2016

Advanced Monte Carlo Methods for the Study of Nucleation

Troy David Loeffler
Louisiana State University and Agricultural and Mechanical College, tloeff1@lsu.edu

ADVANCED MONTE CARLO METHODS FOR THE STUDY OF NUCLEATION

A Dissertation Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Chemistry

by
Troy David Loeffler
B.S. Washington State University 2011
December 2016

Table of Contents

Acknowledgements

I would like to begin by thanking my advisor Dr. Bin Chen at LSU for his support over the last five years as well as my colleague Aliasghar Sepheri for many conversations both casual and work related. For funding I thank the LA-SIGMA grant for supporting me through my first several years and I would like to thank Corning Glass for their support and collaboration on another project for my final semester here at LSU. I would also like to thank my dissertation committee Dr. Butler and Dr. Kumar for their help as well as the GPU team at LSU for teaching me a lot about programming in highly parallel environments. I would also like to thank Dr. Pat Meier and Dr. Rob Ronald at Washington State University for their help and guidance toward getting me into graduate school.

I would like to thank my family for their support in my pursuit of my doctoral degree. My father John, my mother Carol, my brother Robert, my sister Tamar and her husband Steve have been extremely supportive. Lastly I would like to thank my friends both at home and down here in Louisiana. Hanging out each week to talk and play games has been a wonderful way to prevent stress from building up due to graduate school.

And finally. Go Cougs and Geaux Tigers!

List of Figures

List of Tables

Abstract

The process of nucleation is an essential part of understanding and controlling phase changes in a wide array of systems. In the past theories such as Classical Nucleation Theory have been used as a tool to aid experimentalist in the study of phase mechanisms. However, recent studies have shown in detail that theories such as this are not reliable, given that it can mispredict nucleation rates by several orders of magnitude. As a result newer methodologies must be developed into order to improve upon these deficiencies. In this study we use atomistic simulations to examine the non-ideal deviations from classical theory observed in both simple and complex systems. In addition to this we present new algorithms that can be used to improve the rate at which the nucleation properties of these simulations can be sampled. Lastly we apply these new methods to study an atmospherically relevant system that involves the nucleation of water in the presence of multiple charged ionic species.

From these studies it was found that the deviation of more realistic systems from the classical theory can be attributed to both the creation of loosely bound clusters as well as the formation of highly ordered stacking in surface induced systems. The algorithms presented in this work have been shown to quickly and accurately replicate previously published data with very little increase to the computational overhead. Finally the application to the atmospherically relevant system showed an interesting trend where the nucleation rate was more heavily correlated to the number of water molecules that could be successfully bound to the ion pair instead of the quality of the bond.

Chapter 1

Introduction

1.1 An Introduction to Nucleation

The physical phenomenon of Nucleation is a key step in the transition of a material from one phase to another and has been a topic of interest for well over one hundred years.[2,3] The nucleation process is found anywhere from crystallization[4], to the formation of atmospheric aerosols[5], self-assembly[6], etc. Nucleation occurs when the super saturated mother phase begins its reorganization process to begin forming the seed that will eventually grow out and become the new thermodynamically favourable phase. For instance in a gas-liquid system the nucleation process occurs when gas molecules begin attaching to one-another, eventually this aggregation grows and grows until a fully formed liquid droplet is made.

Even though nucleation occurs when the mother phase is no longer the most stable of the two thermodynamic phases, the nucleation process is not spontaneous itself. This is due to the fact that while there is a potential difference between the two bulk phases, the intermediate aggregates are far from bulk-like and as such can have a higher free energy when compared to either bulk phase.[7] Thus for nucleation to fully occur the system must first climb the free energy barrier that is in its way. The properties of this barrier can ultimately determine properties such as the number of atmospheric aerosols formed[8] or even properties such as what kind of crystal structures are formed.[9] As a result of this it is absolutely critical that researchers understand the underlying mechanisms of the nucleation process in order

to be able to predict and control the outcome of phase transitions.

Thanks to modern simulation technology, researchers have had more power than ever to explore theoretical models that would have been otherwise too tedious to perform by hand. However despite these advancements in theoretical power, nucleation has traditionally been very difficult to simulate. The nucleation process very often requires relatively large time scales in order to completely observe. For instance, in gas to liquid nucleation the fastest observable rates at which liquid droplets can be formed[10,11] is on the order of 10^4 droplet nm^{-3} s^{-1}. Even for a Lennard-Jonesium which uses a very simple spherically symmetric potential, a standard molecular dynamic simulation must go through a significant amount of effort to observe the event.[11] The rate at which a system will nucleate is tightly linked to the supersaturation ratio and subsequently the ambient gas pressure. At a constant temperature a small change in the supersaturation ratio can lead to several orders of magnitude change in the rate of nucleation.[12–14] A natural consequence of this is that if the event of interest occurs at a temperature and pressure combination where the nucleation rate is incredibly slow, there is a high chance that it will be impossible to observe this event without the use of specialized methods or conditions. A few methods researchers have used to attempt to circumvent this is by using an exceedingly large number of particles to increase the probability of observing the event[11] or modifying the supersaturation conditions such that the nucleation rate is sufficiently fast.[15,16] Each of these techniques have their own pros and cons. For instance, performing a molecular dynamics simulation at high super saturation with a small number of molecules requires one to take into account that as the system forms liquid clusters the total monomer concentration of the box declines. This in turn can change the thermodynamic conditions of the system which shifts the nucleation rate half way through the simulation. Thus these finite size effects must be compensated for in order to ensure that the correct information is measured. The huge simulation approach does not suffer from the finite size effects quite as easily as the smaller simulation, but at the same time it pays for it in a huge

computational cost.

An alternative way to obtain information related to the nucleation process is by use of biased simulations. In these simulations, techniques such as Umbrella Sampling[17] are used to artificially flatten the free energy barriers found in the system. In this particular case, the key free energy barrier is the nucleation free energy barrier. While biased simulations do have their own share of problems, one of the primary advantages of using them is that important information related to the mechanisms of nucleation can be obtained without great computational cost, which makes them very attractive alternatives to classic "brute force" style simulations. For the purposes of this dissertation we will begin this chapter by outlining the core Monte Carlo methods used in this research, which will serve as the foundation for later chapters in this dissertation.

1.2 Metropolis Sampling

Unlike many stochastic systems that are commonly sampled with Monte Carlo, molecular systems only have a small number of accessible configurations relative to all the possible configurations that could potentially be generated. For instance, in a condensed water system around 300K the most probable states occur when all the water molecules are properly hydrogen bonded with their neighobors. However for any one water molecule the bonded orientation only makes up a small fraction of all possible rotational orientations. Thus the odds of randomly generating a properly oriented system is increasingly small. This problem ultimately increases with system size given that if the probability of orienting a single water molecule is 4% then the probability of orienting N molecules is proportional to 0.04^N. Thus simple "shotgun sampling" techniques are not suitable for efficiently sampling molecular phase spaces. A way to address this issue is to instead use what is known as Metropolis sampling,[18] which is a form of Markov Chain Monte Carlo. Markov Chains are a stochastic model to measure the probability in a system where the probability of an event occurring is dependent which event preceded it. For example in the atmosphere, the probability of

having clear skies on a given day depends on if it was clear or rainy the previous day. The core equation that results from solving a Markov Chain is given by the equation

$$A\pi = \pi \tag{1.1}$$

In this equation A is a transition matrix of the form

$$A = \begin{bmatrix} P_{1\to1} & P_{2\to1} & P_{3\to1} & \dots & P_{n\to1} \\ P_{1\to2} & P_{2\to2} & P_{3\to2} & \dots & P_{n\to2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{1\to n} & P_{2\to n} & P_{3\to n} & \dots & P_{n\to n} \end{bmatrix} \tag{1.2}$$

This naturally assumes that the probabilities contained in these equations are non-zero (i.e. there is a chance of observing these events) and therefore are non-trivial. Each entry of the matrix corresponds to the probability of changing from a given state i to a new state j. For instance $P_{1\to2}$ is the probability of leaving state 1 and entering state 2. In this formalism π is the eigenvector of the matrix A that corresponds to an eigenvalue of $\lambda = 1$.[18] The vector is of the form

$$\pi = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} \tag{1.3}$$

The π vector is known as the steady state probability or in other words this vector

contains the probability of being in a given state once the system has reach equilibrium. Each p_i is the equilibrium probability of being in state i. In many Markov Chain problems, the transition probabilities of the A matrix are known and it is desirable to solve for the steady state probabilities by solving for the eigenvector. However, in the case of molecular sampling the reverse problem is true. It is known ahead of time what the steady state probabilities are since they are simply the Boltzmann probabilities for a given configuration in the chosen ensemble.

$$p_i = \frac{\exp^{-E_i/k_BT}}{\sum_j \exp^{-E_j/k_BT}} \tag{1.4}$$

So instead one must solve for the transition probabilities of the A matrix. In practice there are multiple solutions to this problem given, but for all solutions it can be observed that Eq. 1.1 must be satisfied. If one expands Eq. 1.1 by performing matrix multiplication on the left side, for any given row i one receives the equation.

$$\sum_j p_j P_{j\to i} = p_i \tag{1.5}$$

Thus to satisfy the Markov Chain requirement, this equation must hold for every row. This equation is known as the Global Balance condition and it dictates that the population entering and exiting state i must be equal to the probability of that state i. Or in other words this is a global equilibrium condition saying that the total population of each state must remain static by balancing the incoming and outgoing transitions. This is a critical condition that any choice of transition probabilities must satisfy to guarantee that the correct steady state probabilities are obtained from this Markov Chain. There is a potentially

infinite number of transition probability choices that will satisfy the global balance equation, but in practice many of these solutions will yield unruly equations that will be difficult or outright impossible to use in application to molecular systems. Thus a common choice for the transition probabilities is the simple yet powerful solution known as Detailed Balance.

$$p_j P_{j\to i} = p_i P_{i\to j} \tag{1.6}$$

Detailed Balance is a micro-equilibrium condition where one chooses the the transition probability such that the flow back and forth between given states i and j is equal. It is easy to show via substitution that this choice of transition probabilities satisfies the global balance condition.

$$\sum_j p_j P_{j\to i} = \sum_j p_i P_{i\to j} = p_i \sum_j P_{i\to j} = p_i * 1 = p_i \tag{1.7}$$

This proof takes advantage of the fact that $\sum_j P_{i\to j} = 1$ which can be derived from the normalization condition for the transition probabilities. Thus if we satisfy Detailed Balance we also satisfy Global Balance. Our next step is to now find the exact functional form of the transition probabilities that can be used to solve the Detailed Balance equation. Much like with the global balance there are multiple solutions to this, but it has been found that the Metropolis solution[19] produces the optimal result.

$$P_{ij} = min\left(1, \frac{p_j}{p_i}\right) \tag{1.8}$$

It is almost trivial to show that this will satisfy the Detailed Balance equation

$$If \begin{cases} p_i < pj & \frac{P_{i\to j}}{P_{j\to i}} = \frac{1}{\frac{p_i}{p_j}} = \frac{p_j}{p_i} \\ p_i > pj & \frac{P_{i\to j}}{P_{j\to i}} = \frac{\frac{p_j}{p_i}}{1} = \frac{p_j}{p_i} \\ p_i = pj & \frac{P_{i\to j}}{P_{j\to i}} = \frac{1}{1} = 1 = \frac{p_j}{p_i} \end{cases} \tag{1.9}$$

Thus for all possible cases Detailed Balance is satisfied. This in turn gives an exceptionally convenient equation to use. The Metropolis method only requires that one know the functional form of the probability and for any given transition between states the only knowledge that is required is the statistical weight of the old state and the new state. In addition since one is taking the ratio of the two probabilities, the normalization constant is not required to compute the transition probabilities. These properties along with several others make this algorithm completely desirable for molecular sampling since often the normalization constant is not known a priori. To apply Metropolis sampling the following steps are used:

1. Start the system in a viable configuration.
2. From the current configuration randomly perterb the system using a trial move (i.e. Move a molecule in a random direction).
3. Calculate the statistical weights of the new and old positions.
4. Calculate the transition probability according to the Metropolis formalism.
5. Generate a random number between 0 and 1. If it is less than the transition probability accept the move and transition the system to the new state. Otherwise reject the move and return to the previous state.

This allows the user to generate any arbitrary statistical distribution where the function form of the weight is known by slowly perturbing the system using a set of trial moves.In the case of molecular simulations, these trial moves are very commonly moves such as translating

a molecule or rotating a molecule. Since each new state is generated by making a small change from the previous state the likelihood of generating energetically favourable configurations is significantly higher than "shotgun" methods.

One further caveat that is of importance to note is that the Metropolis scheme shown is only valid for Monte Carlo moves that are symmetrical or that the probability of proposing a move from state i to j is the same probability as proposing the move in reverse from j to i. In other words the way the moves are proposed is unbiased toward any given state. If one wishes to bias the system toward a given set of states without corrupting the correct probability, one must extend the Metropolis formalism to correct for this bias. When one performs the appropriate adjustments to the derivation of the Metropolis algorithm shown above, the Metropolis-Hasting formalism is obtained.[20]

$$P_{i\to j} = min\left(1, \frac{T_{i\to j}}{T_{j\to i}}\frac{p_j}{p_i}\right) \tag{1.10}$$

Where $T_{i\to j}$ is the probability of proposing a trial move from state i to state j and $T_{j\to i}$ is the probability of the reverse proposal. This equation is the backbone of many Molecular Monte Carlo algorithms since it allows a researcher to use any trial move they desire to sample the system. From this, one can generate a massive stable of trial moves that can be applied to solve a problem. However it should be noted that while in theory any valid Monte Carlo move will eventually give the correct probability distribution, they will not converge at the same rate. If the trial moves consistently generate configurations that have a low probability, the overwhelming majority of the proposed moves will be rejected and as a result it will take an exceedingly long time to move through the configurational phase space and thus require a longer simulation to sample the system correctly. The rate at which a Monte Carlo move creates valid configurations is known as the acceptance rate which is

simply given by

$$Rate = \frac{n_{Successes}}{n_{Attempts}} \tag{1.11}$$

The acceptance rate is a measure of how good a given trial move is at generating a valid configuration and very often the acceptance rate is correlated to the rate at which the system converges to the correct probability. Thus one should be very careful in choosing their Monte Carlo moves in order to obtain a feasible answer in a minimal amount of time. In addition one must always know the exact form of the trial probability in order for a given Monte Carlo move to work. Neglecting terms can lead to incorrect probability distribution and by proxy incorrect physical properties for a given system.

1.3 Aggregation-Volume-Bias Monte Carlo

The Aggregation-Volume-Bias Monte Carlo (AVBMC) algorithm[21,22] is a specialized Monte Carlo move that directs the formation and destruction of bonded configurations in molecular systems. The AVBMC move is in the general family of swap moves facilitate the transfer of particles in ensembles such as the Gibbs Ensemble, Grand Canonical Ensemble, etc.

AVBMC is divided into two corresponding moves, an insertion move and a removal move. Alternatively these moves can be refereed to as the swap in and swap out moves respectively. As their names suggest, the moves exchange molecules in and out of the system being studied. The insertion is move is performed by the following

1. If the molecular reservoir is being represented explicitly, choose a molecule from the reservoir to be swapped into the system.
2. Choose a molecule inside of the system to serve as the insertion target.
3. Choose a random position within a maximum distance r_{bond} to insert the molecule.

4. Regrow the molecule.

If all of these steps are performed by uniform random selection, the probability of insertion can be written as

$$P_{insertion} = \frac{1}{N_{reservoir}} \frac{1}{N_{sys}} \frac{1}{V_{bond}} \tag{1.12}$$

Where N_{sys} is the number of particles in the system, V_{bond} the volume given by $V_{bond} = \frac{4}{3}\pi r_{bond}^3$, and $N_{reservoir}$ is the number of molecules in the reservoir. For the removal move a similar procedure is used.

1. Similar to the insertion move, randomly choose a molecule inside of the studied system to serve as the pivot/target molecule.
2. Choose one of the neighbouring molecules of the target molecule. This will be the molecule that is swapped out.
3. Create a new position and configuration in the molecular reservoir and regrow the molecule.

When written out using uniform probabilities, one obtains

$$P_{removal} = \frac{1}{N_{sys}} \frac{1}{N_{nei}} \frac{1}{V_{reservoir}} \tag{1.13}$$

Where N_{nei} is the number of neighbours surrounding the target molecule and $V_{reservoir}$ the volume of the reservoir box. For the purposes of this work the reservoir is represented implicitly and as a result the reservoir terms can be replaced by the chemical potential term found in the Grand Canonical Partition function. When the ideal gas assumption is invoked, the chemical potential reduces to the gas phase density given by $\rho_{gas} = N_{reservoir}/V_{reservoir}$. The gas phase density is thus given as an input parameter which controls the thermodynamic

conditions. When all of the previously mentioned terms are collected and placed into Eq. 1.10 then the acceptance rule for the AVBMC move is given by

$$P_{Acc-In} = min\left(1, \frac{N_{sys} \cdot V_{bond} \cdot \rho_{gas}}{(N_{sys}+1) \cdot (N_{nei}+1)} \cdot e^{-\beta \Delta E}\right) \tag{1.14}$$

$$P_{Acc-Out} = min\left(1, \frac{N_{sys} \cdot N_{nei}}{(N_{sys}-1) \cdot V_{bond} \cdot \rho_{gas}} \cdot e^{-\beta \Delta E}\right) \tag{1.15}$$

It should be noted that the N_{sys} and N_{nei} terms that appear in both the insertion and reverse move must be shifted by 1 in the reverse probability to reflect that after a particle has been inserted or deleted the total number of molecules in the system has increased or decreased by 1.

The primary advantage of using the AVBMC algorithm is that because it targets a small region around a given molecule, the likelihood of generating bonded configurations is greatly enhanced compared to standard swap moves. This both improves the rate of creating bonds as well as the rate of destroying them. Further discussion about the strengths and weaknesses of the AVBMC algorithm can be found in chapter 3

The primary advantage of using AVBMC for the study of nucleation is that the simulation can be designed such that there is no need for a finite volume simulation box. Instead one can construct a boundary condition by using a cluster criteria such as the Stillinger criteria which defines two particles as being neighbors if their center of mass is within a arbitrary distance.[23]. In Monte Carlo simulations enforcing this criteria is exceedingly easy, given that all one needs to do is to ensure that all the members of a given cluster are properly connected to each other. Any move which would result in the destruction of a cluster is simply rejected. Once this criteria is enforced, the AVBMC move can be used to transfer particles into and

out of the cluster giving a simple yet effective method of calculating the thermodynamics of nucleation. The use of the AVBMC method has been shown in previous work to successfully predict nucleation properties such as the onset activity,[24], structural features,[25], and a host of other nucleation properties.[14,26]

While these methods have proven to be very successful, more recent expansions of the method have uncovered problems with the algorithms. For instance in previous work a complicated acid model was used to study the nucleation of water in the presence of an acid defect.[27,28] This model required a significantly greater number of computational cycles to properly calculate all the data of interest. For a molecular cluster with an ionic species in it, the acceptance rate drops sigificantly.

In the work presented in this dissertation, the AVBMC method as well as other commonly used classical theories will be examined in order to understand their weaknesses and strengths in an attempt to figure how to improve the efficiency and accuracy of the theoretical tools that can be used to study the nucleation process.

Chapter 2

Classical Nucleation Theory

2.1 Introduction

As an activated event, the nucleation process can be characterized by a nucleation free energy (NFE) surface expressed as a function of order parameters, e.g., cluster size in the vapor–liquid nucleation, that links the relative thermodynamic stabilities of the two bulk phases via a critical barrier. Evaluating this NFE surface and the related nucleation barrier height(s) that allow for the determination of the nucleation rate has been a major task for a plethora of theoretical methods from the century-old classical nucleation theory (CNT)[29,2,3,30–34] to modern statistical-mechanics based simulation approaches.

In CNT, the thermodynamics of cluster formation is approximated by a bulk-droplet model, in which the formation free energy of a cluster is expressed in terms of properties of a bulk phase (e.g., using the equilibrium density ρ and chemical potential μ of a bulk phase and the interfacial tension γ of an infinite flat surface). More specifically, the formation free energy of a cluster with a radius r in a homogeneous bulk phase is broken into two contributions, one from the surface free energy (simply surface area, A, times the interfacial tension, γ), and the other from the chemical potential difference ($\Delta\mu$) between the two phases involved in a particular phase transition in an m-dimensional space as follows,

$$\Delta G(r) = A\gamma + n\Delta\mu = mC_m r^{m-1}\gamma + C_m r^m \rho\Delta\mu \qquad \text{with} \qquad C_m = \frac{\pi^{\frac{m}{2}}}{\Gamma\left(\frac{n}{2}+1\right)} \tag{2.1}$$

where Γ is the gamma function. When derived for a three dimensional system the equation takes on the form

$$\Delta G(r) = \frac{4}{3}\pi r^3 \rho_l \Delta\mu + 4\pi r^2 \sigma \tag{2.2}$$

This equation while simplistic gives an elementary insight to the underlying physics of cluster formation. The $\frac{4}{3}\pi r^3 \rho_l \Delta\mu$ term corresponds to the gain in energy that is achieved when a particle is taken from the mother phase and inserted into the newly formed cluster. In the case of gas-liquid phase change the $\Delta\mu$ term can be calculated from the equation

$$\Delta\mu = -k_B T \ln \frac{\rho_{gas}}{\rho_{eq}} \tag{2.3}$$

Where ρ_{gas} is the current gas phase density and ρ_{eq} is the gas phase density at equilibrium conditions. Alternatively the ratio of these two quantities can be grouped together into a single term known as the saturation ratio.

$$S = \frac{\rho_{gas}}{\rho_{eq}} \tag{2.4}$$

For saturation ratios greater than 1 the chemical potential is negative which implies that the new phase is thermodynamically favourable while rations less than 1 indicate that the mother phase is still thermodynamically favourable. Thus the first term in the CNT equation primarily corresponds to the relative thermodynamical stability of the two bulk phases. In contrast the $4\pi r^2 \sigma$ term corresponds to the interfacial contribution to the free energy. Since the growth of a cluster requires the formation of an interface between the cluster and the

mother phase, there is a natural free energy penalty to be paid for this.

Thus the overall picture that CNT gives is that the formation of a cluster in the mother phase is a direct competition between the surface and bulk energetics. While the new phase may be thermodynamically favorable, the system must first gain enough momentum to properly establish a stable interface. In a cluster where the surface term is larger than the bulk term the cluster will likely evaporate before additional molecules can be added.

When comparing the theory to simulation work, the radius of the cluster is typically an inconvenient choice of nucleation coordinate since the cluster radius may not be as well defined in simulation as it is in experiment. Thus it is useful to transform the CNT equation in terms of the number of molecules in the cluster instead of radius. Using the spherical assumption of CNT, it is possible to write the liquid density as $\rho_l = \frac{n}{\frac{4}{3}\pi r^3}$ where n is the number of particles. By performing a simple substitution of r for n one obtains the CNT equation written as a function of n.

$$\Delta G(n) = n\Delta\mu + \sigma \left(\frac{36\pi}{\rho_1^2}\right)^{\frac{1}{3}} n^{\frac{2}{3}} \tag{2.5}$$

This formalism provides a direct way to compare simulational work to the theory. In addition this formalism allows one to determine the free energy of addition as a function of the cluster size.

$$\delta\Delta G(n) = \Delta G(n) - \Delta G(n-1) = \Delta\mu + \sigma \left(\frac{36\pi}{\rho_1^2}\right)^{\frac{1}{3}} (n^{\frac{2}{3}} - (n-1)^{\frac{2}{3}}) \tag{2.6}$$

This is known as the delta-delta-G formula which calculates the free energy of adding a monomer to the cluster. According to CNT, if one were to calculate $\delta\Delta G(n)$ as a function of $n^{\frac{2}{3}} - (n-1)^{\frac{2}{3}}$ for a three dimensional system the resulting plot would fall on a completely

straight line. This of course would no longer by true if the system exhibited non-ideal behavior. This gives rise to a useful tool in analysing deviations from CNT which can point researchers toward interesting physical phenomena.

2.2 An Examination of Classical Nucleation Theory in Two Dimensional Nucleation

Unfortunately since CNT is a based on bulk-like assumptions, it is naturally expected that these approximations are clearly problematic for small clusters. Correspondingly deficiencies of CNT are expected. For example, using one of the simplest model systems Lennard-Jonesium, it has been demonstrated from the simulation that CNT provides fairly accurate estimate of the work required to add a monomer to the cluster but deviations between the simulation and CNT are noted toward the smallest clusters leading to the conclusion that the failure of CNT can be traced to its incorrect description of the smallest clusters.[7,35–38] While such analysis (assisted with the simulation results) has been extensively done for three-dimensional (3D) vapor–liquid nucleation and has been shown to be quite useful to extract the source of the errors for the theory, nucleation in other dimensional space hasn't been explored in quite the same detail. However, the change of dimensionality can lead to very different behavior. For example, using density functional theory (DFT) Zeng[39,40] found that CNT can actually predict fairly reasonable free energy barriers (only a small underestimation by the theory) for the 2D nucleation system compared to the large positive errors shown for the 3D system. This better performance can be explained by the much smaller curvature effects for clusters in 2D vs. 3D. On the other hand, using molecular simulation, Santra et al.[1] reported another surprisingly different behavior, that is, CNT underestimates considerably the free energy barrier for 2D in contrast to the overestimation of this property for 3D; both errors are substantial but opposite in sign. While there are clear differences between the conclusions obtained from these two sets of studies, both do agree that the theoretical errors are reversed from 3D to 2D. Questions arise, could these entirely

opposite errors found for the 2D nucleation be accounted by the same origin (i.e., the incorrect description of the smallest clusters by the theory) that was previously concluded for the 3D nucleation? Also what could be the possible causes of this sign switch? Motivated by these questions, we brought the same simulation protocol used for the 3D nucleation systems that has been shown successful for extracting the source of the theoretical errors to 2D here. An important finding from this study is that for both 2D and 3D systems, the theoretical deviations from the simulation can be attributed to the smallest clusters but these deviations show an interesting sign switch from being positive for 3D to being negative for 2D.

2.2.1 Simulation Details

This nucleation simulation was carried out using the aggregation-volume-bias Monte Carlo (AVBMC) method[21,22]coupled with umbrella sampling (US).[17] While the use of AVBMC enables direct particle transfer between the cluster and the mother phase to greatly speed up the sampling of the otherwise slow condensation/evaporation events that are important to nucleation, US is implemented to solve the critical problem of sampling high free-energy clusters by adding artificial biasing potentials to enhance tremendously the probabilities of these clusters being visited by the simulation. The details of this simulation approach for 3D systems can be found in Ref. 14 and the extension of this approach to 2D systems is straightforward. The only difference is that the bonded region used for AVBMC swap moves is defined by an area term (centered on a randomly chosen particle from the cluster phase) in 2D versus a volume term in 3D. Similar to 3D simulations, this area term can be chosen to conform to the Stillinger[23] cluster criterion used here (which is distance based, i.e., any two particles within a certain distance r_c are considered to belong to the same cluster) to make attempts at cluster condensation/evaporation more likely. The grand-canonical ensemble was chosen for the nucleation simulation for computational efficiency and also to be consistent with previous simulation studies performed for 3D systems.[35] In this ensemble, a single cluster is physically isolated from the rest of the system but thermodynamically still

connected with the rest of the gas-phase system. This remaining gas-phase system is represented by a particle reservoir with the chemical potential of this reservoir being specified as one important input parameter for the simulation.

This simulation study was performed for Lennard-Jonesium (LJ). LJ is one of the most basic and popular theoretical models for those species that interact through van der Waals forces. Also as one of the simplest and mostly studied model systems, many of the thermodynamic properties are available or can be obtained with superior accuracy for this model, which makes it convenient to compare the simulation results to CNT. For example, for the 2D LJ system investigated here, extremely long simulations (on the order of 10^8 Monte Carlo cycles here) can be afforded even for very large bulk-phase systems with sufficiently large cut-off values (up to 15 σ here) to accurately calculate the input parameters required by CNT. These include a vapor–liquid phase equilibrium calculation performed in the Gibbs ensemble[41–43], to obtain the needed chemical potential and liquid-phase density data, and a canonical ensemble simulation of a liquid slab for surface tension. Specifically the surface tension is calculated as follows:[44]

$$\gamma = \frac{1}{A} < V_{yy} - V_{xx} >= \frac{1}{2L_x} \left\langle \sum_{i<j} \left(r_{ij} - \frac{2y_{ij}^2}{r_{ij}} \right) U'(r_{ij}) \right\rangle \tag{2.7}$$

where $A = 2L_x$ is the total surface area for 2D systems, V_{xx} and V_{yy} are the virial tensors, r_{ij} and y_{ij} are the distance and its $y-$axis component between particle i and j, respectively, $U'(r_{ij})$ is the first derivative of the potential energy between i and j over their distance (or equivalently the force between them), and the angle brackets denote an ensemble average. The tail corrections for the surface tension were approximated by

$$\gamma_{tail} = \frac{1}{2} \int_{-\frac{L}{2}}^{\frac{L}{2}} \int_{-\frac{L}{2}}^{\frac{L}{2}} \int_{x_{\min}}^{\infty} \rho(y_1)\rho(y_2) \left(r - \frac{2y_{12}^2}{r} \right) U'(r) \mathbf{d}x \mathbf{d}y_1 \mathbf{d}y_2 \tag{2.8}$$

Table 2.1: Vapor–liquid coexistence properties and surface tension for 2D LJ. The subscripts give the statistical accuracy of the last decimal(s).

T^*	ρ^*_{liq}	ρ^*_{vap}	μ^*_{liq}	μ^*_{vap}	γ
0.427	0.7516_1	0.02321_5	-1.7886_{10}	-1.7890_3	0.207_3

where L and ρ are the box dimension and the density profile along the direction normal to the interface (y axis here), respectively, and x_{min} is defined in terms of r_{c}, the potential cutoff distance, as follows:

$$x_{\mathrm{min}} = \left\{ \begin{array}{cc} \sqrt{r_{\mathrm{c}}^2 - y_{12}^2} & |y_{12}| \leq r_{\mathrm{c}} \\ 0 & |y_{12}| > r_{\mathrm{c}} \end{array} \right\} \tag{2.9}$$

Eq. 2.8 was integrated numerically using the density profile obtained at $r_{\mathrm{c}} = 7.5\sigma$. It should be noted that the surface tension value calculated using this approach has been found highly sensitive to r_{c} for the 3D system.[45] For the 2D LJ system, this property was found to be 0.2060±0.00380, 0.2065±0.00549, 0.2074±0.00338, and 0.2073±0.00424, when r_{c} was set at 4, 5, 7.5, and 15 σ, respectively.

2.2.2 Results and Discussion

Listed in Table 2.1 are the bulk-phase properties obtained for 2D LJ. Using these bulk-phase properties, the free energy of cluster formation predicted by CNT can be determined from Eq. 2.1 and compared to those evaluated directly from the nucleation simulation. These results, which were obtained at a reduced temperature $T^* = 0.427$ (the same temperature used in the previous simulation study by Santra et al.[1]) and a supersaturation ratio $S = 1.036$ (or $\Delta\mu = -0.0352 k_{\mathrm{B}}T$), were plotted in Fig. 2.1. Also shown in Fig 2.1 are those free energy data that were obtained previously[35] for a 3D LJ system at $T^* = 0.45$ and $S = 84.6$ (or $\Delta\mu = -4.438 k_{\mathrm{B}}T$). This supersaturation condition is chosen so that both 2D and 3D NFE barrier heights yielded from the simulation match with each other. Thus, if the same

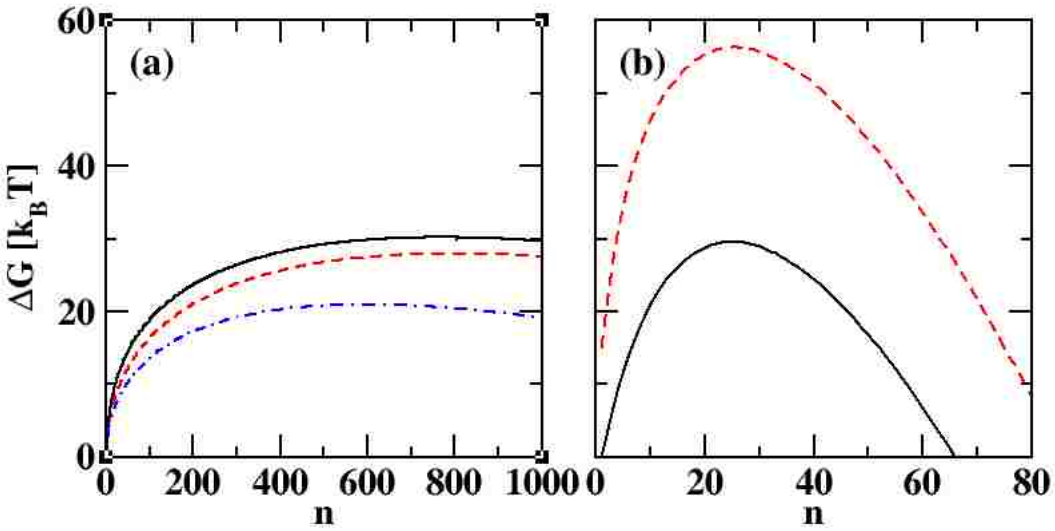

Figure 2.1: NFEs as a function of the cluster size for a 2D LJ system at $T^* = 0.427$ and $S = 1.036$ (in panel a) versus 3D LJ at $T^* = 0.45$ and $S = 89.5$ (in panel b). The simulation results and those predicted by CNT are shown as solid and dashed lines, respectively. Also shown in panel a are the CNT results obtained using the bulk-phase properties reported in Ref. 1 (the dashed-dotted line).

supersaturation condition is used, the NFE barrier height would be significantly higher for 3D than 2D. This result can be directly inferred from the theory. According to Eq. 2.1, the effect of dimensionality on the NFE values can be translated into the different dependencies of the surface free energy (SFE) on the cluster size n. For an m-dimensional system, the SFE term would be proportional to $n^{(m-1)/m}$. Correspondingly, the magnitude of this SFE term would increase with increasing dimensionality. Since the surface free energy is the term contributing to the formation of a nucleation barrier at the beginning stage of the cluster growth, it is expected that at similar supersaturation conditions, the barrier height would increase with increasing dimensionality of the system.

Although CNT can capture such qualitative trends, quantitatively there are clear differences between the NFEs predicted by CNT and those calculated from the simulation. Interestingly, these differences can sometime become opposite between 2D and 3D. As shown from Fig 2.1, for 3D CNT consistently predicts a higher NFE value for all cluster sizes with

the difference between CNT and the simulation growing larger with the increase of the cluster size and reaching a constant positive offset (26.5 $k_{\mathrm{B}}T$) after a certain cluster size. In contrast, for 2D such an overestimation only applies to the very first few clusters (starting from monomer due to the non-zero NFE value predicted by the theory for monomer but zero by our simulation). Then it becomes opposite with the magnitude of the difference also becoming larger with the increase of the cluster size and eventually approaching a stable value (2.1 $k_{\mathrm{B}}T$) after clusters are sufficiently large. This offset, which is considerably smaller than that found for 3D, is roughly equivalent to one order of magnitude difference in terms of nucleation rate, which is greater than that found by Zeng[40] but less than that obtained by Santra et al.[1] The difference between our simulation results and those by Santra et al.[1] could be partly explained by the slightly different thermodynamic properties obtained, particularly surface tension, 0.207 ± 0.003 here vs. 0.178 in their work, which would affect significantly the NFE data predicted by CNT. For example, using the thermodynamic properties obtained from their simulation, much lower NFE values would be obtained for the theory (see Fig. 2.1), which widens the gap between the simulation results and the CNT predictions. This emphasizes the importance of having highly accurate values for these input parameters for CNT. Another source of this discrepancy could be due to a more strict criterion used in their simulations where particles have to be first liquid-like (with the number of neighbors surpassing a certain threshold), then they can be considered whether or not they belong to a certain cluster, modified from the original Stillinger cluster criterion by ten Wolde and Frenkel.[46] This more strict criterion can lead to significantly higher NFE values obtained from the simulation. For example, we noticed that for monomers, their NFE plots show a large non-zero value compared to zero in our simulation since gas-phase monomers are used here.

For 3D systems, the $\delta\Delta G$ curves with $\delta\Delta G(n) = \Delta G(n) - \Delta G(n-1)$, which corresponds to the work required to add a monomer to the cluster, have been conventionally used to

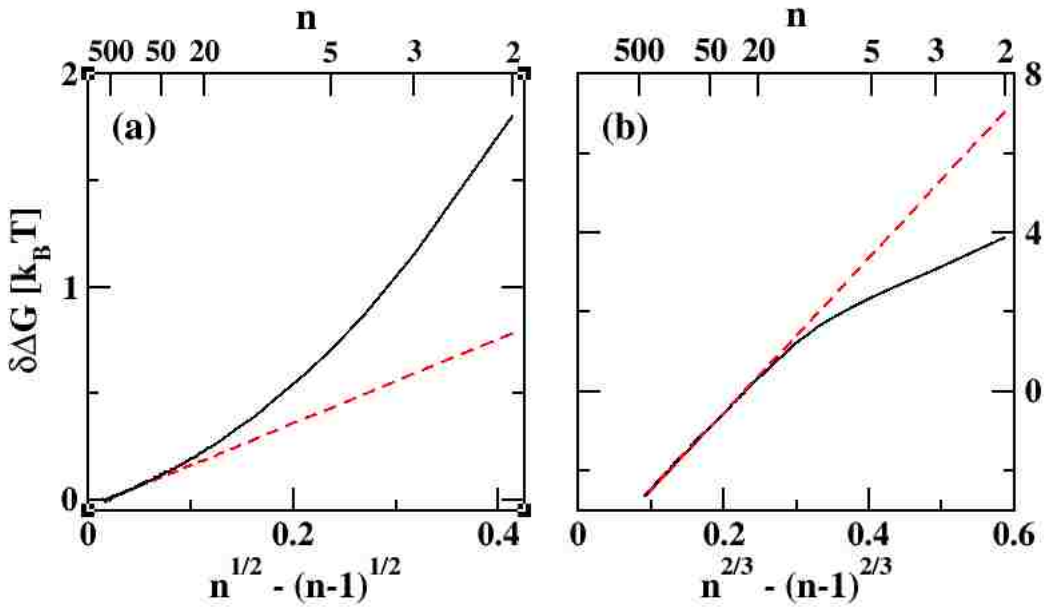

Figure 2.2: $\delta\Delta G(n)$ results analyzed from the NFE profiles shown in Figure 1 for simulation (solid) versus those predicted by CNT (dashed) plotted as a function of $n^{(m-1)/m} - (n-1)^{(m-1)/m}$ for 2D (panel a) and 3D (panel b).

extract the source of the errors for CNT.[7,36,35,37,38] Here we applied this analysis to 2D. In CNT, a plot of $\delta\Delta G$ as a function of $n^{(m-1)/m} - (n-1)^{(m-1)/m}$ would fall onto a straight line. While the slope of this linear line is governed by surface tension, density of the liquid phase, and some other constants, the intercept is determined solely by $\Delta\mu$. As shown in Fig 2.2, this linear behavior works remarkably well for clusters that are sufficiently large for both 2D and 3D systems. In particular, the $\delta\Delta G$ curves obtained from the simulation pretty much follow the CNT-prescribed linear lines for clusters containing more than 50 particles for 2D and more than 30 particles for 3D. However, deviations from this linear line behavior occur for clusters smaller than those sizes for both systems. While positive deviations are found for 2D, negative deviations are found for 3D. These deviations in $\delta\Delta G$ found for the initially formed clusters would contribute to a supersaturation independent offset on ΔG. This offset, combined with the difference in the ΔG value of the monomer (i.e., zero in our simulation versus $A_1\gamma + \Delta\mu$ in CNT), would account for the total deviation in the ΔG curves shown

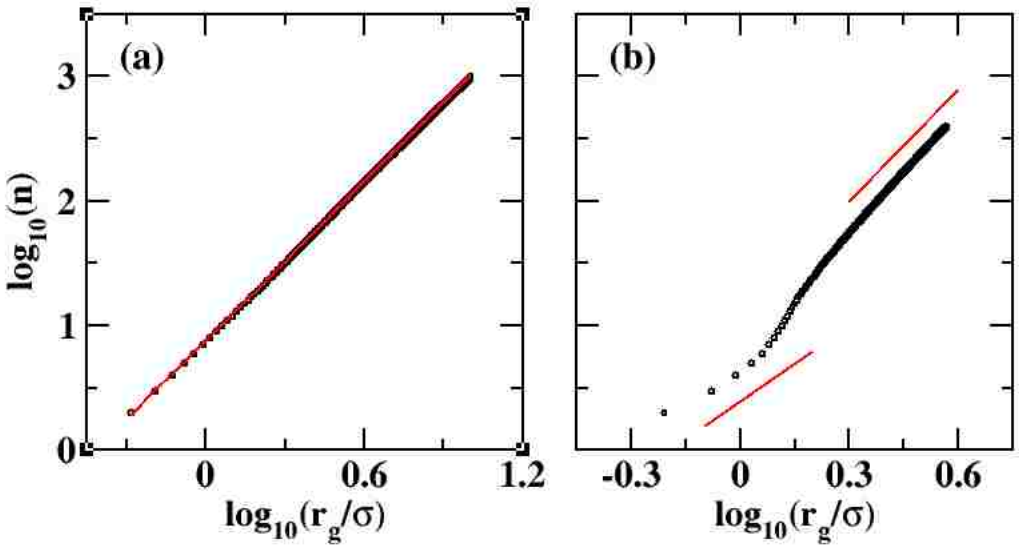

Figure 2.3: Log-log plot of cluster size vs radius of gyration for the 2D system at $T^* = 0.427$ (panel a) and 3D at $T^* = 0.45$ (panel b). While the 2D data can be fit by a linear line (solid) with a slope of about 2, the 3D data exhibit an "S" shape. The two straight lines shown in panel b are drawn as a guide to the eye with the slopes of 2 (lower) and 3 (upper).

in Fig. 2.1. Thus, the theoretical deviations found for those NFE results shown in Fig. 2.1 for both 2D and 3D, although opposite, can be actually explained by the same source, the incorrect description of the theory for the smallest clusters through this $\delta\Delta G$ analysis.

One question remains, that is, why would these theoretical errors switch from being positive for 3D to negative for 2D? In general, the deviations from the CNT behavior found from the simulation for small clusters can be traced to the fact that these small clusters do not behave as a bulk-droplet. From previous simulation studies on 3D systems,[35,14,47] it has been shown that the structure of the clusters can deviate significantly from the bulk-like structure assumed by the theory (that is, clusters of all sizes are compact and perfectly spherical). For example, the radius of gyration ($r_{\rm g}$) data obtained for the 3D system indicates that clusters formed at the beginning are fractal and lower in dimensionality (more or less 2-dimensional like, see Fig. 2.3) and only after a certain size, the clusters start to approach a three-dimensional structure, with the $r_{\rm g}$ curve exhibiting a pronounced S shape. Thus,

the growth of the cluster is accompanied by an increase of dimensionality. However, the extent of this structural change would be limited by the dimensionality of the space in which they grow. For example, for clusters placed in a lower-dimensional space such as 2D, one would expect minimal dimensionality transition effect as clusters containing a few particles are already two-dimensional-like. Indeed, the r_g data obtained for the 2D system shows little sign of this dimensionality transition with nearly all the data points aligning on a straight line with a slope close to 2 (see Fig. 2.3).

This striking structural difference between 2D and 3D could be the cause for the sign switch of the theoretical errors. In particular, the formation of loosely connected clusters is most likely the source of the negative deviation of the simulation results from the theory. These fractal aggregates are stabilized by the entropic factor or the volume term, which has a power-law scaling with the dimensionality, thus making fractal aggregates more stable at higher dimensional space. If one arbitrarily restricts the clusters to perfectly follow the CNT-prescribed structure, i.e., with a dimensionality of a bulk-droplet for all sizes, for small clusters formed in 3D or higher dimensional space, this constraint would lead to a significant increase of the NFE values since they are less stable than the fractal aggregates, possibly approaching or surpassing those predicted by the theory. On the other hand, an energetic argument can be used to explain the positive deviations of the simulation results from the theory. When clusters of all sizes adopt the same shape and dimensionality, small ones are energetically less stable than larger ones since each particle in the cluster has fewer particles to interact with whereas in CNT it is assumed that particles in small clusters can already take the bulk-phase chemical potential (meaning that they are equally stable compared to particles in bulk-phase). One can see this more clearly by working with a one-dimensional system. In that system, the surface term is insignificant and the formation free energy of a 1-D cluster is dominated by the bulk term. Thus this energetic factor directly influences the overall stability of this cluster at different sizes, leading to a more and more stable cluster

as it grows.

It should be noted that other forms of CNT have been developed, particularly to take into account the curvature effects that were found important for 3D clusters. These improved forms often focus on the surface term by treating the surface tension to be dependent on the cluster size.[48,49] These corrections have been shown to capture fairly well the errors found for the traditional form of CNT for 3D LJ clusters. For example, in Ref. 14, it was reported that using the expression provided by McGraw and Laaksonen[48] a positive theoretical error of 19.8 $k_{\mathrm{B}}T$ on the NFE barrier height was estimated, compared to 17.7 $k_{\mathrm{B}}T$ obtained from the simulation at $T^* = 0.8$ for LJ in 3D. In addition, it was shown from previous simulation work[35,14] that this theoretical error decreases with the increase of the temperature, which is in good agreement with the predictions by McGraw and Laaksonen[48] and another theoretical study by Zeng and Oxtoby.[39] In Ref. 39, it was also found that this positive error can become negative when T^* is above 1.08. To examine the temperature effect on the theoretical error for 2D LJ, additional simulations were performed at $T^* = 0.35$ and $T^* = 0.5$. As evident from the $\delta\Delta G$ curves shown in Fig. 2.4, this error remains negative over this entire temperature range. It should be pointed out that the theoretical errors on the NFE barrier heights obtained from the simulation for the 3D LJ system have been shown to be in good agreement with the discrepancies found between the experiment and the theory on the nucleation rates obtained for a LJ-like argon system,[50,51] except when the temperature is far below the triple point. The involvement of crystalline clusters in the nucleation process toward deep undercooling conditions was suggested to be one possible source for the different magnitudes of the theoretical errors observed between the simulation and the experiment, since in the simulation the nucleation rate was computed by assuming that the nucleation process proceeds via liquid-like clusters.[14] In a very recent experimental study on argon nucleation by Sinha et al.[10] that was carried out at much higher supersaturation (with an onset nucleation rate range 10 orders of magnitude larger than previous experimental

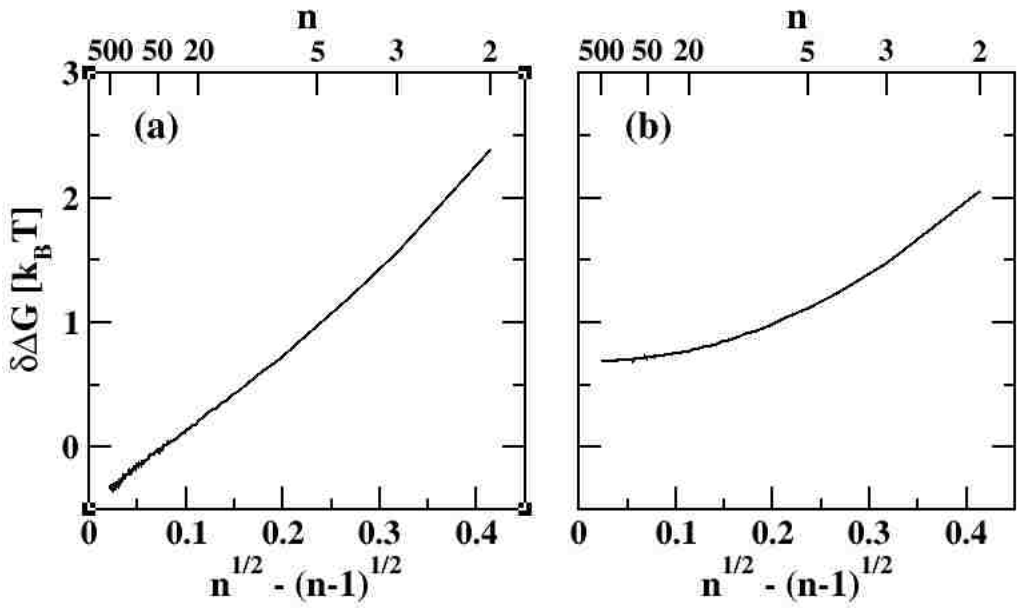

Figure 2.4: $\delta\Delta G(n)$ results obtained from the simulation at $T^* = 0.35$ (panel a) and $T^* = 0.5$ (panel b) for the 2D LJ system.

work[50,51]), the theoretical errors found there are in much better agreement with what we found from the simulation for the 3D LJ system. For 2D systems, it remains to be shown whether the theoretical errors found herein can be verified experimentally.

Acknowledgments

The material and figures presented in this section have been reprinted from from J. Chem. Phys 137(19) p.194304, 2012, with the permission of AIP Publishing. Copyright 2012. Expanded citation may be found in reference 52.

2.3 Surface Induced Theory

Surface nucleation is a form of heterogeneous nucleation where a supersaturated mobile phase condenses onto a solid surface. The surface acts as catalyst in the nucleation process allowing the droplet to form under conditions that would normally be unfavorable for homogeneous nucleation to occur. This area of nucleation is of major importance for applications such as hindering the formation of ice on external airplane equipment[53], atmospheric water droplet formation on the surface of particulates[54], metal cluster formation[55,56], etc. and has

been studied extensively by experiments[57–59] and in part by theoretical methods.[60–65]

The biggest challenge in modeling surface nucleation is the representation of the surface itself. Much like solvation, the surface can be done either implicitly or explicitly. An implicit surface provides a computationally cheap and simple method to represent the surface as a set of equations. This method can quickly provide a great deal of information about the geometry of a surface droplet without greatly increasing the number of calculations per Monte Carlo cycle. The drawback of this method is that for more complex geometries such as hills, valleys, or other complex features that exist on real surfaces the equations used to model such features are either difficult to derive or may be impossible to represent in this form. An explicit surface is expected to be significantly more accurate, but this kind of surface on top of being more computationally expensive provides many additional problems such as defining specific surface geometry, obtaining physically correct interaction potentials, etc.

This area of nucleation has been explored for specific systems such as water on silver iodide which has been studied extensively by Ward, Holdman, and Hale.[66] The silver iodide system was of interest due to its ice inducing properties. It was observed by Ward that the ionic surface allowed water to form a hexagonal pattern across the surface which was a likely aid in ice nucleation. A molecular dynamics study of the surface-induced nucleation of a Lennard-Jones (LJ) system was performed previously by Toxvaerd[67]. Toxvaerd largely focused on the structure of the droplet upon formation; however, not as much attention was given to the free energy of formation or the rate of formation.

Previously there had been several attempts to model surface nucleation through theoretical means. Based on the work by Becker and Döring performed in 1925,[2]the first surface-induced classical nucleation theory was proposed by Volmer[3] for nucleation on a flat featureless surface in 1929. This was later expanded upon in 1950 by Turnbull[68] who derived the equations for vapor-liquid nucleation for a wide array of surface geometries. For a flat

surface the free energy equation is given by

$$\Delta G(r) = f(\theta)\left(\frac{4}{3}\pi r^3 \rho_{\mathrm{l}} \Delta\mu + 4\pi r^2 \gamma_{\mathrm{lg}}\right) \tag{2.10}$$

$$f(\theta) = \frac{1}{2} - \frac{3}{4}\cos(\theta) + \frac{1}{4}[\cos(\theta)]^3 \tag{2.11}$$

Where $\Delta\mu$ is the chemical potential difference between the supersaturated and saturated phase, θ is the surface-droplet contact angle, ρ_{l} is the liquid density, and γ_{lg} is the surface tension of the droplet along the gas-liquid phase boundary. This equation is much like the homogeneous classical nucleation theory (CNT) equation, but differs by the pre-factor which contains the surface contact angle. The value of θ is between 0 and 180 degrees. The pre-factor at a θ value of 0 has a value of 0 while at a θ value of 180 the pre-factor is equal to 1 which returns the equation to the homogeneous CNT equation.

Because this theory is based on a bulk-droplet model it is predicted to have some of the same problems associated with the homogeneous CNT equation. One such problem is that small clusters do not behave like the bulk phase. Correspondingly there are errors in the free energy prediction for small cluster sizes which was shown in previous works.[35,7] While the normal CNT problems would be expected with this theory, it is also possible that the introduction of the contact angle term will provide additional problems. Since the contact angle is a macroscopic quantity, it may also have the small cluster problem that the other thermodynamic quantities in the homogeneous CNT equation have.

To examine the theory we must first determine the predictions made by the theory. In a similar manner to the homogeneous CNT equation, this equation can also be rewritten as a function of cluster size instead of radius by using the density relationship $\rho_{\mathrm{l}} = \frac{n}{\frac{4}{3}\pi f(\theta) r^3}$. This yields the result

$$\Delta G(n) = n\Delta\mu - \gamma_{\mathrm{lg}} \left(\frac{36\pi f(\theta)}{\rho_{\mathrm{l}}^2} \right)^{\frac{1}{3}} n^{\frac{2}{3}} \tag{2.12}$$

According to this theory the primary difference in nucleation free energy is created by the truncated circle geometry the droplet takes upon the surface. We can also show this equation can be used to create a $\delta\Delta G$ plot similar to those which has been used in a multitude of studies.[35,7,69,39] Where

$$\delta\Delta G = \Delta G(n) - \Delta G(n-1) = \Delta\mu + \gamma_{\mathrm{lg}} \left(\frac{36\pi f(\theta)}{\rho_{\mathrm{l}}^2} \right)^{\frac{1}{3}} (n^{\frac{2}{3}} - (n-1)^{\frac{2}{3}}) \tag{2.13}$$

The size of the critical cluster and barrier height predicted by the theory were also calculated by Volmer[3] and found to be

$$\Delta G^* = \frac{16\pi\gamma_{\mathrm{lg}}^3}{3\rho_{\mathrm{l}}^2\Delta\mu^2} f(\theta) \tag{2.14}$$

$$n^* = \frac{32\pi\gamma_{\mathrm{lg}}^3}{3\rho_{\mathrm{l}}^2\Delta\mu^3} f(\theta) \tag{2.15}$$

Therefore when comparing the same liquid-vapor nucleation on two different surfaces, for the same conditions the barrier height ratio as well as the critical cluster size ratio of two different surfaces is given by

$$\frac{\Delta G_2^*}{\Delta G_1^*} = \frac{f(\theta_2)}{f(\theta_1)} = \frac{n_2^*}{n_1^*} \tag{2.16}$$

In the case of a homogeneous system the $f(\theta)$ term is simply replaced with a value of 1. Thus the predicted ratio of a surface-induced system and the homogeneous 3D system is

simply

$$\frac{\Delta G^*_{\mathrm{Sur}}}{\Delta G^*_{\mathrm{Hgn}}} = f(\theta) \tag{2.17}$$

For a bulk droplet, the value of θ is specified by Young's equation as follows.

$$\cos(\theta) = \frac{\gamma_{\mathrm{sg}} - \gamma_{\mathrm{sl}}}{\gamma_{\mathrm{lg}}} \tag{2.18}$$

Where γ_{lg}, γ_{sg}, and γ_{sl} are the surface tensions of the liquid-gas, solid-gas, and solid-liquid interfaces respectively. Since the CNT theory is based on the bulk-droplet model, all clusters are expected to adopt the same contact angle value irrespective of size. Thus the $f(\theta)$ value is expected to be a constant and the barrier height ratios between the homogeneous and heterogeneous systems are also expected to be constant; however, clusters are finite objects, which questions the validity of such bulk-droplet approximations inherent in the classical nucleation theory. For heterogeneous nucleation the theory brings an additional macroscopic term, the contact angle, to the derivation. Much like quantities such as the surface tension and chemical potential, the contact angle can be size dependent.[70] Therefore it is imperative to use more quantitative and direct methods such as molecular simulation based on more detailed microscopic models to verify the validity of the theory built on a macroscopic model in the same manner as previous studies.[35,7,71]

2.3.1 Simulation Details

The simulations were carried out using the aggregation-volume-bias Monte Carlo method (AVBMC) coupled with umbrella sampling.[14] Umbrella sampling[17] is implemented to solve the critical problem of sampling high free-energy clusters by adding artificial biasing potentials to enhance tremendously the probabilities of these clusters being visited by the simulation. AVBMC[21,22] is used to enable the direct transfer of particles between the cluster and the mother phase to greatly speed up the sampling of the otherwise slow condensation/evaporation events that are important to nucleation. This was achieved by dividing

the space surrounding each particle into a bounded and unbounded region and using a swap move to transfer particles between these two regions. The volume of the bounded space can be chosen to conform to the Stillinger[23] cluster criterion used in this study. The in-depth explanation of this simulation approach can be found in Refs. 21 & 22.

The surface used is an implicit 9-3 LJ equation which can be derived by integrating the 12-6 LJ potential over an infinite plane and surface depth.

$$V(z) = \frac{2}{3}\pi\rho_{\mathrm{s}}\varepsilon_{\mathrm{sl}}\sigma_{\mathrm{sl}}^{3}\left(\frac{2}{15}\frac{\sigma_{\mathrm{sl}}^{9}}{z^{9}} - \frac{\sigma_{\mathrm{sl}}^{3}}{z^{3}}\right) \tag{2.19}$$

Where σ_{sl} and $\varepsilon_{\mathrm{sl}}$ are the LJ parameters for the interaction between a single cluster particle and a single surface particle and ρ_{s} is the surface density. σ_{sl} and $\varepsilon_{\mathrm{sl}}$ are defined using the standard Lorentz-Berthelot mixing rules[72,73]

$$\varepsilon_{\mathrm{sl}} = \sqrt{\varepsilon_{\mathrm{ll}}\varepsilon_{\mathrm{ss}}} \tag{2.20}$$

$$\sigma_{\mathrm{sl}} = \frac{\sigma_{\mathrm{ss}} + \sigma_{\mathrm{ll}}}{2} \tag{2.21}$$

Even though there are a total of three control parameters that can vary depending on the type of surface, these parameters can be integrated into a single control parameter which is denoted as ε^{*} as follows.

$$\varepsilon^{*} = \frac{2}{3}\pi\rho_{\mathrm{s}}\varepsilon_{\mathrm{sl}}\sigma_{\mathrm{sl}}^{3} \tag{2.22}$$

$$V(z) = \varepsilon^{*}\left(\frac{2}{15}\frac{\sigma_{\mathrm{sl}}^{9}}{z^{9}} - \frac{\sigma_{\mathrm{sl}}^{3}}{z^{3}}\right) \tag{2.23}$$

For a one-component liquid phase the ε^{*} parameter can be reduced in terms of the liquid phase interaction parameter $\varepsilon_{\mathrm{ll}}$ to create a relative parameter $\varepsilon_{\mathrm{r}}^{*}$

$$\varepsilon_{\mathrm{r}}^{*} = \frac{\varepsilon^{*}}{\varepsilon_{\mathrm{ll}}} = \frac{2}{3}\pi\rho_{\mathrm{s}}\sqrt{\frac{\varepsilon_{\mathrm{ss}}}{\varepsilon_{\mathrm{ll}}}}\sigma_{\mathrm{sl}}^{3} \tag{2.24}$$

The simulation was carried out by varying $\varepsilon_{\mathrm{r}}^{*}$ to values of 1, 2, 3, and 5. The cluster-cluster interaction parameters were set to 1 for both $\varepsilon_{\mathrm{ll}}$ and σ_{ll}. All the pair interactions were included in the computation of the total energy. The simulations were carried out at a reduced temperature of 0.5. This temperature is sufficiently below the freezing point of 3D LJ, which is around 0.69,[14,74] but is above the freezing point of 2D LJ, which is around 0.41,[75,76] and has previously been studied for the homogeneous case.[35] In addition to the heterogeneous systems, a 3D homogeneous and a 2D homogeneous simulation under similar conditions were carried out for comparison purposes. The saturation ratio for the 2D homogeneous was set so that the free energy profile of the 2D homogeneous system would be comparable to the systems with high surface interaction strengths.

2.3.2 Order Parameter

In this study we also employed the Q_6 Steinhardt order parameter to examine the crystallinity of the formed cluster.[77] The Q_6 order parameter is given by

$$Q_6 = \left(\frac{4\pi}{13}\sum_{m=-6}^{6} \mid q_{6m} \mid^{2}\right)^{\frac{1}{2}} \tag{2.25}$$

$$q_{6m} = \frac{1}{N_{\mathrm{b}}}\sum_{n=1}^{N_{\mathrm{b}}} Y_{6m}(\theta_{ij}, \phi_{ij}) \tag{2.26}$$

Where $Y_{6\mathrm{m}}$ is the spherical harmonic angular function, N_{b} is the total number of neighbors, θ and ϕ are the angles between neighboring particles i and j from an arbitrary reference point. Any two particles which are within 1.5σ are defined as being a neighbor to each other.

2.3.3 Contact Angle

The contact angle calculations were performed by generating a droplet on the various surfaces containing 1500 LJ particles. Once the droplet reached equilibrium under translational moves the probability of finding a particle along the Z axis along with the probability of finding a particle at a specific distance r_{X-Y} from the center of mass along the $X-Y$ plane were calculated. From the probability data it was possible to collect the interfacial radius and the droplet height by finding the Z and r_{X-Y} value that yields 95% of the total area when integrated from 0. 95% is used so that a majority of the particles are included in this analysis except those which are loosely bounded and located at the interface. Once these values have been determined it is possible to find the contact angle using the relationship given by Fan and Cağin.[78]

$$\cos(\theta) = 1 - \frac{h}{R} \tag{2.27}$$

Where h is the droplet height and R is the untruncated radius of the droplet. For droplets where the values of the contact angle are greater than 90 degrees the integrated value over the $X-Y$ plane will give an approximated value of R; however, for contact angle values less than 90 degrees, the value given by the integration will be the radius of the circle created by the intersection of the surface plane (Denoted as R_{s}) and the droplet(see Fig. 2.5). This value can still be used to calculate R using the expression

$$R = \frac{R_{\mathrm{s}}^2 + h^2}{2h} \tag{2.28}$$

2.3.4 Results and Dicussion

Shown in Fig. 2.6 are the free energy curves as a function of the cluster size obtained from the simulation. The free energy curves were drawn using the same scale in order to demonstrate the gradual decline of the free energy barrier as surface interaction strength

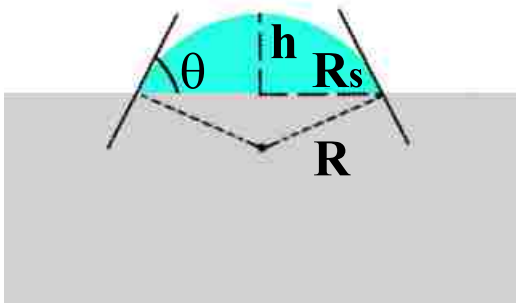

Figure 2.5: Diagram of a surface droplet. Depicted here are the parameters used to calculate the contact angle. h is the height of the droplet, R_s is the radius of the circle generated by the intersection of the surface plane and spherical droplet, and R is the radius of the spherical droplet if it were untruncated by the surface plane.

increases. All free energy curves were obtained using the same gas-phase number density ($n_v = 4.72 \times 10^{-4}$) except for the 2D system which was obtained using a number density of $n_v = 9.45 \times 10^{-5}$. The number density yields a supersaturation ratio of 9.98 which corresponds to a $\Delta\mu$ value of $-2.80\ k_B T$. The barrier for the homogeneous 3D system is estimated to be 240.2 $k_B T$ while the barrier heights for the surfaces with ε_r^* of 1, 2, and 3 have values of 208.1, 129.6, and 36.6 $k_B T$ respectively. The remaining systems showed no barrier. In addition to the decline of the barrier, the critical cluster size shifts to lower values as the surface interaction strength increases. The estimated critical cluster size for the 3D homogeneous system is 334 while the critical cluster size for the surface-induced nucleation was estimated to be 291, 224, and 78 for ε_r^* values of 1, 2, and 3, respectively.

The free energy curves for each system were used to create a $\delta\Delta G$ plot (see Fig. 2.7). According to CNT a plot of the free energy difference between a cluster size of n and $n-1$ will fall on a straight line if plotted against $n^{\frac{2}{3}} - (n-1)^{\frac{2}{3}} \propto n^{-\frac{1}{3}}$. The slope of this line will be governed by the liquid-vapor surface tension and liquid coexistence density while the intercept will be $\Delta\mu$. The heterogeneous case's slope differs by a factor of $[f(\theta)]^{\frac{1}{3}}$. Thus the $\delta\Delta G$ analysis provides a convenient way to examine the theoretical errors. It also provides a robust procedure for extrapolating bulk-phase information using the data

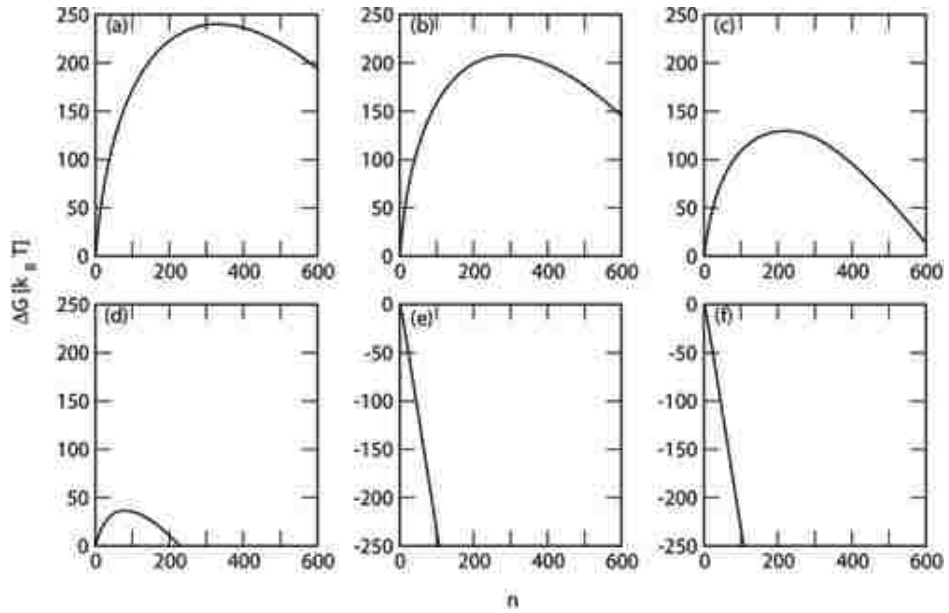

Figure 2.6: ΔG plots for a 3D homogeneous LJ nucleation (panel a) and surface-induced nucleation with ε_r^* values of 1 (panel b), 2 (panel c), 3 (panel d), and 5(panel e). Also included is the curve for the 2D homogeneous nucleation (panel f).

obtained on finite-size clusters. Indeed, previous simulations of many single component nucleation systems have shown that the $\delta\Delta G$ results obey the theoretically prescribed linear behavior for sufficiently large clusters; however, deviations between the simulation and CNT are observed toward the smallest cluster sizes leading to the conclusion that the failure of CNT can be traced to its incorrect description of the smallest cluster[35,7,69]. It is important to note that in previous studies for small cluster sizes the 3D $\delta\Delta G$ plot showed a negative deviation from the predicted CNT results. Conversely the 2D homogeneous LJ showed a positive deviation for the smallest cluster sizes.[52] These characteristic shapes provide a useful guide for analyzing the geometry of the surface-induced systems. As can be seen, for the ε_r^* of 1, the deviation from the predicted line was negative, but as the value of ε_r^* increases the systems show a slow incline. As the interaction strength gradually increases, the $\delta\Delta G$s begin changing from a negative error at small cluster sizes to a positive error. At an ε_r^* of 5 the system finally shows a 2D-like deviation from the theory at small cluster sizes. This

trend is well within reasonable expectations. As the surface-particle interaction increases the droplet prefers to flatten out and form a 2d-like structure, while at low strengths it prefers to be 3D-like in shape. Another interesting trend is the decline in the slope at larger cluster sizes.

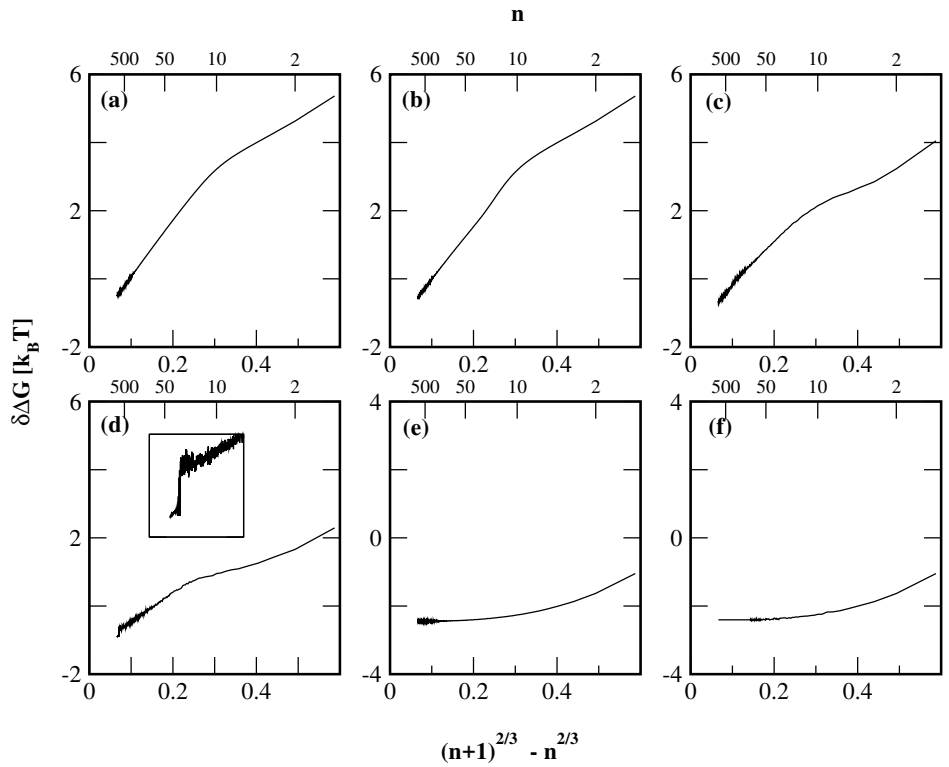

Figure 2.7: $\delta\Delta G$ plots of the nucleation free energy profile for a 3D homogeneous LJ nucleation (panel a) and surface-induced nucleation with $\varepsilon_{\mathrm{r}}^{*}$ values of 1 (panel b), 2 (panel c), 3 (panel d), and 5(panel e). Also included is the curve for the 2D homogeneous nucleation (panel f).

At an $\varepsilon_{\mathrm{r}}^{*}$ value of 1 the $\delta\Delta G$ line was almost identical to the homogeneous 3D nucleation. As the value was increased to 2 and 3 the line showed a gradual decline in slope. At an $\varepsilon_{\mathrm{r}}^{*}$ of 5 the line had a slope value near 0. According to the theory (see Eq. 3) this corresponds to a decrease in the contact angle which implies the droplet prefers to spread out across the surface (also see later discussion in Section III.B). The decline in the contact angle can also be seen directly with the snapshots supplied in Fig. 2.9. The snapshots clearly illustrate

the spreading of the droplets as surface interaction strength increases, which is in agreement with the observed trend from the free energy data. The shift from 3D to 2D is also confirmed by the radius of gyration data (see Fig. 2.8). The radius of gyration plot provides a useful way to analyze the dimensionality of the system. When the log of the cluster size is plotted against the log of the radius of gyration the resulting curve will have a slope proportional to the dimensionality of the system.

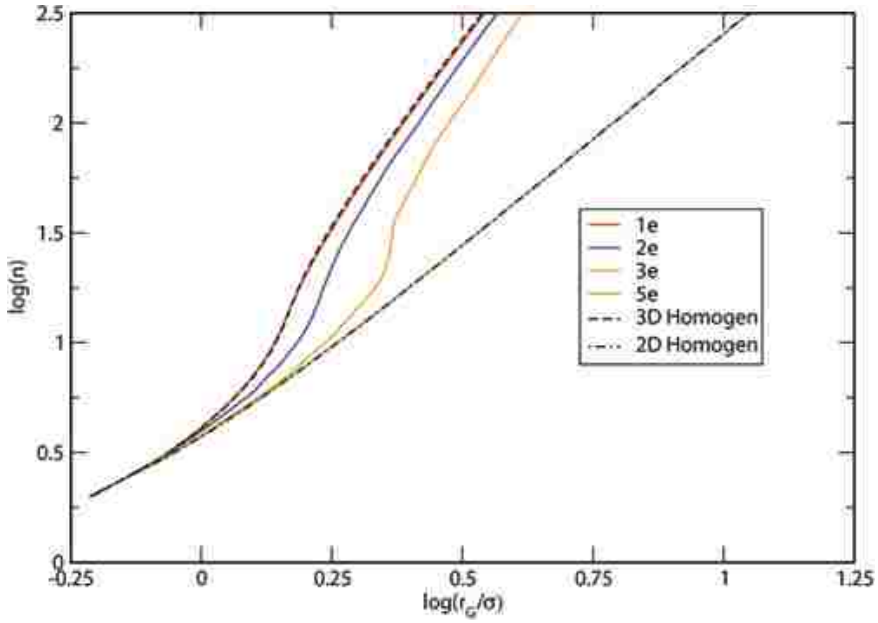

Figure 2.8: Log-Log plot of the radius of gyration of the surface-induced systems and homogeneous systems plotted against cluster size for $\varepsilon_r^* = 1$ (solid red line), 2 (solid blue line), 3 (solid orange line), and 5 (solid green line). Also shown is the radius of gyration for the 2D (dotted-dashed line) and 3D (dashed line) homogeneous systems.

It was noted in previous studies[52] that the 3D homogeneous system initially had a slope of 2 for small cluster sizes which indicates the clusters were 2D-like, but as the clusters grew in size there was a transition from 2D to 3D as shown by the s-curve. In contrast, the 2D slope remained fixed at 2 even at large cluster sizes. The radius of gyration data collected from the $\varepsilon_r^* = 1$ system shows that it follows the 3D homogeneous curve very closely indicating that it follows the same transition pattern as the 3D system; however, for the $\varepsilon_r^* = 2$ and 3 systems the observed transition from 2D to 3D does not occur until the clusters are larger

in size and for the $\varepsilon_r^* = 5$ system it closely follows the 2D homogeneous line indicating the clusters maintain a 2D-like geometry. This supports the observations made in the $\delta\Delta G$ data.

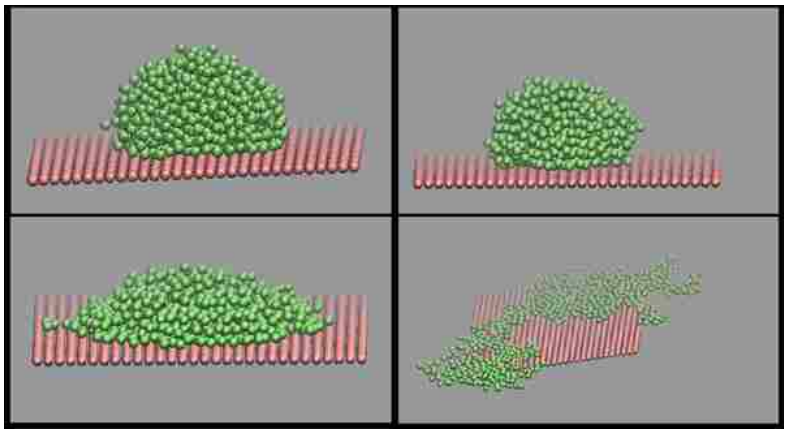

Figure 2.9: Surface snapshots for ε_r^* values of 1 (top-left), 2 (top-right), 3 (bottom-left), and 5 (bottom-right). The red colored surface is added for visualization purposes.

Another interesting result from Fig. 2.7 was the sudden shift in the $\delta\Delta G$ value at high cluster sizes observed for the $\varepsilon_r^* = 3$ system. This sudden change is indicative of another phase transition and in this case a disorder-to-order transition. To show the system was indeed crystallizing, the Q_6 order parameter was employed. The simulation was set to collect the Q_6 value for every cluster of 600 particles. This cluster size was selected since it is beyond the transition point while still small enough to be computationally efficient. The Q_6 parameter mathematically can take on any value between 0 and 1 depending on how ordered the cluster is; however, it is nearly impossible to receive a value near 1 for realistic system. Bulk liquids have a near zero order parameter while bulk solids will have a value between 0.3 and 0.6 depending on their crystal structure. In the case of clusters, values below 0.05 have been noted by Chen[35] as being disordered and closer to a liquid droplet in structure. Conversely if the value is found to be above 0.1, the cluster shows enough order that it can be considered crystalline-like.

As shown in Fig. 2.10 for the $\varepsilon_r^* = 1$ surface the Q_6 distribution was centered around 0.05 and had a very small deviation. This indicates the clusters on the $\varepsilon_r^* = 1$ surface

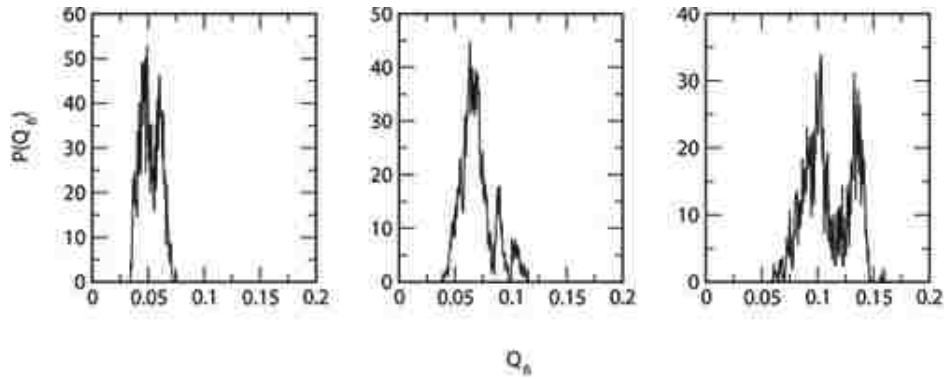

Figure 2.10: Probability distribution of the Q_6 order parameter obtained for clusters containing 600 LJ particles on the $\varepsilon_r^* = 1$ (left), 2 (middle), and 3 (right) systems.

largely behave liquid-like. In contrast the $\varepsilon_r^* = 2$ and $\varepsilon_r^* = 3$ surfaces show a higher Q_6 value. The $\varepsilon_r^* = 2$ surface shows that a small number of the clusters reach values around 0.1 while the $\varepsilon_r^* = 3$ system is completely distributed around 0.1 which indicates these clusters are displaying a near crystalline level of ordering. The implication of this is that crystallization only occurs for strong enough interaction strengths. For lower strengths the droplet is still very close to the bulk liquid in structure. Cluster visualization provides direct evidence on how the surface promotes this type of ordering. In particular, under suitable surface interaction strength, particles close to the surface can adopt an ordered hexagonal arrangement (see Fig. 2.11).

The density profile provides additional insights into the mechanism on how surfaces can promote the formation of ordered structures (see Fig. 2.12). It has been observed before by Toxvaerd[67] that there is a distinct layering effect when particles are placed on top of the implicit surface. This is also confirmed by the results obtained here. In addition, the density profiles obtained for the different surfaces indicate that the magnitude of this effect increases with the interaction strength. It was found for the $\varepsilon_r^* = 1$ surface that the stacking was present, but there was still a high degree of disorder as shown by the high overlap between the peaks of the density profile. As the interaction strength increases from 1 to 2 the peaks become more defined and separated from each other, but there is still a moderate

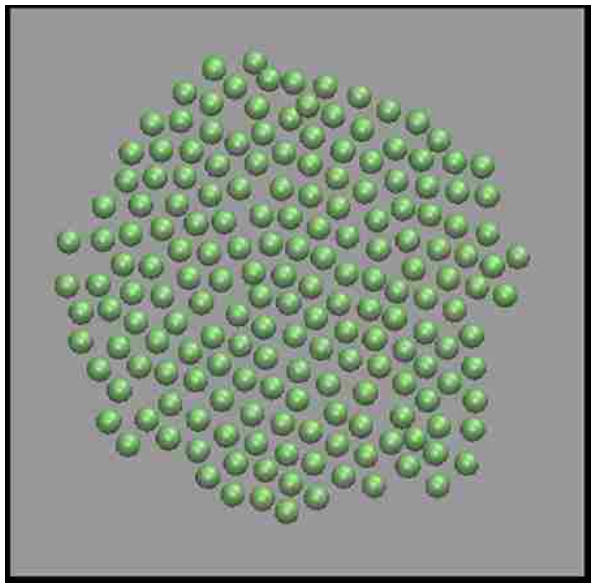

Figure 2.11: Snapshot of the layer closest to the surface that was isolated from a cluster with a Q_6 value of 0.11.

degree of overlap at a surface strength of 2. At a surface interaction strength of 3 the peak becomes distinct and isolated which indicates a much higher degree of ordering than the lower interaction strengths can induce. Based on these results, it can be concluded that this surface-induced layering effect is due to a combination of geometrical constraints and the adhesive forces imposed by the surface which can seed the ordering required to begin the crystallization of the liquid phase. If the interaction strength is not strong enough the ordering effect is not sufficient to begin crystallization. In contrast if the interaction strength is far too strong the particles will begin spreading out to form 2D-like structures which has a significantly lower freezing point.[76] This in turn limits crystallization.

2.3.5 Examination of the Contact Angle Term Introduced by the Heterogeneous CNT

As mentioned in the Introduction, the classical nucleation theory introduces an additional contact-angle dependent term $f(\theta)$ into the description of the surface-induced nucleation. As shown in Eqs. 5&6, under the same supersaturation condition, both the nucle-

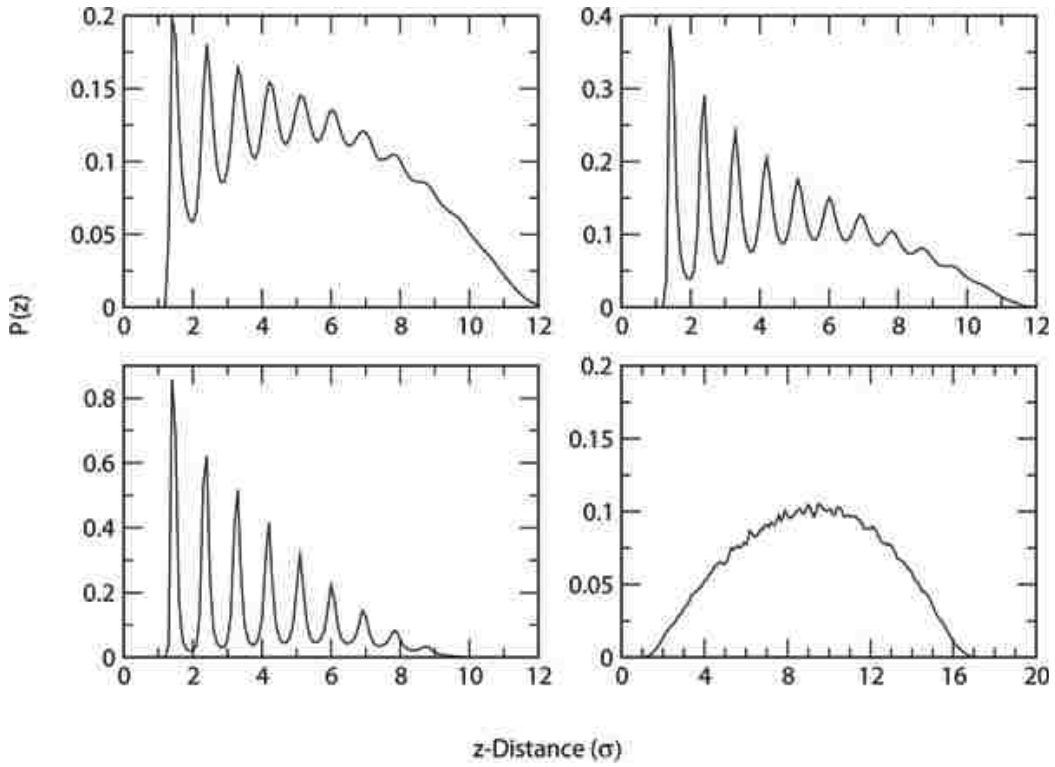

Figure 2.12: Density profiles for surface systems with ε_r^* of 1 (panel a), 2 (panel b), and 3 (panel c) in addition to the 3D homogeneous system (panel d).

ation free energy (NFE) barrier and the critical cluster size for surface-induced nucleation are proportional to the results predicted by CNT for the homogeneous nucleation with the proportionality constant being related to θ (and more exactly $\frac{1}{2} - \frac{3}{4}\cos(\theta) + \frac{1}{4}[\cos(\theta)]^3$). To examine this, at first the NFE barrier heights obtained for the $\varepsilon_r^* = 1$ and $\varepsilon_r^* = 2$ systems are compared to those obtained for the 3D homogeneous system at various supersaturation conditions along with the CNT results for the 3D homogeneous nucleation (see Fig. 2.13). The data obtained for the other systems are not included in this section due to the formation of either crystalline clusters or clusters with 2D geometries.

These results were plotted as a function of $(\ln S)^{-2}$ as according to the CNT, the NFE barrier height is proportional to $(\ln S)^{-2}$. This leads to a linear line for CNT, where the surface tension and the bulk liquid density at coexistence determines the slope of this line. The results for the surface-induced nucleation should also fall onto a straight line, but with

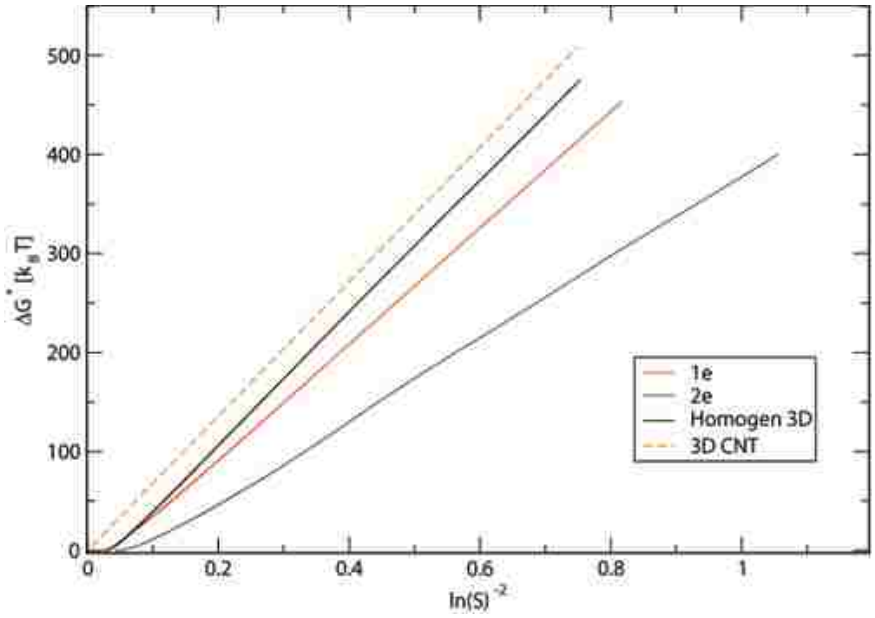

Figure 2.13: NFE barrier heights plotted against $(\ln S)^{-2}$ for the 3D homogeneous system (solid black line), the $\varepsilon_{\mathrm{r}}^{*} = 1$ system (solid red line), the $\varepsilon_{\mathrm{r}}^{*} = 2$ system (solid blue line), and the predicted CNT line for the 3D homogeneous system (dotted orange line)

a reduced slope proportional to the $f(\theta)$ term. In agreement with the CNT, the simulation results for the $\varepsilon_{\mathrm{r}}^{*} = 1$ and $\varepsilon_{\mathrm{r}}^{*} = 2$ systems along with the 3D system show this linear behavior except at high supersaturation conditions. The deviations at high supersaturations are due to the fact that the critical cluster sizes for these supersaturation values are small and fractal, i.e., 2D-like rather than being a compact 3D droplet as assumed by CNT.[7] The discrepancies noted previously on the $\delta\Delta G$ results toward the small cluster size range can be also used to explain why the barrier heights determined from the simulations are consistently lower than the CNT predictions for all supersaturations for the homogeneous 3D nucleation system. Nevertheless, this offset appears to be a constant after supersaturation is sufficiently low. The CNT predictions for the surface-induced nucleation are not included in this figure as the contact angle is also required and this parameter should be derived from a bulk phase simulation or Young's equation which would need the surface tension values of the other two surfaces. Neither approach is practical for the system studied here; however, based on

how the theory compares to the simulation for the homogeneous nucleation and also on how the theoretical (simulation) results would evolve with the inclusion of surface interactions, it is reasonable to expect similar discrepancies between the theory and the simulation for the surface-induced nucleation compared to the homogeneous nucleation. In particular, for both the $\varepsilon_{\mathrm{r}}^{*} = 1$ and $\varepsilon_{\mathrm{r}}^{*} = 2$ systems, it is likely that the theory overestimates the barrier height for all supersaturations and that this difference becomes a constant toward the low supersaturation region when the simulation results start to exhibit the linear behavior in Fig. 2.13. Although such direct comparisons between the simulation and the theory can be troublesome for surface-induced nucleation. Eqs. 7&8 provide a way to examine how the theory performs with this additional contact-angle dependent term, $f(\theta)$. Plotted in Fig. 2.14 are the ratios of the NFE barrier heights obtained for the two surface-induced nucleation systems to the homogeneous system as a function of the supersaturation. According to CNT, this ratio should yield the $f(\theta)$ term which is a constant (i.e. independent of supersaturation) and the value of this constant depends on the surface interaction strength.

In contrast, the ratio estimated from the simulation shows a clear dependence on both supersaturation and the surface interaction strength; however, toward the low supersaturation region for both the $\varepsilon_{\mathrm{r}}^{*} = 1$ and $\varepsilon_{\mathrm{r}}^{*} = 2$ systems, this ratio appears to approach a constant value. Furthermore, for both systems as the supersaturation decreases, this ratio draws closer to the $f(\theta)$ value estimated from the finite droplet containing 1500 particles deposited on the corresponding surface. This implies that the contact-angle term included by the theory for heterogeneous nucleation may work reasonably well for low supersaturation conditions.

Using the method described in Section II.C, the contact angle was estimated to be 127° for the $\varepsilon_{\mathrm{r}}^{*} = 1$ system and 97° for the $\varepsilon_{\mathrm{r}}^{*} = 2$ system. To further check these angles, the density contour profiles were analyzed along both the Z-axis and the radial direction on the $X - Y$ plane. These density data were normalized by the density of the interior region at

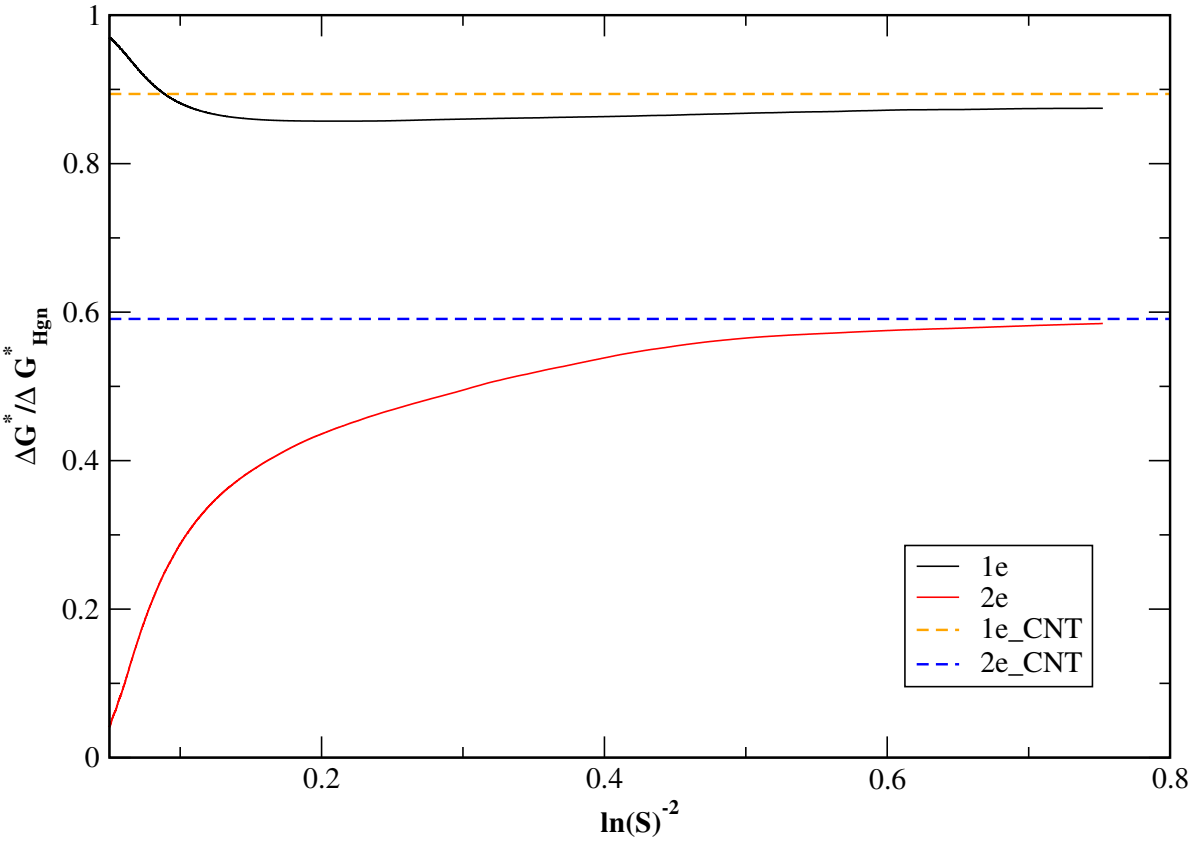

Figure 2.14: Ratio of the barrier heights obtained for the $\varepsilon_r^* = 1$ (solid black line) and $\varepsilon_r^* = 2$ (solid red line) systems with the homogeneous 3D system at the same saturation ratio along with the value predicted by CNT using the finite droplet contact angle calculation for the $\varepsilon_r^* = 1$ (dotted orange line) and $\varepsilon_r^* = 2$ (dotted blue line) systems.

each Z value to remove the density oscillation along the Z direction so that the cluster/vapor interface can be located by a contour line at a constant density. As shown in Fig. 2.15, linear lines with the contact angles obtained above are nearly tangential to this interfacial curve at $Z = 0$. It should be noted that even after this normalization, significant oscillations of the density are still noticed toward the interface, especially for the $\varepsilon_r^* = 2$ system. Also regions with a normalized density value above 1 appear at the interface. These regions are centered around those Z values when the averaged density profile along the Z direction (i.e., the density profiles plotted in Fig. 8) reaches a minimum. These results indicate that particles at the cluster/vapor interface may not exhibit the same ordering that is observed

for the interior particles. This is due to the fact that these interfacial particles are more loosely connected with each other.

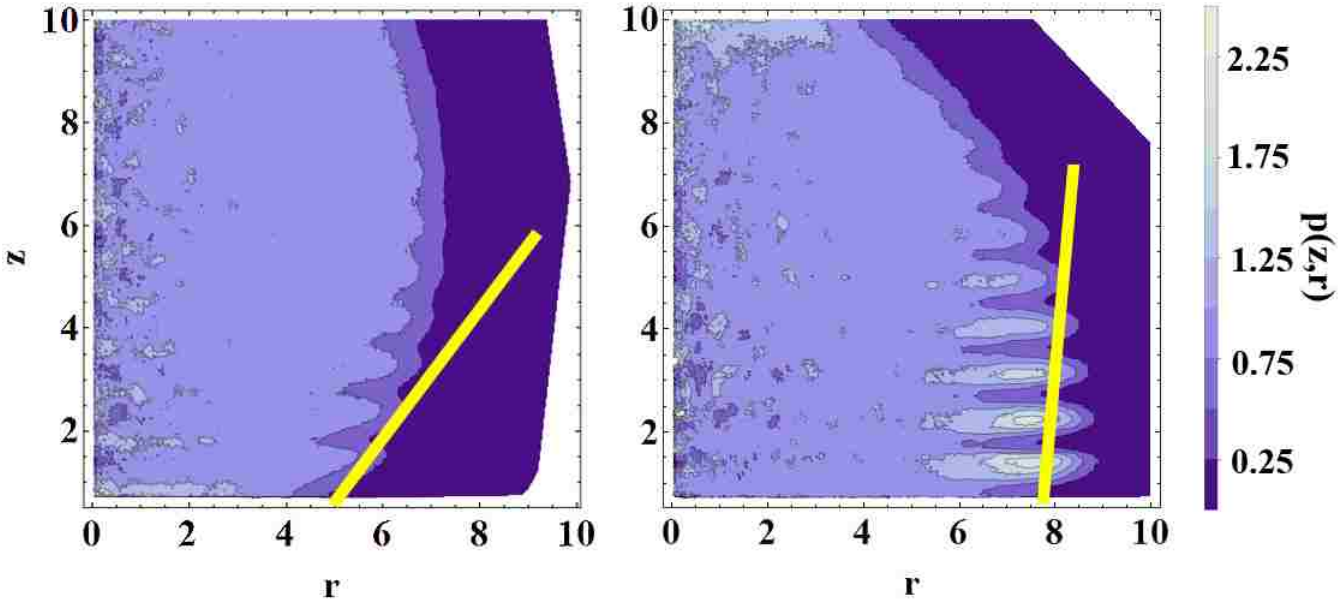

Figure 2.15: Contour density profiles for surface systems with ε_r^* of 1 (left panel) and 2 (right panel) with contact angle lines added. These angles were obtained from the method described in Section II.C

Similar comparison between the simulation and the theory was made on the critical cluster size. However, the critical cluster size estimation from the simulation can be complicated by a few factors. First, the free energy profile can be very flat near the maximum where the critical cluster is located and a wide range of clusters in a size window between 5 and 40, depending on the supersaturation ratio, can have a free energy value close to this maximum, say, within 0.2 k_BT. Even a small uncertainty on the NFE value can lead to a large shift on the location of this maximum. Therefore, instead of using only the maximum position, the critical cluster size was determined by the following: at any given supersaturation, find all clusters with an NFE that differs from the maximum by less than 0.2 k_BT and then obtain an average size for these clusters and use this average size as the critical cluster size

estimated for this condition. Secondly, due to the discrete nature of the cluster size, the critical cluster size determined even by averaging the size over a range of clusters remains as a step function. That is, it would take the same value within a range of supersaturation conditions. In contrast, the theoretical values predicted from Eq. 6 would be continuous. Therefore when comparing to the theory on this property, only the middle point over that range of supersaturation (when the same critical cluster size was estimated) was selected to plot the data shown in Fig. 2.16. Again, the supersaturation axis is represented in a special way (i.e. in $(\ln S)^{-3}$) so that the theoretically predicted critical cluster size would fall onto a straight line. The simulation results obtained for all systems appear to follow this linear behavior especially at low supersaturation conditions. In fact, for the homogeneous system, the simulation curve nearly coincides with the theoretical line. However deviations are noticed toward the high supersaturation when critical clusters are small, consistent with the other results (such as $\delta\Delta G$ and barrier heights).

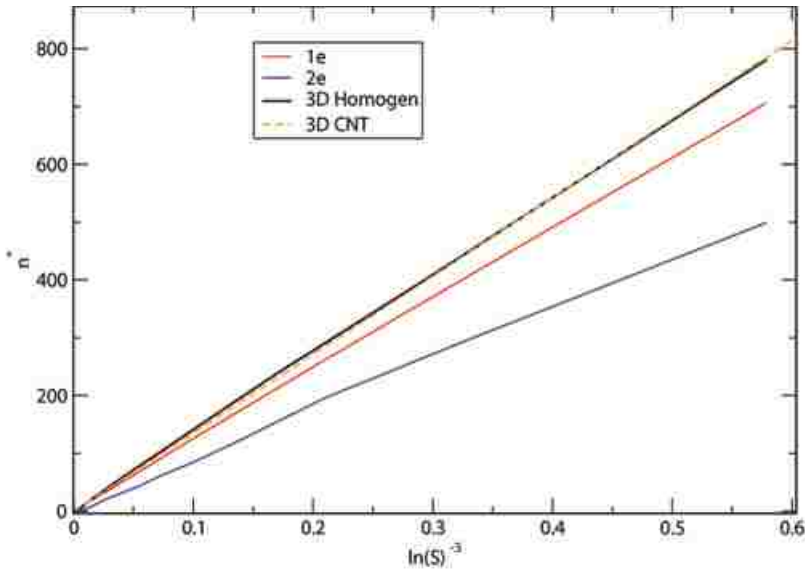

Figure 2.16: Critical cluster size plotted against $-(\ln S)^{-3}$ for the 3D homogeneous system (solid black line), the $\varepsilon_r^* = 1$ systems (solid red line), the $\varepsilon_r^* = 2$ systems (solid blue line), and the predicted CNT line for the 3D homogeneous system (dotted orange line).

Plotted in Fig. 2.17 are the ratios of the critical cluster sizes obtained for the two

surface-induced nucleation systems to the homogeneous system as function of supersaturation. Again, according to CNT, these ratios are predicted to yield the f(θ) term (see Eq. 7) which should be constant, i.e., independent of supersaturation. However, the critical cluster size ratios predicted from the simulation show a clear dependence on supersaturation, similar to the barrier height ratios, especially toward high supersaturation conditions. For both the $\varepsilon_r^* = 1$ and $\varepsilon_r^* = 2$ systems, as the supersaturation decreases, these ratios start to oscillate around a value comparable to the $f(\theta)$ value computed using the contact angle measurement for a finite droplet. These oscillations which are not observed for the barrier height ratio can be due to the larger uncertainties introduced in the estimation of the critical cluster sizes. Also compared to the barrier height ratios, the critical cluster size ratios seem to converge to a slightly higher value for both systems. The constant offset observed between the theory and the simulation on the barrier height (but not on the critical cluster size) toward the low supersaturation conditions can be used to explain this difference.

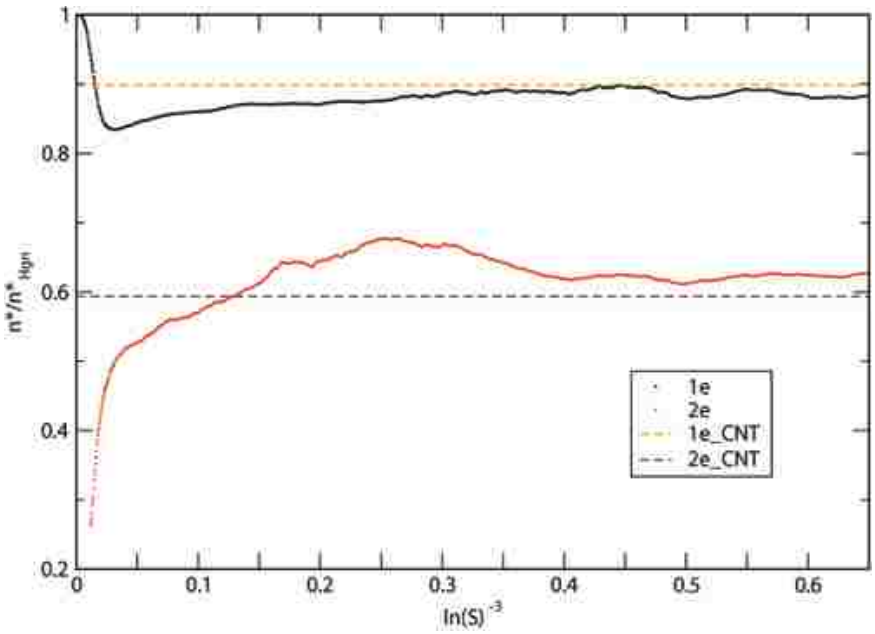

Figure 2.17: Ratio of the critical cluster sizes obtained for the $\varepsilon_r^* = 1$ (solid black line) and $\varepsilon_r^* = 2$ (solid red line) systems with the homogeneous 3D system at the same super saturation ratio along with the values predicted by CNT using the finite droplet contact angle calculation for the $\varepsilon_r^* = 1$ (dotted orange line) and $\varepsilon_r^* = 2$ (dotted blue line) systems.

Acknowledgments

The material and figures presented in this section have been reprinted from from J. Chem. Phys 139(23) p.234707, 2013, with the permission of AIP Publishing. Expanded citation may be found in reference 79.

Chapter 3

Improved Monte Carlo Sampling Methods

3.1 Energy Biased Aggregation-Volume-Bias-Monte-Carlo

As a critical bottleneck encountered by nucleation systems can be often characterized by a large free energy barrier, a common strategy to help speed up the sampling of these systems is to use an artificial bias to drive the system across the free energy barrier. In this regard, techniques such as umbrella sampling[17] or meta-dynamics[80,81] provide very effective ways to allow the system to transverse regions of the phase space with high free energy barriers. The umbrella sampling method is almost trivially implemented into a Monte Carlo simulation given that it only requires the modification of the acceptance rule. However, even with the free energy barrier completely removed via the use of umbrella sampling, the rate at which a simulation can converge the thermodynamic data is still tightly linked to the efficiency of the Monte Carlo moves used to sample phase space. This is especially true when dealing with particle transfer moves, which are another bottleneck for nucleation systems as these moves that are required for cluster growth/destruction can be also time-consuming.

The particle transfer problem is a commonly experienced issue[82,83] when simulating dense or confined systems using the grand canonical or similar ensembles such as the Gibbs ensembles.[84] Among the several techniques introduced in the past,[85–87] the Aggregation-Volume-Bias Monte Carlo (AVBMC) method[21,22] along with its sister algorithm the Unbonding-Bonding (UB) algorithm[88] have both proven very successful at transferring particles into

and out of a wide variety of systems.[89,52,79,90,91,26,92–95] While these moves greatly improve the efficiency of particle transfer, a few key issues still remain. For weakly associating systems such as Lennard-Jones, n-alkane, etc., the overall acceptance rate of the AVBMC algorithm typically reaches upward of 7 to 25% depending on the size of the molecule being studied.[5] This is more than sufficient to accurately sample these systems. However, for strongly associating systems such as water or ethanol the acceptance rate for the AVBMC and UB methods typically drops well below 1%.[5] It was presented in ref 5 that the introduction of a basic energy biasing technique can raise the acceptance rate to around 4 to 6% overall; however, it has since been found that as the number of particles increases the acceptance rate plummets even for this basic energy biasing technique (e.g., to below 1% for a TIP4P system). This problem has in part been linked to the choice of insertion sites by the original algorithms. That is, they use a uniform insertion scheme which simply allows the insertion to occur anywhere in the system despite the fact that the viable insertion sites are located toward the cluster surface. This uniform selection scheme has a glaring flaw that as the cluster grows in size the number of viable insertion sites declines. As a result the odds of randomly picking a valid insertion site via a uniform distribution also decline at a rapid rate (see Section 3.1.3 for more details). This explains why roughly 100 or more attempted moves are required to generate one acceptable configuration for a moderately dense system of strongly associating molecules. Currently the low acceptance rate of these moves is not a limiting factor when using pairwise models since simulations of reasonable length can feasibly be performed. However, recent efforts have been put forward to study much more complicated water models such as reactive and polarizable models for which these inefficiencies can begin to cause problems.[28,96,27] Given their increased computational cost, the low acceptance rate may begin to hinder a researchers ability to study these systems without consuming an unreasonably large amount of computer time. Because of this it is prudent to begin exploring ways to improve the sampling efficiency of current methods.

3.1.1 Background

The AVBMC algorithm[21,22] was originally designed to efficiently aggregate molecules using a set of moves that would transfer molecules in and out of another molecules bound region. This approach had a very similar idea to the UB algorithm developed independently by Wierzchowski and Kofke[88] but differed in how the exchange of particles between the bonding and nonbonding regions is handled. Both methods have very similar moves for the insertion of molecules into a bonded region using the following scheme:

1. Select a molecule from the simulation box to be moved into another molecules bounding region;
2. Select another molecule to act as the target for the previously selected molecule;
3. Choose a random location within the target molecules bound region and insert the previously selected molecule.

These two algorithms differ in the unbinding step. Specifically, the following procedure is used in AVBMC:

1. Pick a molecule that is currently within another molecules bound region;
2. Using this target molecule, select one of its neighbors;
3. Move this neighbor out of the selected molecules bound region to random location within the simulation box.

In UB:

1. Select a molecule that is currently within another molecules bound region;
2. Move this selected molecule to random location within the simulation box.

The UB algorithm selects a molecule for removal from the bonding region by randomly picking from the global pool of bound molecules, while the AVBMC instead first picks a random molecule from the bound molecules and then chooses one of its neighbors for removal.[21,22] While there seems to be only a subtle difference between the two algorithms, this minor difference can have a profound effect on the detailed balance condition. Since the UB

method uses a global approach for the removal step, it must also calculate the global probability of inserting a molecule into the bound region in order to satisfy the detailed balance condition. This includes enumerating all the degenerate insertion pathways that could be used to arrive at the same configuration and their associated probabilities. In contrast the AVBMC methods choice of deletion move allows it to instead rely on the principal of super detailed balance. Since for every possible insertion event there is a corresponding removal via the same path (or target), the degeneracies in the AVBMCs insertion move do not need to be accounted for. For both methods, the random selection of molecules and also the choice of an insertion position in the bound region were performed by using a simple uniform distribution. However, one downfall of a uniform selection scheme is that when these moves are performed on a cluster that is moderate in size (40+ molecules) the acceptance rate begins to decline proportional to the cluster size (see Section 3.1.3 for more details). This is naturally explained by the fact that molecules located in the interior of the cluster are saturated with the maximum or near the maximum number of neighbors possible. Subsequently if these molecules are selected as targets for insertion, the move will likely be rejected since there is no space to insert a molecule without some sort of heavy overlap with one of the existing neighbors. For the deletion move if one of these molecules is selected the move will likely be rejected on the grounds that this removal will result in a massive energy penalty. Particles located closer to an interface serve as better sites for both insertion into and removal from the cluster. However, as the cluster increases in size the number of interior particles will increasingly outnumber the particles near the surface. As a result the odds of selecting an interior molecule rapidly outgrow the odds of selecting a particle from the surface. This in turn causes the simulation to waste valuable computational cycles on moves that have close to no hope of ever being accepted. Therefore, it is desirable to avoid poor choices of insertion/removal targets and instead focus on targets that will likely result in a more energetically favorable configuration.

3.1.2 Derivation of a General Biasing Formula

In order to implement a biasing (or non-uniform) selection scheme, it is important to first derive the general detailed balance rule that will be used. Following the same logic process that was used for the AVBMC algorithm,[22] we can derive a generalized probability for both particle insertion and deletion. For the insertion move, there are two key steps that are generally performed. First, one must select a molecule that will act as the insertion site. Herein we will refer to this as the target molecule. This is typically done by randomly selecting any molecule in the system according to the probability denoted as P_{Target_In}. Second, once a target molecule has been selected, the coordinates of the new molecule must be generated such that the new molecule lies within the target molecules bound region. Typically this is done by first randomly inserting an atom of the new molecule within a predefined distance (this distance is usually defined to conform to a cluster criteria[22,97]). Once the first atom is inserted the remainder of the molecule is regrown using a method such as Configurational-Bias Monte Carlo (CBMC).[98] The probability of generating this new position within the target molecules bonded region will be denoted as P_{Bond}. Since AVBMC relies on the super detailed balance condition, one does not need to account for degeneracies. Therefore, the general probability of proposing an insertion move is given by

$$P_{Insert} = P_{Target-In} \cdot P_{Bond} \tag{3.1}$$

Next for the removal move first we randomly choose a molecule that has at least one neighbor to act as our target, this probability denoted as P_{Target_Out}. Second, once a target molecule has been chosen we need to select a molecule from the target molecules bound region to be removed, i.e., one of the N_{nei} neighbors of the target molecule according to an arbitrary probability denoted as P_{Select}. Therefore, the probability of proposing a move to

remove a particle from a given cluster is given by

$$P_{Remove} = P_{Target-Out} \cdot P_{Select} \tag{3.2}$$

In contrast, generalizing the UB algorithm requires a little more work since one must derive the global insertion probability in order to properly satisfy the detailed balance condition. To do this we first notice that since the base insertion move is identical to the AVBMC insertion we can argue that the probability of generating this new configuration from a given molecule i is identical, i.e.

$$P_i = P_{i,Target-In} \cdot P_{i,Bond} \tag{3.3}$$

However, we must account for degeneracies in the insertion step. If the newly inserted molecule rests within the bound regions of more than one molecule, there is now a situation where any one of these molecules could have been used as the target to generate the same identical configuration. For instance if we move a molecule within the bound regions of molecules i and j, we could have first selected molecule i as a target and oriented the new molecule with respect to i or we could have selected molecule j and generated the same configuration by orienting it with respect to j. Because of this there is a multifold degeneracy in the insertion probability that is proportional to the number of neighbors N_{nei} of the newly inserted molecule. Since the probability of using each pathway is independent of all other pathways, we can find the cumulative probability by summing over the probabilities of all possible pathways. For any given insertion move we have N_{nei} pathways, and the total probability to propose this insertion move is given by

$$P_{Insert} = \sum_{i=1}^{N_{Nei}} P_{i,Target-In} \cdot P_{i,Bond} \tag{3.4}$$

So for every attempted swap move one must determine all potential insertion paths and calculate the probability of each. For instance if the insertion of the first atom of the new molecule is done by randomly generating a point uniformly within a distance r_{max}, then one must search for all insertion sites within this distance. It can be shown if uniform distributions are used, then the original UB insertion probability[88] can be recovered

$$P_{Insert} = \sum_{i=1}^{N_{Nei}} \frac{1}{N} \cdot \frac{1}{V_{in}} = \frac{1}{N} \cdot \frac{N_{nei}}{V_{in}} \tag{3.5}$$

where N is the number of particles contained by the cluster, and V_{in} is the volume of the bonded region used for insertion, more specifically a spherical volume centred on the chosen target molecule with a radius of r_{max}. In comparison to the insertion move, the deletion probability is exceedingly simple. All that needs to be done is to simply select a molecule that is in the bound region of one other molecule according to the arbitrary distribution P_{Select}. Therefore, the removal probability is given by

$$P_{Remove} = P_{Select} \tag{3.6}$$

For both methods depending on the ensemble used there of course are other probability terms to consider such as the probability of selecting an unbound molecule from a given simulation box. However, we will primarily focus on the grand canonical ensemble though extending these ideas to other ensembles is fairly straightforward. With these generalized

probabilities it is now possible to bias the insertion and removal moves with any arbitrary functional form.

3.1.3 Choice of Functional Form

As mentioned in the Introduction, it was found after analyzing the original AVBMC algorithm for mid to large sized TIP4P water clusters (with 40+ water molecules) that the majority of molecules that were added to the cluster were added close to the interface (see Fig. 3.1). Therefore, it is theorized that the overall efficiency of the simulation can be greatly improved if the choice of the insertion site is biased toward the surface of the cluster where there is sufficient empty space. This of course introduces the problem of accurately identifying a molecule that is sufficiently close to the interface. One likely variable that is able to successfully pick out these molecules and will be viable from system to system is the interaction energy of a molecule.

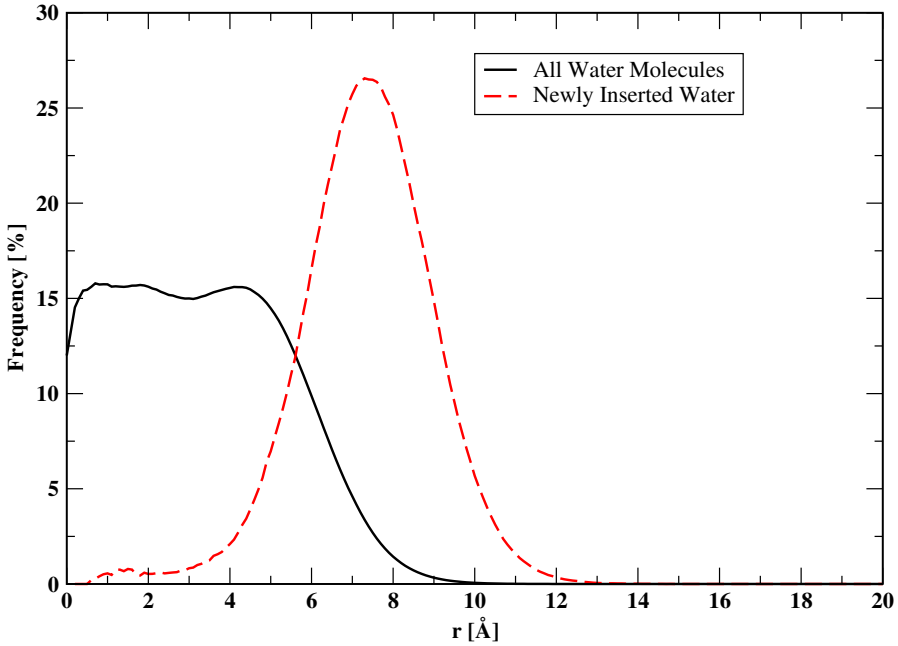

Figure 3.1: Normalized probability distribution of water oxygen atoms of all molecules in a water cluster containing 45 molecules (black solid line) compared to that of accepted oxygen atom positions for the particle insertion as a function of the distance from the center of mass (dashed red line).

A further analysis of the AVBMC algorithm for a TIP4P water system gives credence to this idea as it was found that the molecules which successfully served as a target for insertion tend to have higher energies (see Fig. 3.2). Likewise as shown in previous studies[5] the removal move should also benefit from an energy-based bias to preferentially select particles with the highest energies to be removed from the cluster as these molecules will have the lowest energy penalty for their removal and subsequently will have the highest value for the Boltzmann factor in the acceptance probability. Therefore, we have chosen a scheme that attempts to bias the insertion and removal step based on the molecular energy in order to select candidates that are expected to result in successful insertion and removal moves.

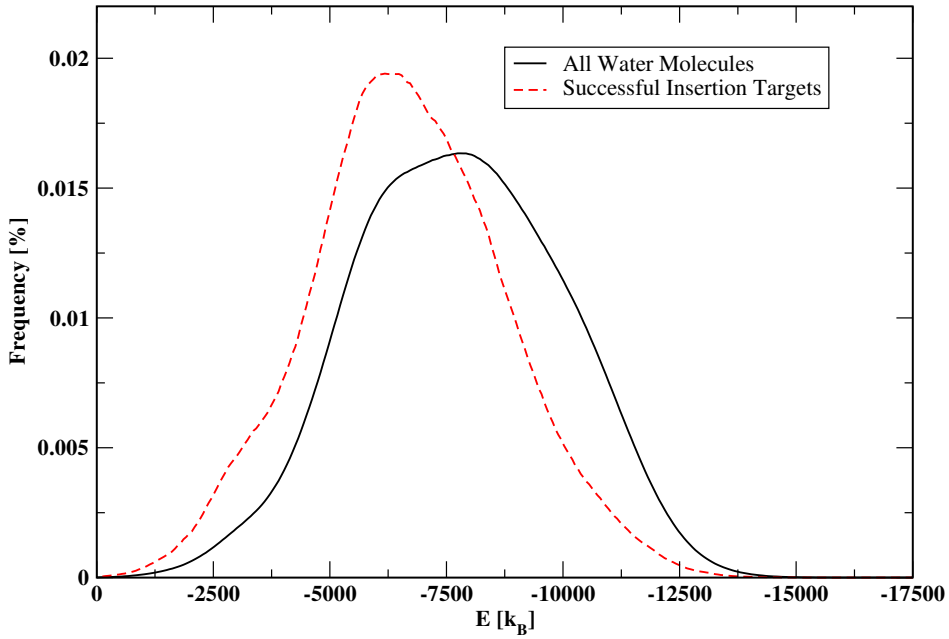

Figure 3.2: Normalized distribution of the molecular energy over all water molecules at a cluster size of 45 (solid black line) compared to that obtained only for the molecules that successfully acted as a target for the particle insertion (dashed red line) in the original AVBMC algorithm.

For simplicity, the $P_{Target-in}$ term in both algorithms was expressed using a similar function form to the one mentioned in ref 37 5

$$P_{i,Target-In} = \frac{e^{\alpha E_i}}{\sum_j e^{\alpha E_j}} \tag{3.7}$$

where E_i is the total interaction energy of molecule i with all other molecules. It is important to note that the exponential constant in this case is not chosen to be the usual $\beta = (1/k_B T)$ constant that is common in the Boltzmann weight. Instead is given as an input parameter that is chosen by the user. This is done because as shown by Fig. 3.2 while the system prefers molecules with higher energy for insertion, there are still some with moderately low energy that can serve as viable insertion targets. Therefore, choosing the exponential constant to be equal to would bias the system too strongly to the most loosely bound molecules. So to avoid this overbiasing, α is chosen on a system to system basis. In contrast for the removal step selecting a molecule i according to the exact Boltzmann factor is desirable since this directly corresponds to the change in the energy of the system. Thus, the P_{Select} term is chosen to have the functional form

$$UB : P_{i,Select} = \frac{e^{\beta E_i}}{\sum_j^N e^{\beta E_j}} \tag{3.8}$$

$$AVBMC : P_{i,Select} = \frac{e^{\beta E_i}}{\sum_j^{N_{i,nei}} e^{\beta E_j}} \tag{3.9}$$

For UB this includes all the molecules that can be removed from the cluster while for AVBMC this only includes the neighboring molecules of the target molecule i. This ensures that the highest energy molecules are chosen as removal targets and subsequently have the highest probability of being successfully removed from a cluster. In the UB style algorithm

this function selects from all the N bound molecules within the system, while for the AVBMC style this is only done with respect to the neighbors of the target molecule. As mentioned in the AVBMC style algorithm there is the additional $P_{Target-Out}$ term to account for that is not found in the UB formulation. The AVBMC style algorithm provides an additional challenge in picking the highest energy molecules for removal since they must be selected indirectly. This was not done in the basic energy biasing scheme mentioned in ref 37 5 which was subject to the same flaw as the original AVBMC in that if a target molecule is chosen from the center (or a highly dense portion) of the cluster, all of its neighbors will likely be low in energy and the move will be rejected due to the energetic factor. To address this difficulty a variable ϵ_i is introduced which corresponds to the energy value of the highest energy neighbor of molecule i to bias the selection of the target molecule as follows

$$P_{i,Target-Out} = \frac{e^{\beta\epsilon_i}}{\sum_j e^{\beta\epsilon}} \tag{3.10}$$

This choice of function ensures that a target molecule with a high energy neighbor will be chosen such that the subsequent molecule selected from its neighbor(s) according to the probability P_{Select} prescribed by eq 3.9 will have a high chance of being removed from the cluster.

The general acceptance rule for these methods can be constructed using the detail balance condition given by

$$P^{Acc}_{A\to B} \cdot P_{A\to B} \cdot P_A = P^{Acc}_{B\to A} \cdot P_{B\to A} \cdot P_B \tag{3.11}$$

where $P^{Acc}_{A\to B}$ is the probability of accepting the transition from state A to state B, $P_{A\to B}$ is the probability of proposing the transition from A to B, P_A is the probability of being in

state A, and all other terms are the equivalent probabilities for the reverse transition. When the appropriate terms are substituted into eq 3.11, the general acceptance rules in the grand canonical ensemble are given by

$$P_{Remove}^{Acc} = \frac{P_{Insert}}{P_{Remove}} e^{-\beta\mu - \beta\Delta E} \tag{3.12}$$

$$P_{Insert}^{Acc} = \frac{P_{Remove}}{P_{Insert}} e^{\beta\mu - \beta\Delta E} \tag{3.13}$$

where μ is the chemical potential of the ideal gas phase reservoir, and ΔE is the energy difference going from the old to the new state. By substituting all the terms previously listed for each algorithm, the acceptance rules for the energy biased AVBMC and UB algorithms are found.

AVBMC Style:

$$P_{Remove}^{Acc} = \frac{\frac{e^{\alpha E_{i,new}}}{\sum_j^{N-1} e^{\alpha E_{j,new}}} \cdot \frac{1}{V_{in}}}{\frac{e^{\alpha \epsilon_{i,old}}}{\sum_j^{N} e^{\alpha \epsilon_{j,old}}} \frac{e^{\alpha E_{i',old}}}{\sum_{j'}^{N_{i,nei}} e^{\alpha E_{j',old}}}} \cdot e^{-\beta\mu - \beta\Delta E} \tag{3.14}$$

$$P_{Insert}^{Acc} = \frac{\frac{e^{\alpha \epsilon_{i,new}}}{\sum_j^{N+1} e^{-\alpha \epsilon_{j,new}}} \frac{e^{\alpha E_{i',new}}}{\sum_{j'}^{N_{i,nei}+1} e^{\alpha E_{j',new}}}}{\frac{e^{\alpha E_{i,old}}}{\sum_j^{N} e^{\alpha E_{j,old}}} \cdot \frac{1}{V_{in}}} \cdot e^{\beta\mu - \beta\Delta E} \tag{3.15}$$

UB Style:

$$P_{Remove}^{Acc} = \frac{\frac{\sum_i^{N_{nei}} e^{\alpha E_{i,new}}}{\sum_j^{N-1} e^{\alpha E_{j,new}}} \cdot \frac{1}{V_{in}}}{\frac{e^{\beta E_{i',old}}}{\sum_j^{N} e^{\beta E_{j',old}}}} \cdot e^{-\beta\mu - \beta\Delta E} \tag{3.16}$$

$$P_{Insert}^{Acc} = \frac{\frac{e^{\beta E_{i',new}}}{\sum_j^{N+1} e^{\beta E_{j',new}}}}{\frac{\sum_i^{N_{nei}} e^{\alpha E_{i,old}}}{\sum_j^{N} e^{\alpha E_{j,old}}} \cdot \frac{1}{V_{in}}} \cdot e^{-\beta\mu - \beta\Delta E} \tag{3.17}$$

All the energy-related variables listed here can be easily tabulated from the normal energy calculations that are carried out during the course of a simulation. The amount of stored data required by each method is O(N), i.e., linear with respect to the system size and with a proper neighbor list the tabulation of this data increases the computational overhead by a trivial amount. All one needs to do is simply update these tables upon the acceptance of any given Monte Carlo move. In addition both schemes can be coupled with CBMC in conjunction with multiple insertion.[98–104] In particular, one can generate multiple proposed trial moves to enhance the chance of having one acceptable trial configuration and then use the so-called Rosenbluth weight[104] for trial selection. For this particular application the Rosenbluth weight is given by

$$W = \sum_{m=1}^{M} w_m \tag{3.18}$$

$$w_m = \frac{e^{-\beta E_m}}{P_{m,gen}} \tag{3.19}$$

$P_{m,Gen}$ is the probability of generating configuration m, w_m is the Rosenbluth weight of trial m, and E_m is the energy of trial m. For insertion $P_{m,Gen} = P_{m,Insert}$, and for removal $P_{m,Gen} = P_{m,Remove}$. This follows the same general scheme as CBMC that has previously been used in AVBMC.[26,98] By selecting one of the trials according to the probability $P_m = (w_m/W)$ and computing the Rosebluth weight for the old configuration (W_{old}), one obtains the following acceptance rules:

$$P_{Insert}^{Acc} = \frac{W_{new}}{W_{old}} \cdot e^{\beta\mu} \tag{3.20}$$

$$P_{Remove}^{Acc} = \frac{W_{new}}{W_{old}} \cdot e^{-\beta\mu} \tag{3.21}$$

It is important to note that for the AVBMC formulation the same target molecule for all trial configurations in a given insertion/removal move was used. This can ensure that the super detail balance condition is not broken by the introduction of the Rosenbluth scheme. In contrast the UB formulation can easily use a different target molecule for each insertion trial since it is not dependent on the super detail balance condition. For small molecules each Rosenbluth trial can be performed by fully growing the molecule (i.e., attempting multiple insertion configurations) although for large chain molecules multiple insertion of the first atom combined with a CBMC regrowth of the rest molecule is more commonly used.

3.1.4 Simulation Details

All nucleation simulations were carried out using the grand canonical ensemble where a cluster is physically separated from but thermodynamically coupled to an ideal gas-phase reservoir whose chemical potential can be specified by a number density. To compare the accuracy and efficiency of the new scheme to the original algorithms, a cluster simulation of

a Lennard-Jones (LJ) and TIP4P Water system was performed for each of the base energy biasing algorithms (denoted as EBias). For the TIP4P system the previously mentioned Rosenbluth scheme was also tested. The LJ simulation was carried out at a reduced temperature of 0.8 and a gas phase number density of $1.1 \cdot 10^{-2}\sigma^{-3}$, while the TIP4P simulations were carried out at 300 K and a gas phase number density of $6 \cdot 10^{-6}$ Å^{-3}. For each system the configurational space was sampled using traditional translational moves in addition to particle swaps. Rotational moves were also used for the water system. These different types of moves were performed with equal frequency (e.g., 1/3 for each of the three different move types for the TIP4P system) except where noted explicitly. The maximum displacements for both the translational and rotational (if applicable) moves were adjusted so that 50% of the moves were accepted. For the LJ system clusters up to 200 particles in size were simulated, while for the TIP4P system clusters containing up to 70 water molecules were simulated. For all simulations the Umbrella Sampling method[17] was used to sample the nucleation free energy of each system. The biasing potential was converged using the method outlined in previous papers via an iterative procedure until uniform sampling is achieved.[14] For all simulations a Stilinger cluster criteria[23] was used. For the LJ simulations a bonding distance of 1.5σ was used, while for the TIP4P simulations a criteria of 4.0 Å was used based on previously published data.[97,105] Any move which would leave a member of the cluster unbound is automatically rejected. To test the rate of convergence a large simulation ($4.5 \cdot 10^{10}$ MC moves for the production run) using the original AVBMC algorithm was performed to calculate the free energies of the TIP4P water clusters with high precision. These results were used as the reference to estimate the rate of convergence of each biasing method. It should be noted that the use of AVBMC or UB for particle transfers combined with the Stillinger cluster criteria avoids the need to define an arbitrary system volume or a simulation cell for the cluster.[14]

3.1.5 Results

As shown in Fig. 3.3 all the methods mentioned here were able to correctly reproduce the nucleation free energy profile for the Lennard-Jones system. For both AVBMC and UB, the energy biasing moves showed a massive improvement in the acceptance rate over their uniform counterparts (see Fig. 3.4).

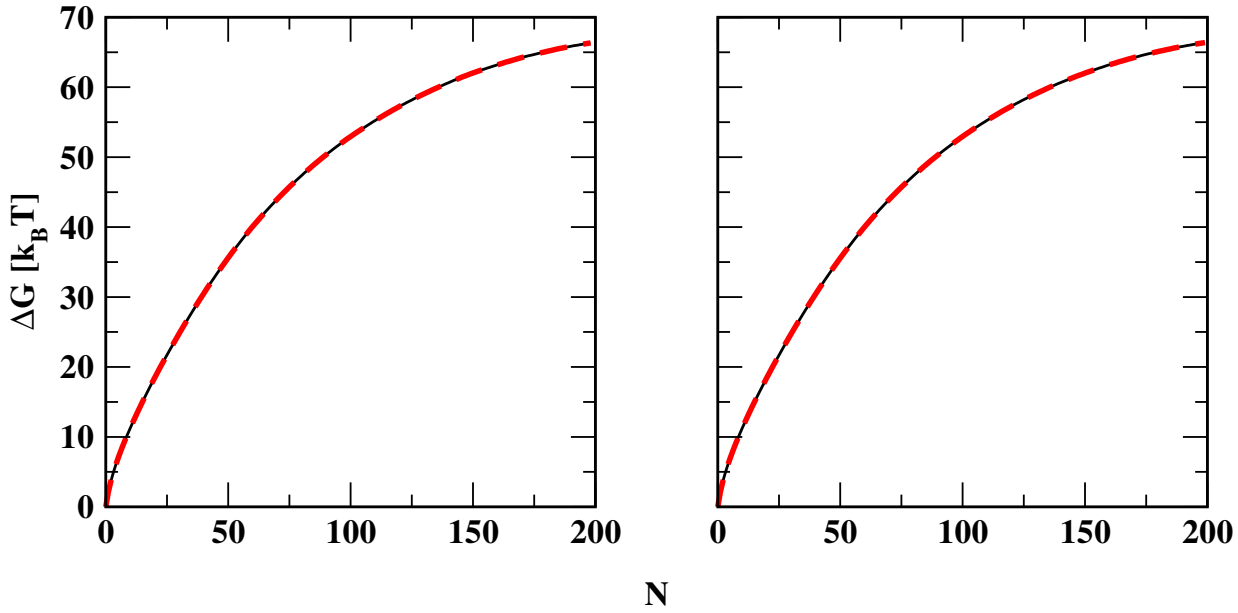

Figure 3.3: Nucleation free energy results obtained for the Lennard-Jones system as a function of the cluster size using the original uniform selection scheme (black) and the energy-biased scheme (red). The results obtained using the AVBMC-based algorithms are shown on the left, while those obtained using the UB-based methods are shown on the right.

In particular, the EBias scheme was able to reach an acceptance rate well above 30% even as the clusters reached larger sizes in comparison to the 10 to 20% range that was typical of the uniform methods. It was also evident from Fig. 3.4) that the UB method produced a slightly higher acceptance rate than the AVBMC method for this particular system no matter whether the EBias scheme is used or not. When analyzing how the remaining 40 to 60% of the attempted EBias moves were rejected, it was found that on average roughly 30% were due to energetic reasons, i.e., largely overlaps with the target molecule or one of its neighbors. It was also found that roughly 5 to 10% of the attempted swap moves

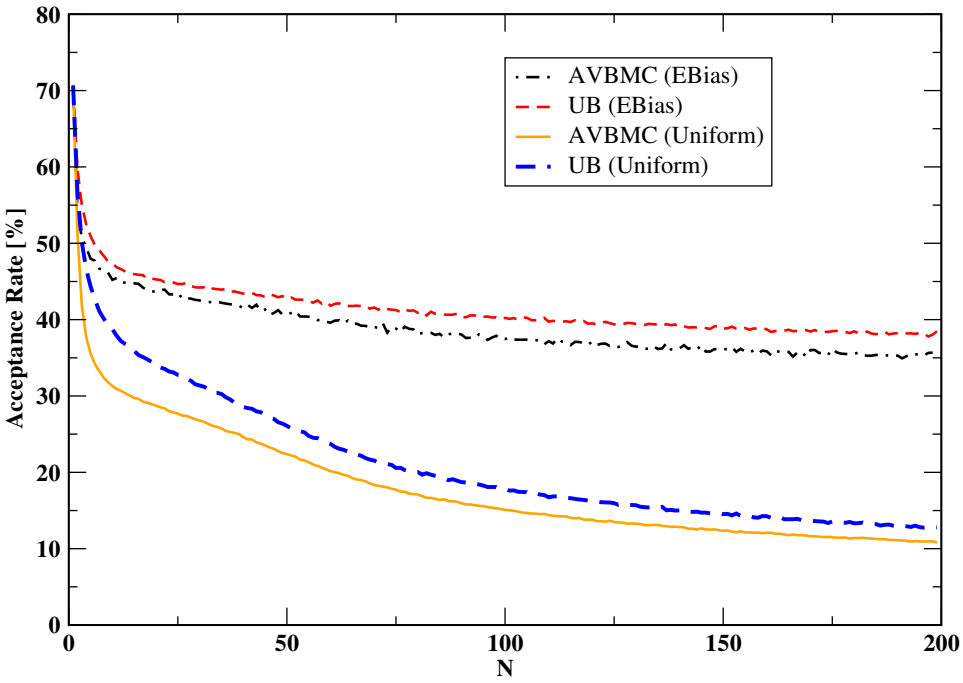

Figure 3.4: Acceptance rate of the particle transfer move as a function of the cluster size in a Lennard-Jones system obtained using the AVBMC-EBias (dashed-dotted black line), UB-EBias (dashed red line), AVBMC (solid orange line), and UB (dashed blue line) methods.

were rejected due to the cluster criteria, namely, in the removal step after the deletion of a particle the molecules remaining in the system no longer belong to the same cluster. These numbers would typically fluctuate depending on the cluster size. For smaller clusters the rejection rate due to the cluster criteria was typically higher, while the rejection rate due to energetic factors would be higher for larger cluster sizes. Since the EBias schemes brought the acceptance rate to nearly 50%, it was not deemed necessary to implement the Rosenbluth sampling scheme for this system.

Similarly the method was able to correctly reproduce the free energy curve of the TIP4P system (see Fig. 3.5). While not as dramatic as the Lennard-Jones system, for this water system the EBias scheme showed a modest improvement over the uniform sampling scheme (see Fig. 3.6). As the cluster grew beyond 30 water molecules in size the uniform scheme showed a sharp decline in the acceptance rate, falling below 1%. In contrast the EBias

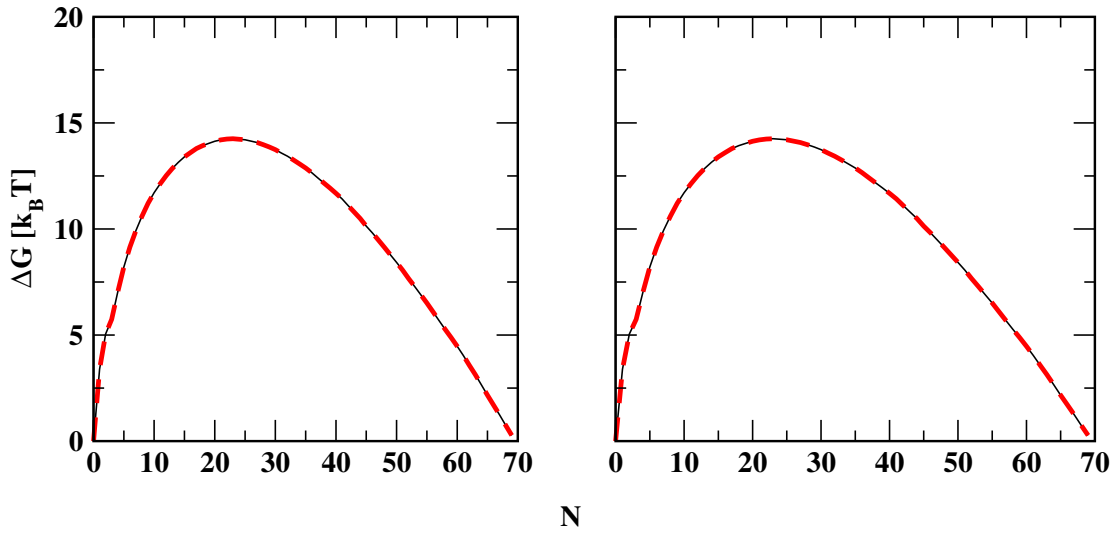

Figure 3.5: Nucleation free energy results obtained for the TIP4P water system as a function of the cluster size using the original uniform selection scheme (black) and the energy-biased scheme (red). The results obtained using the AVBMC-based algorithms are shown on the left, while those obtained using the UB-based methods are shown on the right.

scheme maintained an acceptance rate well above 5% even as the cluster grew larger in size. The overall acceptance rates were around 6.5%. For comparison purposes an equivalent run was performed using the EBias scheme implemented in ref 37 5, and it was found that the overall acceptance rate was approximately 4.2% for the same conditions.

It became apparent that while the smarter selection of target molecules improved the overall acceptance rate, there was still a very large number of moves that were rejected compared to the LJ case because the newly inserted water molecule was not aligned so that it could properly hydrogen bond with the surrounding cluster members. The Rosenbluth scheme described in Section 3.1.3 can be used to deal with this issue. In order to provide a fair one to one comparison with the combined Rosenbluth/EBias method (called EBias-Rosen), the same Rosenbluth sampling method was also implemented with the uniform insertion scheme (denoted as Uniform-Rosen).

As shown in Fig. 3.7, the EBias-Rosen method resulted in a remarkable improvement in the acceptance rate far exceeding that achieved by Uniform-Rosen. It was also found

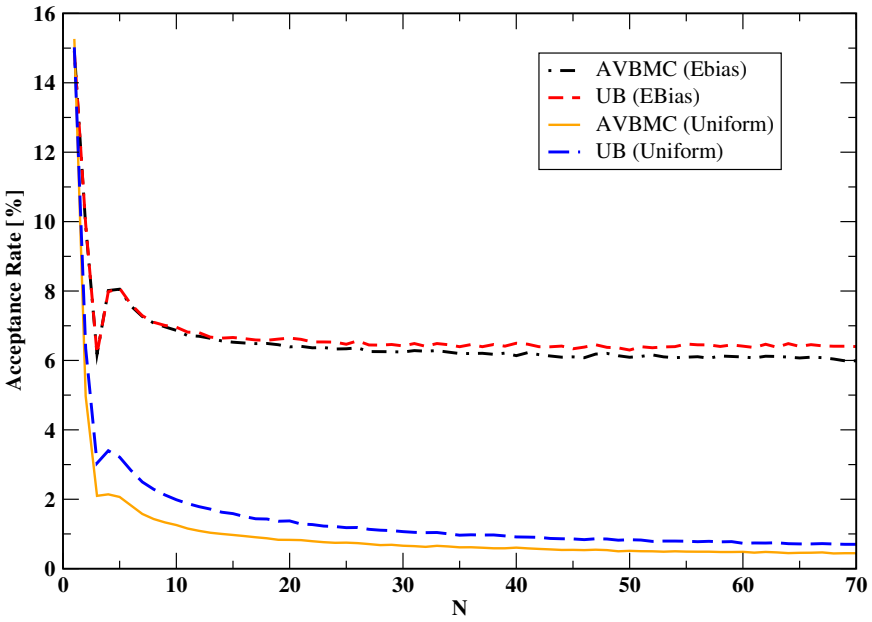

Figure 3.6: Acceptance rate of the particle transfer move as a function of the cluster size for the TIP4P water system obtained using the AVBMC-EBias (dashed-dotted black line), UB-EBias (dashed red line), AVBMC (solid orange line), and UB (dashed blue line) methods.

that EBias-Rosens overall acceptance rate scales very well with the number of trials (see Fig. 3.8). For both the EBias-enhanced AVBMC and UB algorithms, the introduction of the Rosenbluth sampling greatly improved the acceptance rate from 5 to 7% to anywhere between 20 and 35% depending on the number of trials used. It was found that scaling beyond 32 Rosenbluth trials was not desirable given that the acceptance rate only increased by 2-5% from 32 to 64 trials, while simultaneously the computational overhead nearly doubled.

In stark contrast to the EBias-Rosen scheme the Uniform-Rosen method displayed very minor improvements on the absolute magnitude of the acceptance rate. Even for an exceedingly large number of trials the Uniform-Rosen methods did not reach 10% in the acceptance rate. The acceptance rate of the Uniform-Rosen methods much like the standard Uniform methods showed a massive drop toward larger clusters. Thus, one can conclude that the Rosenbluth scheme alone is not sufficient to improve the acceptance rate as the system size grows.

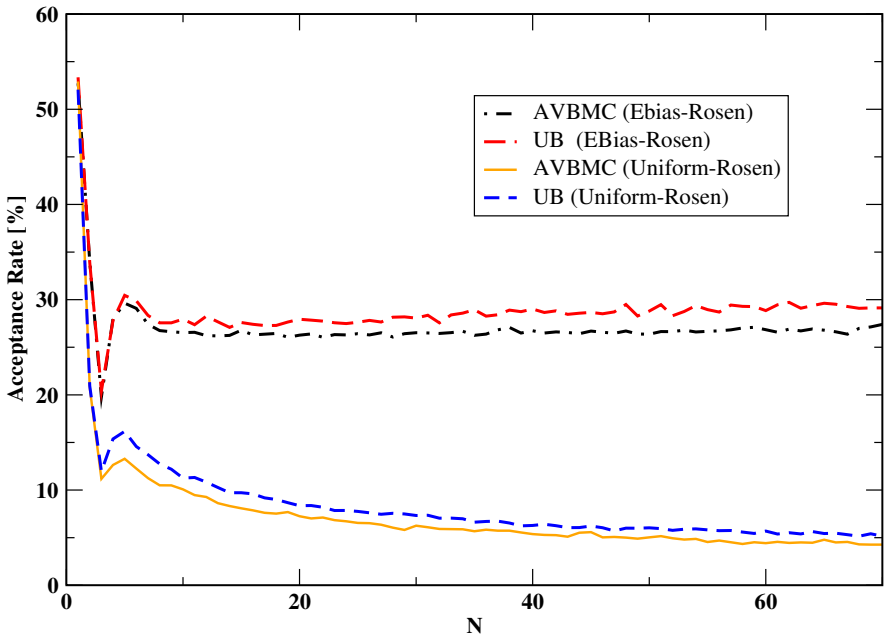

Figure 3.7: Acceptance rate of the particle transfer move as a function of the cluster size for the TIP4P water system obtained using the Rosenbluth coupled version of each algorithm with 32 trials, including AVBMC-EBias (dotted blue line), UB-EBias (dashed-dotted purple line), AVBMC (solid black line), and UB (dashed red line).

An increase in the acceptance rate does not necessarily guarantee that the rate of convergence of the systems thermodynamic properties is improved. To evaluate the convergence rate, two sets of simulations were performed, and these two differ in terms of the input biasing potentials used by umbrella sampling. One starts with unconverged values and the other with well-converged ones. The number of Rosenbluth trials used for the Ebias-Rosen methods were set to 32. For the case using initially unconverged biasing potentials, the starting input biases (see Fig. 3.9) were chosen such that each cluster would have a chance of being sampled so the free energies of all cluster sizes can be evaluated directly from their sampling frequencies. This was done to ensure that the speed-up was due to the method and not to outside factors such as the choice of the extrapolation method.[106]

As shown in Fig. 3.10, the original uniform methods were found to take 4 iterations to accurately converge the free energy curve within a reasonable statistical error, and before that

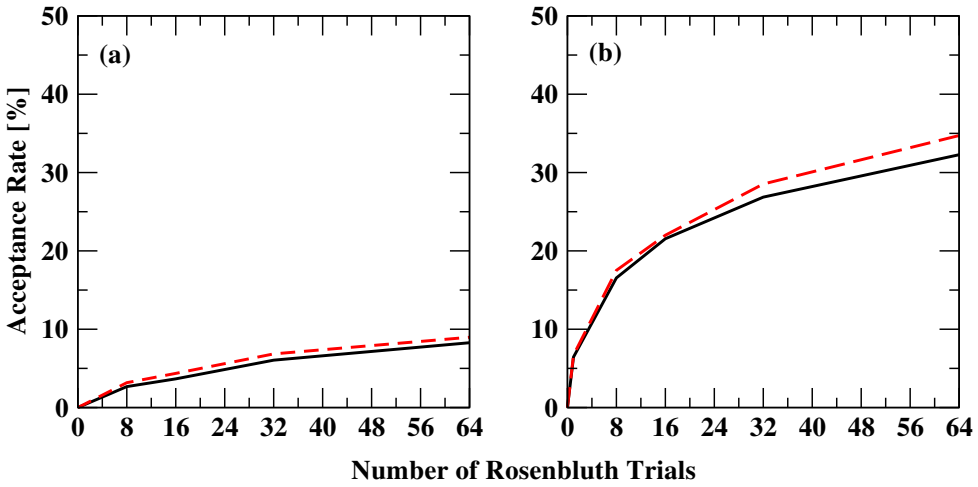

Figure 3.8: Acceptance rate of the particle transfer move as a function of the number of Rosenbluth trials using AVBMC (black lines) and UB (red dashed lines). The results obtained using the standard algorithms are shown on the left, while those obtained using the Ebias enhanced methods are shown on the right.

the statistical error showed a fairly linear decrease with respect to the number of iterations. The EBias scheme needs 3 iterations, and for each step the relative error was smaller than the uniform algorithm. Lastly the EBias-Rosen schemes were able to reach the convergence within 2 iterations. For the second set of simulations using the well-converged biasing potentials, the free energies and also the errors were analyzed every $2 \cdot 10^6$ MC moves. As shown in Fig. 3.11, the EBias and EBias-Rosen methods showed a significant increase in computational efficiency, up to an order of magnitude (measured by the number of moves needed to obtain high-quality free energies) compared to the original uniform selection schemes. It is evident that the increase in the acceptance rate due to these nonuniform algorithms does indeed result in a faster sampling of the conformational space and correspondingly a faster convergence of the thermodynamic properties of the system.

When comparing the overall differences between UB and AVBMC, it was found that on average the various UB methods showed a higher acceptance rate by 1 to 4% compared to their AVBMC counterparts. This difference can largely be attributed to the two-step proposal scheme in the AVBMC algorithm. A two-step selection scheme is inherently more prone

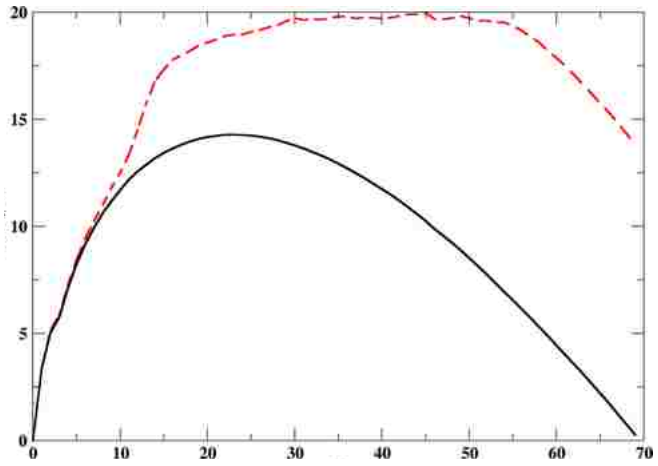

Figure 3.9: Two sets of input biasing potentials used in the convergence evaluation. The solid black curve is the fully converged profile obtained from a very long simulation run, while the dashed red curve is the unconverged one used as the starting point for the iterations.

to selecting a bad candidate for removal because either a poor choice of a target molecule or a poor choice of the subsequent neighbor selection can give this result. Conceptually if there is a 99% chance of generating a good target and a 99% chance of selecting a good neighbor, the cumulative probability of choosing a proper candidate is around 98% or that there is now a 2% instead of 1% chance of selecting a poor candidate. In addition in the uniform case, the two-step algorithm was more likely be rejected due to the cluster criteria. It was found that for the water system UB had a 1.0% rejection rate due to the cluster criteria compared to 1.7% for AVBMC. The two-step selection scheme in the uniform case is statistically more likely to select particles that have a higher number of neighbors (e.g., a particle with two neighbors has two pathways that it can be selected by while a particle with one neighbor only has one pathway to be selected by), whereas in UB all molecules have the same probability of being selected for deletion. However, for the EBias enhanced versions the cluster criteria was no longer an issue. Both algorithms had a similar criteria-induced rejection rate of around 0.05% for the water system.

While UB consistently outperforms AVBMC, the implementation of biased UB algo-

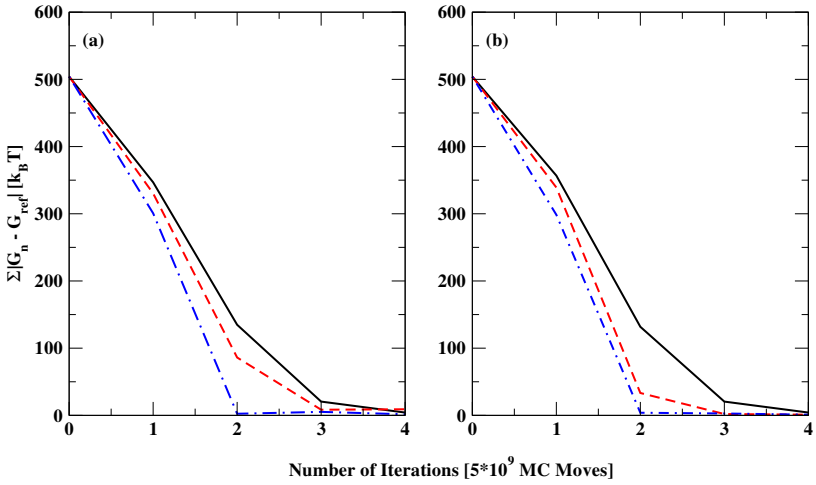

Figure 3.10: Sum of the absolute error of the nucleation free energy over all cluster sizes as a function of the number of simulation iterations for each of the different methods, including the standard method (solid black line), the Ebias method (dashed red line), and the Ebias-Rosen method (dashed-dotted blue line). The results obtained using the AVBMC-based algorithms are shown on the left, while those obtained using the UB-based methods are shown on the right.

rithms is more complicated since it requires additional efforts to take into account the degeneracy-related terms which arise when different pathways can lead to the same configurational change during a particle swap move. In the schemes presented here the increase in algorithmic complexity was reasonable; however, if any sort of orientational bias is introduced (e.g., the insertion of the first bead is biased using an arbitrary orientation angle with respect to the target molecule), the orientation probability must be calculated with respect to all degenerate insertion sites for the UB formalism, whereas in AVBMC this needs to be done only for the chosen insertion target. It should be also pointed out that the optimal value of (a parameter used for the nonuniform selection of the target molecule for insertion) varied significantly between these two formalisms. As shown in Fig. 3.12, for the TIP4P water system the acceptance rate peaked at an value between 0.05β and 0.1β for AVBMC vs a value between 0.1β and 0.15β for UB. Beyond 0.15β both algorithms begin to see a decline in the acceptance rate. In this region, the system begins to suffer from overbiasing or that the probability overwhelmingly favors the highest energy molecules. However, as

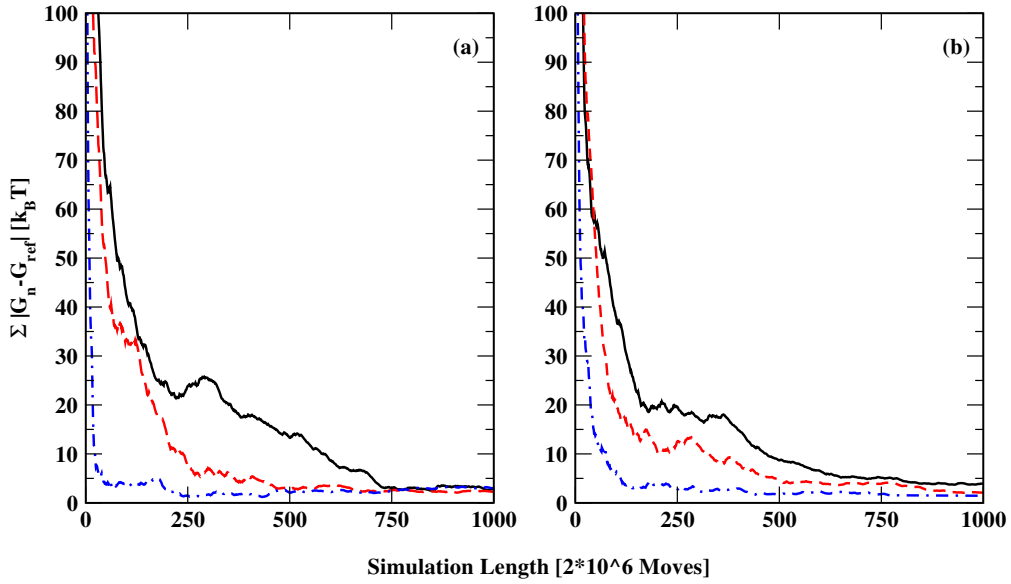

Figure 3.11: Sum of the absolute error of the nucleation free energy over all cluster sizes as a function of the number of Monte Carlo moves using the standard method (solid black line), the Ebias method (dashed red line), and the Ebias-Rosen method (dashed-dotted blue line). The results obtained using the AVBMC-based algorithms are shown on the left, while those obtained using the UB-based methods are shown on the right.

previously shown in Fig. 3.2 this neglects a large population of molecules which can still serve as viable insertion sites. While this affects both algorithms the UB variation of the EBias-Rosen displayed a lower sensitivity to the choice of due to its globalized sampling capabilities. Even though the probability of selecting a moderately high energy molecule was low (often below 5% at an value of 0.3β) over the course of 32 Rosenbluth trials the odds of generating at least one trial using these molecules were high enough that they could be included in the Rosenbluth weight. In contrast the AVBMC due to its restriction to a single target site for all Rosenbluth trials is significantly more sensitive to the choice of the target molecule given that a poor target selection will doom the majority of the Rosenbluth trials. The results reported above were obtained using an optimal value corresponding to each of these two types of methods (i.e., 0.1β for AVBMC and 0.15β for UB). For the TIP4P system, additional simulations were performed by changing the frequency of the swap moves from 1/3 to either 1/6 or 1/2 while splitting the rest of the moves evenly between translations and

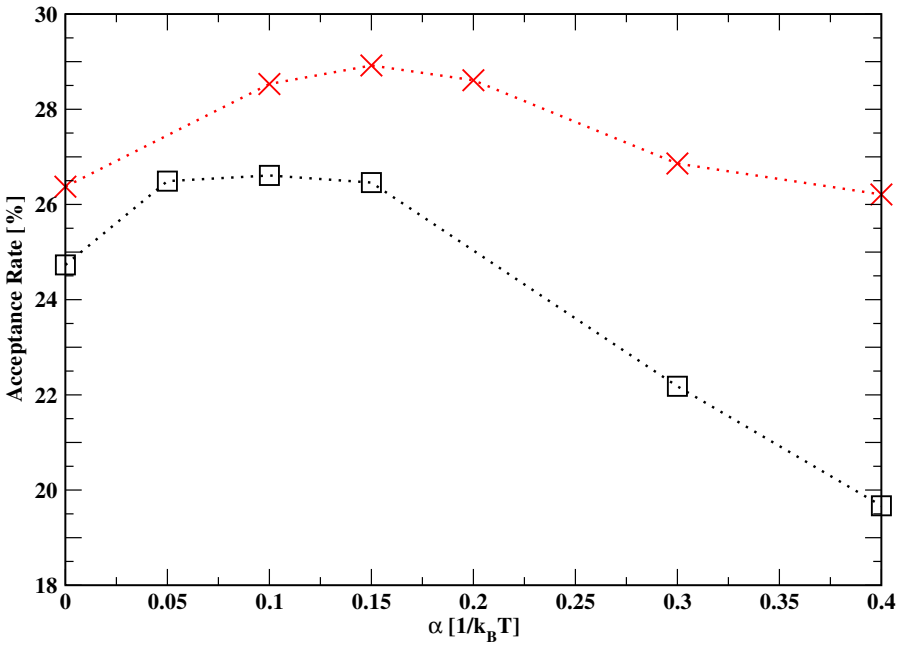

Figure 3.12: Acceptance rate of the particle transfer move as a function of the parameter for the Ebias-Rosen AVBMC (black squares) and the Ebias-Rosen UB method (red s).

rotations to evaluate the influence of this parameter on the rate of convergence. It was found that the use of a lower frequency of the swap moves (i.e., 1/6) led to a slower convergence, while the use of a higher frequency of the swap moves (i.e., 1/2) produced mixed results.

As noted, the extension of these methods into other ensembles such as the Canonical or Gibbs ensembles can be achieved by following the formulations in the original AVBMC and UB papers.[21,22,88] These methods are expected to work for bulk phase systems given that a few additional optimizations are included for bulk systems. For instance since a bulk system has many more interactions compared to a cluster system, using the full energy calculation to compute the Rosenbluth weight could prove to be far too costly. However, this can be solved by weighting according to a local subset of the interactions instead of all the interactions in the system, as what has been done in the dual-cutoff CBMC.[107] While the Rosenbluth sampling assists in overcoming rejections due to the poor choice of an insertion orientation, a more direct approach would be to bias the generation of the trial orientation to the expected

distribution, similar to the idea that was proposed recently for the conformational sampling of chain molecules.[108]

Acknowledgments

Materials and Figures found in this section have been reprinted, with permission, from J. Chem. Theory Comput, 11(9), p.4023-4032, 2015. Copyright 2015 American Chemical Society. (Expanded citation may be found in Reference 109.)

3.2 Self Adaptive Umbrella Sampling

3.2.1 Background and Method

In prior studies, methods such as the AVUS-HR[25,96,90,14,26] method were used to calculate the free energy of nucleation. This method incorporates the Umbrella Sampling method[17] in order to flatten the nucleation free energy landscape so that the system can cross the barrier. Unlike common Umbrella Sampling used in Molecular Dynamic where the Umbrella Sampling bias is given by a function such as a harmonic potential, the Umbrella Sampling potential is represented as a numerical potential where each bin is assigned a free energy value that is used in the detailed balance condition.

$$P_{accept} = min\left(1, \frac{\alpha_{old}}{\alpha_{new}} \cdot e^{-\frac{\Delta E}{k_B T} + \Omega_{new} - \Omega_{old}}\right) \tag{3.22}$$

Where Ω is the umbrella sampling bias for a given bin. The free energy of the system for any given biasing potential can be computed using the equation

$$\Delta G_i - \Delta G_j = \Omega_i - \Omega_j - \ln\left(\frac{p_i^*}{p_j^*}\right) \tag{3.23}$$

Where p^* is the probability calculated from the biased distribution. The goal of this method is to obtain the free energy of a given landscape by modifying the potential such that

the entire system can be sampled. One can naturally see from Eq. 3.23 that if the numerical bias values Ω for bins i and j are equal to the free energy of the bin, the probability of sampling both bins is equal. Thus it is ideal to obtain a uniform biased distribution since this directly relates the free energy to the Umbrella Sampling bias and in turn yields the free energy. This can be done by iteratively by calculating the free energy via equation 3.23 and then using the new estimate for the free energy as the input potential for the next biasing run. This has been shown to correctly converge to the correct free energy.[14,26]

This scheme has produced excellent results when biasing along a small number of reaction coordinates. However, when applying it to a system where it is desirable to bias along multiple reaction coordinates several issues arise that are typically minor in the single reaction coordinate case. Initially when the free energy landscape is not known, the the initial simulations will tend to fall into the natural minima that exist in the landscapes. As more and more iterations are performed the minima are gradually flatted. However in the process of flattening these minima, new minima can be created in the biased distribution during the next simulation run due to any number of statistical and energetic factors such as poor sampling of bins neighboring the first minima, neighboring minima that were not accessible during the first simulation, etc. As a result the simulation will tend to flow toward the newly created minima and get stuck again. As a rule of thumb, the larger the phase space the more times the potential will have to be iterated in order to finally flatten the whole space. However in some cases this can be upwards of 30+ iterations which can prove to be tedious if done manually. In addition it is desirable to not sample each biased distribution longer than is needed to adjust the potential correctly since this will simply result in wasted computational time. In order to solve these issues a self iterative variation of the this algorithm was developed to reduce the need for user intervention and to increase computational efficiency.

1. Start the simulation using an initial biasing potential.

2. After a fixed number of cycles (From here on called segments) halt the simulation. If multiple Markov Chains are used, wait until each chain has reached the break point.
3. Collect the histogram data from each of the Markov Chains and combine them.
4. Using the combined data adjust the potential.
5. Store the current numerical potential and histogram for later use.
6. Send the new potential out to each Markov Chain and resume the simulation.
7. Repeat this process for however many segments is required.
8. Using a method such as the Weighted-Histogram-Analysis Method[110] or any equivalent method, pool the data from each segment into a single free energy landscape.

With this algorithm, there remains one important detail that is undetermined and that is the procedure of how to iterate the potential. This is done by modifying the previous scheme to account for several cases that commonly occur during the iterative process. For each segment, the bin that was sampled the most is first identified. This will serve as the pivot bin. All variables related to this bin will be denoted with a subscripted m. For any given bin i that was sampled during the course of the simulation the new biasing potential is calculated by using a modified version of Eq. 3.23 given by

$$\Omega_{i,new} - \Omega_{m,new} = \Omega_{i,old} - \Omega_{m,old} + \ln\left(\frac{n_i^*}{n_m^*}\right) \tag{3.24}$$

Where n^* is the number of times a given bin was sampled during the current simulation segment. A problem naturally encountered by an algorithm like this is that in the event that a given bin was not sampled during the course of a simulation Eq. 3.24 fails since taking the natural log of 0 gives undefined behavior. Thus for the next simulation segment an arbitrary value must be chosen for this bin's input biasing potential. Failing to change this value can result in the simulation's inability to sample many of these bins. It was found that a simple

yet effective choice is for unsampled bins is given by.

$$\Omega_{i,new} - \Omega_{m,new} = \Omega_{m,old} - \Omega_{m,old} + \ln\left(n_m^*\right) \tag{3.25}$$

This is the equivalent of giving any unsampled bin a n^* value equal to 1/. This ensures that the difference in biasing potential between unsampled and sampled bins remains continuous for each simulation segment. Once the new values of each bin have been calculated, they are re-shifted so that the reference state for each simulation segment is the same. For instance in a nucleation study it is often convenient to choose the gas phase monomer as the reference state. Thus each segment's biasing potential is scaled such that the reference state is equal to $0k_BT$. This is done to ensure that each simulation segment can be pooled together at the end.

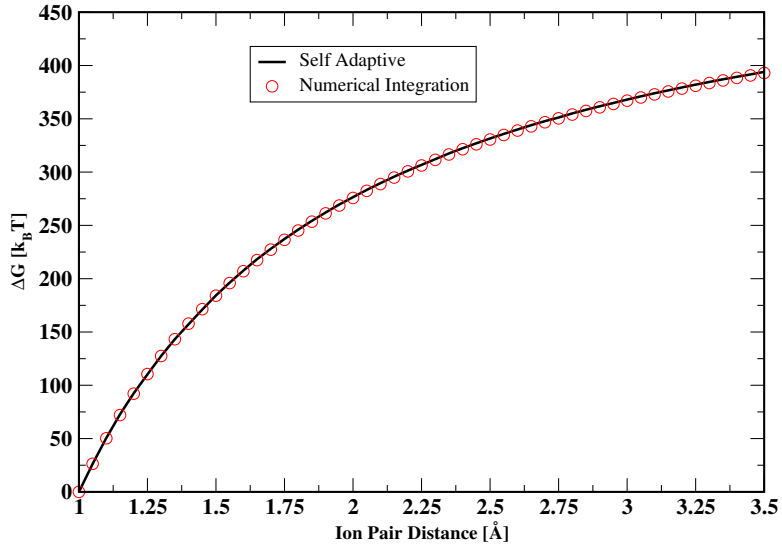

Figure 3.13: Presented here, the free energy of an ion pair in the gas phase is plotted as a function of the ion pair distance. This is done to demonstrate the self-adaptive algorithms ability to generate the free energy correctly for a simple test case.

3.2.2 Results

To first test the algorithm, a simple test case using an ion pair in the gas phase was used. In this test case a ion pair in the gas phase was simulated. In this simple system the energy along the reaction coordinate is given by a simple Columbic potential that it is only dependent on the ion pair distance. As such it is very easy to numerically calculate the free energy of a given bin of r using simple integration. Shown in Fig. 3.13 are the results of the simulation and numerical integration.

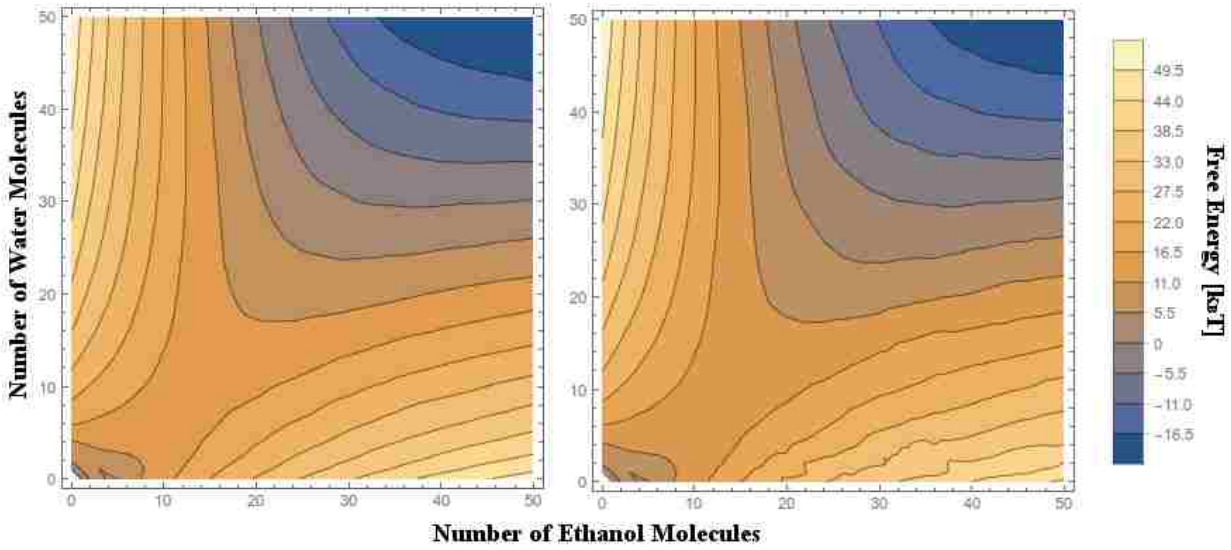

Figure 3.14: Presented here, the free energy of nucleation in a binary system consisting of water and ethanol at 230K. The free energy is plotted as a function of the number of water and the number of ethanol molecules in a given cluster. In the left panel the free energy results from the Self-Adaptive algorithm while in the right panel is the previously calculated results. This is shown to demonstrate that the Self-Adaptive algorithm can converge to the correct free energy value for a complicated two dimensional free energy landscape.

Since the algorithm was successful for a simple test, it was next applied to a previously studied water-ethanol system to ensure that it could properly replicate the free energy and onset activities that were previously published.[24] The results for this test can be found in Fig.

3.14 and Fig. 3.15. The algorithm was successful in reproducing the previously published properties. The total number of Monte Carlo cycles required to converge the free energy was estimated to be $8 \cdot 10^9$. The simulations took roughly about a day and a half of real time on LSU's SuperMic cluster.

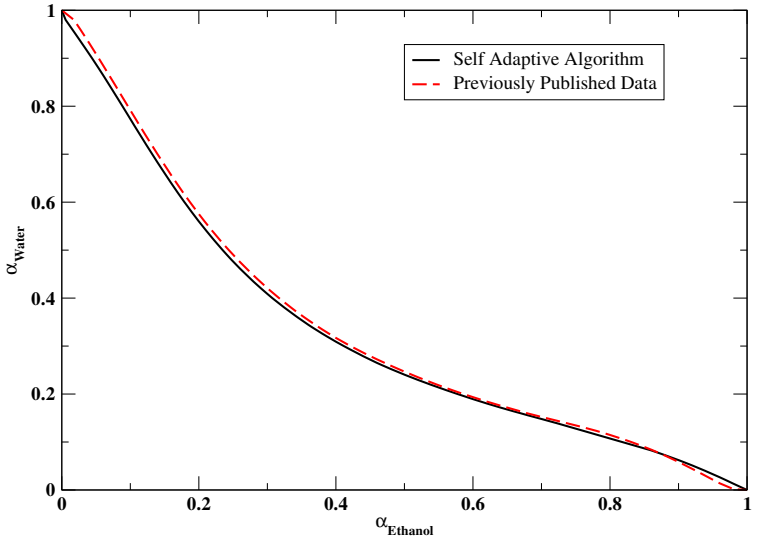

Figure 3.15: Presented here, onset activity of a water/ethanol binary system. This result is generated from the free energy landscape presented in Fig. 3.14. The curve shown represents the various thermodynamic conditions required to observe a nucleation barrier height of 40 k_BT. This in turn can be used to determine if the two components nucleate faster together or apart.

Chapter 4

A Practical Application of the Improved Monte Carlo Methods: Water Nucleation in the Presence of an Ionic Pair

4.1 Background

Atmospheric aerosols have long been suspected to play import roles in several atmospheric processes.[111,112] Ranging from cloud formation,[113], controling solar radiation,[114], and playing an role in atmospheric chemistry.[115] Since these aerosols play roles in both macroscale atmospheric behavior and can also influence things such as human health, understanding these aerosols is absolutely critical to understanding global phenomena.

As a consequence the nucleation of atmospheric aerosols has been a widely studied phenomena, but while there has been significant headway made into the understanding of how these aerosols are formed there still remain a great number of questions about the underlying mechanics that lead to their formation. Measuring the nucleation process experimentally is a very challenging undertaking given that by the time the particles are large enough to observe the critical point in the nucleation process has largely been completed. As a consequence it is very typical that only properties such as the nucleation rate[116] can be obtained. Thus many researchers have instead turned to using other methods such as computer simulations to shine light on the underlying mechanisms. It has been found previously that the presence of ionic species can greatly enhance the nucleation rate of water by a very significant rate.[117,118,96,90,119–121] In realistic systems these ions can be anything from acidic species[122–125]

to salts[126–128]. These species can catalyse the nucleation process of water by several orders of magnitude which leads to a high prevalence of ionic aerosols.

Many of the studies, both computational and experimental,[117,118,96,90,119,121] have uncovered that water displays a preference toward negatively charged ions over that of equivalent positively charged ions. It was argued that the origin of this sign preference was due to the fact that a negative ion can pack far more water hydrogen atoms around it compared to the number of water oxygen atoms that can pack around an equivalent positive ion. However for the many of the computational studies only a single ion was considered at a time. While this yields a large amount of information about how water interacts with given ion, it tends to leave out one very important factor which is the ion-ion interaction. This interaction in a host of systems can be non-negligible. If the ion-ion interaction is much stronger than the ion-water interaction, this interaction could potentially inhibit the ability of a given ion to nucleate water. At the same time the presence of two different ions could in theory nucleate water at a faster rate than a single ion could. To answer this question we have set out to study the nucleation process in the presence of various ion pairs via computer simulation in order to provide insights into the mechanisms related to ion pair nucleation.

4.2 Simulation Method

To study the physics of the nucleation process in the presence of an ionic pair, a series of different ion types were studied. The water was modelled using the TIP4P water model[105] and the ions were defined as a hard sphere ion with fixed charges of $+1$ and -1. The hard sphere criteria was enforced by rejecting any moves which would bring a ion closer than a given r_{min}. This r_{min} was defined using the Berthelot[73] mixing rules given by

$$r_{min} = \frac{\sigma_{ion} + \sigma_i}{2} \tag{4.1}$$

This rule applies for both ion-ion and ion-water interactions. For the water-ion interactions the Lennard-Jones σ parameter for each atom was used in the r_{min} definition with the exception of the TIP4P charge site. The size of both ions, denoted as σ_+ and σ_-, were scanned through sizes to values of 1Å, 3Å, and 5Å. Each combination of these values were simulated giving a total of 9 simulations. From here on each of these simulations will be referenced by their ion sizes via the notation (σ_+, σ_-). For example the system with a 1Å negative ion and a 1Å positive ion will be called the (1,1) system.

All simulations were performed in the grand canonical ensemble using the Energy Biased[129] version of the Aggregation-Volume-Bias Monte Carlo[21,22] method to facilitate the exchange of particles to and from the cluster. The cluster criteria used in this case was an energy based criteria where two molecules are defined as neighbors if their interaction energy is equal to below a -500$k_B \cdot K$ energy threshold similar to what has been used previously.[26,24,96,90]. Following what was done previously, the gas phase was represented using an implicit ideal gas with a fixed gas phase density of $3.5 \cdot 10^{-6}$ Å^{-3}. For all free energy calculations, the Self Adaptive Algorithm outlined in section 3.2 was used to converge the free energy at this gas phase density. Once the free energy landscape was collected, it was rescaled to a wide variety of different gas phase densities using the thermodynamic equation

$$\Delta G_{new}(n, r) = \Delta G_{sim}(n, r) - \ln\left(\frac{\rho_{new}}{\rho_{sim}}\right) \cdot n \tag{4.2}$$

Where $\Delta G_{new}(n, r)$ is the free energy of a given bin at the new gas phase density, $\Delta G_{sim}(n, r)$ is the free energy calculated from the simulation, ρ_{new} is the new gas phase density, ρ_{sim} is the simulation gas phase density, and n is the number of water molecules in the cluster. This was done to collect the barrier height isotherm to show how the barrier scales with the gas phase density.

For the free energy plots that were collected using only the cluster size as thereaction coordinate, clusters consisting of up to 100 water molecules were simulated. For the simulations where the cluster size and ionic distance were used as the reaction coordinates water clusters containing up to 60 water molecules were sampled and the largest sampled ion pair distance was 2Å above the minimum allowed distance for each system. This was done in the interest of saving computational costs given that the majority of the information of interest could be captured within this smaller window.

The neighbor analysis was performed using a fixed cluster size of 60 water molecules. For each system the Ion_+-O and Ion_--H radial distribution functions were computed and then integrated to determine the total number of neighbors. In addition a second analysis where the total number of solvating waters were counted and averaged over a large number of configurations. In this analysis if a water donated either a hydrogen atom to the negative ion or an oxygen atom to the positive ion, it was counted as part of the ionic solvation shell. If a water was bound to both ionic species it would only be counted once in this analysis. The cutoff distance for the solvation shell was determined from the radial distribution function for each ion.

4.3 Results

Presented in Fig. 4.1 are the results from the free energy calculations obtained by biasing the system along the cluster size coordinate. It is observed that the free energy drop from the monomer cluster decreases by increasing the negative ion size where as the increase in the positive ion size has mixed result. This can be explained in part by previous observations[117,118,96,90,119] where the negative ion was preferred over the positive ion given that, due to their size, more hydrogen atoms can favorably pack around a negative ion compared to the number of oxygen atoms that can favorably pack around a positive ion of equivalent size. This observation is consistent with previous literature on the subject. Generally across each of the systems, the free energy curve displayed similar trends to the

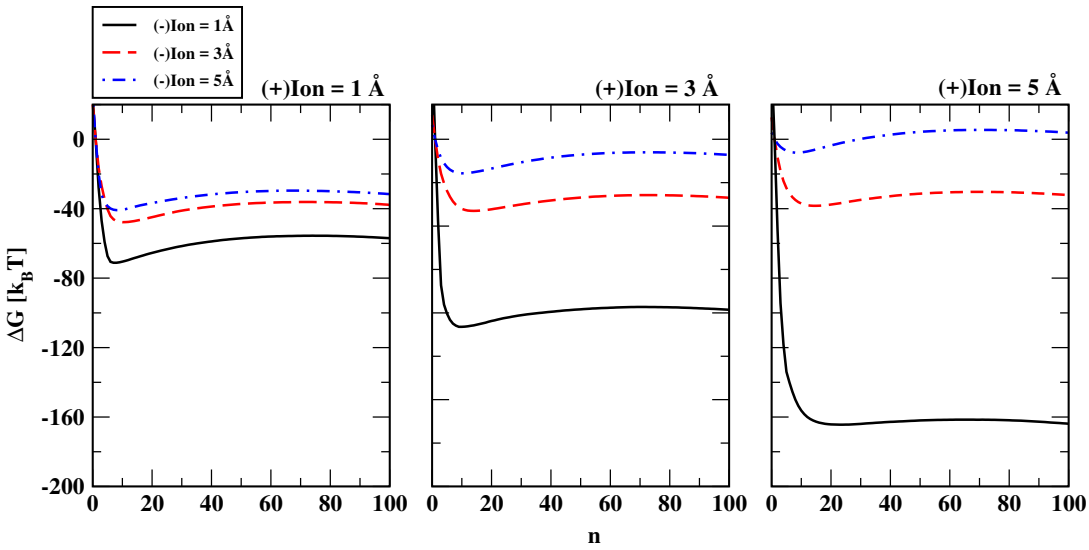

Figure 4.1: Presented here, the free energy of the different systems as a function of the number of water molecules in the cluster. The panels are organized by the size of the positive ion. In the left panel are the results for the (1,1), (1,3), and (1,5) systems, in the middle panel are the results for the (3,1), (3,3), and (3,5) system, and in the right panel are the results for the (5,1), (5,3), and (5,5) systems. The colors and shapes of the curves are designated by the negative ion size. All curves with a negative ion of $\sigma_- = 1$Å are given by a solid black curve, all systems with a negative ion of size $\sigma_- = 3$Å are given by a dashed red line, and all systems with negative ion $\sigma_- = 5$Å are given by a dot-dashed blue line. All plots are plotted on the same x-y scale.

single ion cases.[96,90,28] In the single ion case at a vapor phase density of the water is at mild super saturation contains a large drop in the free energy that corresponds to the creation of the ion's solvation shell. This drop in free energy continues till a minima is reached which corresponds to the saturation of the first solvation shell. However, once the shell has been formed the effects of the ion is now screened by the solvation shell and as such the free energy of addition from this point onward the begins to behave similar to the homogeneous water nucleation curve. This effect can be readily seen in Fig. 4.3 by plotting the results for a homogeneous water simulation at the same conditions on top of the curve of the ion system. Because of the nature of this system, the nucleation barrier can not be simply measured by finding the largest free energy value in the system since this will naturally be the monomer. Instead the free energy barrier must be measured by taking the free energy difference of

the local minima and local maxima present in these systems since this is the rate limiting region. Thus to obtain information related to the nucleation barrier and consequentially the nucleation rate, one must rescale the free energy such that the local minima is the reference state. When this is done, Fig. 4.2 is obtained. From here a very different trend from Fig. 4.1 is observed. For the 1Å positive ions it was observed that the barrier decreases with respect to increasing negative ion size. However for both the 3Å and 5Å positive ions this trend was completely reversed from the 1Å case or that the barrier increased with increasing ion size.

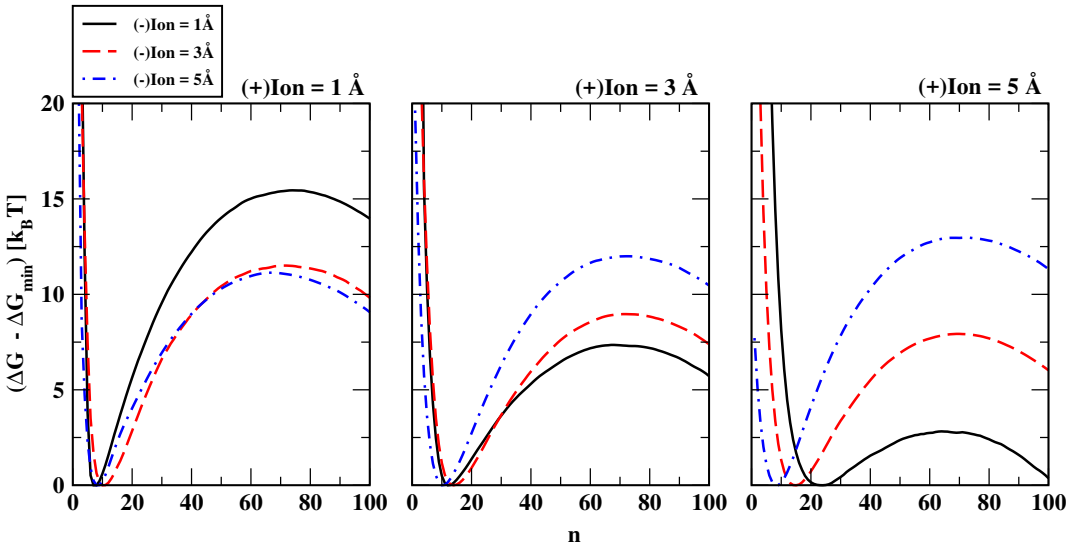

Figure 4.2: Presented here, the nucleation free energy of the different systems as a function of the number of water molecuels in the cluster. All plots have been shifted such that the local minima is equal to 0 k_BT. The panels are organized by the size of the positive ion. In the left panel are the results for the (1,1), (1,3), and (1,5) systems, in the middle pannel are the results for the (3,1), (3,3), and (3,5) system, and in the right panel are the results for the (5,1), (5,3), and (5,5) systems. The colors and shapes of the curves are designated by the negative ion size. All curves with a negative ion of $\sigma_- = 1$Å are given by a solid black curve, all systems with a negative ion of size $\sigma_- = 3$Å are given by a dashed red line, and all systems with negative ion $\sigma_- = 5$Å are given by a dot-dashed blue line. All plots are plotted on the same x-y scale.

To further analyse this effect the barrier height over a range of gas phase densities were computed to see if this trend was consistent. Shown in Fig.4.4 are the results when the nucleation barrier is calculated across many different vapor phase densities. From these results

a couple trends can be observed. The smallest ion pair, the (1,1) system, showed the largest nucleation barrier . A barrier was observable up untill a gas phase density of approximately $6.0 \cdot 10^{-6}$ Å^{-3}, which around this density every other system had no observable barrier. The (1,3) and (1,5) systems all displayed comparable barrier heights, which dissipated in a range between $5.0 \cdot 10^{-6}$ Å^{-3} to $5.5 \cdot 10^{-6}$ Å^{-3}, but for the most part the 5Å ion showed a slightly smaller barrier consistently. In contrast to these systems the (3,3) system and the (5,3) system showed a significantly lower barrier that dissipated around $4.5 \cdot 10^{-6}$ Å^{-3}. However the systems which stood out from nearly all the others was the (5,1) and (3,1) systems. The (5,1) and (3,1) systems showed a remarkably smaller barrier compared to every other system. When these systems are compared directly to the (3,3) and (5,3) systems it was found that there was between a 1.7-4.5 k_BT difference in barrier height at corresponding gas phase densities in addition to showing a barrier dissipation point for of around $4.0 \cdot 10^{-6}$ Å^{-3} and $4.3 \cdot 10^{-6}$ Å^{-3} for the (5,1) and (3,1) respectively.

It was hypothesized based on examining configurations from the system that two important coordinates to examine are the ion-ion pair distance r and the location of the ion pair in the cluster. To examine this, a second set of simulations that were biased along both the cluster size coordinate and the ion pair distance coordinate was performed. This yields the free energy as a two dimensional plot. The results of these simulations can be found in Fig. 4.5. From these plots it is very easy to see that systems such as the (1,1), (1,3), (3,3), (3,5), and (5,5) systems have their minimal energy pathway located along the smallest allowed r value.

This indicates that as these systems nucleate the ions remain firmly in direct contact with each other. Each of these systems have sizeable free energy barriers in the center of the free energy landscape that prevents solvent separation. Systems such as the (1,5) and (5,3) showed a barrier in the center of the landscape, yet these barriers were only about a 2-5k_BT difference from the contact pair which seems to indicate that this barrier is potentially

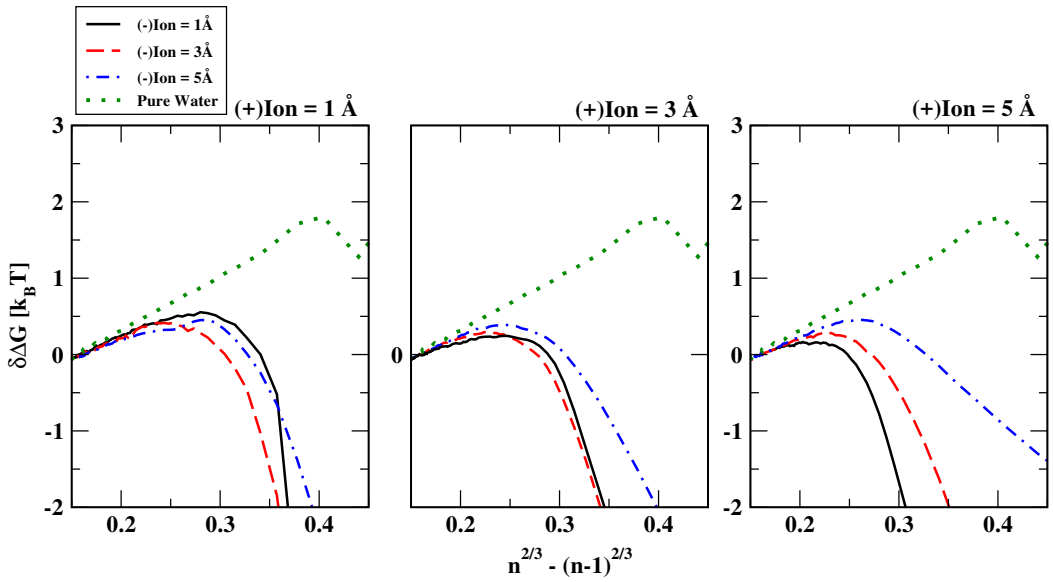

Figure 4.3: Presented here, the $\delta\Delta G$ plots for each system as a function of the number of water molecules in the cluster. The panels are organized by the size of the positive ion. In the left panel are the results for the (1,1), (1,3), and (1,5) systems, in the middle panel are the results for the (3,1), (3,3), and (3,5) system, and in the right panel are the results for the (5,1), (5,3), and (5,5) systems. The colors and shapes of the curves are designated by the negative ion size. All curves with a negative ion of $\sigma_- = 1$Å are given by a solid black curve, all systems with a negative ion of size $\sigma_- = 3$Å are given by a dashed red line, and all systems with negative ion $\sigma_- = 5$Å are given by a dot-dashed blue line. In addition, a curve obtained from a homogeneous water simulation with no ions present is added to highlight important trends in the ion system. All plots are plotted on the same x-y scale.

scalable especially at conditions where the nucleation barrier is above $5k_BT$. However, systems such as the (3,1), (5,1) there is very little question that these systems can not only solvent separate, but that their miminmal free energy pathway is located at larger r values.

In addition to the free energy plots, the average ion pair distance as a function of cluster size was calculated. The results for these can be seen in Fig. 4.6. It was found that the (1,5), (3,1), (5,3), and (5,1) systems showed appreciable ion pair separation while nearly all other systems remained at the smallest ion pair distance possible. It was observed that dissociation for the (1,5),(3,1), (5,1), and (3,5) occurred at cluster sizes of approximately 13, 9, 9, and 25 respectively. This trend co-insides with the barrier height trend in that the systems which dissociate sooner also have the lowest barrier. The (3,5) system shows a small

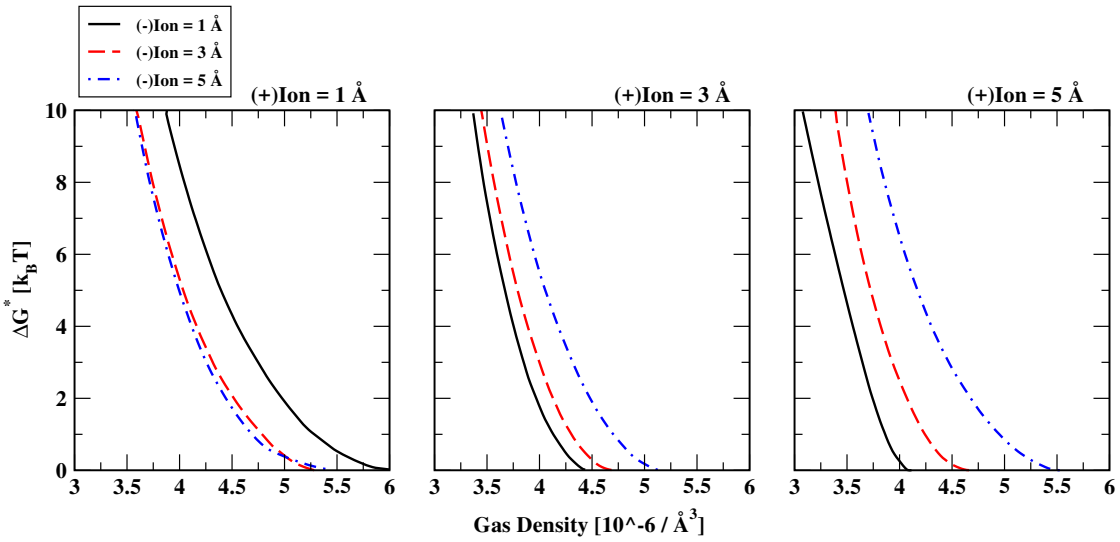

Figure 4.4: Presented here, the nucleation barrier heights of the different systems as a function of the vapor phase density of water. The panels are organized by the size of the positive ion. In the left panel are the results for the (1,1), (1,3), and (1,5) systems, in the middle panel are the results for the (3,1), (3,3), and (3,5) system, and in the right panel are the results for the (5,1), (5,3), and (5,5) systems. The colors and shapes of the curves are designated by the negative ion size. All curves with a negative ion of $\sigma_- = 1$Å are given by a solid black curve, all systems with a negative ion of size $\sigma_- = 3$Å are given by a dashed red line, and all systems with negative ion $\sigma_- = 5$Å are given by a dot-dashed blue line. All plots are plotted on the same x-y scale.

yet noticeable dissociation.

It becomes clear that the ability to dissociate the ions pairs early on in the cluster formation is linked to a significant drop in the nucleation free energy barrier. To further understand this it is necessary to examine the water-ion interactions for each systems. To do this the radial distribution function for a fixed cluster size of 60 water molecules was calculated and then integrated to find the number of nearest neighbors in both ions' solvation shells. The results of these calculations can be found in Table 4.1. It is observable that the total number of hydrogen atoms in the first solvation shell of the negative ion was between 5-7 for all systems except the (1,1) system, where the total number of hydrogen neighbors was half that of most other systems. In contrast to this, the number of oxygen neighbors varies significantly with the size of the positive ion. This number can range from 3.1 all

the way up to 11 neighbors. In addition the number of neighbors for the positive ion is highly dependent on both the the size of the positive and negative ion. For a positive ion size of 3 Å, the average number of oxygen neighbors can range for 6.9 up to 9.5. While this neighboring analysis provides some basic insight, the analysis can be taken a step further by performing a number of unique neighbor analysis. This analysis looks to see how many waters in total are bonded to one or both ions. This type of analysis can account for water molecules that are bound to both ions (i.e. A water whose hydrogen atom is facing the negative ion while at the same time the oxygen atom on the same water is connected to the positive ion). Any water which has an oxygen or hydrogen within either ion's first solvation shell is counted in this analysis. The results from this can be seen in Column 2 of Table 4.2. From this trend one may observe that the systems that have shown lower free energy barriers also have a high number of unique water molecules in their solvation shells. When the barrier height is plotted as a function of the number of solvation neighbors a correlation is observed. (See Fig. 4.7) While the relationship is not perfectly linear, it does imply there is a correlation between the number of water molecules that can be successfully packed onto both ion's solvation shell. From instance the (1,5) system relative free energy is sigificantly higher than the (1,1), (1,3), (3,3), and (3,5) systems yet the (1,5) system shows a lower or comparable free energy barrier compared to these systems. It is likely that the ion-water bonds in the (1,5) system are sigificantly weaker. However, this ion combiniation is capable of separating reasonably early into the nucleation cycle which allows it to stablize a greater number of water molecules compared to ion pairs such as the (1,1) pair.

These results seem to dictate that the quality of the ion-water bonds is not the most important factor to the nucleation process. Instead the most important factor appears to be the total number of water molecules that can be stably added to the ion pair. If this true, then so long as the water-ion interaction is sufficiently strong to prevent the water system from simply ejecting the ions, it becomes largely a problem of being able to

Table 4.1: Presented here, the average number of nearest neighbors of the positive and negative ions for each system at a fixed cluster size of 60 water molecules. These numbers were obtained by integrating the radial distribution function up through the first solvation shell. For the positive ion these numbers represent the number of water oxygen atoms present in the first solvation shell while for the negative ion it represents the number of hydrogen atoms in the first solvation shell.

System	Pos/Oxygen	Neg/Hydrogen
(1, 1)	3.1	3.1
(1, 3)	4.7	6.7
(1, 5)	4.9	5.7
(3, 1)	9.5	5.4
(3, 3)	7.3	6.1
(3, 5)	6.9	5.1
(5, 1)	10.6	6.0
(5, 3)	10.6	6.9
(5, 5)	7.9	5.2

Table 4.2: Presented here, the total number of water molecules found in either ion's solvation shell for each system is presented along side their corresponding barrier heights.

System	Unique Water Neighbors	Barrier $[k_BT]$
(1, 1)	6.0	15.5
(1, 3)	8.4	11.5
(1, 5)	10.7	11.3
(3, 1)	11.7	7.3
(3, 3)	10.9	9.0
(3, 5)	7.0	12.0
(5, 1)	15.5	2.8
(5, 3)	14.5	7.9
(5, 5)	7.9	13.0

create a large enough surface area for water to aggregate around. This can explain why the systems which underwent solvent separation had significantly smaller nucleation free energy barriers compared to the majority of the contact pair systems given that separating the two ions maximized the total ionic surface thus giving the (1,5), (3,1), (5,1), and (5,3) ion pairs dramatically lower barriers. This can also explain why the (3,3) showed a reasonably low nucleation barrier enough though it largely existed in the contact pair since the (3,3) configuration could still accommodate a reasonably high number of water molecules. If these trends hold true for polyatomic species, this can also potentially explain why acidic species such as H_2SO_4 are very good nucleating agents in the atmosphere since it is known that many strong acids can undergo dissociation for a small number of waters.[130,131]

At this time there is further work to be done on this system such as an energetic/entropic analysis, temperature dependence, or other assorted tests that can provide additional information. It is expected that this work will be finished by the end of the year and will be published. In addition further studies into more complicated ion cases where the charge, number of ions, and type of ion (e.g. molecular ion vs single atom ion) can be conducted to examine if these trends remain true.

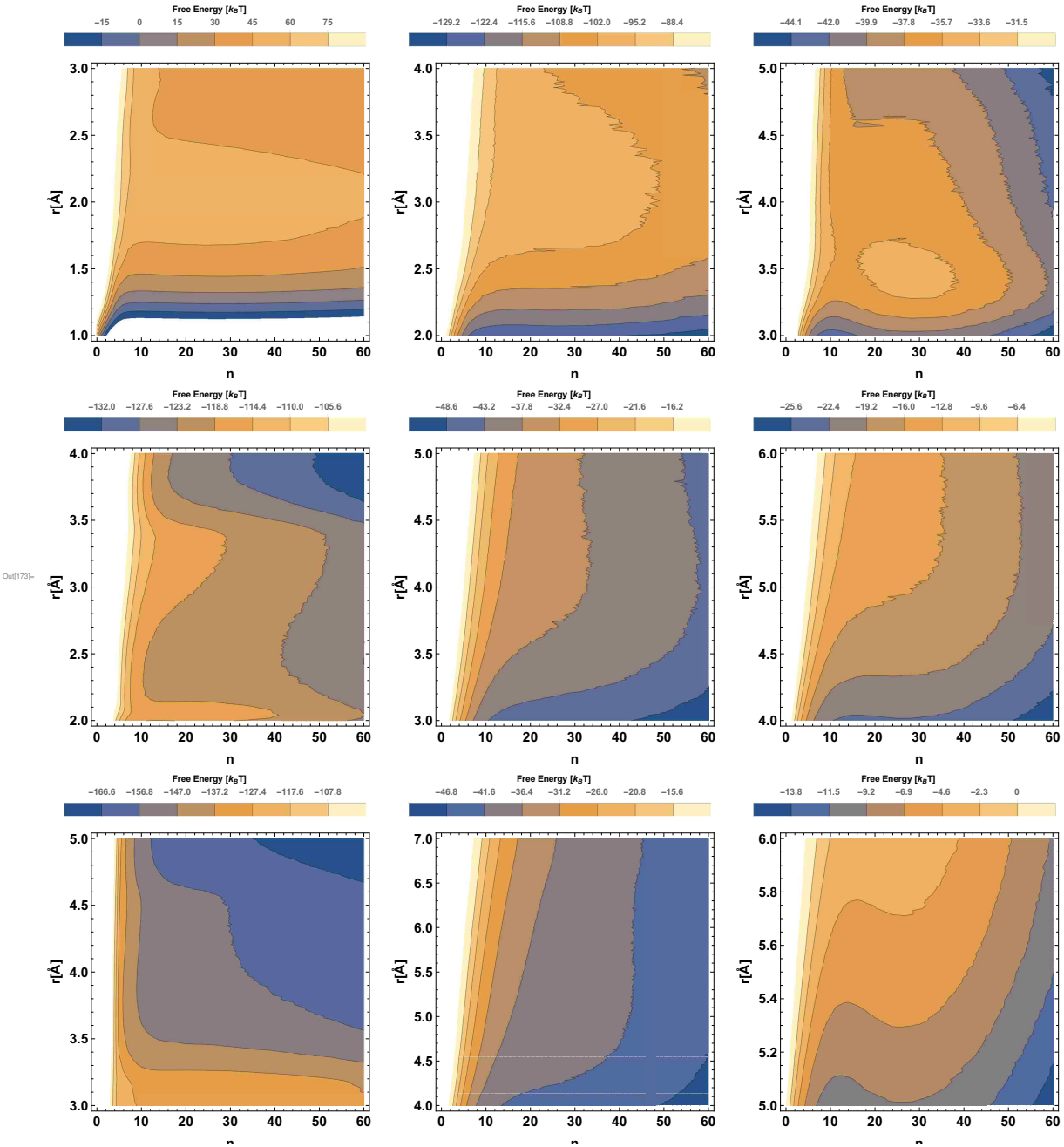

Figure 4.5: Presented here, the 2D ΔG plot given as a function of the cluster size and ion pair distance. The columns from left to right correspond to a negative ion sizes of 1Å, 3Å, and 5Å while the rows from top to bottom correspond to positive ion sizes of 1Å, 3Å, and 5Å.

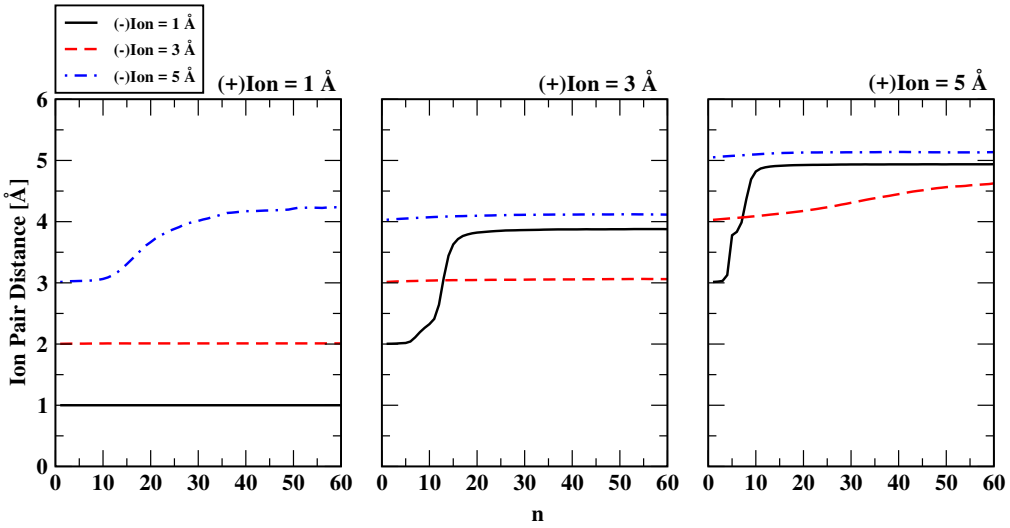

Figure 4.6: Presented here, the average ion pair distance as a function of the number of water molecuels in a given cluster. The panels are organized by the size of the positive ion. In the left panel are the results for the (1,1), (1,3), and (1,5) systems, in the middle pannel are the results for the (3,1), (3,3), and (3,5) system, and in the right panel are the results for the (5,1), (5,3), and (5,5) systems. The colors and shapes of the curves are designated by the negative ion size. All curves with a negative ion of $\sigma_- = 1$Å are given by a solid black curve, all systems with a negative ion of size $\sigma_- = 3$Å are given by a dashed red line, and all systems with negative ion $\sigma_- = 5$Å are given by a dot-dashed blue line. All plots are plotted on the same x-y scale.

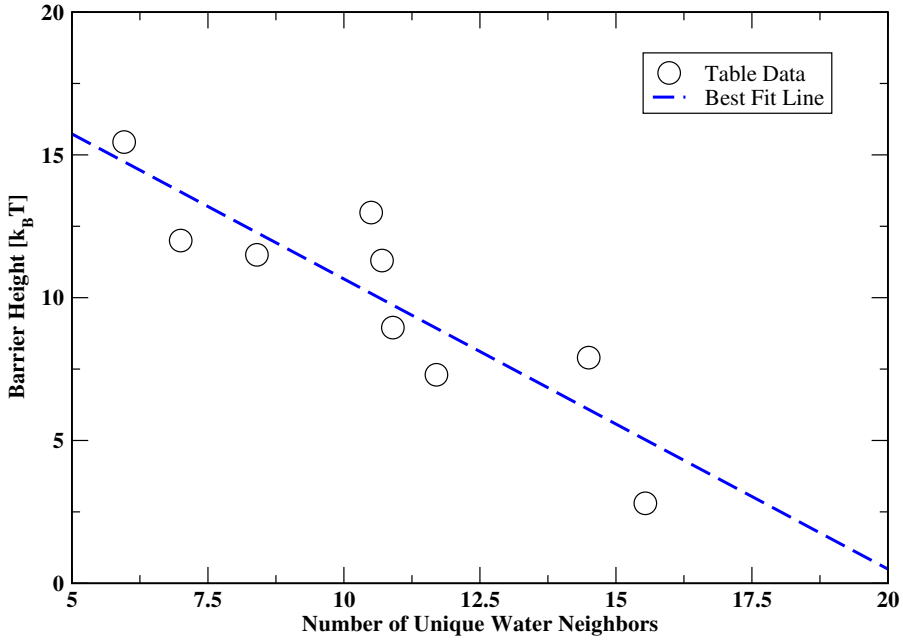

Figure 4.7: Presented here, the data from Table 4.2 shown as black circles is plotted along side its best fit line which is given by the dashed blue line.

Chapter 5

Conclusion

In this dissertation we have examined the Classical Nucleation Theory equations to study the weaknesses of the theory as well as use the deviations from the theory to uncover interesting catalytic phenomena. In addition we have developed new formulations to the Aggregation-Volume-Bias method and the Umbrella Sampling Algorithm that has significantly improved the computational efficiency of sampling complicated systems. Lastly we have put these methods to use to sample an atmospherically relevant ionic system in order to uncover the underlying trends in the nucleation behavior. It is hoped that this research will pave the way so that others may use these techniques in order to advance knowledge on the topic of nucleation.

Chapter 6

Appendix

6.1 Copyright Information

6.1.1 Chapter 2.2

This Agreement between Troy D Loeffler ("You") and AIP Publishing LLC ("AIP Publishing LLC") consists of your license details and the terms and conditions provided by AIP Publishing LLC and Copyright Clearance Center.

- License Number 3971420542479
- License date Oct 17, 2016
- Licensed Content Publisher AIP Publishing LLC
- Licensed Content Publication Journal of Chemical Physics
- Licensed Content Title Vaporliquid nucleation in two dimensions: On the intriguing sign switch of the errors of the classical nucleation theory
- Licensed Content Author Troy D. Loeffler,David E. Henderson,Bin Chen
- Licensed Content Date Nov 21, 2012
- Licensed Content Volume Number 137
- Licensed Content Issue Number 19
- Type of Use Thesis/Dissertation
- Requestor type Author (original article)
- Format Electronic

- Portion Excerpt (> 800 words)
- Will you be translating? No
- Title of your thesis dissertation Advanced Monte Carlo Methods for the Study of Nucleation

- Expected completion date Oct 2016
- Estimated size (number ofpages) 110
- Requestor Location (redacted)
- United States
- Attn: Troy D Loeffler

AIP Publishing LLC – Terms and Conditions: Permissions Uses AIP Publishing hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material:"Reprinted from [FULL CITATION], with the permission of AIP Publishing." For an article, the credit line and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP Publishing and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP Publishing, with permission, from another source. Under no circumstances does AIP Publishing purport or intend to grant permission to reuse material to which it does not hold appropriate

rights. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes.

3. The foregoing license shall not take effect unless and until AIP Publishing or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
4. AIP Publishing or Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP Publishing, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
5. AIP Publishing makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP Publishing, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
6. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP Publishing. This license may not be amended except in a writing signed by the party to be charged.
7. If purchase orders, acknowledgements or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State

of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

6.1.2 Chapter 2.3

This Agreement between Troy D Loeffler ("You") and AIP Publishing LLC ("AIP Publishing LLC") consists of your license details and the terms and conditions provided by AIP Publishing LLC and Copyright Clearance Center.

- License Number 3971420060745
- License date Oct 17, 2016
- Licensed Content Publisher AIP Publishing LLC
- Licensed Content Publication Journal of Chemical Physics
- Licensed Content Title Surface induced nucleation of a Lennard-Jones system on an implicit surface at sub-freezing temperatures: A comparison with the classical nucleation theory
- Licensed Content Author Troy D. Loeffler,Bin Chen
- Licensed Content Date Dec 19, 2013
- Licensed Content Volume Number 139
- Licensed Content Issue Number 23
- Type of Use Thesis/Dissertation
- Requestor type Author (original article)
- Format Electronic
- Portion Excerpt (> 800 words)
- Will you be translating? No
- Title of your thesis / dissertation Advanced Monte Carlo Methods for the Study of Nucleation
- Expected completion date Oct 2016
- Estimated size (number of pages) 110 Requestor Location Troy D Loeffler (redacted)

BATON ROUGE, LA 70810 United States

- Attn: Troy D Loeffler 1 of 3 Billing Type Invoice Billing Address Troy D Loeffler (redacted)

AIP Publishing LLC – Terms and Conditions: Permissions Uses AIP Publishing hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material:"Reprinted from [FULL CITATION], with the permission of AIP Publishing." For an article, the credit line and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP Publishing and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP Publishing, with permission, from another source. Under no circumstances does AIP Publishing purport or intend to grant permission to reuse material to which it does not hold appropriate rights. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes.
3. The foregoing license shall not take effect unless and until AIP Publishing or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.

4. AIP Publishing or Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP Publishing, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
5. AIP Publishing makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP Publishing, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
6. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP Publishing. This license may not be amended except in a writing signed by the party to be charged.
7. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

6.1.3 Chapter 3.1

- Title:Improved Monte Carlo Scheme for Efficient Particle Transfer in Heterogeneous Systems in the Grand Canonical Ensemble: Application to VaporLiquid Nucleation
- Author:Troy D. Loeffler, Aliasghar Sepehri, Bin Chen
- Publication: Journal of Chemical Theory and Computation

- Publisher: American Chemical Society
- Date: Sep 1, 2015

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

References

[1] Mantu Santra, Suman Chakrabarty, and Biman Bagchi. Gas-liquid nucleation in a two dimensional system. *The Journal of Chemical Physics*, 129(23):234704, 2008.

[2] R. Becker and W. Dring. Kinetische behandlung der keimbildung in bersttigten dmpfen. *Annalen der Physik*, 416(8):719–752, 1935.

[3] Max Volmer et al. Kinetik der phasenbildung. 1939.

[4] J. Schneider, A. Schneider, A. Sarikov, J. Klein, M. Muske, S. Gall, and W. Fuhs. Aluminum-induced crystallization: Nucleation and growth process. *Journal of Non-Crystalline Solids*, 352(920):972 – 975, 2006. Amorphous and Nanocrystalline Semiconductors - Science and TechnologyProceedings of the 21st International Conference on Amorphous and Nanocrystalline Semiconductors - Science and Technology21st International Conference on Amorphous and Nanocrystalline Semiconductors.

[5] Ricky B. Nellas, Matthew E. McKenzie, and Bin Chen. Probing the nucleation mechanism for the binary n-nonane/1-alcohol series with atomistic simulations. *The Journal of Physical Chemistry B*, 110(37):18619–18628, 2006. PMID: 16970491.

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

[7] Joonas Merikanto, Evgeni Zapadinsky, Antti Lauri, and Hanna Vehkamäki. Origin of the failure of classical nucleation theory: Incorrect description of the smallest clusters. *Phys. Rev. Lett.*, 98:145702, Apr 2007.

[8] R Makkonen, A Asmi, H Korhonen, H Kokkola, S Järvenoja, P Räisänen, KEJ Lehtinen, A Laaksonen, V-M Kerminen, H Järvinen, et al. Sensitivity of aerosol concentrations and cloud properties to nucleation and secondary organic distribution in echam5-ham global circulation model. *Atmospheric Chemistry and Physics*, 9(5):1747–1766, 2009.

[9] Benjamin J Murray, Daniel A Knopf, and Allan K Bertram. The formation of cubic ice under conditions relevant to earth's atmosphere. *Nature*, 434(7030):202–205, 2005.

[10] Somnath Sinha, Ashutosh Bhabhe, Hartawan Laksmono, Judith Wlk, Reinhard Strey, and Barbara Wyslouzil. Argon nucleation in a cryogenic supersonic nozzle. *The Journal of Chemical Physics*, 132(6):064304, 2010.

[11] Raymond Anglil, Jrg Diemand, Kyoko K. Tanaka, and Hidekazu Tanaka. Properties of liquid clusters in large-scale molecular dynamics nucleation simulations. *The Journal of Chemical Physics*, 140(7):074303, 2014.

[12] Barbara N. Hale. Temperature dependence of homogeneous nucleation rates for water: Near equivalence of the empirical fit of wlk and strey, and the scaled nucleation model. *The Journal of Chemical Physics*, 122(20):204509, 2005.

[13] David Brus, Vladimr dmal, and Hermann Uchtmann. Homogeneous nucleation rate measurements in supersaturated water vapor ii. *The Journal of Chemical Physics*, 131(7):074507, 2009.

[14] Bin Chen, J. Ilja Siepmann, Kwang J. Oh, and Michael L. Klein. Aggregation-volume-

bias monte carlo simulations of vapor-liquid nucleation barriers for lennard-jonesium. *The Journal of Chemical Physics*, 115(23):10903–10913, 2001.

[15] Kyoko K. Tanaka, Akio Kawano, and Hidekazu Tanaka. Molecular dynamics simulations of the nucleation of water: Determining the sticking probability and formation energy of a cluster. *The Journal of Chemical Physics*, 140(11):114302, 2014.

[16] Jan Wedekind, Judith Wlk, David Reguera, and Reinhard Strey. Nucleation rate isotherms of argon from molecular dynamics simulations. *The Journal of Chemical Physics*, 127(15):154515, 2007.

[17] Glenn M. Torrie and John P. Valleau. Monte carlo free energy estimates using non-boltzmann sampling: Application to the sub-critical lennard-jones fluid. *Chemical Physics Letters*, 28(4):578 – 581, 1974.

[18] Daan Frenkel and Berend Smit, editors. Academic Press, San Diego, second edition edition, 2002.

[19] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[20] W. K. HASTINGS. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[21] Bin Chen and J. Ilja Siepmann. A novel monte carlo algorithm for simulating strongly associating fluids: applications to water, hydrogen fluoride, and acetic acid. *The Journal of Physical Chemistry B*, 104(36):8725–8734, 2000.

[22] Bin Chen and J. Ilja Siepmann. Improving the efficiency of the aggregationvolumebias

monte carlo algorithm. *The Journal of Physical Chemistry B*, 105(45):11275–11282, 2001.

[23] Frank H. Stillinger. Rigorous basis of the frenkelband theory of association equilibrium. *The Journal of Chemical Physics*, 38(7):1486–1494, 1963.

[24] Bin Chen, J. Ilja Siepmann, and Michael L. Klein. Simulating the nucleation of water/ethanol and water/n-nonane mixtures: mutual enhancement and two-pathway mechanism. *Journal of the American Chemical Society*, 125(10):3113–3118, 2003. PMID: 12617679.

[25] Ricky B. Nellas, Bin Chen, and J. Ilja Siepmann. Dumbbells and onions in ternary nucleation. *Phys. Chem. Chem. Phys.*, 9:2779–2781, 2007.

[26] Bin Chen, J. Ilja Siepmann, Kwang J. Oh, and Michael L. Klein. Simulating vaporliquid nucleation of n-alkanes. *The Journal of Chemical Physics*, 116(10):4317–4329, 2002.

[27] Yujie Wu, Hanning Chen, Feng Wang, Francesco Paesani, and Gregory A. Voth. An improved multistate empirical valence bond model for aqueous proton solvation and transport. *The Journal of Physical Chemistry B*, 112(2):467–482, 2008. PMID: 17999484.

[28] Revati Kumar, Chris Knight, Collin D. Wick, and Bin Chen. Bringing reactivity to the aggregation-volume-bias monte carlo based simulation framework: Water nucleation induced by a reactive proton. *The Journal of Physical Chemistry B*, 119(29):9068–9075, 2015. PMID: 25289603.

[29] Farid Abraham. *Homogeneous nucleation theory: the pretransition theory of vapor condensation*, volume 1. Elsevier, 2012.

[30] Howard Reiss. The kinetics of phase transitions in binary systems. *The Journal of Chemical Physics*, 18(6):840–848, 1950.

[31] Ari Laaksonen, Vicente Talanquer, and David W. Oxtoby. Nucleation: Measurements, theory, and atmospheric applications. *Annual Review of Physical Chemistry*, 46(1):489–524, 1995. PMID: 24341941.

[32] D. Stauffer. Kinetic theory of two-component (hetero-molecular) nucleation and condensation. *Journal of Aerosol Science*, 7(4):319 – 333, 1976.

[33] Gerald. Wilemski. Revised classical binary nucleation theory for aqueous alcohol and acetone vapors. *The Journal of Physical Chemistry*, 91(10):2492–2498, 1987.

[34] A. JaeckerVoirol, P. Mirabel, and H. Reiss. Hydrates in supersaturated binary sulfuric acidwater vapor: A reexamination. *The Journal of Chemical Physics*, 87(8):4849–4852, 1987.

[35] Bin Chen, Hyunmi Kim, Samuel J. Keasler, and Ricky B. Nellas. An aggregation-volume-bias monte carlo investigation on the condensation of a lennard-jones vapor below the triple point and crystal nucleation in cluster systems: an in-depth evaluation of the classical nucleation theory. *The Journal of Physical Chemistry B*, 112(13):4067–4078, 2008. PMID: 18335920.

[36] Barbara N Hale. Monte carlo calculations of effective surface tension for small clusters. *Australian Journal of Physics*, 49(2):425–434, Apr 1996.

[37] Shawn M. Kathmann, Gregory K. Schenter, and Bruce C. Garrett. Understanding the sensitivity of nucleation kinetics: A case study on water. *The Journal of Chemical Physics*, 116(12):5046–5057, 2002.

[38] Ricky B. Nellas, Samuel J. Keasler, J. Ilja Siepmann, and Bin Chen. Exploring the discrepancies between experiment, theory, and simulation for the homogeneous gas-to-liquid nucleation of 1-pentanol. *The Journal of Chemical Physics*, 132(16):164517, 2010.

[39] X. C. Zeng and David W. Oxtoby. Gasliquid nucleation in lennardjones fluids. *The Journal of Chemical Physics*, 94(6):4472–4478, 1991.

[40] X. C. Zeng. Gasliquid nucleation in twodimensional fluids. *The Journal of Chemical Physics*, 104(7):2699–2704, 1996.

[41] Athanassios Z. Panagiotopoulos. Direct determination of phase coexistence properties of fluids by monte carlo simulation in a new ensemble. *Molecular Physics*, 61(4):813–826, 1987.

[42] A.Z. Panagiotopoulos, N. Quirke, M. Stapleton, and D.J. Tildesley. Phase equilibria by simulation in the gibbs ensemble. *Molecular Physics*, 63(4):527–545, 1988.

[43] B. Smit, Ph. De Smedt, and D. Frenkel. Computer simulations in the gibbs ensemble. *Molecular Physics*, 68(4):931–950, 1989.

[44] John Shipley Rowlinson and Benjamin Widom. *Molecular theory of capillarity*. Courier Corporation, 2013.

[45] Matthias Mecke, Jochen Winkelmann, and Johann Fischer. Molecular dynamics simulation of the liquidvapor interface: The lennard-jones fluid. *The Journal of Chemical Physics*, 107(21):9264–9270, 1997.

[46] Pieter Rein ten Wolde and Daan Frenkel. Computer simulation study of gasliquid nucleation in a lennard-jones system. *The Journal of Chemical Physics*, 109(22):9901–9918, 1998.

[47] Shawn M. Kathmann, Gregory K. Schenter, Bruce C. Garrett, Bin Chen, and J. Ilja Siepmann. Thermodynamics and kinetics of nanoclusters controlling gas-to-particle nucleation. *The Journal of Physical Chemistry C*, 113(24):10354–10370, 2009.

[48] Robert McGraw and Ari Laaksonen. Interfacial curvature free energy, the kelvin relation, and vaporliquid nucleation rate. *The Journal of Chemical Physics*, 106(12):5284–5287, 1997.

[49] A. E. van Giessen and E. M. Blokhuis. Determination of curvature corrections to the surface tension of a liquidvapor interface through molecular dynamics simulations. *The Journal of Chemical Physics*, 116(1):302–310, 2002.

[50] Alexander Fladerer and Reinhard Strey. Homogeneous nucleation and droplet growth in supersaturated argon vapor: The cryogenic nucleation pulse chamber. *The Journal of Chemical Physics*, 124(16):164710, 2006.

[51] Kristina Iland, Judith Wlk, Reinhard Strey, and Dimo Kashchiev. Argon nucleation in a cryogenic nucleation pulse chamber. *The Journal of Chemical Physics*, 127(15):154506, 2007.

[52] Troy D Loeffler, David E Henderson, and Bin Chen. Vapor–liquid nucleation in two dimensions: On the intriguing sign switch of the errors of the classical nucleation theory. *The Journal of chemical physics*, 137(19):194304, 2012.

[53] Azar Alizadeh, Masako Yamada, Ri Li, Wen Shang, Shourya Otta, Sheng Zhong, Liehui Ge, Ali Dhinojwala, Ken R. Conway, Vaibhav Bahadur, A. Joseph Vinciquerra, Brian Stephens, and Margaret L. Blohm. Dynamics of ice nucleation on water repellent surfaces. *Langmuir*, 28(6):3180–3186, 2012. PMID: 22235939.

[54] C. M. Archuleta, P. J. DeMott, and S. M. Kreidenweis. Ice nucleation by surrogates

for atmospheric mineral dust and mineral dust/sulfate particles at cirrus temperatures. *Atmospheric Chemistry and Physics*, 5(10):2617–2634, 2005.

[55] W. Knauer. Formation of large metal clusters by surface nucleation. *Journal of Applied Physics*, 62(3):841–851, 1987.

[56] S Rousset, V Repain, G Baudot, H Ellmer, Y Garreau, V Etgens, J.M Berroir, B Croset, M Sotto, P Zeppenfeld, J Ferr, J.P Jamet, C Chappert, and J Lecoeur. Self-ordering on crystal surfaces: fundamentals and applications. *Materials Science and Engineering: B*, 96(2):169 – 177, 2002.

[57] J. L. Holbrough, J. M. Campbell, F. C. Meldrum, and H. K. Christenson. Topographical control of crystal nucleation. *Crystal Growth & Design*, 12(2):750–755, 2012.

[58] G. W. Arnold. Nearsurface nucleation and crystallization of an ionimplanted lithiaaluminasilica glass. *Journal of Applied Physics*, 46(10):4466–4473, 1975.

[59] A. Tabazadeh, Y. S. Djikaev, P. Hamill, and H. Reiss. Laboratory evidence for surface nucleation of solid polar stratospheric cloud particles. *The Journal of Physical Chemistry A*, 106(43):10238–10246, 2002.

[60] Stefan Frieder Hopp and Andreas Heuer. Kinetic monte carlo study of nucleation processes on patterned surfaces. *The Journal of Chemical Physics*, 133(20):204101, 2010.

[61] Amanda J. Page and Richard P. Sear. Heterogeneous nucleation in and out of pores. *Phys. Rev. Lett.*, 97:065701, Aug 2006.

[62] Gregory Brown, Amitabha Chakrabarti, and J. F. Marko. Surface-induced nucleation. *Phys. Rev. E*, 50:1674–1677, Aug 1994.

[63] Ma Qian and Jie Ma. Heterogeneous nucleation on convex spherical substrate surfaces: A rigorous thermodynamic formulation of fletchers classical model and the new perspectives derived. *The Journal of Chemical Physics*, 130(21):214709, 2009.

[64] John D. Hoffman and Robert L. Miller. Surface nucleation theory for chain-folded systems with lattice strain: curved edges. *Macromolecules*, 22(7):3038–3054, 1989.

[65] T. V. Bykov and X. C. Zeng. Heterogeneous nucleation on mesoscopic wettable particles: A hybrid thermodynamic/density-functional theory. *The Journal of Chemical Physics*, 117(4):1851–1868, 2002.

[66] Richard C. Ward, Jon M. Holdman, and Barbara N. Hale. Monte carlo studies of water monolayer clusters on substrates: Hexagonal agi. *The Journal of Chemical Physics*, 77(6):3198–3202, 1982.

[67] S. Toxvaerd. Molecular dynamics simulation of heterogeneous nucleation at a structureless solid surface. *The Journal of Chemical Physics*, 117(22):10303–10310, 2002.

[68] D. Turnbull. Kinetics of heterogeneous nucleation. *The Journal of Chemical Physics*, 18(2):198–203, 1950.

[69] Ricky B. Nellas, Samuel J. Keasler, J. Ilja Siepmann, and Bin Chen. Exploring the discrepancies between experiment, theory, and simulation for the homogeneous gas-to-liquid nucleation of 1-pentanol. *The Journal of Chemical Physics*, 132(16):164517, 2010.

[70] Erik E. Santiso, Carmelo Herdes, and Erich A. Mller. On the calculation of solid-fluid contact angles from molecular dynamics. *Entropy*, 15(9):3734, 2013.

[71] Barbara N. Hale, Jerry Kiefer, Sergio Terrazas, and Richard C. Ward. Theoreti-

cal studies of water adsorbed on silver iodide. *The Journal of Physical Chemistry*, 84(12):1473–1479, 1980.

[72] HA Lorentz. Ueber die anwendung des satzes vom virial in der kinetischen theorie der gase. *Annalen der Physik*, 248(1):127–136, 1881.

[73] D. C. R Berthlot. Sur le mlange des gaz. *Comptes rendus hebdomadaires des sances de lAcadmie des Sciences*, 126:1703–1855, 1898.

[74] Rupal Agrawal and David A. Kofke. Thermodynamic and structural properties of model systems at solid-fluid coexistence. *Molecular Physics*, 85(1):23–42, 1995.

[75] J.A. Barker, D. Henderson, and F.F. Abraham. Phase diagram of the two-dimensional lennard-jones system; evidence for first-order transitions. *Physica A: Statistical Mechanics and its Applications*, 106(1):226 – 238, 1981.

[76] James M. Phillips, L. W. Bruch, and R. D. Murphy. The twodimensional lennardjones system: Sublimation, vaporization, and melting. *The Journal of Chemical Physics*, 75(10):5097–5109, 1981.

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

[78] Cun Feng Fan and Tahir Cain. Wetting of crystalline polymer surfaces: A molecular dynamics simulation. *The Journal of Chemical Physics*, 103(20):9053–9061, 1995.

[79] Troy D Loeffler and Bin Chen. Surface induced nucleation of a lennard-jones system on an implicit surface at sub-freezing temperatures: A comparison with the classical nucleation theory. *The Journal of chemical physics*, 139(23):234707, 2013.

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

[81] Bernd Ensing, Marco De Vivo, Zhiwei Liu, Preston Moore, and Michael L. Klein. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Accounts of Chemical Research*, 39(2):73–81, 2006. PMID: 16489726.

[82] Wei Shi and Edward J. Maginn. Atomistic simulation of the absorption of carbon dioxide and water in the ionic liquid 1-n-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide ([hmim][tf2n]. *The Journal of Physical Chemistry B*, 112(7):2045–2055, 2008. PMID: 18217747.

[83] Hyung-June Woo, Aaron R. Dinner, and Benot Roux. Grand canonical monte carlo simulations of water in protein environments. *The Journal of Chemical Physics*, 121(13):6392–6400, 2004.

[84] Athanassios Z. Panagiotopoulos. Direct determination of phase coexistence properties of fluids by monte carlo simulation in a new ensemble. *Molecular Physics*, 61(4):813–826, 1987.

[85] Mihaly Mezei. A cavity-biased (t, v,) monte carlo method for the computer simulation of fluids. *Molecular Physics*, 40(4):901–906, 1980.

[86] N. A. Busch, M. S. Wertheim, Y. C. Chiew, and M. L. Yarmush. A monte carlo method for simulating associating fluids. *The Journal of Chemical Physics*, 101(4):3147–3156, 1994.

[87] J. K. Brennan*. Cavity-bias sampling in reaction ensemble monte carlo simulations. *Molecular Physics*, 103(19):2647–2654, 2005.

[88] Scott Wierzchowski and David A. Kofke. A general-purpose biasing scheme for monte carlo simulation of associating fluids. *The Journal of Chemical Physics*, 114(20):8752–8762, 2001.

[89] Scott Wierzchowski and David A. Kofke. {UB} association bias algorithm applied to the simulation of hydrogen fluoride. *Fluid Phase Equilibria*, 194197:249 – 256, 2002. Proceedings of the Ninth International Conference on Properties and Phase Equilibria for Product and Process Design.

[90] Samuel J. Keasler, Hyunmi Kim, and Bin Chen. Sign preference in ion-induced nucleation: Contributions to the free energy barrier. *The Journal of Chemical Physics*, 137(17):174308, 2012.

[91] D.J. Naresh and Jayant K. Singh. Virial coefficients and inversion curve of simple and associating fluids. *Fluid Phase Equilibria*, 279(1):47 – 55, 2009.

[92] Sandip Khan and Jayant K. Singh. Prewetting transitions of one site associating fluids. *The Journal of Chemical Physics*, 132(14):144501, 2010.

[93] Ganesh Kamath, Feng Cao, and Jeffrey J. Potoff. An improved force field for the prediction of the vaporliquid equilibria for carboxylic acids. *The Journal of Physical Chemistry B*, 108(37):14130–14136, 2004.

[94] Ali Reza Mehrabi and Muhammad Sahimi. Cluster conformations and multipole distributions in ionic fluids. i. two-dimensional systems of mobile ions. *The Journal of Chemical Physics*, 128(23):234503, 2008.

[95] G. R. Birkett and D. D. Do. Simulation study of water adsorption on carbon black: the effect of graphite water interaction strength. *The Journal of Physical Chemistry C*, 111(15):5735–5742, 2007.

[96] Samuel J. Keasler, Hyunmi Kim, and Bin Chen. Ion-induced nucleation: The importance of ionic polarizability. *The Journal of Physical Chemistry A*, 114(13):4595–4600, 2010. PMID: 20235500.

[97] Bin Chen, J. Ilja Siepmann, and Michael L. Klein. Simulating vaporliquid nucleation of water: a combined histogram-reweighting and aggregation-volume-bias monte carlo investigation for fixed-charge and polarizable models. *The Journal of Physical Chemistry A*, 109(6):1137–1145, 2005. PMID: 16833423.

[98] Jrn Ilja Siepmann and Daan Frenkel. Configurational bias monte carlo: a new sampling scheme for flexible chains. *Molecular Physics*, 75(1):59–70, 1992.

[99] Fernando A. Escobedo and Juan J. de Pablo. Extended continuum configurational bias monte carlo methods for simulation of flexible molecules. *The Journal of Chemical Physics*, 102(6):2636–2652, 1995.

[100] Marcus G Martin and J Ilja Siepmann. Novel configurational-bias monte carlo method for branched molecules. transferable potentials for phase equilibria. 2. united-atom description of branched alkanes. *The Journal of Physical Chemistry B*, 103(21):4508–4517, 1999.

[101] Collin D Wick and J Ilja Siepmann. Self-adapting fixed-end-point configurational-bias monte carlo method for the regrowth of interior segments of chain molecules with strong intramolecular interactions. *Macromolecules*, 33(19):7207–7218, 2000.

[102] J. C. Shelley and G. N. Patey. A configuration bias monte carlo method for water. *The Journal of Chemical Physics*, 102(19):7656–7663, 1995.

[103] Michele Vendruscolo. Modified configurational bias monte carlo method for simulation of polymer systems. *The Journal of Chemical Physics*, 106(7):2970–2976, 1997.

[104] Marshall N. Rosenbluth and Arianna W. Rosenbluth. Monte carlo calculation of the average extension of molecular chains. *The Journal of Chemical Physics*, 23(2):356–359, 1955.

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

[106] Bin Chen, Hyunmi Kim, Samuel J Keasler, and Ricky B Nellas. An aggregation-volume-bias monte carlo investigation on the condensation of a lennard-jones vapor below the triple point and crystal nucleation in cluster systems: an in-depth evaluation of the classical nucleation theory. *The Journal of Physical Chemistry B*, 112(13):4067–4078, 2008.

[107] T. J. H. VLUGT, M. G. MARTIN, B. SMIT, J. I. SIEPMANN, and R. KRISHNA. Improving the efficiency of the configurational-bias monte carlo algorithm. *Molecular Physics*, 94(4):727–733, 1998.

[108] Aliasghar Sepehri, Troy D. Loeffler, and Bin Chen. Improving the efficiency of configurational-bias monte carlo: A density-guided method for generating bending angle trials for linear and branched molecules. *The Journal of Chemical Physics*, 141(7):074102, 2014.

[109] Troy D Loeffler, Aliasghar Sepehri, and Bin Chen. Improved monte carlo scheme for efficient particle transfer in heterogeneous systems in the grand canonical ensemble: Application to vapor–liquid nucleation. *Journal of chemical theory and computation*, 11(9):4023–4032, 2015.

[110] Alan M. Ferrenberg and Robert H. Swendsen. Optimized Monte Carlo data analysis. *Physical Review Letters*, 63(12):1195–1198, September 1989.

[111] Meinrat O. Andreae and Paul J. Crutzen. Atmospheric aerosols: Biogeochemical sources and role in atmospheric chemistry. *Science*, 276(5315):1052–1058, 1997.

[112] Ulrich Pschl. Atmospheric aerosols: Composition, transformation, climate and health effects. *Angewandte Chemie International Edition*, 44(46):7520–7540, 2005.

[113] Jasper Kirkby, Joachim Curtius, João Almeida, Eimear Dunne, Jonathan Duplissy, Sebastian Ehrhart, Alessandro Franchin, Stéphanie Gagné, Luisa Ickes, Andreas Kürten, et al. Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation. *Nature*, 476(7361):429–433, 2011.

[114] JT Kiehl and BP Briegleb. The relative roles of sulfate aerosols and greenhouse gases in climate forcing. *Science*, 260(5106):311–314, 1993.

[115] Manabu Shiraiwa, Kathrin Selzle, and Ulrich Pschl. Hazardous components and health effects of atmospheric aerosol particles: reactive oxygen species, soot, polycyclic aromatic compounds and allergenic proteins. *Free Radical Research*, 46(8):927–939, 2012. PMID: 22300277.

[116] J. Almeida, J. Curtius, J. Kirkby, and CLOUD Collaboration. Multi-species nucleation rates in cloud. *AIP Conference Proceedings*, 1527(1):326–329, 2013.

[117] C. T. R. Wilson. Condensation of water vapour in the presence of dust-free air and other gases. *Philosophical Transactions of the Royal Society of London A*, 189:265–307, 1897.

[118] C. T. R. Wilson. On the comparative efficiency as condensation nuclei of positively and negatively charged ions. *Philosophical Transactions of the Royal Society of London*, 193:289–308, 1900.

[119] Elena Brodskaya, Alexander P. Lyubartsev, and Aatto Laaksonen. Investigation of water clusters containing oh- and h3o+ ions in atmospheric conditions. a molecular dynamics simulation study. *The Journal of Physical Chemistry B*, 106(25):6479–6487, 2002.

[120] K. J. Oh, G. T. Gao, and X. C. Zeng. Nucleation of water and methanol droplets on cations and anions: The sign preference. *Phys. Rev. Lett.*, 86:5080–5083, May 2001.

[121] Elena Brodskaya, Alexander P. Lyubartsev, and Aatto Laaksonen. Molecular dynamics simulations of water clusters with ions at atmospheric conditions. *The Journal of Chemical Physics*, 116(18):7879–7892, 2002.

[122] Mohammad Solimannejad, Mohaddeseh Rabbani, and Mehdi D. Esrafili. Toward understanding the role of water molecules in the uptake of nitrosyl hydride by sulfuric acid aerosols: A computational study. *Computational and Theoretical Chemistry*, 1017:78 – 84, 2013.

[123] S. Schobesberger, A. Franchin, F. Bianchi, L. Rondo, J. Duplissy, A. Kürten, I. K. Ortega, A. Metzger, R. Schnitzhofer, J. Almeida, A. Amorim, J. Dommen, E. M. Dunne, M. Ehn, S. Gagné, L. Ickes, H. Junninen, A. Hansel, V.-M. Kerminen, J. Kirkby, A. Kupc, A. Laaksonen, K. Lehtipalo, S. Mathot, A. Onnela, T. Petäjä, F. Riccobono, F. D. Santos, M. Sipilä, A. Tomé, G. Tsagkogeorgas, Y. Viisanen, P. E. Wagner, D. Wimmer, J. Curtius, N. M. Donahue, U. Baltensperger, M. Kulmala, and D. R. Worsnop. On the composition of ammoniasulfuric-acid ion clusters during aerosol particle formation. *Atmospheric Chemistry and Physics*, 15(1):55–78, 2015.

[124] Renyi Zhang, Paul J. Wooldridge, Jonathan P. D. Abbatt, and Mario J. Molina. Physical chemistry of the sulfuric acid/water binary system at low temperatures: stratospheric implications. *The Journal of Physical Chemistry*, 97(28):7351–7358, 1993.

[125] T. Kurtén, V. Loukonen, H. Vehkamäki, and M. Kulmala. Amines are likely to enhance neutral and ion-induced sulfuric acid-water nucleation in the atmosphere more effectively than ammonia. *Atmospheric Chemistry and Physics*, 8(14):4095–4103, 2008.

[126] Antony D. Clarke, Steven R. Owens, and Jingchuan Zhou. An ultrafine sea-salt flux from breaking waves: Implications for cloud condensation nuclei in the remote marine atmosphere. *Journal of Geophysical Research: Atmospheres*, 111(D6):D06202, 2006. D06202.

[127] S. L. Gong, L. A. Barrie, and J.-P. Blanchet. Modeling sea-salt aerosols in the atmosphere: 1. model development. *Journal of Geophysical Research: Atmospheres*, 102(D3):3805–3818, 1997.

[128] C. GENTHON. Simulations of desert dust and sea-salt aerosols in antarctica with a general circulation model of the atmosphere. *Tellus B*, 44(4):371–389, 1992.

[129] Troy D. Loeffler, Aliasghar Sepehri, and Bin Chen. Improved monte carlo scheme for efficient particle transfer in heterogeneous systems in the grand canonical ensemble: Application to vaporliquid nucleation. *Journal of Chemical Theory and Computation*, 11(9):4023–4032, 2015. PMID: 26575898.

[130] Chengteh Lee, Carlos Sosa, Marc Planas, and Juan J. Novoa. A theoretical study of the ionic dissociation of hf, hcl, and h2s in water clusters. *The Journal of Chemical Physics*, 104(18):7081–7085, 1996.

[131] Angela Smith, Mark A. Vincent, and Ian H. Hillier. Mechanism of acid dissociation in water clusters: electronic structure studies of (h2o)nhx (n = 4, 7; x = oh, f, hs, hso3, ooso2h, oohso2). *The Journal of Physical Chemistry A*, 103(8):1132–1139, 1999.

Vita

Troy Loeffler graduated from Washington State University in 2011 with a bachelor's degree in chemistry and a second bachelor's in theoretical mathematics. He has been a part of the LSU Chemistry department for the last five years and has been working toward finishing his doctoral degree with a specialization in chemical theory and computations.