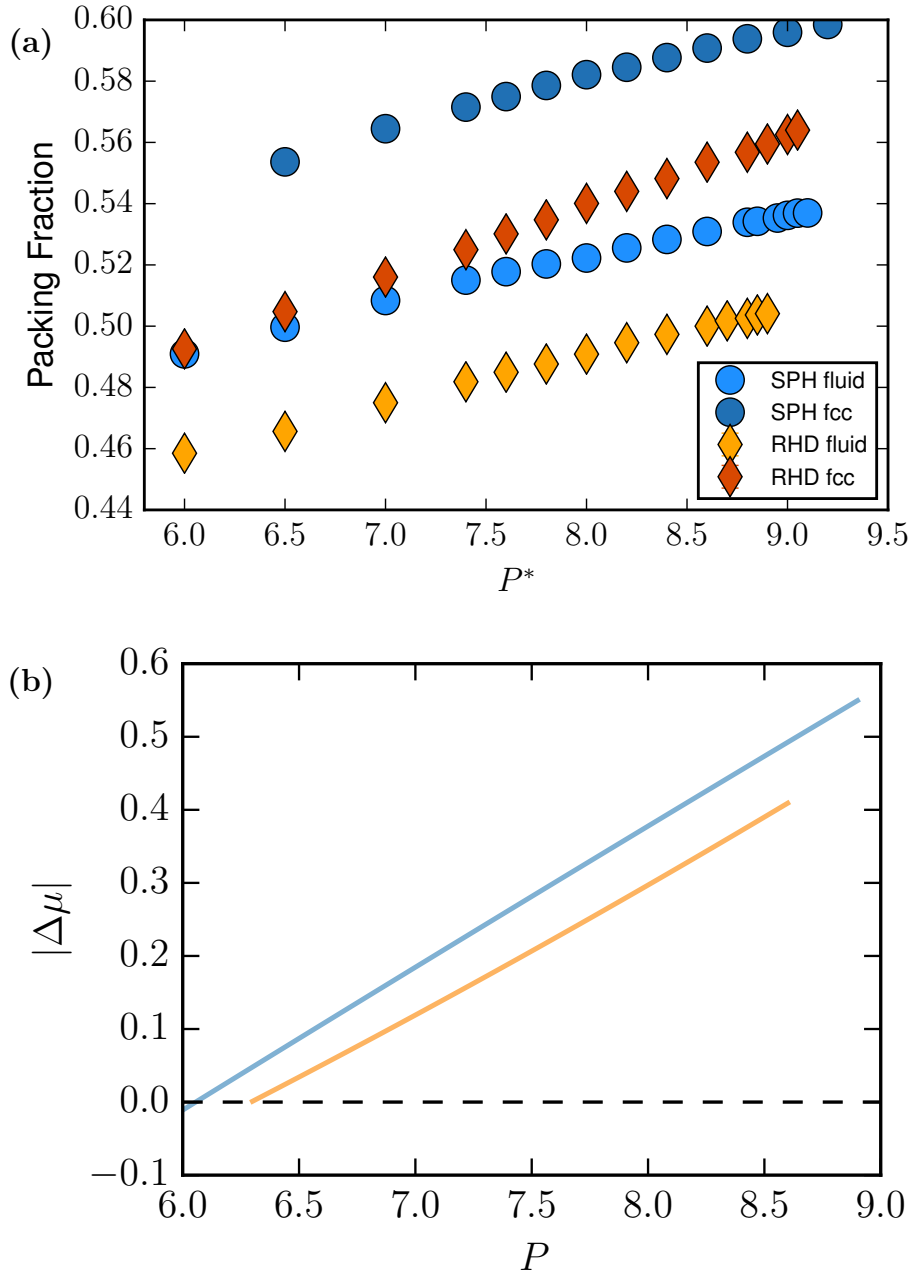
Figure 3.1: (a) Fluid and solid equations of state and (b), the chemical potential plotted *vs.* pressure. In (a) at least 10k samples from well-equilibrated systems are used, so standard deviation error bars are smaller than markers on chart. The dashed horizontal line in (b) at $|\Delta\mu| = 0$ highlights the coexistence pressures for fluid and solid phases of each shape.

21

these larger clusters as pre-critical nuclei, which in turn raises the question of what stabilizes them preferentially in the fluid of rhombic dodecahedra.

Fig. 3.2(e) shows that both shapes begin to crystallize at a measurable rate only after large clusters are seen to form in the fluid. This happens at a significantly higher driving force for spheres than for rhombic dodecahedra. The larger cluster sizes correspond to a density where the emergent entropic bonding directs them more strongly into local arrangements congruent with the FCC crystal, which is demonstrated in Fig. 3.3 and will be discussed below.

The free energy barriers for the formation of nuclei of spheres and rhombic dodecahedra are plotted in Fig. 3.2(b) and 3.2(c). The difficulty of nucleation can be characterized by the free energy required to form the critical nucleus from a fluid, *i.e.*, the location at which $\beta\Delta G$ is maximal. For comparison, these maxima are shown in Fig. 3.2(d), as functions of chemical potential for both particle shapes. These calculations demonstrate that at constant driving force rhombic dodecahedra nucleate far more easily than spheres The nucleation barriers reported here differ from those found in a previous study [54], which were reported to be as low as $\beta G \simeq 5$. Our results agree qualitatively in that rhombic dodecahedra nucleate much more easily than spheres under similar conditions, and propose that barrier heights can be significantly understated if the largest cluster size is used as an order parameter in umbrella sampling in regions where small clusters are common. Tracking the largest cluster size in the system is only valid when large clusters are rare [7, 10] and doing so when clusters are frequent leads to a false free energy minimum at the mean small-cluster size in the fluid. In the small-clusters regime we track the full cluster probability distribution.

To better understand why rhombic dodecahedra self-assemble with a significantly lower driving force (*i.e.*, at a lower chemical potential), as well as having larger local cluster size distributions in the fluid, we calculate the potential of mean force and
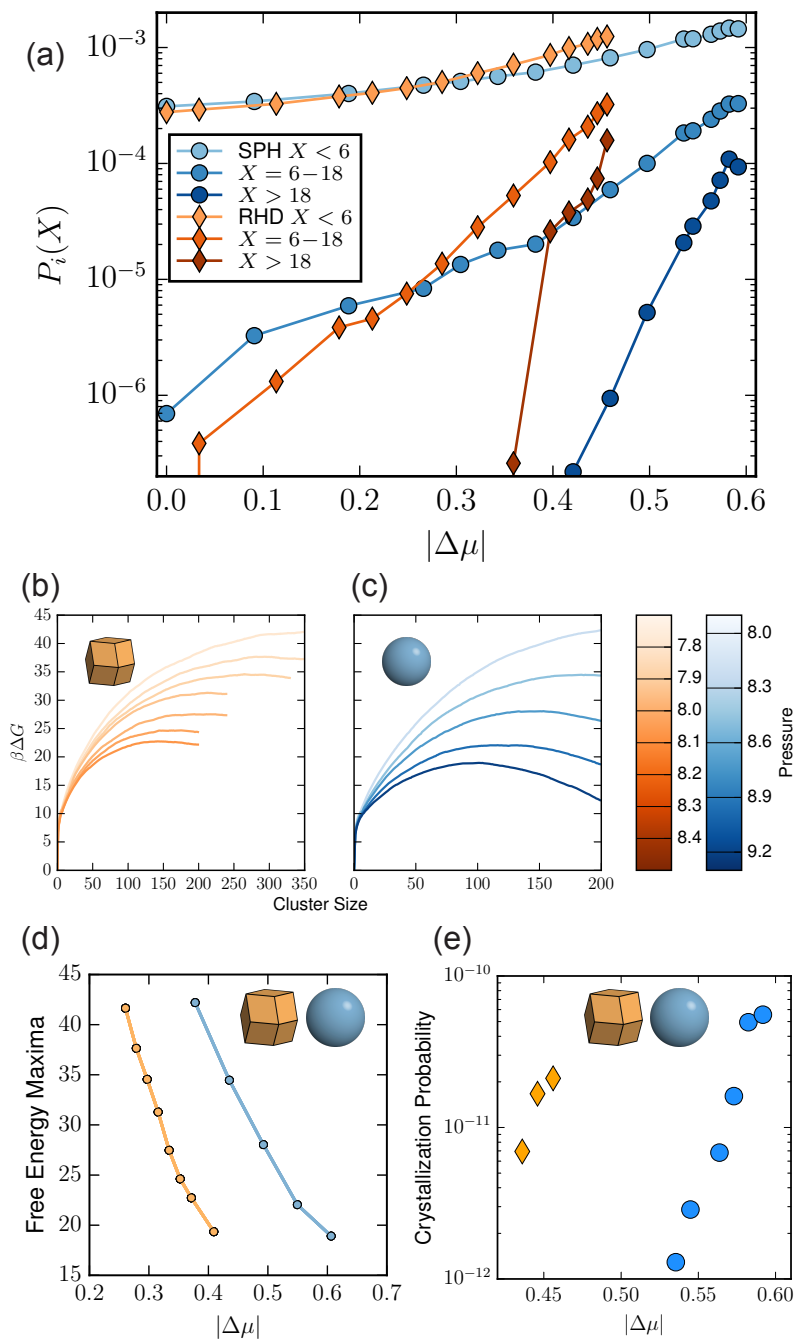
Figure 3.2: (a) Cluster size distributions, (b) – (c) free energy barriers from umbrella sampling, and (d) free energy maxima plotted against the chemical potential and pressure from interfacial pinning of hard rhombic dodecahedra and spheres, (e) directly measured nucleation rates. For plots of the free energy barriers, graphs of different colors represent free energy curves at different pressures (exact values are given by horizontal lines and tick labels on the color bars).

torque (PMFT) to identify particle-neighbor bonds. The PMFTs of the system of spheres, as shown in Fig. 3.3(a) and 3.3(b), are expectedly isotropic and correspond to the well-known radial distribution functions for hard spheres. To illustrate the PMFT of the system of rhombic dodecahedra, cross-sections are performed through the center of the particle perpendicular to the 4-fold and the 3-fold symmetry axes (see Fig. 3.3(c) and 3.3(d)). The crystal PMFTs for both shapes exhibit regions of high free energy (shown in red in Fig. 3.3) corresponding to interstitial positions in the FCC structure. In the fluid, the boundaries between the first and second coordination shell are poorly defined, whereas in the crystal the interstitial positions are geometrically 'forbidden'.

The anisotropic shape of the rhombic dodecahedron causes FCC-like pairwise local arrangements in the first coordination shell with respect to particle orientation, even in the disordered fluid phase. In both the fluid and solid, we observe that rotations of a rhombic dodecahedron around a 3-fold axis are more entropically favorable than rotations about the 4-fold. This corresponds to the previously reported fact that the 4-fold vertices of rhombic dodecahedra are more restrictive to rotational motion than the 3-fold ones [61], and thus their truncation leads to an greater improvement of the assembly-behavior into the FCC crystal structure.

In Fig. 3.3(g) and 3.3(h), we show a three-dimensional representation of the PMFTs in the fluid and crystalline systems of rhombic dodecahedra. The lower hemisphere in each image illustrates the anisotropic PMFT around the polyhedron; additionally, isosurfaces at a value of $\beta F = 0.5$ are superimposed. While the first coordination shells have a similar appearance in both the fluid and the solid and therefore similar local order, the most entropically favorable positions are much more diffuse, $i.e.$, the low energy basins of the PMFT are broader for the fluid and narrower for the crystal of rhombic dodecahedra.

Directional entropic forces favoring face-to-face alignment of the rhombic dodec-

ahedra cause the local structure of the fluid to resemble that of the crystal. This makes fluid and solid more similar and therefore lowers the $\Delta\mu$ between them. This shape-anisotropy-driven ordering is what we believe stabilizes clusters of rhombic dodecahedra, which more easily form from the fluid at lower $\Delta\mu$, and which lead to lower nucleation barriers. Because this effect is due to the directions of preferred face-to-face contacts, it cannot occur for isotropic spheres.

3.4   Discussion

With increasing $\Delta\mu$ values the probability of finding particles in a cluster increases more sharply for RHD systems than for spheres and eventually crosses the probability of spherical particles to be found within clusters of the same sizes. This is a direct consequence of entropic bonding – the preference of faceted particles to arrange themselves with face-to-face contacts to maximize the free volume in the system [60]. In the fluid, this ordering of the local environment is strong (see Fig. 3.3), but does not extend to the second neighbor shell. Face-to-face alignment is measurable in the first coordination shell in the fluid of rhombic dodecahedra, which means that pairwise ordering occurs. However, despite individual pairs being arranged accordingly, the neighbor shell is far from complete at lower pressures and therefore does not correspond to a solid-like local environment. For both shapes the crystallization process leads to high-energy, forbidden regions between the first and second neighbor shells as the particles order. In the case of rhombic dodecahedra, the anisotropy of the FCC local environment is further imposed on the system *via* deep minima in the angular distribution corresponding to the crystal structure and disfavoring the occupation of interstitial particle positions.

The local particle symmetry of the rhombic dodecahedron is commensurate with the local bond-orientational order of the cubic close-packed crystal structure. Therefore we observe that directional entropic forces resulting from the polyhedral facets
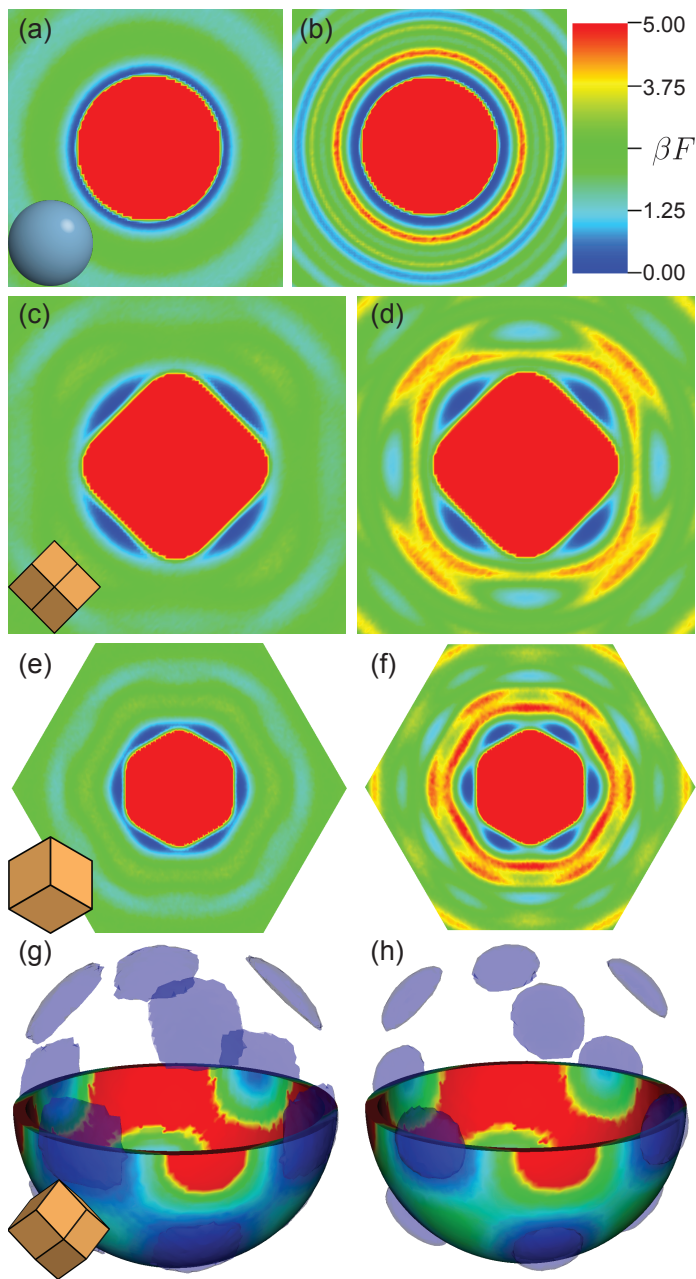
Figure 3.3: Potentials of mean force and torque (PMFTs) for (a) sphere fluid at $\phi = 0.52$, (b) sphere crystal at $\phi = 0.58$, (c), (e), and (g) rhombic dodecahedra fluid at $\phi = 0.50$, (d), (f), and (h) rhombic dodecahedra crystals at $\phi = 0.58$. All color maps set the deepest minimum at $\beta F = 0$. Second and third rows show cuts perpendicular to the 4-fold and 3-fold symmetry axes, respectively. The bottom row shows a thin hemisphere at the radius of the first neighbor shell, along with transparent isosurfaces at an energy of $\beta F = 0.5$. The insets depicting the particle illustrate the orientation of the respective set of PMFTs.

lead to a face-to-face alignment in the fluid of rhombic dodecahedra that corresponds to the crystalline state. We believe that this pre-existing local order is what increasingly stabilizes the crystal nuclei of rhombic dodecahedra as compared to spheres as the packing fractions increase and more and more neighbors are located within first coordination shells. The emergent local entropic interactions stabilize the pathway from a supersaturated fluid to an ordered crystal. By increasingly positioning neighboring particles in arrangements commensurate with the crystal structure, the need for rearrangement decreases, leading to the much lower free-energy pathways to crystallization at given supersaturations shown in Fig. 3.2(b) – 3.2(d) and the larger fluid-phase cluster fluctuations given in Fig. 3.2(a) and 3.2(e).

Previously, it had been shown that self-assembling fluids and crystals of polyhedra exhibit almost identical coordination numbers [17]. The same study showed that measures of shape anisotropy – in particular the isoperimetric quotient – and dense fluid coordination numbers could be used to predict the expected equilibrium phase behavior. Here we demonstrated that anisotropic particles nucleate with lower driving forces than do isotropic spheres. We propose that this arises because the local coordination in the fluid has a greater probability to exhibit pairwise bonds corresponding to the crystal structure. Here we also suggest that directional entropic forces in the fluid caused by face-to-face interactions are beneficial to stabilizing the nucleation pathway.

CHAPTER IV

Metastable fluids of sphere-like polyhedra

4.1 Abstract

In this chapter we present work in preparation for publication with co-authors Julia Dshemuchadse and Sharon C. Glotzer. We compare the metastable equilibrated fluids of several polyhedra known to assemble into the cubic close packing of spheres, the face centered cubic crystal. We report the fluid-solid coexistence pressures, the cluster size distributions and anisotropic local order in the fluids, and the probabilities of crystallizing at high pressures. We find that some polyhedra crystallize more readily and at lower packing fractions than spheres do. We also find that polyhedra form larger clusters than spheres in the metastable fluid as the driving force to crystallize increases. The affinity of particles for aligning face-to-face in either an icosahedral or hexagonal local dense packing arrangement, and the strength of local orientational constraints, all play roles in the crystallization behavior of hard particle systems, and in the local order and stability of the metastable fluids. We analyze local order in the metastable fluids and relate that local structure to crystallization probabilities and to particle shapes.

4.2   Introduction

In recent decades, hard spheres have been simulated and studied extensively as a model for the nucleation of colloidal systems. Hard spheres are ideal spheres that do not attract, overlap, or deform, interacting only at contact by volume exclusion. They model much of the behavior seen in colloidal systems and have been studied both computationally and experimentally [19, 51, 50, 36]. The nucleation behavior of hard sphere systems was well researched by the Frenkel group in the last decade, including investigating variations to the system by adding parameters such as polydispersity and softness [9]. Of more interest to us here is the entropic contribution to crystal formation [24, 21], and the complexity of the free energy landscape and the "path" through phase space from one state to another [65, 42], both of which have recently been subjects of increasing interest.

Additional interactions can be introduced to the hard sphere model to explore specific types of systems, such as softness and "stickiness" to model surface interactions of coated nanoparticles [33, 38], or patchiness to model strong directional interactions [64]. Another important aspect of systems not represented in the hard sphere model is the entropic contribution of particle orientation and rotational freedom in their local environment. Local anisotropy can be introduced to hard particle systems by faceting the particles, using hard polyhedra rather than spheres.

Hard polyhedron systems are of particular interest because they form structures that are stabilized only by entropy, allowing us to isolate the effect of shape alone. In addition to recent investigations to find their densest packings [14], hard polyhedra have been found to self-assemble into crystal structures ranging from relatively simple face centered cubic (FCC) crystal to surprisingly complex structures, due only to their shapes [16, 54, 19]. In dense fluids, local order due to emergent directional entropic forces has been quantified [59]. These studies have shown that facets are important for determining what, if any, ordered structures will form in hard particle systems.

Sphericity is measured by the isoperimetric quotient (IQ) where $0 < IQ \leq 1$ and $IQ = 1$ for a perfect sphere. The IQ is a measure of the curvature of an object, and in the case of polyhedra (flat faces with discontinuous edges) is reduced to the squared ratio of the volume V of a polyhedron to the volume V_{sph} of the sphere with the same surface area S: $IQ = (\frac{V}{V_{sph}})^2 = \frac{36V^2}{\pi S^3}$. Previous work by Damasceno *et al.* studied the crystal structures formed by 145 polyhedra and identified many FCC formers. They proposed that the IQ and the coordination number (CN), the average number of neighbors each particle has in the fluid, are predictive of what structures monodisperse systems of polyhedra form in the solid state.

Here we study the crystallization conditions for spheres and a selection of five polyhedra known to form FCC crystals. We build on the work of Damasceno *et al.*, and examine how variations in sphericity and symmetry affect the local environment in the fluid and conditions at which they crystallize. We choose five FCC forming shapes with a variety of sphericities: DDT, TGR, TIC, RHT, and RHD, with IQ values $IQ_{\text{DDT}} = 0.958$, $IQ_{\text{TGR}} = 0.939$, $IQ_{\text{TIC}} = 0.903$, $IQ_{\text{RHT}} = 0.887$, and $IQ_{\text{RHD}} = 0.740$, respectively.

4.3 Results

Figure 4.1 shows the equation of state (EOS), i.e., the packing fraction ϕ measured in the fluid as a function of pressure. For this selection of shapes the more spherical ones have a higher ϕ at a given pressure, however the difference in ϕ is not proportional to the difference in IQ (see Figure S4.10 in the supplemental information) [16]. Qualitatively we understand that pressure at this scale is a measure of collisions, and less spherical shapes are able to collide at greater distances due to protruding vertices. Collisions can occur between particles that are farther apart, so the count of collisions, and therefore the pressure, is on average higher for faceted shapes than for spheres at any given ϕ.
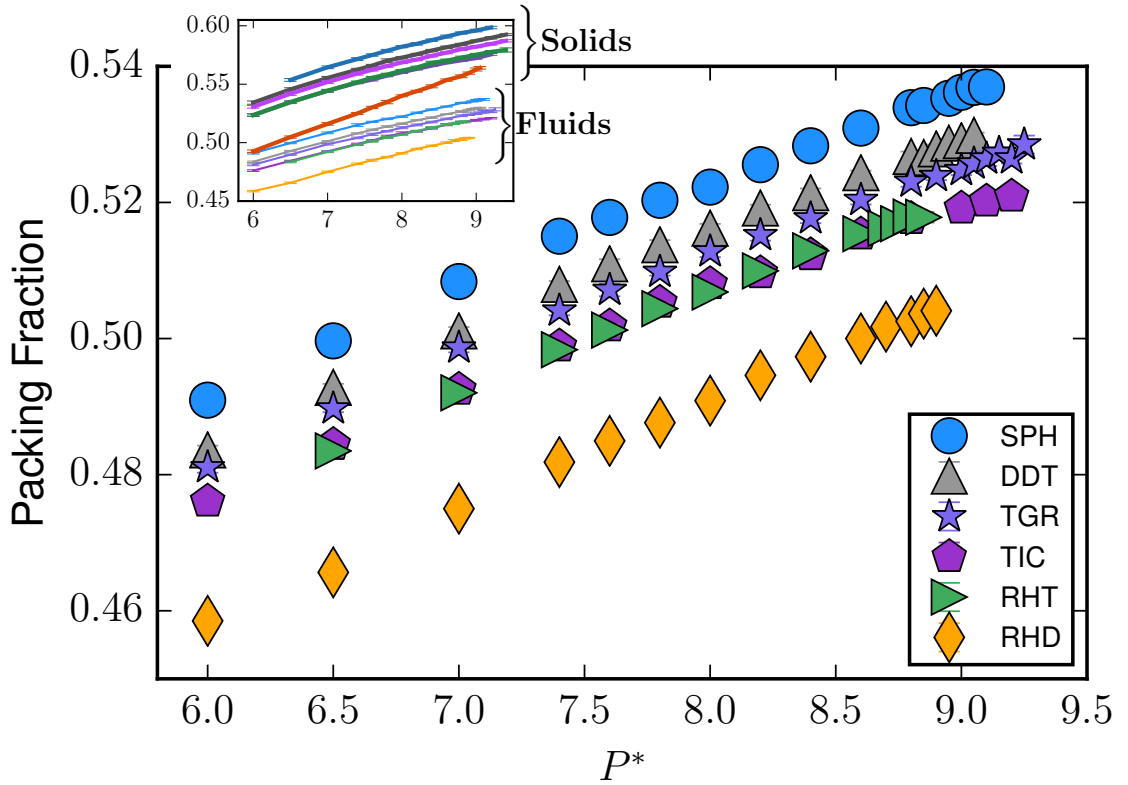
Figure 4.1: The equations of state, i.e., the packing fractions of the metastable fluids as a function of normalized pressure measured from NPT simulations, for all six shapes. The inset shows the same fluid EOS curves along with the respective FCC EOS curves for comparison. The standard deviation of the packing fraction is plotted on the graph as error bars that are smaller than the markers in most cases.
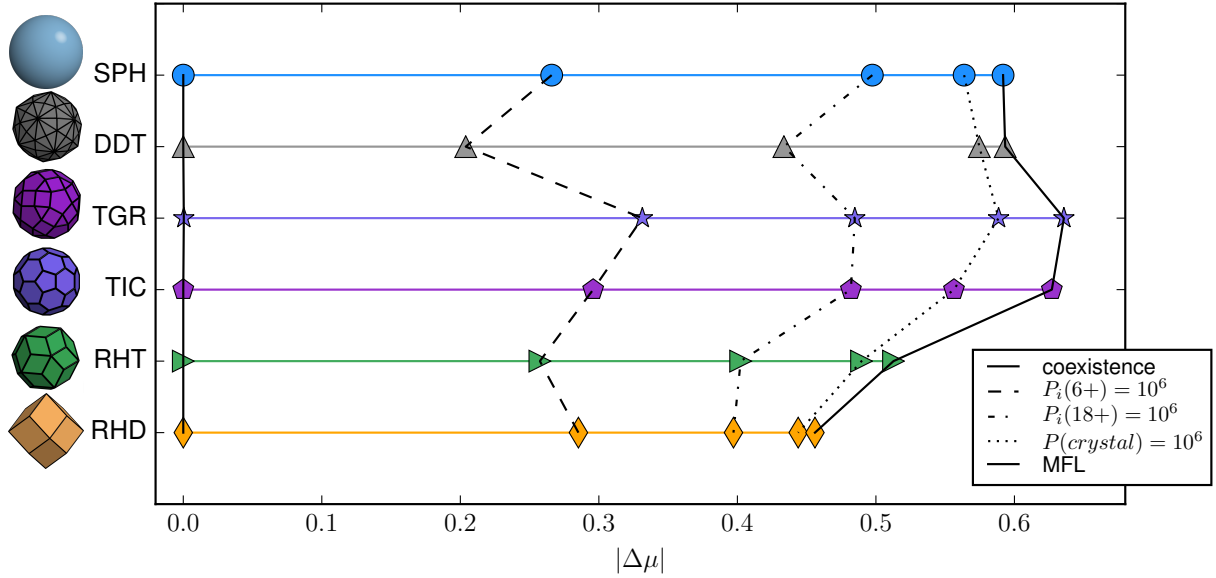
Figure 4.2: Five crucial points in the metastable fluid for each shape, as a function of $\Delta\mu$: the fluid-solid coexistence ($\Delta\mu \equiv 0$), the point where particles have probabilities $P_i(X > 5) = 10^{-6}$ and $P_i(X > 18) = 10^{-6}$ of being in a cluster greater than $X = 5$ or 18 respectively, the point where the probability of crystallizing (per particle) is 10^{-11}, and the metastable fluid limit above which the fluid is not stable enough to be sampled thoroughly. The shapes are arranged top to bottom in order of decreasing IQ.

In addition to showing the EOS for each fluid, the inset in Figure 4.1 shows the fluid EOS along with the EOS for FCC crystals of each shape. The area between the fluid and solid EOS curves, integrated from the coexistence pressure to a pressure of interest, gives the difference in free energy between the fluid and the solid. This difference in energy is the work needed to convert a particle from a crystal to a fluid, the driving force for crystallization. We will use this driving force, expressed per particle as the absolute value of the chemical potential, $\Delta\mu$, as a means of comparing systems of different shapes.

In order to understand differences in how these systems behave we compare points of specific behavior in the metastable fluid between systems. Figure 4.2 shows five useful points for each of these shapes, starting with the point of coexistence between fluid and solid where $\Delta\mu \equiv 0$. Also shown are the points where they start forming

clusters with more than 5 particles, start forming clusters with more than 18 particles, start crystallizing spontaneously (i.e., where the crystallization probability is 10^{-11}), and last is the point where they are effectively no longer metastable (MFL). Several distinctions are quickly apparent, and can be related to different shape attributes.

RHD, RHT, and DDT, the three Catalan solids, all have a significantly narrower range where crystallization is measurable but the fluid is stable enough to be characterized, with spheres closely resembling RHT. TGR and TIC both have a wider stable range, with TIC having roughly twich the range of spheres, whilc RHD has a much narrower range. This difference is not explained by the sphericity of these particles, nor by the equations of state. To break specific differences, we will consider comparisons between pairs of shapes along with spheres.

Of particular interest are the comparisons between RHT and TIC, and between TGR and DDT. In each case we are considering a pair of polyhedra with similar IQ values, but different driving forces associated with clustering and crystallization. We will examine the differences between the shapes and how they relate to differences in local environments seen in the fluids.

To describe the local order and examine differences we analyze the metastable fluids in two ways: cluster size distribution and PMFT. Figure 4.3 shows cluster size distributions by plotting the probabilities $P_i(X) = (Obs_X * X)/(N * Obs_{tot})$ for a particle i to be in a cluster of size X, for small, moderate, and large X ranges as described in Section 2.2.

Sphere systems show greater $P_i(X)$ for all measurable X near the coexistence point, however there are crossover points where RHD systems are more likely than spheres to produce clusters, $P_i^{RHD}(X) > P_i^{sph}(X)$ at $\Delta\mu > 0.25$ for both $X < 6$ and $6 \leq X \leq 18$. All polyhedra studied here show this behavior of having fewer clusters than spheres near to coexistence but overtaking spheres prior to the MFL. This crossover point is less distinct for TGR and DDT where the cluster probabilities
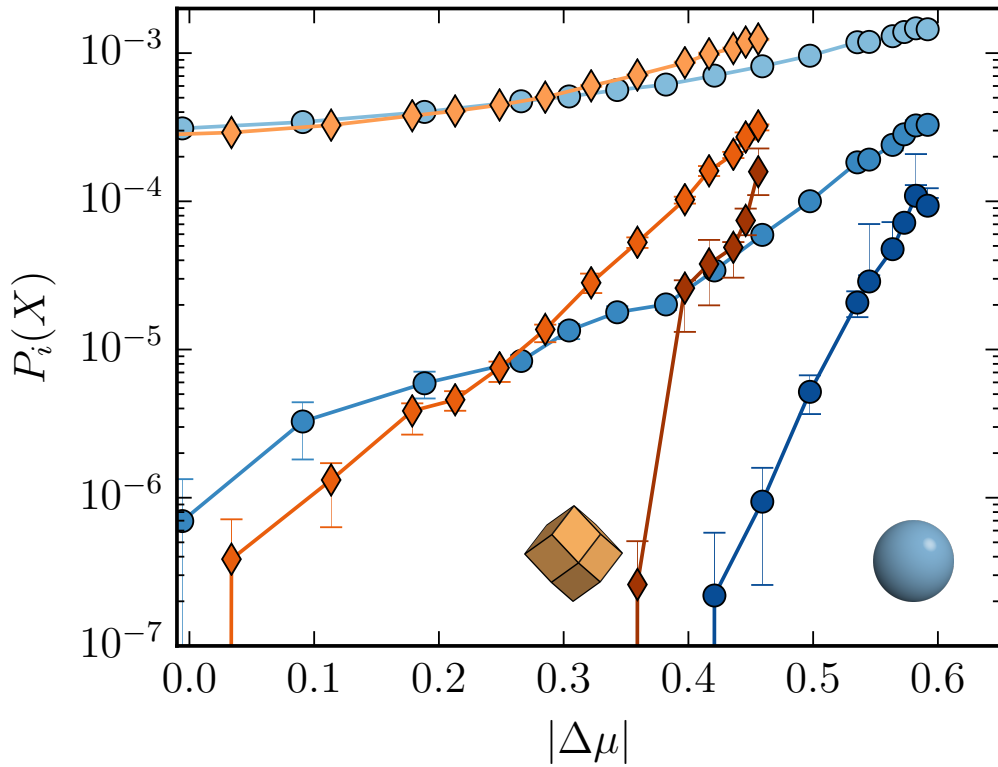
Figure 4.3: Occurrence of clusters of different sizes in fluids of spheres and fluids of RHD. The probability of a particle to be in a small cluster of $X < 6$, a medium cluster of $6 \leq X \leq 18$, and in large clusters $X > 18$. Error bars are the standard error of the mean $P_i(X)$ calculated for esch separate fluid trajectory. The probability of small clusters is significant even at the coexistence pressure, and differs between these shapes only at high $\Delta\mu$. RHD systems show marked increase in moderate cluster probability, the change between pure fluid behavior occurring over a narrower range of $\Delta\mu$ than for spheres.
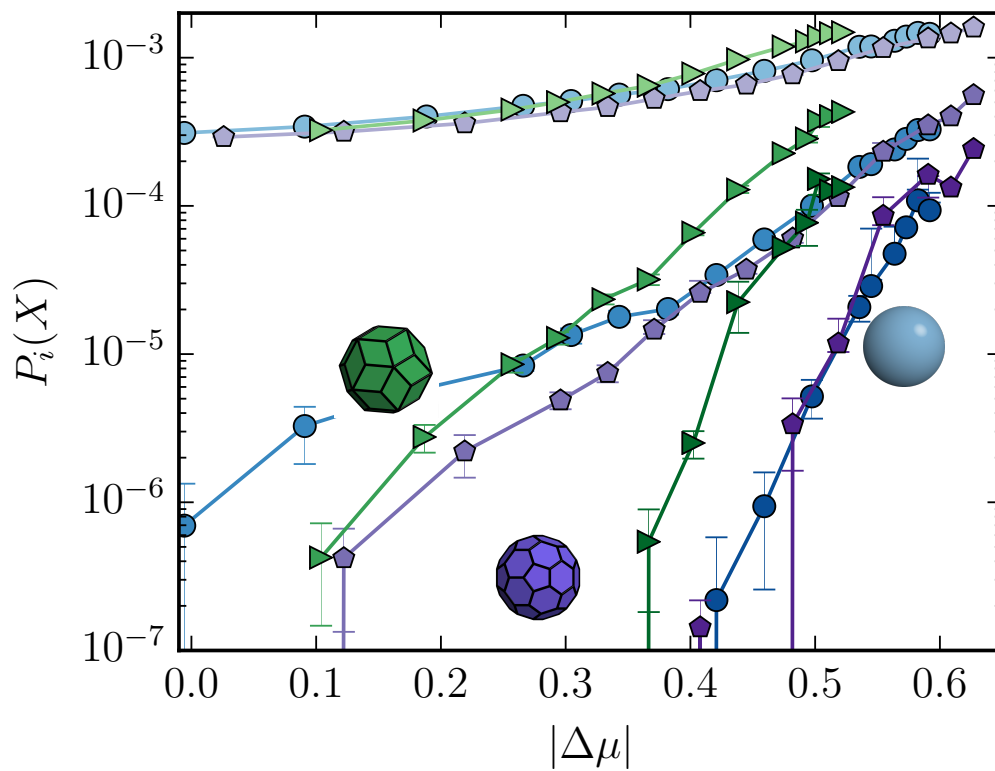
Figure 4.4: Occurrence of clusters of different sizes in fluids of spheres, RHT, and TIC. Shows the probability of a particle to be in a small cluster of $X < 6$, a medium cluster of $6 \leq X \leq 18$, and in large clusters $X > 18$.
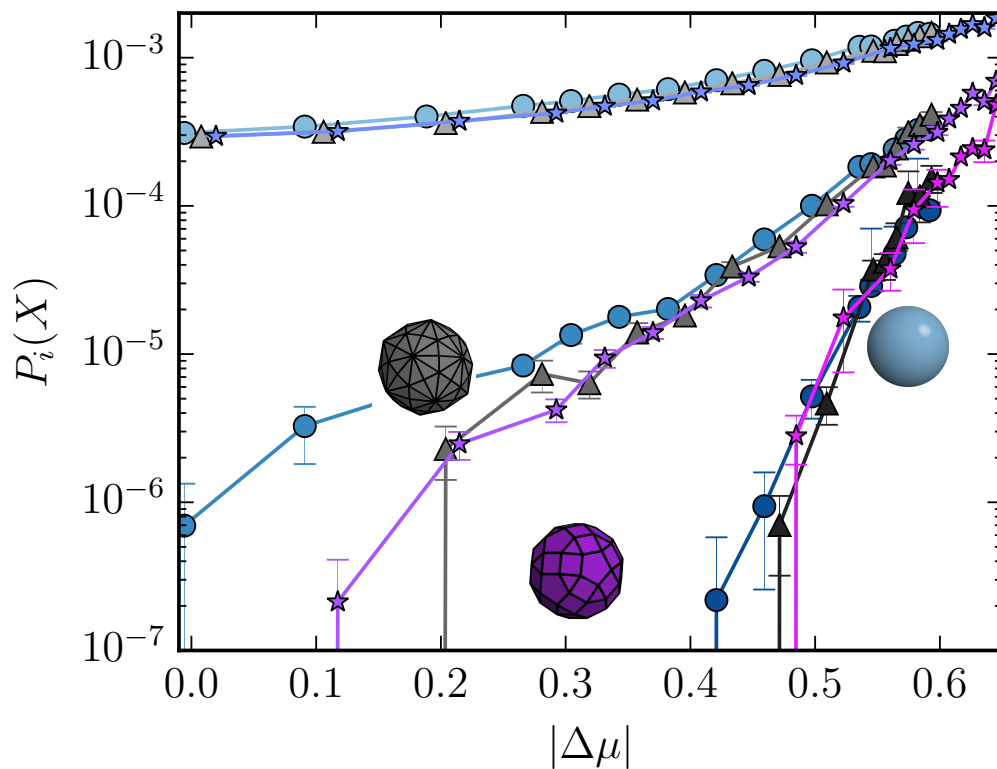
Figure 4.5: Occurrence of clusters of different sizes in fluids of spheres, DDT, and TRG. Shows the probability of a particle to be in a small cluster of $X < 6$, a medium cluster of $6 \le X \le 18$, and in large clusters $X > 18$.

are similar to spheres. TGR and DDT do show a drop in $P_i(X > 18)$ relative to spheres at $\Delta\mu < 0.35$, implying that closer to coexistence the subtle deviations from sphericity become relevant, while at higher pressures the sphere-like interactions dominate, obfuscating any anisotropic effects. The observation of large clusters ($X > 18$) precedes crystallization for all shapes, but to different extents. RHT forms more clusters than TIC at all states and for all cluster sizes, hinting at the extent to which factors other than IQ alone must be influencing cluster distribution in these systems.

For spheres, TGR, and DDT there is a range in driving force from $\Delta\mu \sim 0.5$ to 0.6 where there is little difference between the probabilities of moderate-sized clusters, though the crossover behavior does appear to occur. Closer to coexistence, moderate clusters disappear from the polyhedra fluids while still appearing in the sphere fluid.

The densest local packings for spheres exhibit 12 neighbors in contact with a central particle, giving a local coordination number of $CN = 12$. All the shapes chosen for this study have a CN of approximately 12 in the fluid state. There are three equally dense arrangements of 12 neighbors around a central sphere, two that are consistent with stackings of hexagonal layers, and one with 5-fold, icosahedral symmetry. The icosahedral arrangement, unlike the other two, cannot be extended to fill space. All of these systems form FCC, an ABCABC stacking of hexagonal layers rather than HCP, an ABAB stacking of hexagonal layers, however the cutoff values we choose here do not clearly distinguish between the two hexagonal local environments, and we do not consider occasional HCP-like stacking faults to be significant to this study.

Our method for defining clusters considers the positions of particle centers, and does not take into account the orientation of particles relative to their environment or relative to their neighbors' orientations and positions. We use the potential of mean force and torque (PMFT) in the fluid to understand how the directionality of the particle facets relates to the location of neighbors [60].

PMFTs measure the relative probability for a particle to have a neighbor at any specific location in the surrounding volume [61]. It is a three dimensional extrapolation of the radial distribution function, taking into account the angular location of each neighbor as well as its distance. The PMFT measurements shown are taken of fluids where clusters of more than 18 particles have begun to form.

Differences in the local order stand out immediately between RHT and DDT (Figures 4.6d. and 4.6f.) and the rest of the systems. RHT and DDT both exhibit continuous networks of favorable neighbor locations, while the other three shapes have discrete favored locations. For RHD, shown in Figures 4.6(a)-(c), the favored locations for neighbors are positioned correctly for an FCC structure.

The relative probabilities of forming a crystal from the metastable fluid are shown in Figure 4.7. For each shape there is a narrow range of pressures where the fluid is stable enough to be sampled thoroughly, but crystallization occurs often enough for us to measure a crystallization probability without the use of advanced sampling techniques. This probability is the count of crystallization events observed per MC step in the fluid, per particle.

It is intuitively expected that RHD will crystallize readily, since the rhombic dodecahedron is the Voronoi cell of the FCC structure, filling space perfectly with full face-to-face contact between neighboring particles in the FCC configuration. The recently established emergent directional entropic forces [59] associated with face directions of RHD would promote the correct alignment of particles in the fluid. This makes the fluid more similar to the solid and lowers the driving force for the transition.

In the case of RHT we see that it crystallizes and clusters far more readily than spheres; it also crystallizes more readily than TIC despite their IQ and equations of state being remarkably similar. The differences between TIC and RHT in cluster size distribution are even more interesting when the local environment, as expressed in the PMFTs in Figures 4.6(d) and 4.6(e), is taken into account. The first thing that stands
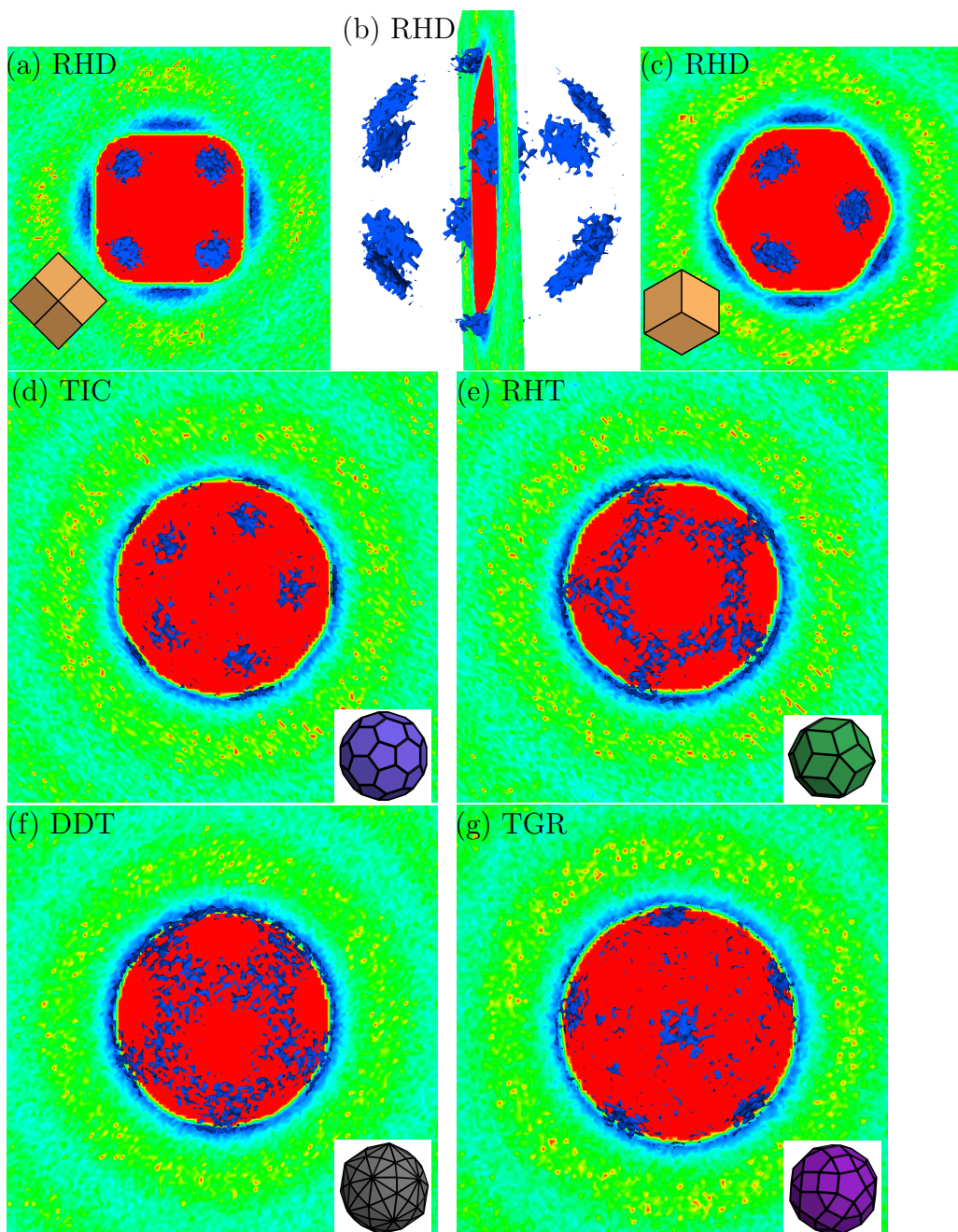
Figure 4.6: Potentials of mean force and torque (PMFTs), showing the relative energy of neighbor positions to the most probable neighbor position at $0k_BT$Images show an isosurface at $k_BT = 0.5$ (blue patches) superimposed on a planar cut through the particle center. Panel (b) is rotated to better show the three-dimensional nature of the PMFTs. Each shape is sampled at the states where large clusters have started to form (middle point of Figure 4.2). The patterns of local energy minima are consistent over different $\Delta\mu$ for each shape (see Supplementary Information).
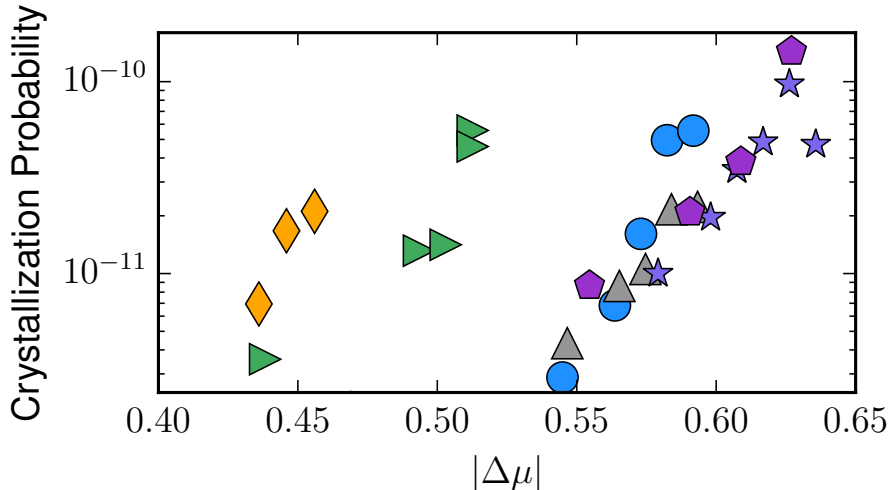
Figure 4.7: Crystallization probabilities, i.e. crystals observed/steps simulated of the metastable fluid, as a function of $\Delta\mu$. This plot shows the necessary driving force for the onset of crystallization for different shapes, up to the metastable fluid limit (MFL).

out is that RHT, like DDT, has a network of lowest energy locations for neighboring positions, indicating a freedom to rotate with constrained but connected orientations with little penalty. This freedom, while still including the neighbor arrangements congruent with FCC, seems to lead to clustering and crystallization probabilities not quite as high as for RHD where the local order is specifically matched to FCC, but still quite significant. TIC does have pentagonal faces oriented such that they could contact face-to-face while in FCC, but TIC can also have significant face-to-face contact in an icosahedral configuration, which is locally as dense and is entropically favored over FCC, but cannot fill space. Previous work has suggested that icosahedral, i.e. 5-fold, arrangements in dense fluids are a significant inhibitor of nucleation in metastable colloidal systems [35, 20, 32]. This means that we have competition between multiple locally close packings, both of which are promoted but only one of which can lead to a crystal.

The sphere, DDT, TGR, and TIC all exhibit a similar trend in the range $0.55 < \Delta\mu < 0.6$, with TGR and TIC remaining fluid up to a slightly higher $\Delta\mu$. The

differences in the PMFTs between DDT and TGR mimic those between RHT and TIC, but we propose that the sphere-like isotropic interactions dominate at higher ϕ, leading to the similar crystallization probabilities.

4.4 Conclusions

The shape of hard polyhedra is seen to influence the crystallization and clustering behaviors of the metastable fluids by two mechanisms, both related to emergent directional forces in the system.

Emergent directional entropic forces are visualized via the PMFTs, highlighting differences in local order, both in where neighbors are most likely to be and in how readily neighbors can move. One influence of shape is that directional effective attraction due to a preference for face-to-face alignment can directly promote specific local arrangements, which can promote crystalline local order, non-space filling local order, or both. For shapes where the order in the fluid is inconsistent with the crystal, we still find crystallization and clustering with higher probability than for spheres at some $\Delta\mu$ values, but find that for TIC and TGR the upper limit of metastability is greater than for spheres despite the presence of larger clusters. This suggests that the icosahedral tendencies in local order do not strongly discourage moderate hexagonal clustering, but promote 5-fold order which resists the growth of clusters.

In the cases of RHT and DDT we see in the PMFT the local environment can shift between configurations easily due to the connectivity between preferred neighbor locations (energy minima shown in the PMFT). This corresponds to crystallization occurring readily at driving forces only slightly greater than those that cause significant clustering, indicating that the availability of local rearrangements promotes crystallization. Additional parameters describing the local fluctuations in the parent phase are necessary to describe the entire crystallization process. This is consistent with recent work describing pre-nucleation heterogeneity in metastable phases [36, 65]

41

and highlights the importance of orientation in describing the local environment and pre-nucleation "clouds" in systems during crystallization.

4.5   Supplementary Information

To determine the density of the system and the distribution of cluster sizes we must allow a fluid to equilibrate thoroughly. In order to do that we must determine a relaxation time, i.e. a number of MC sweeps between two samples of a trajectory for those samples to be uncorrelated. At pressures where nucleation events occur on a time scale accessible using MC simulations, it is correspondingly difficult to obtain sufficient samples of the metastable fluid. This is shown in Figure 4.8, where the system packing fraction is plotted over time for many trajectories. Samples taken every 20 thousand steps are still correlated, allowing us to see drift in overall system density. We see in Figure 4.8(a) that at $P^* = 9.00$ the system ϕ ranges between 0.5345 and 0.537 while remaining stable. In Figure 4.8(a) it is apparent that the system has reached a metastable equilibrium around a packing fraction of 0.536, with long-term fluctuations of ± 0.0015.

Figures 4.8(b) and 4.8(c) appear similar, but closer inspection shows that in the case of (c) fewer of the trajectories remain fluid after equilibration, while 4.8(b) shows all 16 of the trajectories as equilibrated metastable fluid after equilibration. The 16 trajectories in 4.8(b) do all crystallize, but they do so over the course of over 15 million MC sweeps. This means that we cannot trivially sample the metastable fluid thoroughly, so 4.8(c) shows that $P^* = 9.25$ is above the MFL for spheres.

Figure 4.9 shows the same 5 points as Figure 4.2, now plotted in terms of P^* rather than $\Delta\mu$. This allows us to see that the differences in coexistence pressure and eos, even taken together, do not correspond to the trends in behavior between these shapes. TGR and TIC fluids are stable to a higher pressure as well as a higher driving force. RHT, which we noted required a lower driving force than the more
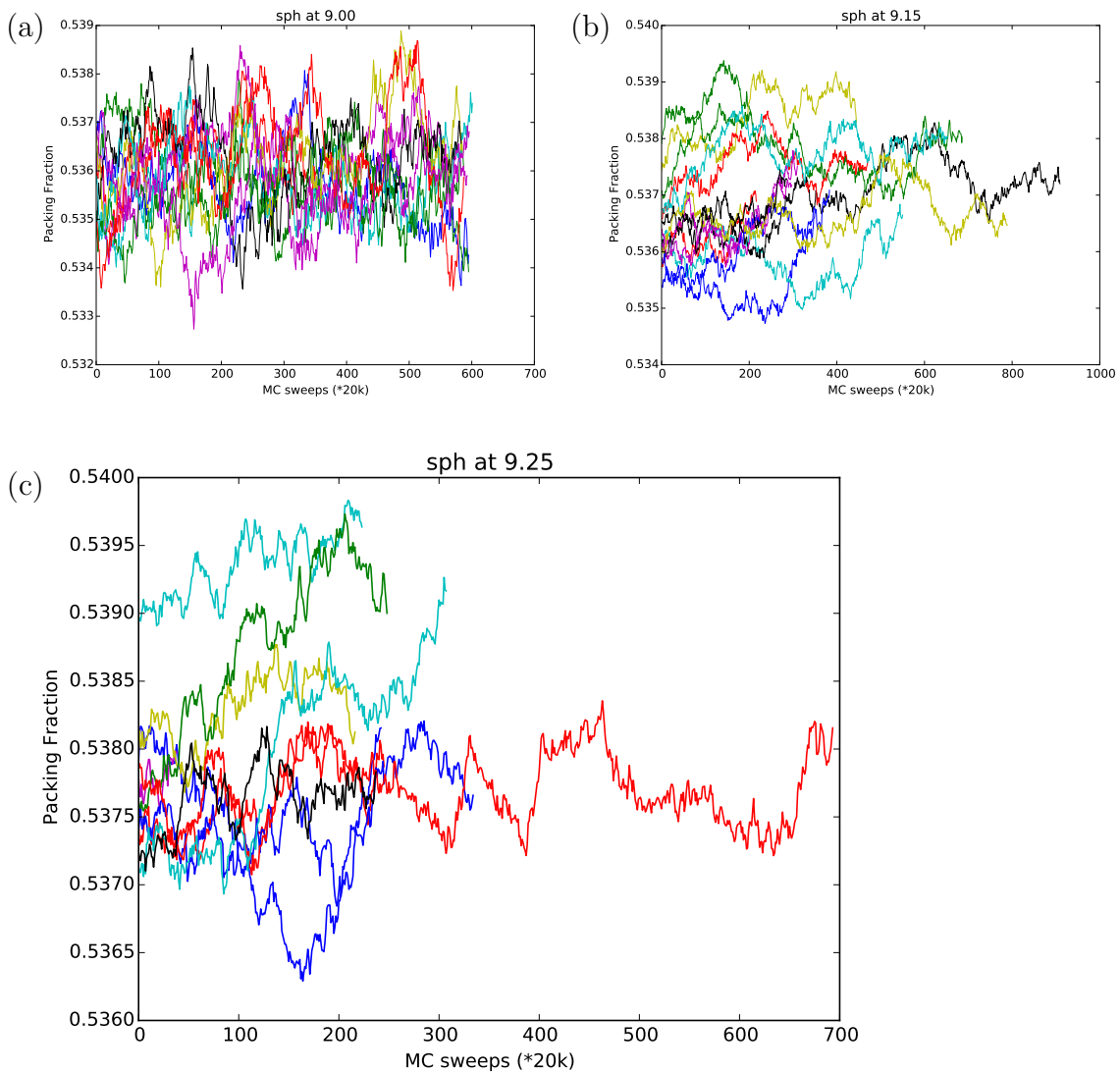
Figure 4.8: Packing fraction *vs.* time at higher pressures measured for systems that remain fluid after initialization and equilibration for one million MC sweeps. Fluids are measured every 20k MC sweeps. In (a) sphere fluids at $P^* = 9.00$ sample a metastable range of ϕ for several million steps without crystallizing, showing that crystallization is a rare event. For $P^* = 9.15$ in (b) the fluid remains metastable for several million steps for the majority of trajectories, allowing good sampling of the metastable fluid, but nucleation is not a rare event. Shown in (c), at $P^* = 9.25$ some trajectories remain fluid for millions of timesteps, but only approximately half are still fluid after the initialization and equilibration routine, putting it above the metastable fluid limit for this study.
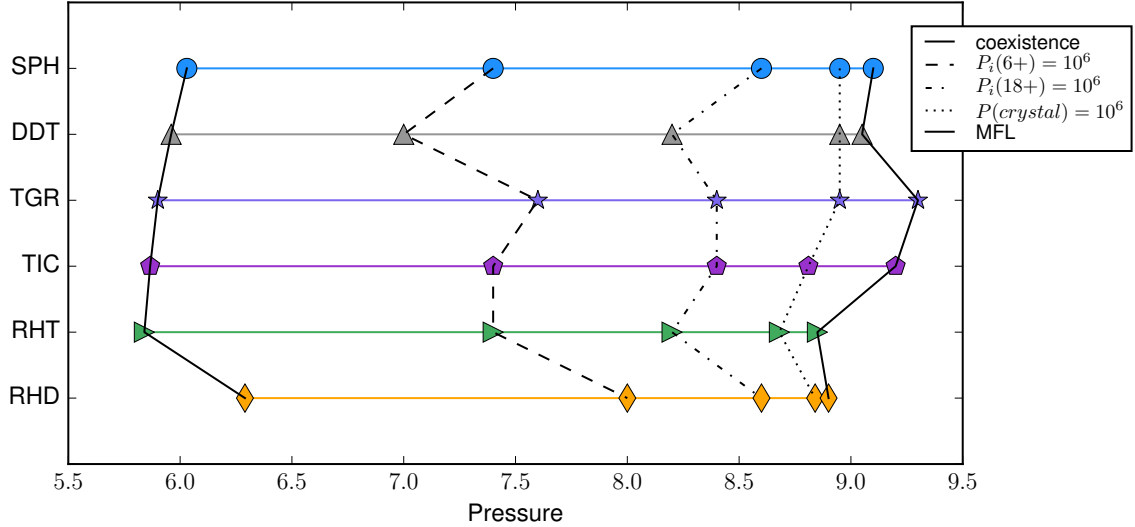
Figure 4.9: For each shape we show five important pressures: The fluid-solid coexistence pressure, the pressure where particles have probabilities $P_i(X > 5) = 10^{-6}$ and $P_i(X > 18) = 10^{-6}$ of being in a cluster greater than 5 or 18 respectively, the pressure where the probability per particle of crystallizing is 10^{-11}, and the MFL pressure at which the fluid is not stable enough to sample thoroughly.

spherical fluids to nucleate, stands out even more because it has a coexistence pressure slightly lower than that of spheres while RHD has a higher coexistence pressure. It is RHT, rather than RHD, that crystallizes readily at the lowest pressure, though RHD requires less driving force. DDT is found to form moderate sized clusters at significantly lower pressures than other shapes, though it does not crystallize easily pressures quite similar to spheres.

Figure 4.10 shows the results of scaling ϕ by IQ. It was noted that the equations of state increase with increasing IQ, and it was speculated that an isotropic measure of sphericity could be used to account for the difference in ϕ at a given pressure between shapes, a measure that could compensate for some effective volume taken up by the object. While the general trend for this set of shapes follows ordering by IQ, Figure 4.10 shows that IQ overcompensates for the difference. No isotropic measure of sphericity, such as IQ, $\sqrt{IQ}$, r_{out}/r_{in} where r_{out} and r_{in} are minimum
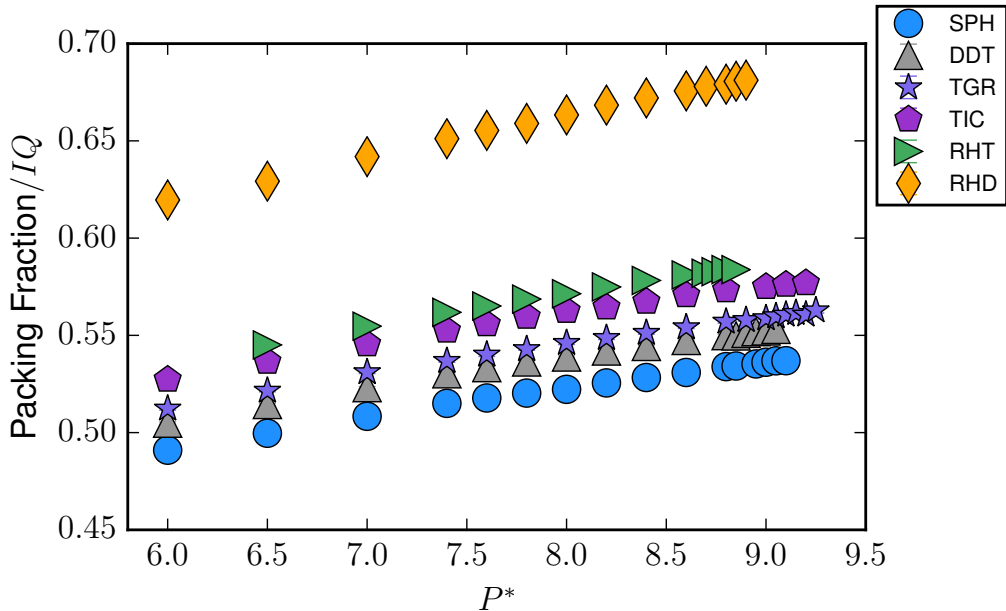
Figure 4.10: A modified EOS, plotting ϕ/IQ for each shape as a function of pressure. IQ overcompensates for the difference in ϕ between shapes.

and maximum interaction distances, has been found to accurately compensate for the difference in fluid density caused by shape and collapse the particle equations of state onto a single master curve.

To further understand the relevance of sphericity, we can consider the trends of the five points of interest we considered in Figure 4.2and Figure 4.9: coexistence, moderate and large sized cluster appearance, directly observable crystallization, and MFL, but plot them in terms of IQ. Figure 4.11 show that the coexistence points of all shapes - except for the special case of RHD - follow a trend of decreasing coexistence pressure.

As mentioned previously in terms of $\Delta\mu$, the narrow range of pressure where crystallization probabilities are high enough to be measured directly from straightforward MC simulations but nucleation is still a rare enough event that the metastable fluid can be sampled thoroughly is wider for TGR and TIC, and narrow for RHD. That easy crystallization range is relatively narrower for RHD, which promotes hexagonal
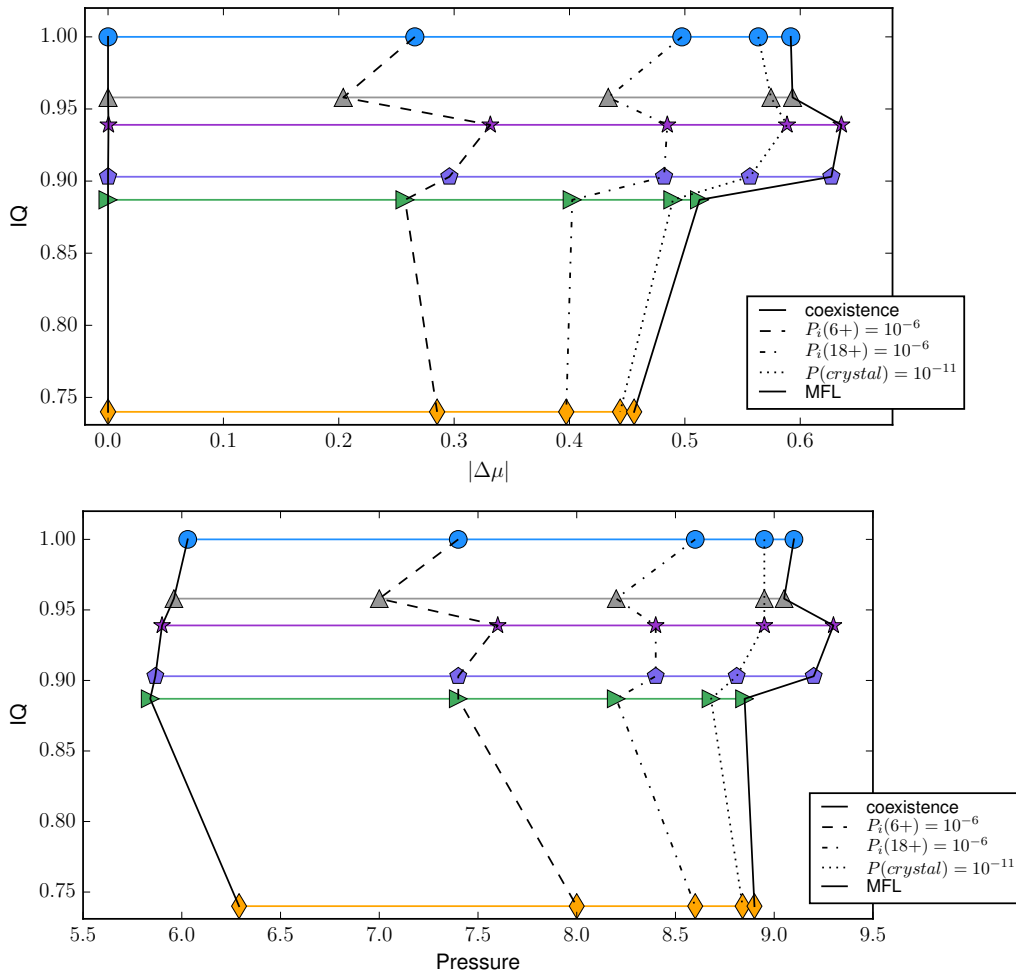
Figure 4.11: Additional views of the five points of interest shown in Figures 4.9 and 4.2. The same data is presented in terms of $\Delta\mu$ and P^*, with shapes on the Y-axis plotted by their IQ values.

order, and wider for shapes that promote other local order. RHT systems meet both of those criteria, and have a resulting range of driving force where crystallization is easy to observe from metastable fluid similar to that of spheres. If we consider the same five points in terms of IQ *vs.* ϕ, we see a trend emerge that removes the seemingly strange behavior of the MFL for the more spherical shapes. This indicates more consideration of ϕ may be in order, but the trend in the MFL breaks down for the least spherical shapes, and trends for moderate cluster formation in the dilute metastable fluid, where crystallization events are rare, are not explained by considering ϕ. Quantitative analysis of these emerging trends is awaiting sampling of additional shapes, but we can see from figures 4.9, 4.11, and 4.12 that driving force and geometric constraints related to ϕ are both important as we study emergent behavior caused by particle shape.

All of these observations serve to reinforce that the behavior of the fluids are strongly dependent on subtle changes in shape.

4.5.1 Properties of FCC forming shapes

Table 4.1: Properties of the investigated polyhedra: Disdyakis Triacontahedron (DDT), Trigyrate Rhombicosidodecahedron (TGR), Truncated Icosahedron (TIC), Rhombic Triacontahedron (RHT), Rhombic Dodecahedron (RHD)

ID	IQ	ϕ_{max}	Vertices	Faces	Point group
Sphere	1.000	0.741	-	-	∞m
DDT	0.958	0.766	$12_{10}+20_6+30_4$	120_3	$\bar{5}\bar{3}m$
TGR	0.939	0.805	60_4	$20_3+30_4+12_5$	$3m$
TIC	0.903	0.785	60_3	$12_5 + 20_6$	$\bar{5}\bar{3}m$
RHT	0.887	0.802	$12_5 + 20_3$	30_4	$\bar{5}\bar{3}m$
RHD	0.740	1.000	$6_4 + 8_3$	12_4	$m\bar{3}m$

The shapes of these polyhedra are well defined as a convex hull surrounding a set of vertices. Table 4.1 shows properties of the shapes. Of interest are the differences in the numbers of faces and vertices, and the different associated symmetries. Generally
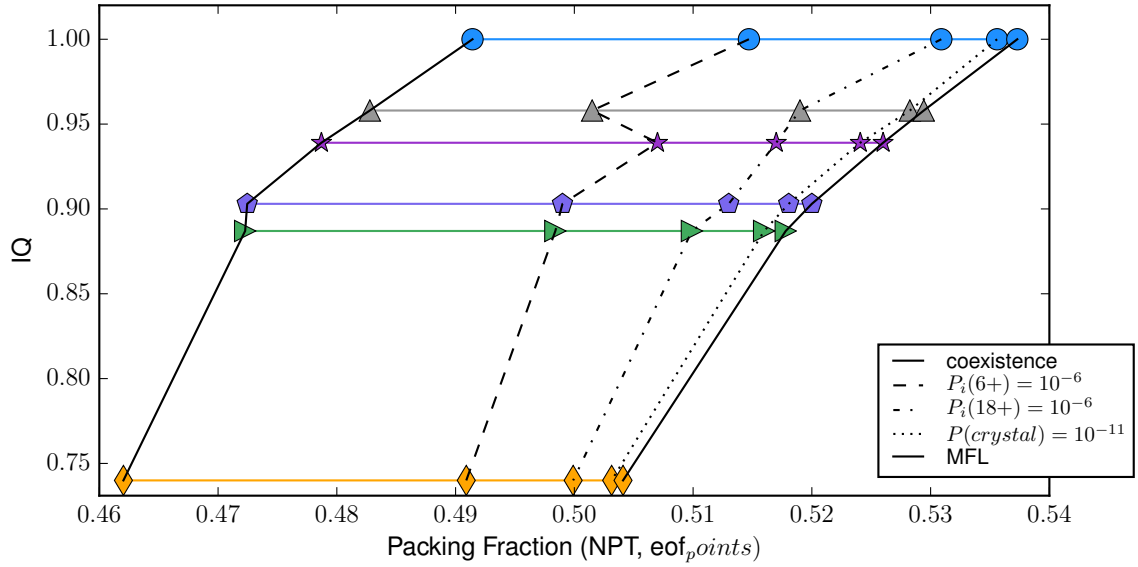
Figure 4.12: Five points for each particle as shown in Figures 4.2, 4.9, and 4.11, *vs.* ϕ. These highlight the joint role of IQ and ϕ in coexistence and fluid stability in the trends of the first and last points. It also shows that additional properties must be considered to understand the clustering behavior in the metastable fluids.

the more faces a shape has the smaller that shape will be. The different faces can have different sizes, for instance TIC has two types of faces, one with five edges and the other with six, but all of its vertices are 3-fold.

CHAPTER V

Rare event sampling of transition path states

5.1 Abstract

This chapter studies in detail formation of crystalline nuclei in regimes where nucleation is both a rare event (meaning that the time between events is orders of magnitude longer than the duration of an event), and a non-equilibrium process. When the barrier to nucleation is much larger than the fluctuations in the metastable fluid, that fluid can be sampled as a stable, equilibrated state. Deviations from the equilibrated metastable fluid, such as the local ordering needed to form a stable nucleus, are rare. Sampling rare crystallization requires that we sample rare, out-of-equilibrium states and determine their relationship to transition states that lead to the lower energy minimum crystal structure. We have developed a method for generating ensembles of these out-of-equilibrium states along with their probabilities of crystallizing, relating them to transition pathways. We have generated such ensembles for spheres and several highly spherical polyhedra.

This chapter is divided into three sections. First, I review existing rare event sampling methods. I then discuss work on nucleation rates in highly spherical polyhedra using forward flux sampling (FFS)-type rare event sampling[1]. Thirdly I discuss

---

[1] "Nucleation of FCC Crystals from Hard Polyhedra of Various Shapes", Samanthule Nola and Sharon C. Glotzer, Presented at the AiChe Conference in Atlanta, GA, in Anisotropic Particles: Synthesis, Characterization, Modeling, Assembly, and Applications II, Nov. 19 2014

Adaptive Segmented Transition Sampling (ASTS), a modified FFS-type rare event sampling method I developed for further study of nucleation processes in hard polyhedra, and present briefly some new data coming generated by ASTS sampling.

5.2 Background: Rare event sampling methods

We want to sample unlikely things, to understand how unlikely they are, and to understand how unlikely they are in relation to other systems. Nucleation is a rare event, meaning that the time between events is at least an order of magnitude longer than the event duration. A wide variety of methods have been proposed and used for sampling rare events in different systems, many derived from Transition State Theory (TST), some tied to investigations of energy landscapes and ergodicity. Many reviews have been published discussing uses, strengths, and flaws of different methods, *e.g.*, E and Vanden-Eijnden [18], Bolhuis et al. [12].

These methods fall broadly into two families: methods that make rare events more likely by imposing an explicit energetic bias on the system, and methods that employ preferential sampling to pick out states of interest from unbiased simulations. An example of a biasing method is umbrella sampling, which generates short trial trajectories and then accepts or rejects the entire trial trajectory in a Metropolis MC-type manner based on the change in energy ΔU_{trial} of the system after the trial trajectory. The bias is applied by adding an order parameter (OP)-based term to the calculation of ΔU, so that the acceptance probability $P_{keep} \propto exp(-\Delta U/k_B T)$ is greater for trial trajectories that bring the system closer to the state promoted by the OP. This is a useful method to understand energy barriers when the transition path is well known, and a useful way to explore the energy landscape along an order parameter [57].

Homogeneous nucleation occurs due to fluctuations in the supercooled fluid, but there are many sorts of fluctuations in a supercooled fluid. In this work we wanted

to make minimal *a priori* assumptions regarding the transition path between fluid and solid, so we chose to look at the unbiased family of methods . This family of sampling schemes, including Transition Path Sampling (TPS) [12], Forward Flux Sampling (FFS) [1], Branched Growth Sampling, Rosenbluth Method [40], Transition Interface Sampling (TIS) [63], Partial Path Transition Interface Sampling (PPTIS) [62], and related methods implemented for specific situations, have been developed to sample rare transitions efficiently, without biasing or constraining the nature of the transition.

From this family, we follow the work of Filion et al. [19], implementing a FFS-type method. After completing initial sets of simulations we saw the need for advancements such as PPTIS developed from TIS, and pruning schemes used by Allen et al. [2]. We adapted our methods in analogous ways, which will be detailed below.

In general, FFS-type methods sample a transition using a set of interfaces $[\lambda]$ and committers, which I will explain here. The interfaces $[\lambda]$ are defined in phase space with λ_A defining the boundary of phase A, and the final interface λ_B defining the boundary of phase B. The interfaces are non-intersecting surfaces in phase space, each defined by a specific value of an order parameter (OP). The OP need not be a proper reaction coordinate that tracks the path and progress from A to B, but must clearly specify limits of A and B, and must be defined at all points between them. This allows interfaces to be defined between A and B that must be crossed sequentially to transition, and can be spaced to suit a particular system. The interface setup is shown schematically in Figure 5.1. I will discuss considerations for the spacing of interfaces in detail in the final section of this chapter.

This method does not rely on applying an energetic bias to the system, but tracks the progress through phase space between the initial and final configurations by noting crossings of λ. It is interesting to note that these interfaces describe what is fundamentally a $6N$ dimensional system when the position and orientation of all N
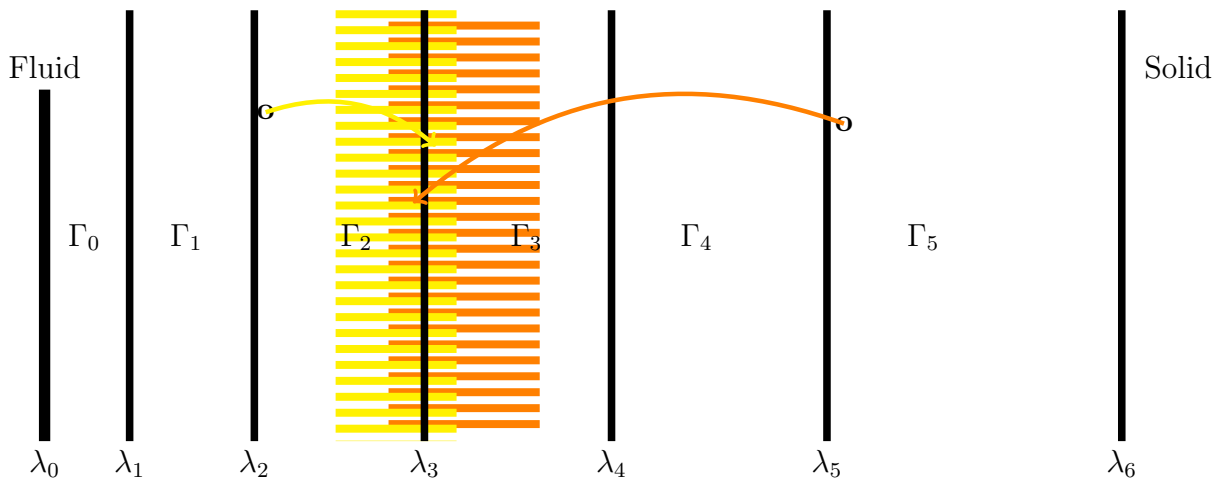
Figure 5.1: Schematic diagram of interfaces separating initial and final states. Shaded regions schematically show the effective locations of clusters when crossing λ_3 from below (yellow) or above (orange), both of which are considered part of the Γ_3 node when using Markov chain analysis to calculate nucleation rates.

particles in 3-space is considered, but project the problem into a more tractable one dimensional space. The progress of the system in the transition is measured along the OP axis, as shown in Figure 5.1.

Committers, the other primary component shared by these methods, are sets of trajectories started from the same "snapshot" (a specific configuration of the system with position and orientation of all particles, *i.e.* a microstate) of a system. Each of these trial trajectories is tracked either for a prescribed duration or until an end condition is met. The results of the set of trial trajectories are used to determine the relative probabilities of different system behaviors. Consider a committer that starts from a state at λ_i and runs 20 trajectories. If 12 of the trial trajectories reach λ_{i+1} and 8 melt, this results in a probability $P(\lambda_n|\lambda_1) = 0.6$. For systems in a rare event regime it is impractical to measure entire transitions from state A at λ_1 to B at the final interface λ_n, because $P(\lambda_n|\lambda_1)$ is, by the definition of "rare event", immeasurably low. Interfaces are spaced to make each $P(\lambda_n|\lambda_1)$, called P_i for convenience, measurable,

then $P(\lambda_{i+1}|\lambda_i)$ (P_i) is measured for each interface.

The definition of P_i depends subtly on how melting is defined. Most FFS-type algorithms consider a system to melts when it reaches λ_0, and measure P_i as the relative probability of growing to the next interface vs. melting all the way to the fluid at λ_0. Some implementations introduce a pruning algorithm to end some trajectories that have melted significantly without waiting for them to reach the pure fluid [2]. Pruning can significantly reduce computational investment, especially in cases such as trial trajectories starting near the top of a flat energy barrier, where melting can be slow. We address an explicit implementation analogous to pruning in section 5.4.6, shown schematically by the orange and yellow markings in Figure 5.1.

In the common case where P_i gives the probability of growing to the next interface before melting to a fluid, the probability of crystallizing from the penultimate interface $C_{n-1} = P_{n-1}$. The probability of crystallizing from the interface below is then $C_{n-2} = P_{n-2} \times P_{n-1}$. Following from that the probability of crystallizing from the first interface, $P(B|A)$, can be found as

$$P(\lambda_n|\lambda_1) = \prod_{i=1}^{n-1} P(\lambda_{i+1}|\lambda_i), \qquad (5.1)$$

where the product is over all n interfaces. The nucleation rate is then $P(\lambda_n|\lambda_1)$ multiplied by the probability of being at the first interface (or the flux across the first interface in the TIS method of van Erp and Bolhuis [62]).

There are subtle differences between the different methods, some because of the specific question under investigation, and some to minimize the use of compute time, data storage, or other resources. Forward flux sampling uses a committer to find each probability $P(\lambda_{i+1}|\lambda_i)$, then chooses only one successful trajectory from the committer's set of successful trial trajectories and uses the end point of that trajectory as a "child" snapshot to start the $i+1$ committers. The short trajectories can later be stitched together and used as a single complete transition path. The Branched

Growth method, as the name may suggest, keeps all successful trial trajectories and generates branched transition paths.

Transition interface sampling and partial path transition interface sampling [62] use the committers to measure mean first passage times (MFPT), and frame their data in terms of rates rather than probabilities. This gives additional insight into the behavior of the system, but requires molecular dynamics, kinetic MC, or some other understanding of time throughout the simulation process.

All of these methods are subject to a competition between generating sufficient samples of the unlikely events they are designed to investigate, generating an intractably large dataset, and accurately weighting probabilities and stitching together partial trajectories.

The method we used in our initial sampling is an implementation of FFS with pruning explicitly built in, considering trial trajectories starting from λ_i to "melt" once they fall below λ_{i-2}, two interfaces below their starting state, rather than wait for them to melt completely.

Calculating $P(B|A)$ as described in Equation 5.1 makes the assumption that the only options for a trajectory are to grow to the next interface or to melt completely.

An overview of a general FFS-type workflow is shown here, noting specifics of our initial implementation:

1. Choose interfaces (based on hard sphere literature).

2. Run simulations of phase A (metastable fluid).

   Collect N_{root} generation 0 snapshots from the fluid with $OP = \lambda_1$.

3. For each λ:

   Run committers (sets of trial trajectories to determine relative probabilities of different types of system evolution from that starting state) on each of the

$N_{frames_{\lambda}}$ snapshots available in λ. This generates the next generation of snapshots from successful trial trajectories and produces local statistics ($P(\lambda_{i+1}|\lambda_i)$ in our case).

Each generation corresponds to an interval in phase space between interfaces; statistics are averaged for each interval.

Resolve pruning of trial trajectories while running (a trial is considered to have melted if it crosses below λ_{i-2}).

4. Use local statistics to resolve overall system behavior.

One main advantage of FFS-type sampling methods is that the validity of the results do not depend on how well the OP describes the transition or generally on how well the set of interfaces, $[\lambda]$ are chosen. The efficiency of the process, however, depends strongly on the choice of $[\lambda]$. Consider that the number of frames generated by the running process in each interval is the product of the number of initial states and the average number of children generated per frame in each interval, ks_{λ}. The total number of frames is the sum of these:

$$N_{frames_{total}} = N_{root} \times \sum_{i=1}^{n} \prod_{j=1}^{i} ks_j. \tag{5.2}$$

If each committer uses ten trial trajectories, P_i will be known to a precision of 0.01 for each committer, and each frame can generate up to 10 "children". The children are each saved as single frames, but the rest of the trial trajectory is not saved. Each of those children can generate up to 10 children, and so on, though most generate fewer than 10 children. The maximum number of frames generated in an interval is the number of initial frames times the number of trials per committer to the power of the number of interfaces below. The number of interfaces depends on the system, ranging generally between 6 and 14. The purpose of the process is to determine the likelihood

of progressing from any given state (generating child frames), and it varies greatly. A rough estimate of the number of snapshots at an interface is given by using Equation 5.2: assuming each committer generates three child snapshots, over ten intervals, with 50 initial frames, the final interface would have $50 \times 3^{10} = 2,952,450$ frames.

The wide variation in P_i between samples at the same i, as well as the subjectivity of selecting $[\lambda]$ for systems of different shapes, led us to design and implement further modifications to our sampling method. We will discuss details of our advanced implementation in the last section of this chapter. First we will present the results of preliminary investigations on two highly spherical polyhedra along with spheres, where we found surprising differences in the rare event regime crystallization behavior due only to small differences in shape. We well then describe adaptations to the sampling method, briefly show new results given by those adaptations, and finally discuss ongoing use of the new method.

5.3   Nucleation of hard sphere-like polyhedra

Hard spheres have been studied thoroughly and are a major part of understanding the contribution of entropy to crystallization [21, 34]. The solidification process for hard particle systems is known to follow a nucleation-and-growth type process in many cases. Ordered clusters exist spontaneously and transiently in the dense fluid. These clusters, pre-critical nuclei, have some probability of becoming large enough to be stable, and a related probability of growing to encompass the entire system. The critical nucleus is defined here as one that has a 50% chance of growing to span the system.

Within the known framework of hard particle crystallization, here I address the question: what is the contribution of shape to the rate and process of forming critical nuclei? In order to isolate nuclei we need to sample systems where multiple nuclei are unlikely to form at the same time. This restricts us to studying systems closer
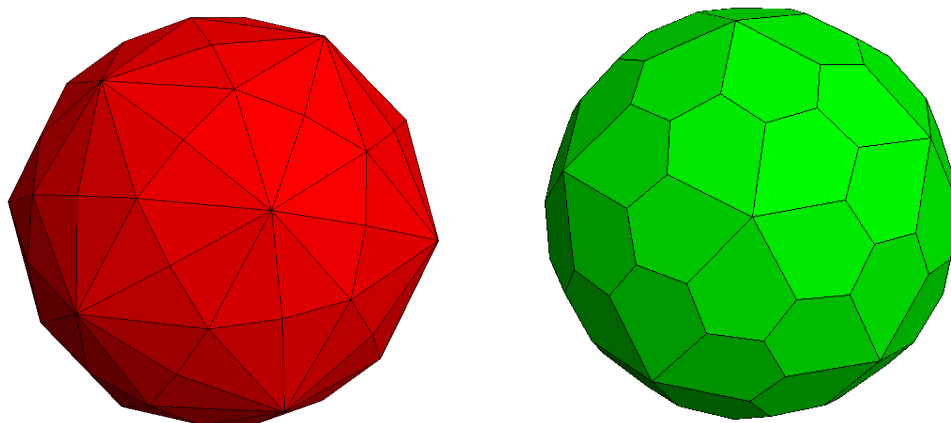
Figure 5.2: Two sphere-like polyhedra, the disdyakis triacontahedron (C12, red) and the pentagonal hexicontahedron (C13, green).

to the fluid-solid coexistence point, where nucleation is a rare event and the driving force for crystallization is small. Here we compare monodisperse systems of spheres to two faceted shapes, the dysdyakis triacontahedron (C12, called DDT in previous chapters) and the trigyrate hexicontahedron (C13), shown in Figure 5.2.

Simulations were run on systems of 4096 particles at fixed temperature and volume (NVT) using the Hard Particle Monte Carlo (HPMC) implementation developed with HOOMD-blue [3, 5]. We considered each of the three shapes at four packing fractions. We found that the range of packing fractions where FFS was effective for measuring nucleation rates was different for the three shapes, although there was some overlap.

With the expectation that highly spherical particles would behave similarly to spheres, the same FFS parameters were initially set for all systems. The interfaces were chosen following Filion et al. [19]'s implementation of TIS/FFS sampling to validate our implementation for spheres before extending it to compare with polyhedra.

5.3.1 Nucleation probabilities

We used our FFS implementation to run committers and calculate crystallization probabilities. These probabilities of forming a crystal out of a dense fluid, as calculated by converging the Markov-like chain using the unweighted averaging of the
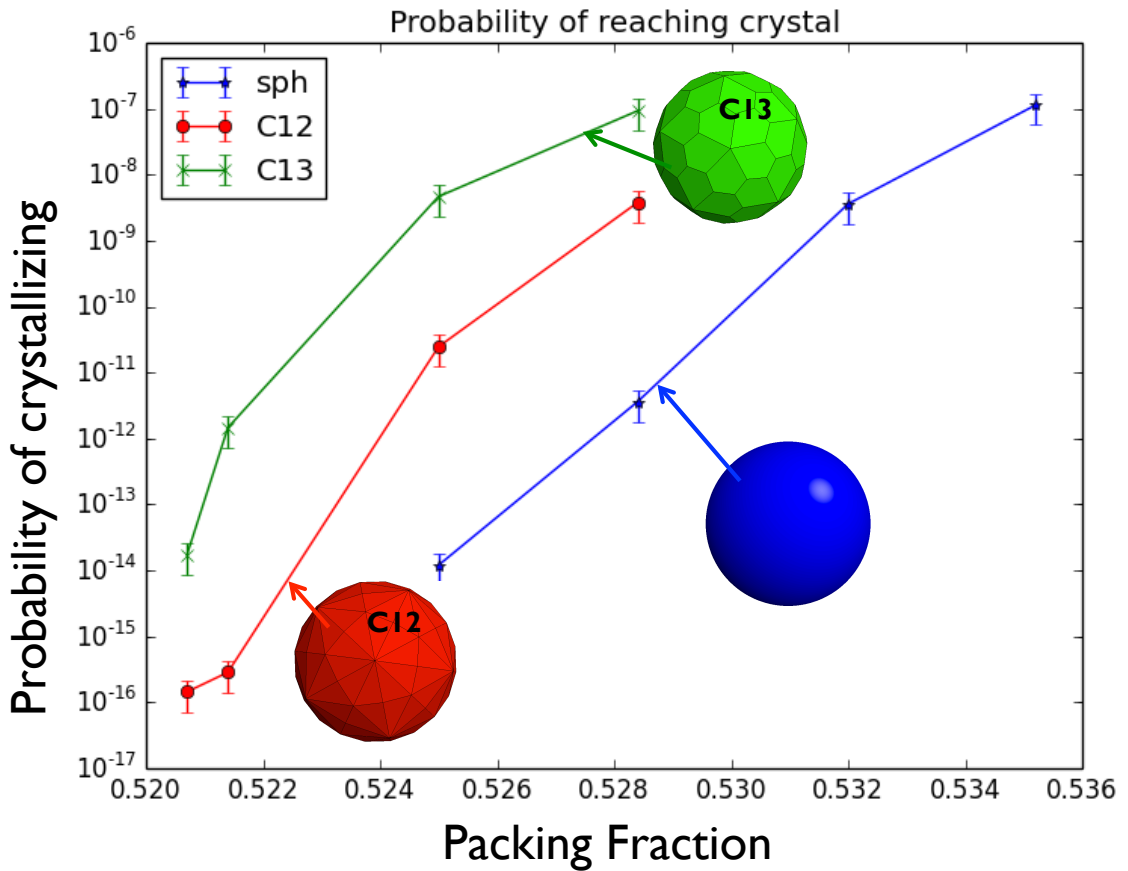
Figure 5.3: Crystallization probabilities for spheres, C12, and C13. Error bars indicate variation between correlated branches.

progress probabilities of all trials starting in a given interval, are plotted in Figure 5.3.

We see in Figure 5.3 the unexpected differences between crystallization probabilities of these three shapes. Not only are the polyhedra systems different than the spheres, they are notably different from each other. Below a packing fraction (ϕ) of 0.525 we were unable to sample enough crystallizing trajectories to calculate a crystallization probability for spheres, while both polyhedra were able to assemble the crystal eventually down to a $\phi = 0.5207$. On the other end, at $\phi > 0.529$, the two polyhedra systems crystallized readily, and nucleation was not a rare enough event to need advanced sampling techniques. At more moderate ϕ values where nucleation and

crystallization could be sampled for all three systems we found that there were orders of magnitude differences in the probabilities of crystallizing between the shapes, with C13 crystallizing most readily at each ϕ studied.

This indicates that the difference in crystallization behavior can not be explained entirely by the mere introduction of rotational entropy, but depends on additional properties of the shapes. It is not sufficient to consider only the overall probability of the system to crystallize; we want to know more about the path to crystallization. We want to understand what parts of the path were changed, and how the shape of the particles led to those changes.

5.3.2 Cluster growth

The local growth probabilities measured from individual committers are a means to calculate the overall transition probability, however there is a great deal of additional information to be found from the probabilities of a cluster growing vs. shrinking from each interface.

Figure 5.4 shows the probabilities P_i of cluster growth from clusters of different sizes i, in various system states. Each subplot in Figure 5.4 shows individual committer values *vs.* the size of the largest cluster in the initial snapshot. Along with the individual committer values (black dots), each plot shows the average probability of progressing for all samples at that interface, connected by a colored line as a guide to the eye. The average probabilities in these plots are unweighted, considering each snapshot individually rather than averaging "sibling" snapshots together to weight branches equally (branch weighting will be discussed further in Section 5.4.3. The top row shows spheres, the middle row shows C12, and the bottom row, C13.

The interfaces were chosen manually based on hard sphere literature for spheres at each ϕ, then used for all systems with the intent of comparing growth probabilities of same-sized clusters between shapes. Indeed there was a significant difference in

59
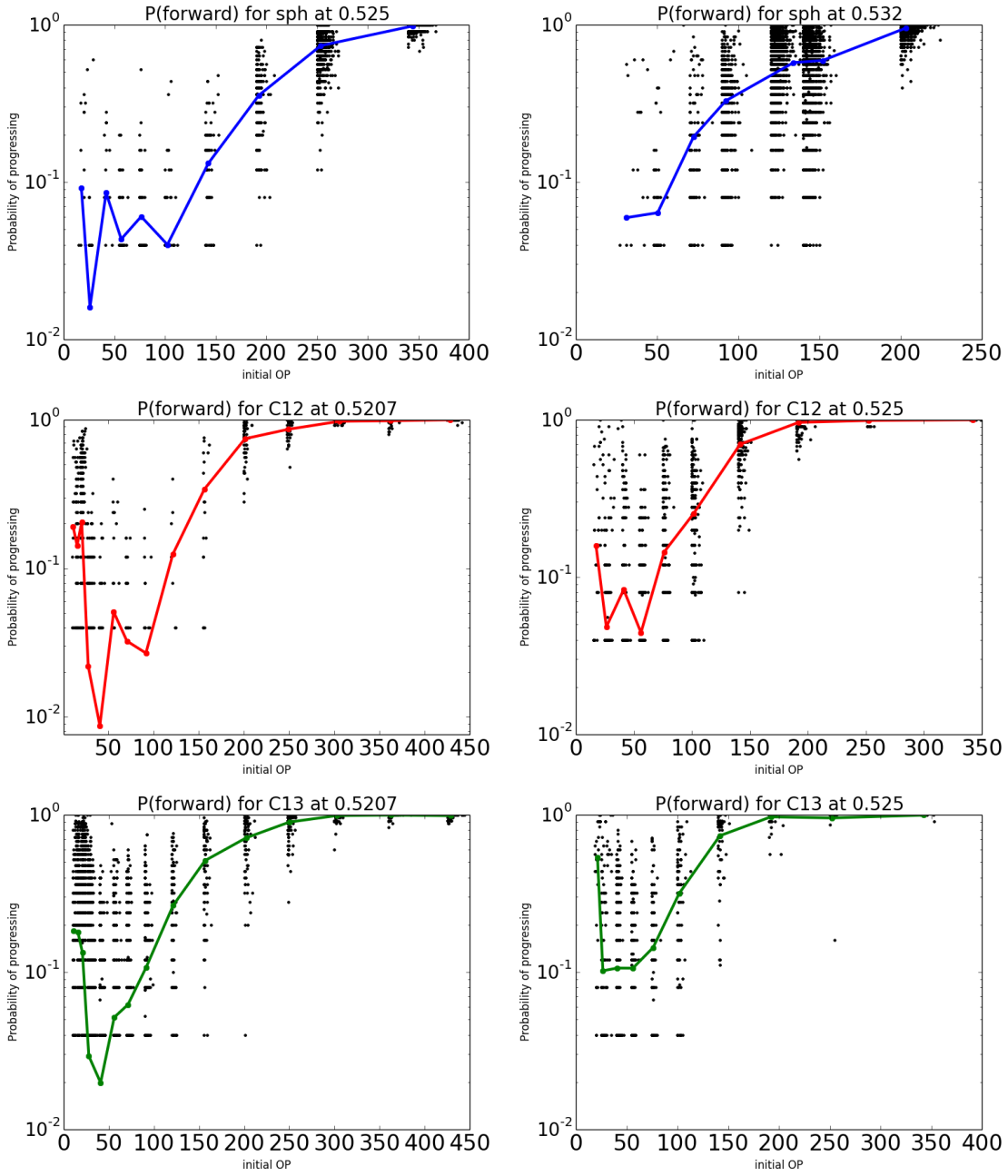
Figure 5.4: Plots of P_i, the probability of reaching the next interval, λ_{i+1}, without regressing below λ_{i-2}. Green line is an eye guide for average, black dots show individual non-zero values to highlight the distribution (zero values do not appear on the log plot). Shown for three shapes at two packing fractions each.

the cluster size distribution between shapes. So much so that it was clear a more adaptive and objective way of setting interfaces was needed.

The top row of Figure 5.4 shows the growth probabilities of sphere systems at packing fractions $\phi = 0.525$ and $\phi = 0.532$. The probabilities are in increments of 0.05 since each committer used 20 trial runs so the minimum probability of progressing was 1/20. In the first image, the interfaces are spaced so that the average probability of progressing fluctuates around tractable values, despite significant variation between individual committers. There is some variation in the exact initial cluster size, but there are clear bands showing that the snapshots were taken from around an interface.

The first panels in the bottom two rows are for $\phi = 0.5207$, and show that growth to subsequent interfaces was not at all rare. For C13 at $\phi = 0.5207$ we see that cluster sizes of up to 30 particles are sampled freely, not confined to the interfaces. This is the case for the ϕ values sampled for C13, showing that the interfaces are spaced too closely. This does not necessarily mean that the data is invalid, proper branch weighting will correct for the high population of descendants from early successful states. It does raise a question of how unbiased FFS methods are in cases where interfaces are spaced closely enough for transitions to happen more often than the sampling rate. In that case the probability of passing from one interface to another may not be measured accurately. This is another concern we will address in Section 5.4.

For systems like the spheres at $\phi = 0.5207$ there is a temptation to decrease the interface spacing until new samples become common. This leads to a situation where new samples can be made that are not identical to the parent but are within fluctuations of the OP. The number of particles in a cluster can change slightly by the inclusion or exclusion of surface particles that have an ambiguous local environment.

Indeed, the trends seen for C13 in Figure 5.4 are corroborated when we look at the distribution of cluster sizes in the fluids. Figure 5.5 shows the rates of observing
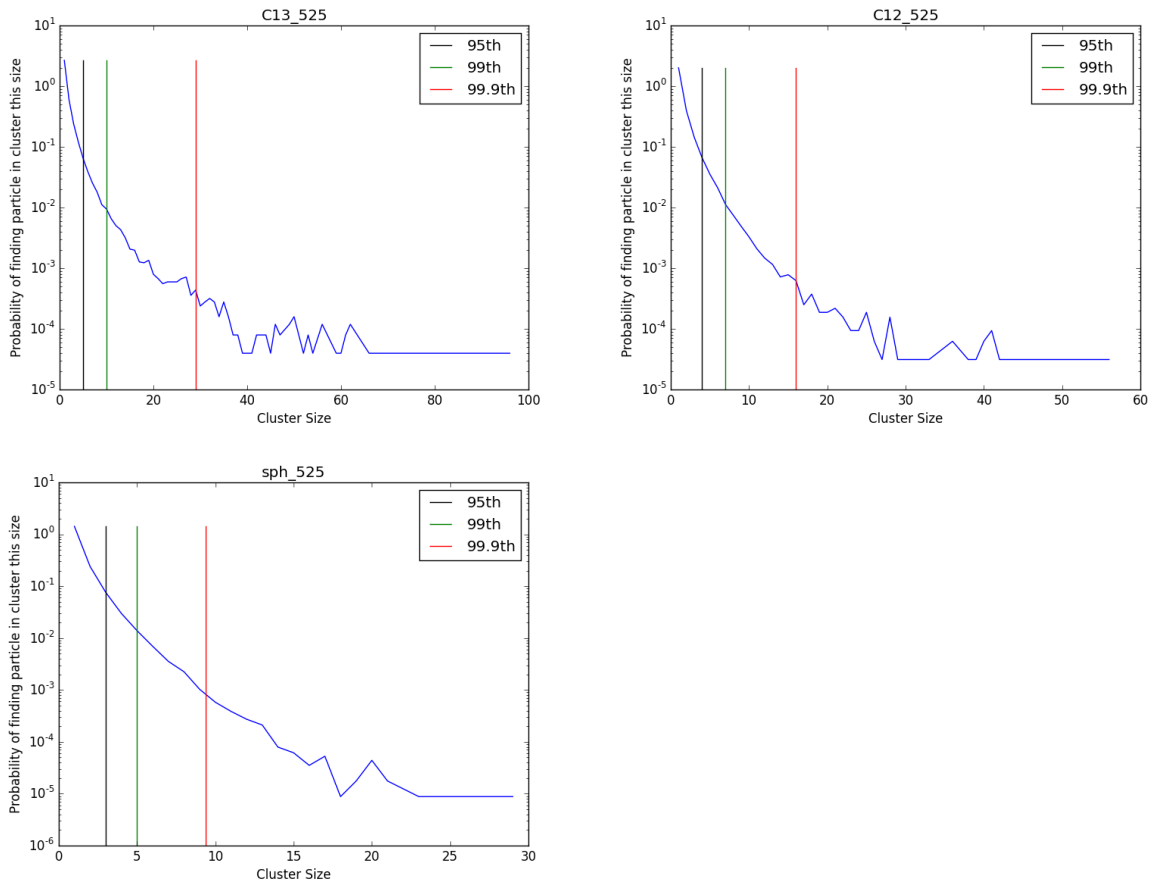
Figure 5.5: Expectation value of finding clusters of increasing size in the dense fluid at $\phi = 0.525$ for spheres, C12, and C13. Vertical lines indicate cluster sizes encompassing 95%, 99%, and 99.9% of observed clusters.

clusters in metastable fluids of each shape at $\phi = 0.525$, with vertical lines denoting what cluster size is seen in 5%, 1%, and 0.1% of observations. Note that these are direct measurement rates of cluster sizes, not only the largest cluster in the system as is used for the OP, so it is valid that multiple small clusters can be seen in an average snapshot for the polyhedra. What is most notable in Figure 5.5 is that for spheres, less than one snapshot in 1000 has a cluster greater than 10, but for C12 one in 1000 snapshots has a cluster greater than 16. C13 has clusters of at least 10 particles roughly ten times as often as sphere systems do, a surprisingly dramatic difference for such a subtle difference in the shape.

The inconvenient placement of interfaces did not invalidate these results, but did highlight several important factors to improve on. Specifically, it highlighted the need to properly weight branches and ensure broad sampling without overloading any single branch, and the need to set the interfaces objectively based on system behavior in a specific state and region of phase space.

5.4 Adaptive segmented transition sampling

Our adaptive sampling method closely follows the FFS method, and makes modifications similar to those made by van Erp and Bolhuis [62], Metzner et al. [41] when developing PPTIS from TIS, and to pruning adaptations used, for instance, by Allen et al. [2]. The overall workflow is

1. Initial stage: sample metastable fluid.

   Quantify OP distribution in fluid.

2. Define λ_0 and λ_1 based on OP cluster size distribution in the fluid.

   Select generation 0 snapshots from fluid states between λ_0 and λ_1.

3. While the final interface has not been set:
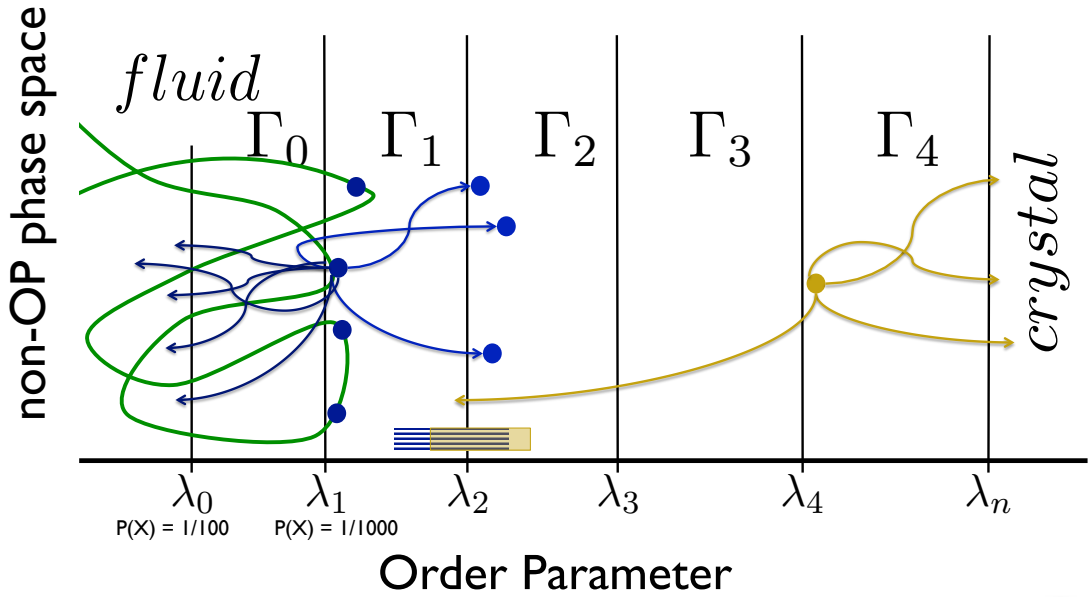
Figure 5.6: Schematic of Adaptive Segmented Transition Sampling (ASTS).

Run setup committers to find local OP noise and set next interface.

Run committers to generate samples at new interface.

4. Complete committers on remaining active frames.

5. Post processing: calculate overall nucleation and crystallization probabilities.

5.4.1 Markov processes

As with other FFS-based sampling schemes, an initial "brute force" simulation explores the portion of phase space corresponding to the metastable fluid, thoroughly sampling this region to generate a stable distribution of OP values. This is done as described previously in the methods (Section 2.2), and used for the fluid sampling discussed in chapters 3 and 4. For ASTS we use the OP distribution in the fluid to set λ_0 and λ_1, and to measure the probability of a system occupying the segment of phase space in between.

At relevant packing fractions we can take thousands of snapshots of the metastable fluid and count how often clusters of a specific size X are observed $P(X)_{sys}$. From

that we can calculate the distribution of cluster sizes and the probability for a particle to find itself in a cluster of a given size X, $P_i(X)$. Here $P(X)_{sys}$ is distinct from the order parameter in that it includes all clusters, not only the largest in the system for each observation:

$$P(X)_{sys} = \frac{Obs_X}{Obs_{tot}} = P(P_{ClX}) * \frac{N}{X}, \qquad (5.3)$$

$$P_i(X) = \frac{Obs_X * X}{N * Obs_{tot}}. \qquad (5.4)$$

Here Obs_X is the count of clusters of size X observed, Obs_{tot} is the count of observations of the system, and N is the number of particles in the system. For small X, $P(X)_{sys}$ can be greater than 1 since multiple small clusters ($X \approx 1 - 2$) can be observed at the same time. Since we are only interested in rare clusters, where the likelihood of multiples is negligible, $P(X)_{sys}$ corresponds to the observation probability.

The cluster size that is found in less than 5% of fluid observations will serve as a 0th interface λ_0, meaning that a system is defined as purely fluid if it has no cluster that is greater than the 95th percentile in cluster sizes. The 95th percentile is consistent with the general practice of choosing a cutoff for "pure" fluid, so that we can sample different systems hard particle systems in a consistent way. To begin our sampling we identify frames that have a largest cluster between the 99th and 99.9^{th} percentile found in the fluid, setting interface one (λ_1) to this value. Choosing λ_0 and λ_1 for a single system is based on the distribution of the OP in the parent phase, balancing between a need for broad sampling of the fluid and a wish to minimize wasted compute time on snapshots that are uninteresting. The OP distributions shape is sufficiently consistent across our systems we can use the 95th and 99th percentiles. In systems such as freezing water [27] where the OP distribution in the fluid, representing slopes of the parent phase energy well, are sufficiently different, different methods could be
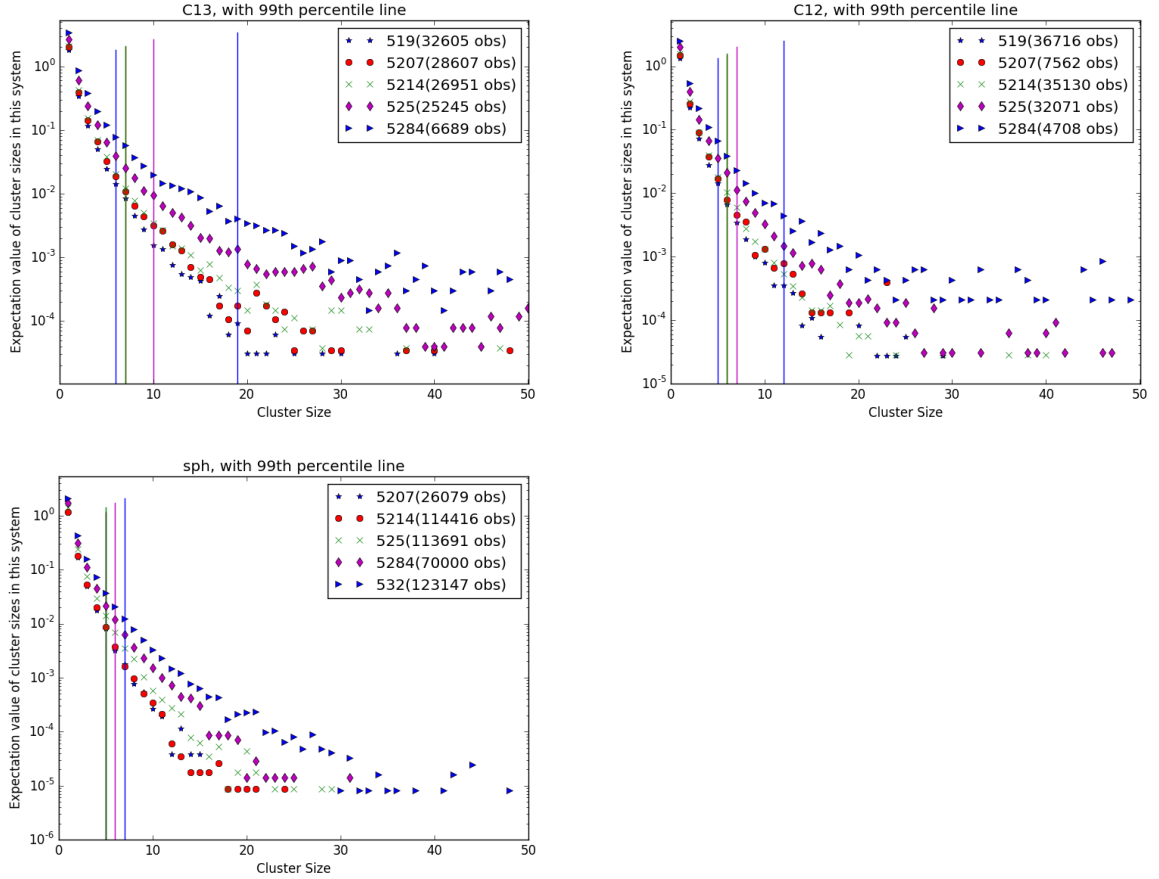
Figure 5.7: Expectation value of finding clusters of increasing size for each shape in the dense fluid at various packing fractions. Vertical lines represent the cluster size for 99% of observed clusters at each ϕ.

used to λ_0 and λ_1. The region of phase space between λ_0 and λ_1 is noted as Γ_0, shown in Figure 5.1, and the probability of a system visiting this region is measured in these brute force simulations of the metastable fluid as discussed thoroughly in Chapters 3 and 4.

5.4.2 Setting interfaces

The initial work showed the need to make sure that λ_1 was not within the range of OPs sampled often by the parent phase. To determine how closely subsequent interfaces could be set, we introduce a setup phase into the work flow, where at least ten independent snapshots (here independent means that they have descended from
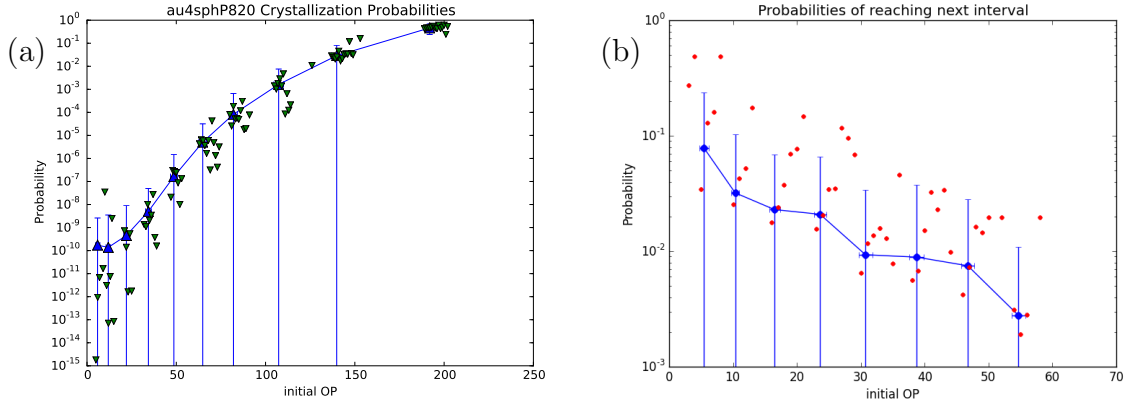
Figure 5.8: Growth probabilities at interfaces set using ASTS. (a) An example of expected behavior with the average growth probability remaining fairly flat until clusters become large, then increasing as the interface spacing increases approaching the critical nucleus size. (b) A state that has $P(B|A)$ below the accessible range; interfaces cannot be close enough to obtain sufficient samples without being within OP noise.

different fluid samples) were used as initial frames for trial trajectories similar to committers. At the start of each of these trial trajectories, ten samples were taken quickly to estimate the noise in the OP measurement at that state, before the system had evolved significantly. The trial trajectories were then allowed to continue until they either reached a time limit, reached phase B, or melted. Reaching phase B, *i.e.* crystallizing, only occurs for later interfaces, and is handled manually. The majority of cases melted. For each trial trajectory the largest OP found before melting was noted. For each set of trial trajectories starting from a snapshot the OP noise was averaged, and the largest clusters reached before melting used to calculate the cluster size that was reached by 10% of trials. For each starting snapshot the 10% cluster size OP value was compared to the starting OP + measured noise, and the greater of the two was taken as that snapshot's prediction for the next interface. The next interface was then set as the predicted next interface that would be reached by 30% of the independent starting frames 10% of the time (unless increased to avoid bias from the OP measurement noise).

This method generates interfaces that avoid issues of uncontrolled bias, and ideally will give us a sampling setup that will generate ensembles of intermediate states with flat growth probabilities. Figure 5.8(a) shows this working well with an NPT simulation of C12 using ASTS. Figure 5.8(b) shows the case where setting the interfaces outside noise results in cluster growth probabilities too small to maintain a sufficient population of independent samples to successfully generate transition trajectories. This could possibly be overcome by running more trials in each committer, or by using a more nuanced OP, if there was a need to expand the range of states accessible by ASTS sampling.

The size of a cluster fluctuates rapidly with small fluctuations in a system. We set meaningful interfaces to be outside the noise in the OP, but near enough that enough systems can reach them regularly enough to thoroughly sample the transition. Initial tests, like previous FFS-like sampling methods, relied on manually observing the system and setting interfaces based on trial and error. Here we have developed code to automatically measure the noise around the highest interface that was previously set, and adaptively determine a spacing for the next segment that is accessible to the system if possible, but is still outside small variations. Sampling within the noise of the OP leads to a bias that is difficult to quantify or correct for.

5.4.3 Committer weights

Collecting statistics on only some of the generated frames, and changing the fraction of frames that are tested, necessitates an individual weighting factor for each frame in its contribution to the crystallization probability of its parent, and to the system overall. In principle this individual weighting is straightforward. The weight for each trial run from a parent frame, W_k, is $1/n_{trys}$. Its weight in the overall ensemble is:

$$W_k = 1/n_{tries} * W_{parent}. \tag{5.5}$$

In practice this is somewhat demanding due to the need for associated data on each trial run, and we have implemented a document-based database using MongoDB in order to properly calculate and associate relative weights[2]. In addition to the changing weights, the precision of each progression probability depends on the number of attempts. This means that we are averaging sub-branches with different precisions, which also requires detailed information propagation, and takes advantage of the database.

5.4.4 Transition sampling

In the fluid we measure $P(X)_{sys}$ and $P(P_{ClX})$ over observed values of X as described above, and determine cluster sizes that are suitable early interfaces.

The second stage of the process explores transition rates between out-of-equilibrium states. Having defined $[\lambda]$ and obtained a set of frames F_1 between λ_1 and λ_2, we start K independent "trial" simulation runs from each state in F_1 and run forward until it either progresses to have an OP value beyond λ_2 or "melts" to an OP below λ_0, either of which terminates the trial run. "Successful" trials that progress past the next interface generate child frames, which are then put in queue for the same process, with the distinction that the lower termination condition is set by a B parameter specifying how many interfaces have to be crossed in the melting direction to be considered melting. This approach is similar to FFS and other techniques. A significant difference between this method and other similar methods is the use of the parameter B to cleanly and explicitly describe finite growth and shrinking of clusters. Our probability of growing is relative to shrinking, rather than relative to melting completely. This is expressed in $P(\Gamma_{i+1}|\Gamma_i) + P(\Gamma_{i-B}|\Gamma_i) = 1$, meaning that each cluster either must grow or shrink.

---

[2]There have been recent efforts to use a shared database systems to organize, maintain, and share data within the Glotzer research group, and eventually with outside collaborators. The design of my database is compatible with the new framework "Signac", and is being integrated into it to facilitate convenient use of new analysis tools, as well as future use of the data.

For convenience we abbreviate the probabilities as $P_i = P(\Gamma_{i+1}|\Gamma_i)$ and $Q_i = P(\Gamma_{i-B}|\Gamma_i)$.

The specification of the melt condition is not a trivial one. When coming from Γ_{i+1} or above, an interval Γ_i is "reached" if the OP of the system is observed to be below interface λ_i. When coming from below, Γ_i is reached if the OP exceeds λ_i.

There are two compelling reasons to terminate the trial trajectories at a known lower OP value rather than require them to return all the way to the fluid. One reason is purely pragmatic: it takes more time and compute power to continue a trajectory until it melts completely. Once the system is close to the critical point, at or near the top of the energy barrier, there may be little or no driving force for the system to evolve, making the computation time needed to melt each trial trajectory significant. This leads to extremely long trial runs, which we would like to avoid.

The other reason is more nuanced; it aims to understand the driving forces for nuclei to change size without necessarily proceeding all the way to the crystal or fluid. If, as is common, a system is considered to melt only if it returns all the way to the fluid, the probability of reaching the final state n from an earlier state a can be calculated straightforwardly by the product of all transition probabilities:

$$P(\Gamma_a) \prod_{i=a}^{n} P_i. \tag{5.6}$$

However this allows systems to move and loop within the entire region between the fluid and λ_{i+1}, obscuring the probability of growing directly from that configuration. Understanding how the system behaves in these out-of-equilibrium intermediate states requires additional information about short time fluctuations (*i.e.*, noise) in the OP, as well as probabilities for physically meaningful changes in the system as measured by the OP. This conceptualization of nucleus growth as nodes rather than fluxes requires careful handling of memory in the system.

We are measuring transitions between intervals, and using our measured probabilities of occupancy for the first interval. The final interface is chosen such that there is no measurable probability of returning from it. This lends itself to analysis like a Markov chain with one boundary having a fixed occupancy probability and at the other end an absorbing boundary. We can think of the intervals as nodes and converge this chain as a Markov chain to determine the probabilities of every interval to be occupied. We must be cautious here, since the probability of a cluster growing is not known to be independent of how large the cluster had previously been. In other words, we must be cautious of converging a non-Markovian system as if the nodes were independent of memory. We consider the probability of occupying an interval O_i = rate of reaching Γ_i from below plus the rate of reaching Γ_i, calculated as

$$O_i = O_{i-1} \times P_{i-1} + O_{i+b} \times Q_{i+b}. \tag{5.7}$$

Converging the set of these equations makes the assumption that P_i are independent of how the system reached λ_i. In other words it treats the state as Markovian, as having no relevant memory, for the calculation of $[O]$. This assumption can be validated by a comparison between subsets of branches, and by observing the local short- and medium- term fluctuations.

5.4.5   Efficiency analysis

The number of trial trajectories run in any interval is found by:

$$O_0 * \sum_{i=0}^{n} \prod_{j=0}^{i} C_j, \tag{5.8}$$

where C_j are the average numbers of children produced at each interval j, found as $P_j * k_j$, where k_j is the number of trials and P_j is the probability of progressing from interface j. If the values of C_j are greater than one for several intervals, especially if

there are many intervals, the result is a potentially unmanageable number of samples. In this case we want a large number of independent samples to serve as an ensemble of transition and pre-transition states, but we need to make sure we are not wasting an excessive amount of compute time or memory running on post-critical states, or running many correlated samples from branches of the same fluid root.

Our primary goal is to efficiently generate an inclusive ensemble of transition paths with known weights, including those paths with low weights. The number of trial runs initiated in an interval i is k_i, their average forward transition rate is $\bar{P}_i$. The number of frames $N_{F_{i+1}}$ we will generate in the interval Γ_{i+1} above interface λ_{i+1} is then

$$N_{F_{i+1}} = N_{F_i} * k_i * \bar{P}_i = N_{F_1} * \prod_{j=1}^{i}(k_j * \bar{P}_i) \tag{5.9}$$

For a system with 10 interfaces, where frames generated an average of 4 children and we started with 100 samples in the first interval, this would give us

$$\sum_{j=1}^{10}(100 * 4^j) \approx 140,000,000$$

total frames. In order to have a tractable number of frames to investigate, while still generating sufficient samples at all stages, $[\lambda]$ must be chosen with care to maintain $N_{F_{i+1}}$ at a reasonable number for all intervals.

In addition to choosing $[\lambda]$ with care, we generate sufficiently detailed output data to individually adapt k for every frame and maintain the ability to correctly weight each branch. Adapting k for each frame means setting a maximum and minimum k, but not running additional tries if P is above a cutoff value of interest. In addition, we can limit the number of 'sibling' frames that will be used for generating the next generation by selecting up to NS of those siblings at random and setting aside the rest (those frames set aside are still available for additional analysis, but are not added to the queue for the following interval).

These adaptations make the maximum number of frames in an interval then:

$$N_{F_{i+1}} = \sum_{a=1}^{N_{F_i}} \max(k_a \times P_a, NS), \qquad (5.10)$$

where a are all the samples in Γ_i, and k_a will be limited (additional tries will not be attempted) if the product $k_a \times P_a$ is above a value of interest, here set at 70%, because we are less concerned with precision of the growth rate of post-critical snapshots.

5.4.6 Resolving transition rates from local growth probabilities

The committers give us hundreds to thousands of individual growth probabilities, each for a specific state with a cluster, for that cluster to grow or shrink.

Iterating the process of generating a new interface and running committers from the previous interface until the interfaces have been defined through the transition region generates a set of local transition probabilities P_i. The probability for the system to occupy each state can then be found by converging Equation 5.7, or by using Equation 5.6. To perform the Markov chain convergence, the set of P_i is generated from the measured growth probabilities. The first interval, just outside the fluid region, is held at a fixed occupancy as measured with brute force initial simulations. The final interface is absorbing to represent the negligible probability of a crystalline system melting under these conditions. The occupancy of each interval i is found by converging the Markov chain.

These methods make different assumptions about the behavior of the system, either that a cluster that shrinks significantly will then have a negligible probability of growing again, or that a cluster's probability of growing or shrinking is independent of the system's history. The difference between crystallization probabilities calculated by the two methods, as well as differences caused by using different branch weighting schemes shown in Figure 5.9, has been less than the variation between trees of

correlated samples as shown in Figure 5.3. The variation between independent trajectories, based on physical differences in the cluster that are not captured by the OP, is of great interest for future work. From that variation we can examine different out of equilibrium transition states.

In order to be more nuanced about options for a given transition window, we can modify Equation 5.7 to resolve after a certain number of sweeps, and add a probability of remaining in the same region of phase space. We would then have:

$$O_i = O_{i-1} * P_{i-1} + O_{i+b} * Q_{i+b} + O_i * P(i|i), \tag{5.11}$$

where O_i is the probability of a system being in interval i, P_i is the measured probability of a system in interval i reaching interval $i+1$ before reaching $i-b$, and Q_i is the probability of returning to the less crystalline state below λ_{i-b}.

5.5 Error estimation

When sampling equilibrated systems we find well-defined distributions of measured values. Sources of error and variation in studying out-of-equilibrium data are more complex. We can still use bootstrapping-type methods on data sets, and take the standard error of the mean from subsets of data as sampling error. Because we are considering out-of-equilibrium states, and values such as the local growth probability that can have non-normal distributions, it is interesting to consider sources of varition.

Comparing different ways of calculating nucleation rates from committer data, Figure 5.9 shows how the probability of occupying a state i, O_i, relates to the probability of progressing as calculated by Equation 5.6, with and without branch weighting. The resulting overall crystallization probability is not drastically effected in well-behaved systems. Plots showing the local noise verification can be found in
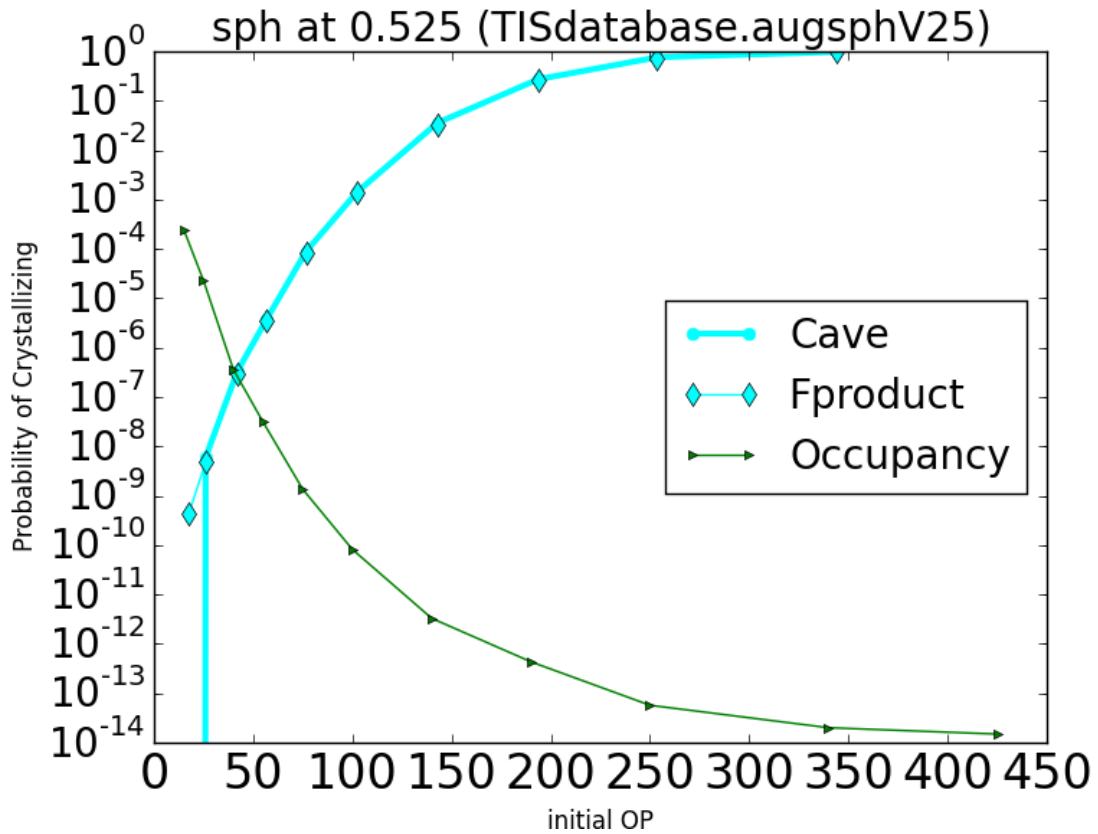
Figure 5.9: Several methods of calculating nucleation rates for a sphere NVT system at a packing fraction of 0.525. Two branch weighting methods are shown in blue to agree well, Markov chain convergence method is shown in green. These methods agree well for well behaved systems such as shown here.

Appendix B.1.

The P_i values used in both 5.6 and 5.11 can be averaged over all samples take in an interval, but can weight the resulting value in a non-physical way. Either because the sampling was designed to sample separate branches evenly even when their probabilities of occurring differ, or because branches that were successful early on can be more probable despite efforts to balance the sampling.
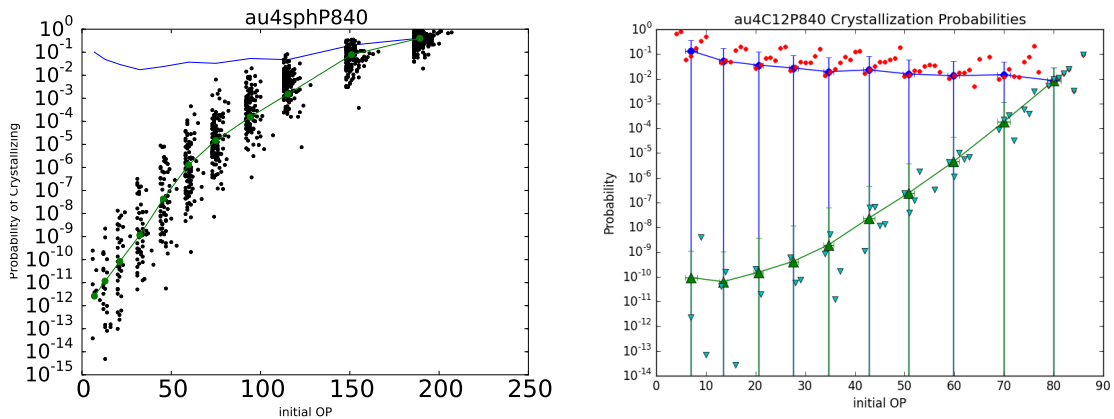
Figure 5.10: Plots of P_i, the probability of reaching the next interval, λ_{i+1}, without regressing below λ_{i-2}, and of crystallization. Shown for two shapes at two packing fractions each.

5.6 New findings

Using ASTS, we have begun sampling the more spherical polyhedra in the intermediate region of the metastable fluid, where we suspected based on clustering that crystallization probability could could differ between spheres and highly spherical polyhedra. Figure 5.10 shows well behaved intermediate data for spheres and DDT (labeled C12). Figure 5.11 shows crystallization probabilities, showing that initial results do show more probable crystallization for DDT than spheres at pressures where clustering differences were noted.
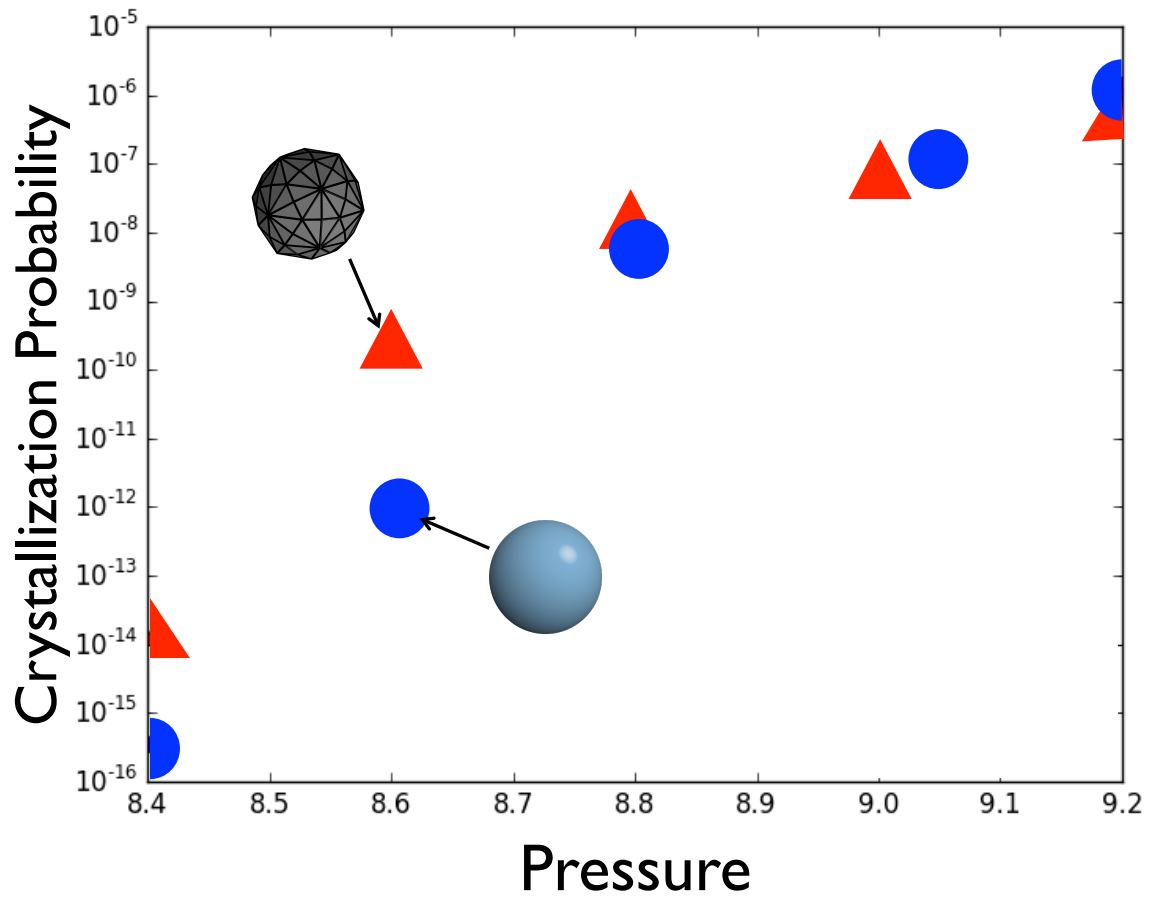
Figure 5.11: Crystallization probabilities for spheres and DDT from rare event regime of metastable fluid.

CHAPTER VI

Concluding remarks & outlook

In this work we have shown differences in metastable fluids due to subtle differences in particle shape. We have shown that small variations in particle shape can promote subtle changes in local motifs in a fluid, which can then lead to variations in how readily clusters and FCC crystals form. Specifically we have seen that promoting icosahedral arrangement of neighbors and having discrete preferred neighbor locations inconsistent with the FCC lattice is correlated with slower increases in crystallization rate with driving force. FCC motifs in the fluid promote clustering and lower the driving force needed to crystallize the fluid, and the ability to access FCC-like motifs while limiting the accessible range promotes crystallization at least in one case, and begs for further investigation. To better understand the process of transitioning from fluid to crystal we have developed a method for sampling the out-of-equilibrium states while carefully tracking their probability of forming.

From these findings and the data that lead to them, there are several studies that come to mind. In addition to continuing the study on dilute regime behavior of spherical polyhedra, two new studies in particular seem promising for giving us better understanding of the role of well defined properties of shapes in the process of crystallization *via* nucleation. Both the ASTS method and the ensemble of transition snapshots have potential for further use.

6.1 Isolating facet influence on crystallization

My findings on the influence of faceting on fluid clustering in the limited set of shapes studied could be extended and clarified by choosing additional shapes to test the mechanisms I have proposed. For example, a set of spheres with flattened faces oriented icosahedrally, oriented hexagonally, and oriented to promote neither of these could be compared, as well as comparisons of how large the facets must be to influence the system. A framework for selecting or designing shapes for disambiguation of the importance of specific shape properties is found in recent work by *van Anders et. al.*, where they specify anisotropy dimensions [58]. A set of spheres with flattened surfaces that did and did not connect, altering the freedom for neighbors to sample intermediate locations, could, for example, show us more about the apparent effect of connectedness of preferred neighbor locations.

Future studies such as these could help us understand both the relationship between preferred neighbor locations and crystallization, and between preferred neighbor locations and particle shape. This could help translate this work to real systems.

6.2 Machine learning to recognize nuclei

One of our motivations in developing ASTS was to generate a broad ensemble of transition states, capturing as much of transition space as we could access, *i.e.* as much variation in the transition path possible. In doing this we have generated thousands of system snapshots, each describing the positions and orientations of 4096 objects, and each with a rare cluster and a known probability for that cluster to grow or shrink.

One difficulty in the study of nucleation events is the need to recognize and watch pre-critical nuclei before they become critical and enter a growth phase. Distinguishing nuclei with a significant probability of growing from other fluctuations in a fluid is

not a solved problem. There have been many efforts in the hard sphere literature as well as other fields to better identify these important fluctuations by defining different order parameters, including multi variable order parameters. I propose that the set of transition states I have generated is sufficiently uninfluenced by my choice of OP that it would serve as a useful training set for a machine learning study. There are sufficient snapshots to act as both a training and testing set for an initial state. The learned recognition could then be tested on different shapes at an equivalent state, as well as on states of the same shape at different state points. Observing the reliability of a learned nuclei recognition algorithm across states and systems has great potential to give us additional information about the influence of particle shape on the crystallization of colloidal systems.

APPENDICES

APPENDIX A

Metastable fluid analysis

Measuring phase transitions from a metastable state via a critical event has different challenges depending on how common the critical events are. Rare event sampling techniques work in the regime where events are rare, but when multiple events can occur simultaneously rare event sampling methods can behave unexpectedly leading to corrupted data. Metastable fluids also pose a challenge in some cases in defining clearly what is equilibrated. Relaxation and autocorrelation times are easily measurable when transition events are rare, but as the metastable state becomes less stable there are difficulties resulting from transition events happening within relaxation times, as well as with fluctuation size increasing.

A.1 Equilibration of metastable states

As described in Section 2.2, we run NPT simulations where the box size is able to fluctuate while maintaining a constant pressure. The equilibrium packing fraction for that pressure is measured, along with fluctuation range. We also measure relaxation times for the box volume, but note that these are useful only as a tool for planning simulation runs because the physical meaning of the box size fluctuation time is

obfuscated by the method of choosing box resize moves, which are set to find the equilibrated state and range of states but can limit how quickly out of equilibrium states evolve, and how quickly systems equillibrate.

Initial simulations established an approximate packing fraction for each pressure, then sets of longer runs were started to thoroughly sample the metastable state. Showing the packing fraction as a function of time, as well as comparing packing fraction distributions between trajectories, helps us assess the stability of a fluid.

A.2   Cluster size distribution

Figures 4.3, 4.4, and 4.5 show probabilities of a particle being part of a cluster of several size ranges. These size ranges are chosen from a full distribution of cluster sizes found in metastable fluids.

Below, in Figure A.2, we show full distributions of cluster sizes for different systems. We first show each pressure separately for C12 for clarity, and then show composite plots of several pressures for each shape in Figure A.3.
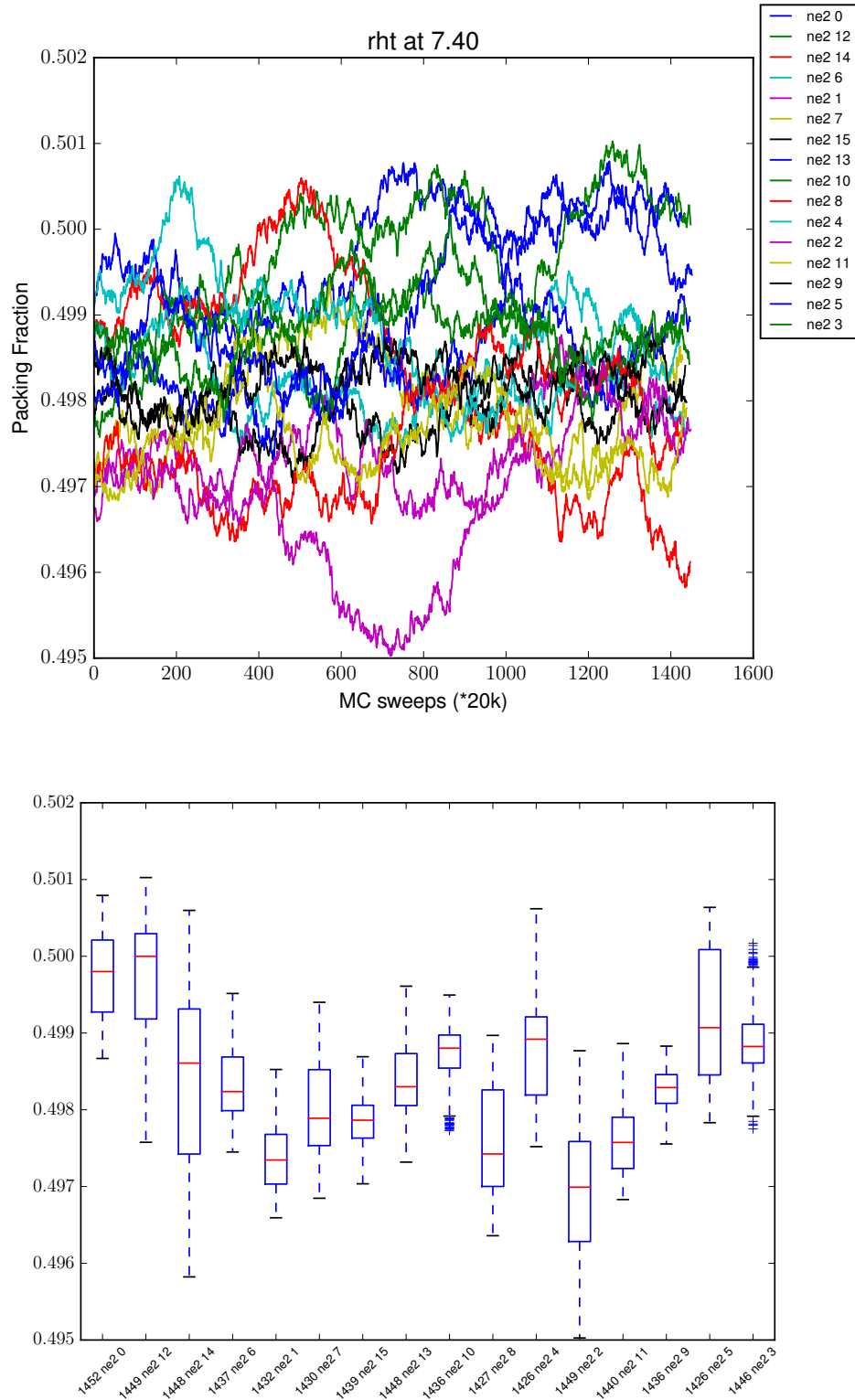
Figure A.1: Examples of metastable fluid, the same set of individual trajectories are plotted over time and with box plots. The temporal plots were needed to establish that equilibrium had been reached, the box plots to quantitatively show that ϕ was stable but varied over a notable range for each trajectory.
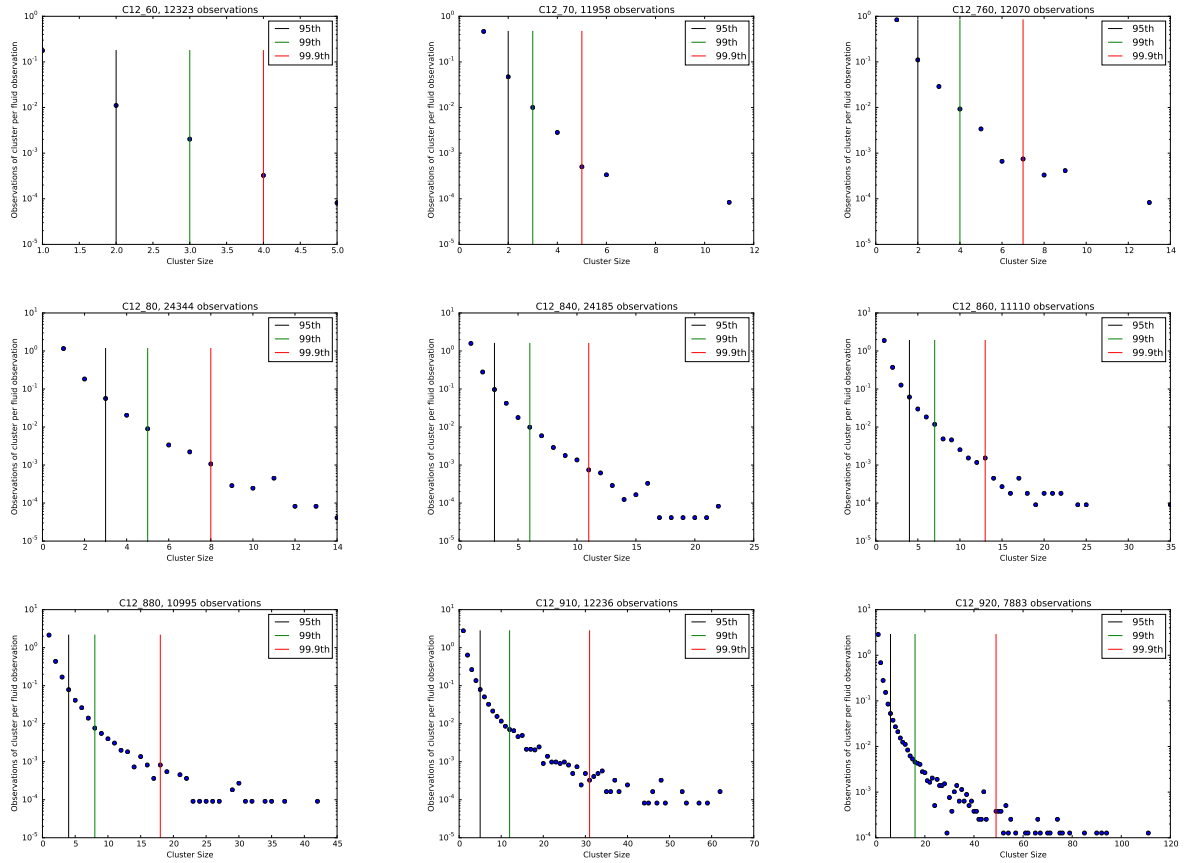
Figure A.2: Cluster size distribution of C12 fluids at pressures shown on plots. Vertical lines note cluster sizes exceeded by 5%, 1%, and 0.1% of clusters observed.

Figure A.3: Cluster size distributions for each shape at representative pressures, with vertical lines indicating the 99th percentile cluster size for a pressure. Each color corresponds to a pressure. Legends show how many snapshots of fluid system were included in the statistics shown.

APPENDIX B

Order parameter fluctuation analysis for interface placement

There are two main conditions for setting subsequent interfaces (segmenting phase space) in Adaptive Segmenting Transition Sampling: sufficient spacing to avoid noise in the OP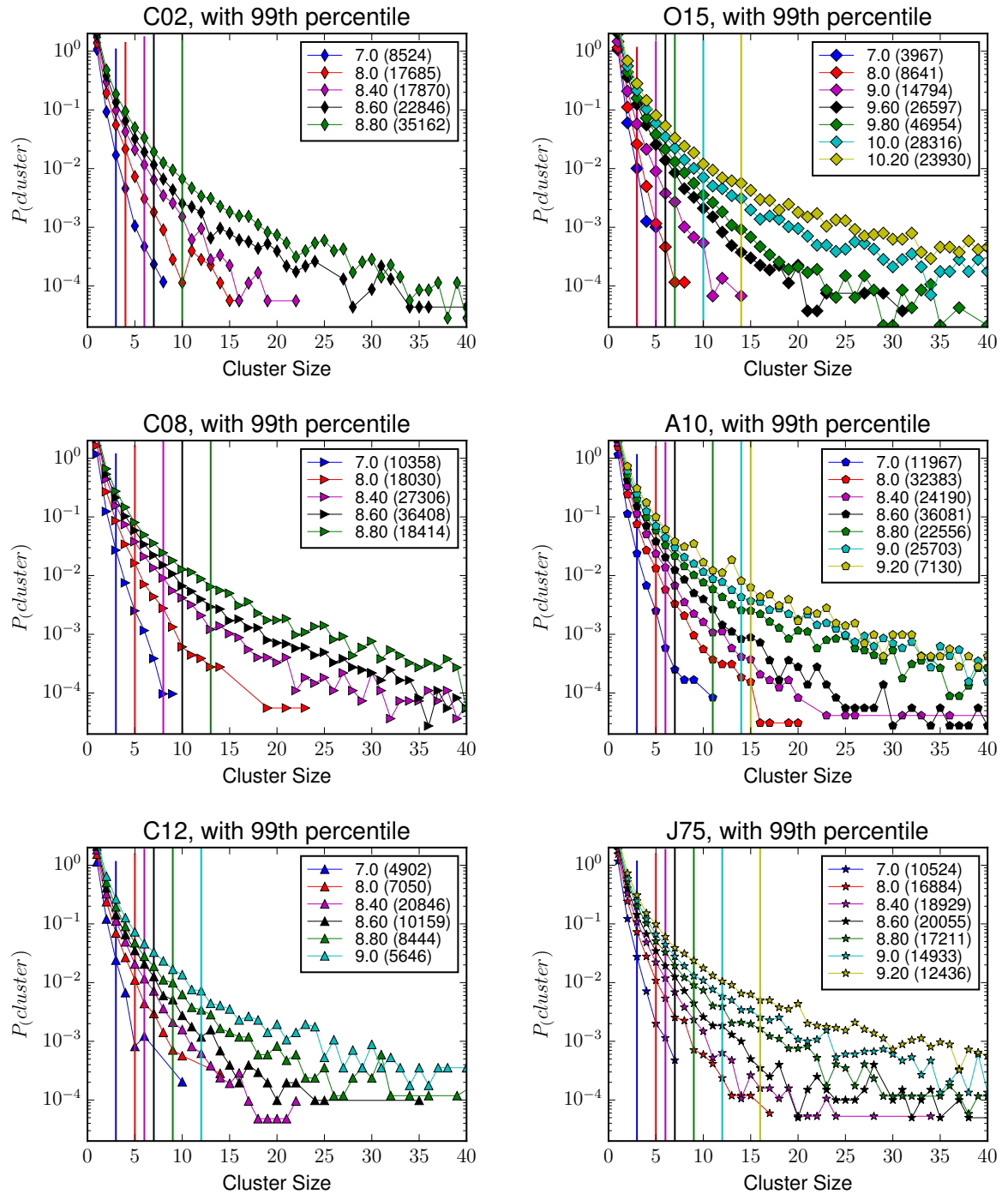, and transition probabilities on the order of 5-10%. Additional details of the ASTS are specified in order to balance sampling between branches descended from independent fluid samples, here called independent samples, by adaptively selecting what descendent frames to run committers on.

To set the interfaces we considered fluctuations in the metastable fluid at short times. This analysis was initially done manually, and here we show how the automatic segmenting algorithm was developed, and catalogue the intermediate data used.

We will then present the algorithm and underlying intermediate data used to develop the adaptive branch priority algorithm.

B.1 Order parameter fluctuations

We want to have the interfaces mark meaningful changes in the system that occurred without external bias, so we want to make sure that small fluctuations in the
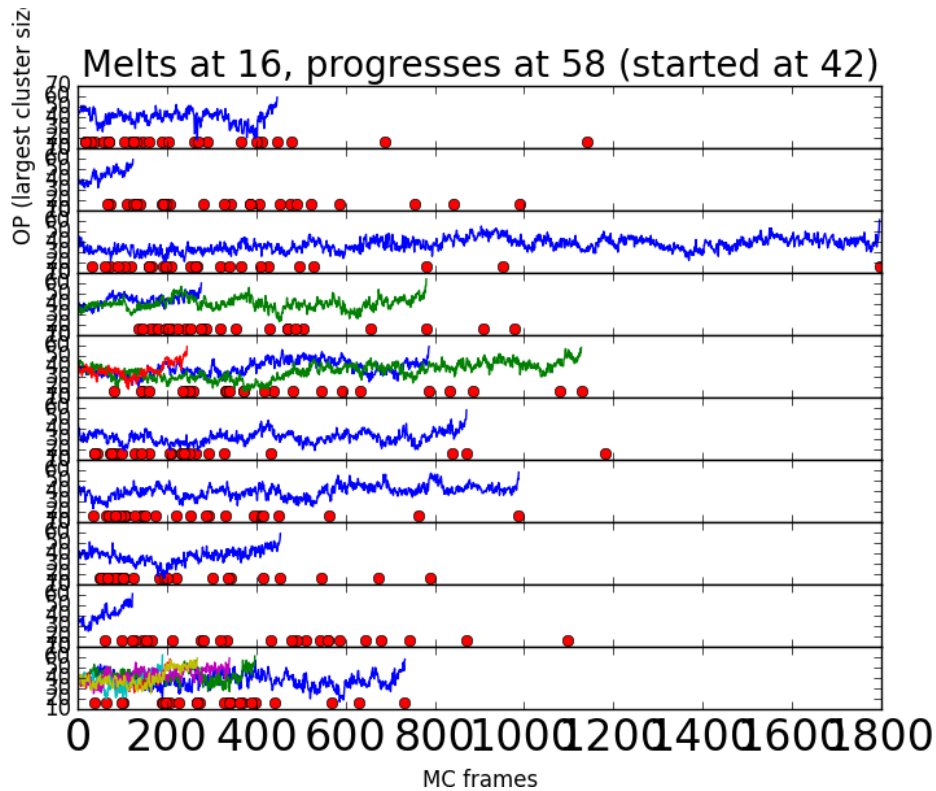
Figure B.1: Fluctuation in OP values for successful trial trajectories for C12 at volume fraction 0.525 from NVT run of the first ASTS version, developing the automation of interface placement. Each row is a set of trial trajectories from a snapshot with a largest cluster size of 42, where each trial trajectory was run until it either grew to 58 or melted to 16. Successful trial trajectories are shown by lines, melting trajectories are shown only by a red dot marking the number of frames they took to melt.

OP due to small shifts in surface particle positions altering exactly which particles at the surface of a cluster are counted. In developing an explicit algorithm for setting interfaces we verified that the interfaces were reasonable outside of noise.
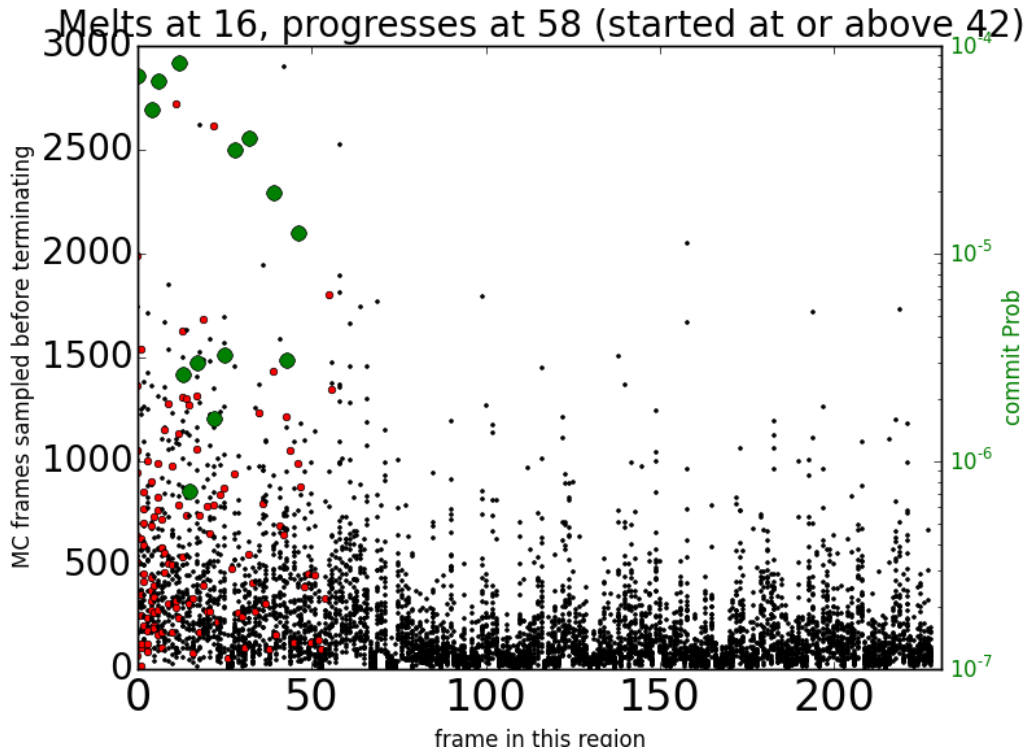
Figure B.2: Number of steps before melting (black) or progressing (red) for each trial run from a given interface. Each point on the x axis is a starting simulation snapshot, with a point plotted above for each trial run in its committer. If a descendent of that snapshot crystallized completely, a green point is plotted at the commit probability on the right y axis. Green points show measured non-zero probabilities of reaching a crystal from the starting interface shared by all committers represented in this plot.
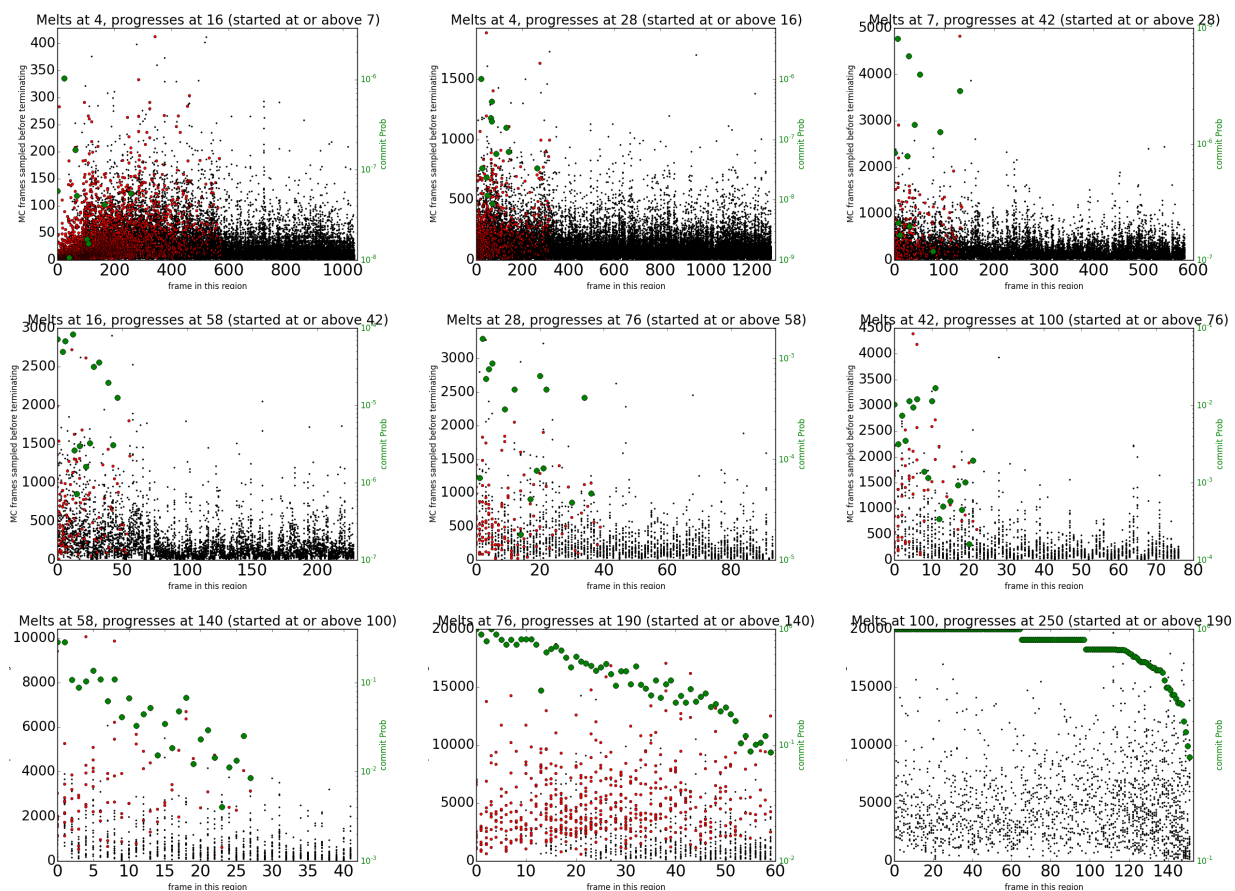
Figure B.3: Committer trial run length for ASTS NVT test set of shape C12 at packing fraction $\phi = 0.525$. Length of successful (red) and unsuccessful (black) trial trajectories of all committers for each interface. Each panel is an interface. The X-axis has individual committers sorted by growth probability, the Y-axis on the left is number of MC sweeps before the trial trajectory terminated. The right Y-axis is the crystallization probability (green) for all committers that had a decedent reach the crystal. It can be seen in the first panel that there were nearly 1000 initial frames taken from the fluid, of those only 9, see by their green dots, ever reached a crystalline state. That means that all the crystallization events and transition states observed from this set of data are derived from 9 fluid states, of which the most likely to crystallize had only a probability of 1 in a million.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] LeSar, R., *Introduction to Computational Materials Science Fundamentals to Applications*, Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK, 3 2013.

[2] Thornton, K., Nola, S., Edwin Garcia, R., Asta, M., and Olson, G. B., "Computational materials science and engineering education: A survey of trends and needs," *JOM*, Vol. 61, No. 10, 2009, pp. 12–17.

[3] Auer, S., *Quantitative Prediction of Crystal Nucleation Rates for Spherical Colloids: A Computational Study*, Ph.D. thesis, Universiteit van Amsterdam, 2002.

[4] Auer, S. and Frenkel, D., "Numerical prediction of absolute crystallization rates in hard-sphere colloids." *The Journal of Chemical Physics*, Vol. 120, No. 6, mar 2004, pp. 3015–29.

[5] Haji-Akbari, A., Engel, M., and Glotzer, S. C., "Phase diagram of hard tetrahedra," *Journal of Chemical Physics*, Vol. 135, 2011, pp. 1–12.

[6] Schultz, B. A., Damasceno, P. F., Engel, M., and Glotzer, S. C., "Symmetry Considerations for the Targeted Assembly of Entropically Stabilized Colloidal Crystals via Voronoi Particles," *ACS Nano*, Vol. 9, No. 3, 2015, pp. 2336–2344, PMID: 25692863.

[7] Frenkel, D., "Order through entropy," *Nature Materials*, Vol. 14, No. 1, Dec. 2015, pp. 9–12.

[8] Personick, M., Langille, M., Zhang, J., and Mirkin, C., "Shape Control of Gold Nanoparticles by Silver Underpotential Deposition," *Nano letters*, Vol. 11, 2011, pp. 3394–3398.

[9] Auer, S. and Frenkel, D., *Advanced Computer Simulation: Approaches for Soft Matter Sciences I*, chap. Numerical Simulation of Crystal Nucleation in Colloids, Springer, Berlin, Heidelberg, 2005, pp. 149–208.

[10] Zahn, D., "Thermodynamics and Kinetics of Prenucleation Clusters, Classical and Non-Classical Nucleation," *ChemPhysChem*, Vol. 16, No. 10, 2015, pp. 2069–2075.

[11] Auer, S. and Frenkel, D., "Prediction of absolute crystal-nucleation rate in hard-sphere colloids." *Nature*, Vol. 409, No. 6823, 2001, pp. 1020–3.

[12] Cacciuto, A., Auer, S., and Frenkel, D., "Onset of heterogeneous crystal nucleation in colloidal suspensions," *Nature*, Vol. 428, No. March, 2004, pp. 2003–2005.

[13] Compton, O. C. and Osterloh, F. E., "Evolution of size and shape in the colloidal crystallization of gold nanoparticles," *Journal of the American Chemical Society*, Vol. 129, No. 30, 2007, pp. 7793–7798.

[14] Keys, A. S. and Glotzer, S. C., "How do Quasicrystals Grow?" *Phys. Rev. Lett.*, Vol. 99, Dec 2007, pp. 235503.

[15] ten Wolde, P. R., *Numerical Study of Pathways for Homogeneous Nucleation*, Ph.D. thesis, Universtieit van Amsterdam, 1998.

[16] Kawasaki, T. and Tanaka, H., "Formation of a crystal nucleus from liquid," *Proceedings of the National Academy of Sciences*, Vol. 107, No. 32, 2010, pp. 14036–14041.

[17] Heitkam, S., Drenckhan, W., and Fröhlich, J., "Packing Spheres Tightly: Influence of Mechanical Stability on Close-Packed Sphere Structures," *Physical Review Letters*, Vol. 108, No. 14, April 2012, pp. 1–5.

[18] Filion, L., Hermes, M., Ni, R., Dijkstra, M., and Megen, V., "Crystal nucleation of hard spheres using molecular dynamics, umbrella sampling, and forward flux sampling: a comparison of simulation techniques." *The Journal of Chemical Physics*, Vol. 133, No. 24, Dec. 2010, pp. 244115.

[19] Karayiannis, N. C., Malshe, R., Kröger, M., de Pablo, J. J., and Laso, M., "Evolution of fivefold local symmetry during crystal nucleation and growth in dense hard-sphere packings," *Soft Matter*, Vol. 8, No. 3, 2012, pp. 844.

[20] Pusey, P. N. and W. van Megen, "Phase behvior of concentrated suspensions of nearly hard spheres," *Nature*, Vol. 320, No. March, 1986, pp. 340–342.

[21] Hales, T. C. and Ferguson, S. P., *The Kepler Conjecture: The Hales-Ferguson Proof*, Springer-Verlag New York, 2011.

[22] Bolhuis, P. G., Frenkel, D., Mau, S.-C., and Huse, D. a., "Entropy difference between crystal phases," *Nature*, Vol. 388, No. 1, jul 1997, pp. 235–236.

[23] Henzie, J., Grünwald, M., Widmer-Cooper, A., Geissler, P. L., and Yang, P., "Self-assembly of uniform polyhedral silver nanocrystals into densest packings and exotic superlattices," *Nature Materials*, Vol. 11, No. 2, Nov. 2011, pp. 131–137.

[24] Gao, Y. and Tang, Z., "Design and Application of Inorganic Nanoparticle Superstructures: Current Status and Future challenges." *Small (Weinheim an der Bergstrasse, Germany)*, Vol. 7, No. 15, May 2011, pp. 2133–2146.

[25] Glotzer, S. C. and Solomon, M. J., "Anisotropy of building blocks and their assembly into complex structures." *Nature Materials*, Vol. 6, No. 8, aug 2007, pp. 557–62.

[26] Damasceno, P. F., Engel, M., and Glotzer, S. C., "Crystalline assemblies and densest packings of a family of truncated tetrahedra and the role of directional entropic forces." *ACS nano*, Vol. 6, No. 1, Jan. 2012, pp. 609–14.

[27] Damasceno, P. F., Engel, M., and Glotzer, S. C., "Predictive self-assembly of polyhedra into complex structures." *Science*, Vol. 337, No. 6093, July 2012, pp. 453–457.

[28] Odriozola, G., "Revisiting the phase diagram of hard ellipsoids," *Journal of Chemical Physics*, Vol. 136, No. 13, Feb. 2012, pp. 8.

[29] Anderson, J. A., Irrgang, M. E., and Glotzer, S. C., "Scalable Metropolis Monte Carlo for simulation of hard shapes," *accepted for publication in Computer Physics Communications*, 2016.

[30] Anderson, J. A., Lorenz, C. D., and Travesset, A., "General purpose molecular dynamics simulations fully implemented on graphics processing units," *Journal of Computational Physics*, Vol. 227, 2008, pp. 5342–5359.

[31] Glaser, J., Nguyen, T. D., Anderson, J. A., Lui, P., Spiga, F., Millan, J. A., Morse, D. C., and Glotzer, S. C., "Strong scaling of general-purpose molecular dynamics simulations on GPUs," *Computer Physics Communications*, Vol. 192, 2015, pp. 97–107.

[32] Anderson, J. A. and Glotzer, S. C., "HOOMD-blue," http://codeblue.umich.edu/hoomd-blue, 2009–2016.

[33] Steinhardt, P. J., Nelson, D. R., and Ronchetti, M., "Bond-orientational order in liquids and glasses," *Physical Review B*, Vol. 28, Jul 1983, pp. 784–805.

[34] ten Wolde, P. R., Ruiz-Montero, M. J., and Frenkel, D., "Numerical Evidence for bcc Ordering at the Surface of a Critical fcc Nucleus," *Physical Review Letters*, Vol. 75, No. 14, 1995, pp. 2714–2717.

[35] Pedersen, U. R., "Direct calculation of the solid-liquid Gibbs free energy difference in a single equilibrium simulation." *The Journal of Chemical Physics*, Vol. 139, No. 10, sep 2013, pp. 104102.

[36] Thapar, V. and Escobedo, F. A., "Extensions of the interfacial pinning method and application to hard core systems," *The Journal of Chemical Physics*, Vol. 141, 2014, pp. 124117.

[37] Thapar, V. and Escobedo, F. A., "Localized Orientational Order Chaperones the Nucleation of Rotator Phases in Hard Polyhedral Particles," *Physical Review Letters*, Vol. 112, Jan 2014, pp. 048301.

[38] van Anders, G., Klotsa, D., Ahmed, N. K., Engel, M., and Glotzer, S. C., "Understanding shape entropy through local dense packing," *Proceedings of the National Academy of Sciences*, Vol. 111, No. 45, nov 2014, pp. E4812–E4821.

[39] Ni, R., Gantapara, A. P., de Graaf, J., van Roij, R., and Dijkstra, M., "Phase diagram of colloidal hard superballs: from cubes via spheres to octahedra," *Soft Matter*, Vol. 8, No. 34, 2012, pp. 8826.

[40] Torrie, G. M. and Valleau, J. P., "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *Journal of Computational Physics*, Vol. 23, No. 2, 1977, pp. 187 – 199.

[41] Auer, S. and Frenkel, D., "Suppression of crystal nucleation in polydisperse colloids due to increase of the surface free energy," *Nature*, Vol. 413, No. 6857, oct 2001, pp. 711–3.

[42] Grossfield, A., "WHAM: the weighted histogram analysis method," version 2.0.9, http://membrane.urmc.rochester.edu/content/wham, 2013.

[43] Hales, T. C., "A proof of the Kepler conjecture," *Annals of Mathematics*, Vol. 162, 2005, pp. 1065–1185.

[44] van Anders, G., Klotsa, D., Karas, A. S., Dodd, P. M., and Glotzer, S. C., "Digital Alchemy for Materials Design: Colloids and Beyond," *ACS Nano*, Vol. 9, No. 10, 2015, pp. 9542–9553.

[45] Tan, P., Xu, N., and Xu, L., "Visualizing kinetic pathways of homogeneous nucleation in colloidal crystallization," *Nature Physics*, Vol. 9, No. 11, Dec. 2013, pp. 1–7.

[46] Taffs, J., Williams, S. R., Tanaka, H., and Royall, C. P., "Structure and kinetics in the freezing of nearly hard spheres," *Soft Matter*, Vol. 9, No. 1, 2013, pp. 297.

[47] Moroni, D., ten Wolde, P., and Bolhuis, P. G., "Interplay between Structure and Size in a Critical Crystal Nucleus," *Physical Review Letters*, Vol. 94, No. 23, June 2005, pp. 235703.

[48] Hoy, R. S., Harwayne-Gidansky, J., and OHern, C. S., "Structure of finite sphere packings via exact enumeration: Implications for colloidal crystal nucleation," *Physical Review E*, Vol. 85, No. 5, may 2012, pp. 051403.

[49] Largo, J., Miller, M. a., and Sciortino, F., "The vanishing limit of the square-well fluid: the adhesive hard-sphere model as a reference system." *The Journal of Chemical Physics*, Vol. 128, No. 13, apr 2008, pp. 134513.

[50] Wang, Y., Wang, Y., Breed, D. R., Manoharan, V. N., Feng, L., Hollingsworth, A. D., Weck, M., and Pine, D. J., "Colloids with valence and specific directional bonding," *Nature*, Vol. 491, No. 7422, nov 2012, pp. 51–55.

[51] Chen, E. R., Klotsa, D., Engel, M., Damasceno, P. F., and Glotzer, S. C., "Complexity in Surfaces of Densest Packings for Families of Polyhedra," *Phys. Rev. X*, Vol. 4, Feb 2014, pp. 011024.

[52] van Anders, G., Ahmed, N. K., Smith, R., Engel, M., and Glotzer, S. C., "Entropically Patchy Particles : Engineering Valence Through Shape Entropy," *ACS Nano*, Vol. 8, No. 1, 2014, pp. 931–940.

[53] Frank, F. C., "Supercooling of liquids," *Proceedings of the Royal Society of London, Series A*, Vol. 215, No. 1120, 1952, pp. 43–46.

[54] Hirata, A., Kang, L. J., Fujita, T., Klumov, B., Matsue, K., Kotani, M., Yavari, A. R., and Chen, M. W., "Geometric frustration of icosahedron in metallic glasses." *Science*, Vol. 341, No. 6144, 2013, pp. 376–9.

[55] E, W. and Vanden-Eijnden, E., "Transition-path theory and path-finding algorithms for the study of rare events." *Annual review of physical chemistry*, Vol. 61, No. 1, mar 2010, pp. 391–420.

[56] Bolhuis, P. G., Chandler, D., Dellago, C., and Geissler, P. L., "Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark," *Annual Review of Physical Chemistry*, Vol. 53, No. 1, oct 2002, pp. 291–318.

[57] Allen, R. J., Frenkel, D., and ten Wolde, P. R., "Simulating rare events in equilibrium or nonequilibrium stochastic systems." *The Journal of chemical physics*, Vol. 124, No. 2, Jan. 2006, pp. 024102.

[58] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., "Equation of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics*, Vol. 21, No. 6, 1953, pp. 1087.

[59] van Erp, T. S., "Efficient path sampling on multiple reaction channels," *Computer Physics Communications*, Vol. 179, No. 1-3, jul 2008, pp. 34–40.

[60] van Erp, T. S. and Bolhuis, P. G., "Elaborating transition interface sampling methods," *Journal of computational Physics*, Vol. 205, No. 1, 2005, pp. 157–181.

[61] Allen, R. J., Frenkel, D., and ten Wolde, P. R., "Forward flux sampling-type schemes for simulating rare events : Efficiency analysis," *The Journal of chemical physics*, Vol. 194111, 2006.

[62] Karayiannis, N. C., Malshe, R., de Pablo, J. J., and Laso, M., "Fivefold symmetry as an inhibitor to hard-sphere crystallization," *Physical Review E*, Vol. 83, No. 6, June 2011, pp. 061505.

[63] Metzner, P., Schütte, C., and Vanden-Eijnden, E., "Transition Path Theory for Markov Jump Processes," *Multiscale Modeling & Simulation*, Vol. 7, No. 3, jan 2009, pp. 1192–1219.

[64] Haji-Akbari, A. and Debenedetti, P. G., "Direct calculation of ice homogeneous nucleation rate for a molecular model of water," *Proceedings of the National Academy of Sciences*, Vol. 112, No. 34, aug 2015, pp. 10582–10588.

[65] van Anders, G., Ahmed, N. K., Smith, R., Engel, M., and Glotzer, S. C., "Entropically Patchy Particles: Engineering Valence through Shape Entropy," *ACS Nano*, Vol. 8, No. 1, jan 2014, pp. 931–940.