12-2018

Molecular Dynamics Investigation of the Thermodynamics of Nanoscale Droplets and Kinetics of Weakly Correlated Systems

Kai Yang Leong
University of Arkansas, Fayetteville

Molecular Dynamics Investigation of the Thermodynamics of Nanoscale Droplets and Kinetics of Weakly Correlated Systems

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Chemistry

by

Kai-Yang Leong
Boston University
Bachelor of Science in Chemistry, 2012

December 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Feng Wang, Ph.D.
Dissertation Director

_____ _____
Colin Heyes, Ph.D. Jingyi Chen, Ph.D.
Committee Member Committee Member

_____
Peter Pulay, Ph.D.
Committee Member

Abstract

The surface tension of nanoscale droplets of water was studied with molecular dynamics simulations using the BLYPSP-4F water potential. The internal pressure of the droplet was measured using a correlation between the pressure and density, established through a series of bulk simulations performed at pressures from 1 to 1000 bar. Such a procedure allows for reliable determination of internal pressure without the need to calculate the local Virial. The surface tension, estimated with the Young-Laplace relation, shows a good agreement with the Tolman equation with a Tolman length of -0.48 Å. The interface of a liquid water droplet is shown to be around 1.1 to 1.3 nm thick depending on radii. The fairly thick interface region put an upper limit on the size of droplets that still have a bulk-like interior.

The effect for removing weak longtime correlation is studied using a model system that contains a driven atom at liquid density under strong thermal fluctuations. The force that drives the tagged particle is about 1% the average random force experienced by the particle. The tagged particle is allowed to assume a range of masses from 1/8 to 800 times that of a surrounding particle to study the effects of inertia. The driving force is indefinitely correlated but much weaker than “random” fluctuations from the environment. From this study, it is shown that the environmental influence is not fully random leading to the force autocorrelation function being a poor metric for detecting the correlated driving force. For systems with small inertia, our study reveals that discarding longtime correlation has negligible influence on the first passage time (FPT) estimate, whereas for particles with large inertia, the deviation can indeed be appreciable. It is interesting that the Markov State Model (MSM) still produces reasonable estimates on the FPT even when a very short lag time that clearly violates the Markovianity assumption is used. This is likely a result of favorable error cancellations when the MSM transition matrices were constructed with trajectories much longer than the lag time.

Acknowledgements

I believe, someday, I will look back on these 5 ½ years and realized it was a turning point in my life. I'm leaving Arkansas a very different man than when I first came. The 3 drivers for this tempering are Dr. Wang, my mother and God. Thank you, Dr. Wang, for taking a chance on me despite having no quantitative and prior programming experience. You have taught me how to be an independent thinker and you have pushed me to develop the quality of my thinking and my technical prowess. I'm very grateful towards you for helping to prepare me for the private sector. Mom, thank you for believing in me and supporting me even when you did not understand or agree with the rationale behind my actions sometimes. And finally, thank you Jesus for carrying me through the darkest moments in my life during this period. Your faithfulness came through. Your love overwhelming.

I would also like to thank my committee members for their support. A special note of appreciation to Jicun who has been a tremendous blessing to me. My friends, both far and near, who have been very encouraging and supportive to me all these years. Thank you.

Table of Contents

List of Published Papers

1. Chapter 1: A molecular dynamics investigation of the surface tension of water nanodroplets and a new technique for local pressure determination through density correlation

2. Chapter 2: On approximating a weak Markovian process as Markovian: are we justified when discarding longtime correlations

Thesis Introduction

In the last paper published by fame Chemist/Physicist, Richard C. Tolman, he postulated that the surface tension of small droplets or bubbles could be express as a function of radius. In the last 70 years, we have observed a myriad of seminal works contributed by the scientific community in studying the properties of these tiny droplets. These droplets occur in many natural phenomenon, biological processes and are involve in various cutting edge industrial procedures. In this work, we show the evolution of our investigation in relating the surface tension of nanodroplets to its radius. We start from first principles in establishing certain important known physical properties of the water model used in the simulation such as its planar surface tension and then we move on to finding an order parameter that correlates strongly to its internal pressure. Using its internal pressure, we are able to use the Young-Laplace Equation to compute the surface tension for a given droplet size and its corresponding curvature, thus, shedding light on Tolman's original hypothesis several years ago.

It is of much interest both in academic circles and in the private sector to be able to develop a statistical formulism for large data set problems. Such statistical approach often requires certain tenets to be fulfilled or at least to be weak enough such that its effects can be assumed to be negligible hence allowing for the application of well-developed statistical tools to be applied to derived ensemble observables. One such condition is the long-time correlation of a deterministic system is assumed to be negligible even though the system may still be weakly correlated in the application of certain stochastic models such as the popular Markov State Model. Here, we show that discarding long-time correlation cause by a driving force smaller than thermal fluctuations can introduce a significant error in ensemble kinetic observables in a model system. This is meant to direct greater attention to the treatment of long-time correlation for certain important processes that do discard long-time correlation.

A molecular dynamics investigation of the surface tension of water nanodroplets and a new technique for local pressure determination through density correlation

1.1. Introduction

Nanoscale droplets play an important role in nucleation processes, such as cloud formation.[1, 2] The fascinating chemistry of nano-reactors[3] and nano-lithography requires an understanding of the properties and formation of nanoscale droplets. Structure of proteins studied by free electron X-ray laser[4] could potentially be influenced by the internal pressure of micrometer to nanometer sized droplets that solvate the proteins. The surface tension of the droplets directly influences the stability of the critical nucleus according to the classical nucleation theory,[5-9] the equilibrium vapor pressure through the Kelvin equation,[10-12] and the internal pressure through the Laplace equation.[10, 13, 14] It is thus very important to have a detailed understanding of the surface tension of the highly curved surfaces.[10]

While the tension of planar surface has been studied extensively,[15, 16] the surface tension of highly curved nanodroplets is more challenging.[17] Highly convex surfaces have a high vapor pressure. Nanodroplets thus evaporate quickly even at the dew point, making it challenging to investigate experimentally. The experimental surface tension of small droplets is sometimes inferred from classical nucleation theory and is subjected to other approximations.[18, 19] Theoretical calculations are challenging due to the nontrivial definition of a local pressure for a microscopic system. For example, the frequently used Irving-Kirkwood[20] and Harasima[21] approaches show instability and ambiguity,[22, 23] with measured local pressure depending strongly on the choice of parameters.[24] Although the slice-average pressure method[25] and several versions of test area methods[26-28] show improved stability, a conclusive understanding regarding the surface tension of nanodroplets is far from achieved.

The curvature dependent surface tension is frequently described by the Tolman equation,[29]

$$\sigma(R) = \sigma_{\infty}\left(\frac{1}{1+2\delta/R}\right), \quad (1)$$

where the Tolman length, δ, is the distance between the Gibbs surface, also referred to as the equimolar surface, and the surface of tension. While modeling based on the classical density functional theory[17, 30, 31] predicts a negative Tolman length for Lennard Jones droplets and liquid water, simulations gave differing predictions with either positive or negative Tolman lengths[32, 33] depending on the way the local pressure is determined. By measuring the free energy of mitosis,[34] Joswiak et al. determined a negative Tolman length for water,[35] Lau[26] concluded the Tolman length is positive with a test area method. Interestingly, the Lau and Joswiak studies reached differing conclusions even with the same water model, namely TIP4P/2005.[36] Such disagreement further indicates challenges in the methods of determination rather than simply differences in water models used. Furthermore, some studies predicted a breakdown of a first order Tolman's equation as shown in Eq. 1,[37] while others insisted that the surface tension should be independent of the curvature leading to a Tolman length of zero.[38]

In this work, we use a unique method to measure the internal pressure of a droplet through a proxy. Our pressure determination is intuitive, robust, and numerically stable. Thus our investigation of the curvature dependent surface tension will shed light on the hotly debated and important topic of the physical properties of nanoscale droplets. Although we will restrict our study to liquid water at the ambient temperature of 298 K, our pressure determination method can be readily applied to other systems, such as Lennard-Jones droplets, and oil-droplets in water.

The chapter is organized in 9 main sub-sections. The selection of the chosen force field for this particular study will be investigated in Section 1.2. The direct virial pressure measurement method will be described in Section 1.3. The Finite Difference method in the determination of the droplet's internal pressure will be investigated in section 1.4. and the intramolecular virial method will be explored in Section 1.5. Finally, the method for our

determination of the droplet internal pressure by density correlation will be described in Section 1.6. Computational details for the nanodroplets are provided in Section 1.7., and the results and discussion will be presented in Section 1.8. The summary of the chapter will be consolidated in Section 1.9.

1.2. Surface Tension of Slab Water using BLYPSP-4F model

In this study, the BLYPSP-4F potential predicts the slab surface tension of liquid water to be 67.1 ± 0.3 m Nm^{-1} at 298 K. The measurement was performed with the mechanical method through the equation

$$\gamma = \frac{L_Z}{2}(P_Z - \frac{P_X + P_Y}{2}) \quad (2)$$

Where L_Z is the Z dimension of the box and the factor of 2 on the left is due to the presence of two liquid-vapor interfaces in a slab. The water slab contains 5000 molecules and is continuous in the X-Y plane. The X and Y dimensions were held at 5.3053 nm with the Z dimension being 15.0 nm. This allows a vacuum region of approximately 9.6947 nm. In order to use a 1 fs time-step, the hydrogen isotope of Tritium was chosen with a mass of 3.016 $g.mol^{-1}$. This choice of a heavier isotope should not influence the surface tension in a Newtonian MD simulation. The van der Waals interactions were truncated beyond 1.75 nm and the long range electrostatics were treated with the particle-mesh Ewald method. Each simulation trajectory is 6 ns long and statistics were collected only after the first ns. Up to 10 independent trajectories were performed in each simulation condition to reduce the error bar.

The BLYPSP-4F model was developed by force-matching electronic structure calculations without fitting to any experimental property and its potentials were fitted to coupled cluster quality forces obtained with the BLYPSP method. Only liquid configurations were used in the parameterization of the BLYPSP-4F potential. Finally, the BLYPSP-4F water model gives

better Quantum ΔH_{vap}, D and ε_s than the TIP4P/2005 model. The BLYPSP-4F water model does give a surface tension comparable and marginally better than the TIP4P/2005, underestimating the experimental value by 4/5 % at 300 K. At temperature below T_B, BLYPSP-4F is arguably the best model for liquid water.

1.3. Direct pressure measurement of Trialsphere by virial calculation

In molecular dynamics, the equilibrium pressure of a system is generally obtained through the Virial pressure,[39] which for a three dimensional system is defined as

$$p = \frac{nRT}{V} + \frac{< \sum_i F_i \cdot r_i >}{3V}, \qquad (3)$$

where the angle bracket represents the ensemble average and the sum is over the dot product of the atomic force F_i, and atomic coordinate r_i vectors. Eq. 3 can be derived by taking the derivative of the Helmholtz free energy with respect to volume through a homogenous affine scaling factor *s*, where $r_i(s) = s \cdot r_i$ and $V(s) = s^3 V$. Complications arise for periodic systems, where for pair-wise potentials, Eq. 2 becomes

$$p = \frac{nRT}{V} + \frac{< \sum_{i<j} F_{ij} \cdot r_{ij} >}{3V} - < \frac{\partial U}{\partial V} > \qquad (4)$$

where r_{ij} is the shortest distance between two atoms *i* and *j*. F_{ij} are the pairwise forces acting through r_{ij} and the last term arises from a periodic treatment of long range interactions. With the typical Ewald summation of infinite electrostatics, the last term is the contribution to pressure from the reciprocal space sum.

If one further assumes isotropic central forces and ignores the contribution from the reciprocal space electrostatics, a local definition of the pressure can be written as

$$p_{IK} = \rho_n RT - \frac{2\pi}{3} \rho_n^{\,2} \int_0^\infty \frac{\partial V}{\partial r} \cdot r \cdot g(r) \cdot r^2 dr, \qquad (5)$$

where ρ_n is the local number density and $g(r)$ is the radial distribution function. Eq. 5 is generally referred to as the Irving-Kirkwood[20] definition of local pressure. We note that in bulk liquid, the pressure is isotropic. For a slab or near a droplet interface, a tensor form of Eq. 4 has to be used to calculate the normal and transverse components of p_{IK}.

Although the tensor form of Eq. 5 replaces $4\pi r^2 g(r)$ with a full two-body density correlation that includes the orientation information, the tensor form of Eq. 5 based on Irving and Kirkwood's 1950 classic[20] still assumes intermolecular forces to have central symmetry and is thus problematic for treating more complex systems, such as liquid water. For non-isotropic molecules, the relative orientations influence the energy derivatives and will require a treatment that involves explicit considerations of angular gradients of the energy. Even for systems with isotropic central potentials, such as Lennard Jones fluids, the tensor version of Eq. 5 shows strong dependence of the normal and transverse pressure on the cutoff used for the integration.[24]

Other methods for estimating pressure based on variants of Virial based approaches exist,[20, 21, 40] also internal pressures have been measured with infinitesimal local deformations[41] through the free energy of droplet formation,[42] or through coarse graining.[43] In our investigation, we will examine first a direct virial approach.

In this particular study, we will utilize the Gromacs Simulation Package due to its speed. All our methods can be used with any MD package that support NVT simulations of a droplet in a periodic cell, and is able to calculate the system Virial.

For a droplet at equilibrium in a periodic cell, a direct virial measurement of its pressure will yield an average of zero. This is because the force exerted by its internal pressure is canceled out by the restoring action of its surface tension. An intuitive and simple approach is to remove the interface layer of the droplet until we reach the interior of the droplet. This removal

of the interface layer will reveal a sphere, almost perfectly spherical, which we denote as a trialsphere.

The trialsphere is defined with these properties:

I. The center of the trialsphere is defined as the same position as the center of mass of the entire droplet.

II. The radius of the trialsphere extends from its center as defined in **I.** outwards towards its surface

III. A point on the water molecule is chosen to be the location of water. If the distance between this point and the center of the trialsphere is less than the defined trialsphere radius, this water molecule is included in the definition of the trialsphere. If the distance between this point and the center of the trialsphere is larger than the defined trialsphere radius, the molecule is removed from the trialsphere.

Such a formal definition of a trialsphere allows one to formally define a cutoff radius for each trialsphere size. Applying the virial calculation to each configuration of individual trialsphere would enable us to calculate its instantaneous pressure.

The results of the calculations produce a negative pressure for all trialsphere sizes. This is because when the interface layer is removed leaving the trialsphere expose, it forces the trialsphere to reform a new interface. This reformation of a new interface layer causes the droplet to implode upon itself giving rise to the massive negative pressure.

An alternative way to investigate the droplet's internal pressure without removing its interface layer exposing its trialsphere is by keeping the entire droplet intact and scaling the coordinates of the water molecules within the trialsphere. This small change in its calculated energies can be express as a taylor's expansion and by use of a small change in expansion and

contraction in volume of its trialsphere within a whole droplet, an expression for its internal pressure can be derived.

1.4. The Finite Difference method

The Finite Difference Method stems from the application of a Taylor's series expansion to the definition of the potential energy E of a system with respect to its volume V. For the volume expansion of a system, its energy can be express as:

$$E(V^{+}) = E_o + (-P)\,dV + \frac{1}{2}\frac{d^2E_o}{dV^2}dV^2 + \frac{1}{6}\frac{d^3E_o}{dV^3}dV^3 + \cdots \quad (6)$$

$$\text{Since } P = -\frac{dE_o}{dV}$$

In order to find a simple relationship between the configuration pressure of the system at a particular radius within the droplet and its potential energy, we manipulate the Taylor's series expansion of the potential energy of the system with respect to its volume by the following series of steps:

i) Finding an expression for both the expansion and contraction of volume in the expression.

ii) Looking for cancellation of terms by solving simultaneous equation to produce a simple equation expressing the configuration pressure of the droplet as a function of potential energy and small change in volume.

Taylor series expansion of volume

$$E(V^{+}) = E_o + (-P)\,dV + \frac{1}{2}\frac{d^2E_o}{dV^2}dV^2 + \frac{1}{6}\frac{d^3E_o}{dV^3}dV^3 + \cdots$$

Taylor series contraction of volume

$$E(V^-) = E_o + (-P)(-dV) + \frac{1}{2}\frac{d^2E_o}{dV^2}(-dV)^2 + \frac{1}{6}\frac{d^3E_o}{dV^3}(-dV)^3 + \cdots$$

Combining both equations

$$dE = E(V^+) - E(V^-) = -2P\,dV + 2[\frac{1}{6}\frac{d^3E_o}{dV^3}dV^3] + \cdots \quad (7)$$

A scaling factor of 1.00001 and 0.99999 was chosen in the expansion and contraction of the atom coordinates within the trialsphere. The small change in volume, dV, for terms with $3rd$ order and higher can be approximated to 0. We arrive at a new relationship following the approximation:

$$E(V^+) - E(V^-) \approx -2P\,dV \quad (8)$$

where the difference between the potential energy of the scaled volume expansion and contraction is approximately proportional to the pressure at a given small change in trialsphere volume. We note that the small change in volume, dV, is defined based on the coordinates of the atoms scaled within the trialsphere.

In order to treat potential problems of the droplet possibly moving across the periodic boundary during MD simulation, each water molecule in the droplet is first translated by the coordinates of the entire droplet's approximate center of mass. This will shift the entire droplet such that the new center of mass of the droplet is now at the origin. The trialsphere radius is align to originate at the origin of the box and all water molecules whose location of water fall within the trialsphere radius is kept within the trialsphere. Water molecules that fall outside the trialsphere radius is then separated and held temporarily called the excess layer. The coordinates of all atoms within the trialsphere is then scaled by a factor of 1.00001 and 0.99999

in two different sets. Both the excess layer and the trialsphere are recombined to reform the droplet. The droplet is now ready for a 0 step Mdrun to compute its potential energy for the scaled trialsphere expansion and contraction within the droplet.

A pressure profile can be set up by employing the finite difference method to a series of different trialsphere sizes. The pressure profiles for 4, 6, 8 and 10 nm diameter nanodroplets were calculated. Regardless of nanodroplet sizes where the finite difference was employed, it gave consistently the same pressure of approximately 1500 bar.

This internal pressure calculated is massive and uniform despite different droplet sizes. This is unphysical because when the droplet size increases, its surface tension should converge towards its planar surface tension. A uniform internal pressure across different droplet size is contrary to what we know experimentally. This illustrates the immense challenge for determining accurate internal pressure for nanodroplets.

The Finite Difference Method investigates how the potential energy of the system is affected by atomistic scaling of particle's coordinates in the trialsphere. Hence, the forces which dominate such interactions are mostly intermolecular forces. Potentially, an alternative method to determine the internal pressure of the droplet could be found in the intramolecular interactions within the droplet.

1.5. Intramolecular Virial method

From equation 3, we can represent the terms of the pressure equation into something more intuitive:

$$P_{\mathrm{NET}} = P_{\mathrm{KINETIC}} + P_{\mathrm{CONFIGURATION}}$$

where P_{NET} is the system pressure of the MD system of interest, P_{KINETIC} is the kinetic energy contribution to the pressure and $P_{\mathrm{CONFIGURATION}}$ is the potential energy contribution to the pressure.

For an NVT system, P_{KINETIC} is constant, that means that P_{NET} is directly proportional to $P_{\mathrm{CONFIGURATION}}$ and $P_{\mathrm{CONFIGURATION}}$ is defined as:

$$P_{\mathrm{CONFIGURATION}} = \frac{2}{V}\Xi$$

where $\Xi = \sum_{i<j}^{N}\langle r_{ij} \,.\, f_{ij}\rangle$, f includes both intramolecular and intermolecular forces and r is the distance between atom *i* and *j*.

The virial can be separated into its intramolecular and intermolecular terms

$$P_{\mathrm{CONFIGURATION}} = \frac{2}{V}\,[\Xi(\mathrm{intramolecular}) + \Xi(\mathrm{intermolecular})]$$

Based on an earlier attempt to calculate the internal pressure using the Finite Difference method, it was concluded that the intermolecular virial is not viable in determining the internal pressure of the droplet at very small nanometer scale. Therefore, we assume that the intermolecular virial can be treated as a constant. Hence,

$$P_{\mathrm{CONFIGURATION}} = \frac{2}{V}\,[\Xi(\mathrm{intramolecular}) + \mathrm{constant}]$$

where $\Xi(\mathrm{intramolecular})$ is the aggregate of all the intramolecular virial per water molecule within the trialsphere. As a sum total, $\Xi(\mathrm{intramolecular})$ does not provide any insights into the physical properties that the water molecule is going through under varying pressures. Experimentally, water will compress by only 46.4 parts per million for every unit increase in atmospheric pressure (bar) since water is not a very compressible fluid. At around 4000 bar (58,000 psi) of pressure, at room temperature, water only experiences an 11% decrease in

volume. Hence, if there is a change in intramolecular virial per water molecule as the droplets get smaller, we can infer that the pressure is significantly larger than expected as the size of the droplet decreases.

We are interested in finding the relationship between the intramolecular virial of a trialsphere and how it correlates to its internal pressure, we propose using a series of bulk water simulations under varying barostat pressure to establish the intramolecular virial – pressure parameterization. This would potentially work because if the focus is on its intramolecular physics, then the configuration of the droplet's interior(trialsphere) is not required in establishing such a correlation.

From Figure 1, the tabulated bond energy and Figure 2, the bond angle energy is plotted against varying bulk water barostat pressures. From both plots, it can be seen that there is a better bond angle energy and pressure isotherm. It can be argued that such an isotherm can be parameterized and used to predict the droplet's internal pressure. However, we differ to experimental evidence in the literature to expand our search for a more robust and intuitive proxy for pressure calibration.

1.6. Determination of the droplet internal pressure by density correlation

In this work, we measure the internal pressure by establishing a correlation similar to the way pressures are measured in an experiment.

In an experiment, the internal pressure of a reaction vessel is typically measured through a proxy. For example, a probe molecule with a stable linear response that is correlated to the pressure can be used as the proxy. The nuclear quadrupole resonance (NQR) frequency of Cu_2O or the R1 line of the Ruby florescence is known to have a linear response with the pressure.[44] In a MD simulation, we fortunately do not have to explicitly add a foreign probe. We could use a property of water that shows a simple monotonic response to the internal pressure

and use it as a proxy to indirectly determine the pressure. A few possible order parameters, such as the HOH angle or the intramolecular contribution to the Virial, have been tested, we found the liquid density provide the simplest order parameter for pressure determinations.

In order to do this, an analytical expression of pressure as a function of density has to be fitted first. Such a calibration of the pressure-density isotherm can be accomplished with a series of MD simulations performed in the bulk phase, where both the pressure and density can be known precisely.

It is important that the Coulombic and van der Waals interactions are treated in a consistent manner for both the calibration and the final pressure measurements. In this study, the BLYPSP-4F water model[45] was used for our simulations with a 1 fs time step and a hydrogen isotope mass of 3.016 g/mol. The larger isotope mass allows the use of the 1 fs time step size and does not affect any thermodynamic properties within classical statistical mechanics.[46, 47] The BLYPSP-4F water model was chosen since it gives good surface tension, dielectric constant, radial distribution functions, and other properties of liquid water.[48] In addition, it is worth noting that the BLYPSP-4F model was created with a fully first-principles based parameterization process without fitting to any experimental properties.[45]

For both the calibration and the final pressure measurements, the coulombic interactions were treated with a 6th order particle mesh Ewald method[49] using a Fourier spacing of 0.156 nm and a real space Coulombic cutoff of 1.3 nm. A simple cutoff scheme was used for modeling the van der Waals interactions with a cutoff of 1.75 nm without correction for long-range energy and stress.[50] The use of a large van der Waals cutoff had been shown to be important for surface tension studies[51] and the long-range correction to the stress tensor[50] is not considered since it would be inappropriate to use such a correction for the droplet simulation due to the lack of translational symmetry.

For the determination of the pressure and density correlation, we used a cubic box with 5000 water molecules at a series of seven pressures: 1 bar, 150 bar, 300 bar, 450 bar, 600 bar,

750 bar, and 1000 bar. The pressure for the box was regulated by the Parrinello-Rahman barostat[52, 53] with a relaxation constant of 5.0 ps. At each pressure, 6 ns of MD simulations were performed and one configuration is saved every 0.5 ps. The density and pressure calibrations were performed with the 12,000 saved configurations only. Table 1 reports the Virial pressure calculated for the saved configurations and that regulated by the thermostat showing good agreement between the two pressures.

From Figure 3, it is interesting to note that the density and pressure isotherm is not completely linear, even though the linearality of the isotherm is still better than either the R1 shift of the Ruby fluorescence or the NQR frequency of Cu_2O used experimentally to infer internal pressure.[44] We decided to fit the pressure-density correlation to a quadratic equation to account for the slight reduction of the isothermal compressibility as pressure increases.

The fit of the density to the pressure leads to the following equation,

$$P / bar = 58.2143(\rho_n \cdot nm^3)^2 - 3391.22\rho_n \cdot nm^3 + 48290.2 \qquad (9)$$

where ρ_n is the number density in molecules per nm^3, and the pressure is in bar.

As clear from Figure 3 and Table 1, the fitted pressure is in close agreement with the measured reference pressure showing a root mean square error (RMSE) of approximately of 2.5 bar, which is smaller than the standard error of the measured pressure of about 4 bar. Thus, the uncertainty of the fit is smaller than the uncertainty of a typical pressure determination using 12,000 uncorrelated frames.

1.7. Computational Details

Most phenomenological equations on droplets assume that the liquid inside a droplet is identical to that of the bulk liquid. Some validation regarding this assumption will be provided below. In our investigation, we will assume that sufficiently deep inside the droplet, water behaves like bulk water. With an isotropic internal pressure, the Young-Laplace equation,

$$\Delta P = \frac{2\sigma}{R} \tag{10}$$

can be used to measure the surface tension, σ. In this equation, ΔP is the pressure difference between the bulk-like center and the vapor, and R is the radius of the droplet.

In this study, we created a total of 6 nanodroplets in the range from 4 nm to 16 nm in diameter. Assuming an equilibrium density of 54.82 mol/L at 1 bar, the number of water molecules to include in each droplet was chosen so that a hypothetical sphere with a 1 bar internal pressure and an infinitely sharp spherical surface will have the target radius. Starting from a water cube, each droplet is equilibrated for at least 10 ns until the density profile no longer changes.

We note that the target radii are slightly different from the true Gibbs radii since the internal pressure of the sphere is a function of the droplet size and will not be 1 bar for the small droplets investigated in this work. Table II shows the number of water molecules, the length of the production simulation, and the radius of the Gibbs surface for all the droplets investigated in this study.

Figures 4a and 4c show the density of the 16 nm and 4 nm diameter droplets as a function of the distance from the center of the sphere. It can be seen that the radial density is uniform in the center of the droplet and drops at the liquid-vapor interface. The thickness of the interface layer can be seen clearly by plotting the difference of the radial density to the average density at droplet center. (Figure 4b and 4d) For the 16 nm diameter droplet, the interface is around 1.3 nm thick and this value reduces to approximately 1.15 nm for the 4 nm diameter droplet. Assuming deep inside each droplet, the water is bulk-like, the density profile indicates that as far as density is concerned, water is bulk-like up to 1.3 nm from the interface. Obviously, other properties of water could potentially be influenced more by the presence of the interface.[54]

At finite temperature, the surface of the droplet fluctuates according to the thermodynamics of capillary waves. The surface of nanoscale droplets will not be a perfect

sphere. Close to the liquid-vapor interface of the droplet, the deviation of the radial density from the bulk values is most likely a result of such surface fluctuations. We note that the surface tension can be obtained by measuring capillary wave-fluctuations,[55, 56] the smaller apparent thickness of the interface for the smaller droplets is consistent with smaller droplets having larger surface tension as reported in Sec. 1.6.

Since the density profile is uniform up to at least 1.3 nm from the Gibbs surface, we will measure the density using all the water in the center up to 1.5 nm from the surface. Technically an even smaller sphere closer to the center can be used due to the uniform density profile. However, we want to use as many water molecules as possible without getting too close to the interface in order to reduce statistical noise.

Since the density-pressure calibration is performed on bulk water, the reliability of the method depends on the extent water in the interior of the droplet sufficiently resemble that of the bulk water. It is worth noting that the rate at which density changes as a function of pressure

$$\frac{\partial \rho_n}{\partial P} = \rho_n \cdot \kappa_T \,, \tag{11}$$

is proportional to the isothermocompressibility,

$$\kappa_T = -\frac{1}{V}\frac{\partial V}{\partial P} \tag{12}$$

which can be calculated with the fluctuation-dissipation theorem,

$$\kappa_T = \frac{V}{k_B T} \cdot \frac{< \mathrm{N}^2 > - < \mathrm{N} >^2}{< \mathrm{N} >^2} = \frac{1}{k_B T} \cdot \frac{< \mathrm{V}^2 > - < \mathrm{V} >^2}{< \mathrm{V} >} \tag{13}$$

With the pressure-density correlation in Eq. 9, the isothermal compressibility can be calculated with Eq. 11 at any pressure between 1 to 1000 bar. At the internal pressure of the 8 nm droplet of 170.1 bar, the bulk water compressibility should be 5.46 x 10^{-5} bar^{-1} based on the correlation. The measured compressibility of this droplet is 5.58 ± 0.2 x 10^{-5} bar^{-1} indicating that water in the droplet does respond to pressure changes in a way similar to bulk water. We note

that the determination of κ_T requires the measurement of the variance of particle number in a small interior volume of a droplet, which offer significantly less significant numbers than the volume. It is fairly difficult to converge κ_T for small droplets. Thus the comparison was only performed for the largest droplet.

We further validated the bulk interior assumption by confirming the presence of translational symmetry and the lack of spatial correlation in water orientations. In order to show that there is translational symmetry close to the droplet center, we calculated the radial density using both the center of the mass of the droplet and an arbitrary point 1 nm away from the center in the positive z direction as origins. As shown in Figure 5, the interior radial density of the droplet is the same regardless of the origin of the measurement. Since the distance to the interface depend on the choice of origin, the measured radial density will start to deviate from the interior value when the density is measured close to the interface.

In order to check whether there is any spatial correlation between the orientation of the water and the radius of the droplet, we measured the density of water using three different reference points for the 4 nm and 16 nm droplets. Namely, we measured the number density of water using the location of the oxygen, the center of geometry, and one of the two hydrogen atoms as the point of reference. If there is any preferential water orientation in the center, the measured density using the three methods should differ. For example, if the water has a preferential orientation of pointing the hydrogens outwards toward the interface, a small difference will be expected when comparing the water density based on oxygen locations and hydrogen locations. As shown in Figure 6, all three density profiles show identical densities up to 1.5 nm from the interface. Thus, there is no spatial correlation between the water orientations and the droplet center.

Thus, we argue that as far as density is concerned, the interior of a droplet is similar to the bulk solution. We, thus, assume the pressure-density correlation calibrated for the bulk phase is applicable to deduce the pressure in the center of a droplet.

Using the Young Laplace equation to calculate the surface tension requires knowledge of the radius of the droplet. Although this radius should technically be the location of the surface of tension, the surface of tension is unknown. The most convenient definition, which is generally used by classical nucleation theory and other phenomenological equations, is the location of the Gibbs equimolar surface. We argue that since the droplet size is generally approximated by the Gibbs surface in practice, the tension at the Gibbs surface would be the most useful definition in practice.

1.8. Results

Table III, reports the internal pressure and the surface tension at the equimolar surface of each droplet calculated using the Young-Laplace equation. As a droplet becomes smaller, the surface tension increases, thus consistent with a negative Tolman length. It is clear in Figure 7 that the surface tension fits well to the Tolman equation even for droplets as small as 2 nm in radius. At this size, there are only slightly more than 10 water molecules across the 4 nm diameter of the droplet. The fit leads to a Tolman length of -0.48 Å and a planar surface tension of 66.4 mN/m. We note that with a infinite slab simulation, the BLYPSP-4F model gives a planar surface tension of 67.0 mN/m using the 1.75 nm van der Waals cutoff. Thus the agreement from the fit to the Tolman equation and the direct measurement using infinite slab is within the margin of error.

The negative Tolman's length of -0.48 Å is consistent with predictions based on the classical density functional theory for Lennard Jones liquids,[17, 30] where a negative Tolman length approximately half of the particle sizes was predicted. When comparing with literature values, care has to be taken since the precise value of the Tolman length might be model

dependent and will also depend on the way Tolman length is determined.[57] The small Tolman's length indicates that the use of bulk surface tension for small droplets is likely to be good approximation. Even for 4 nm diameter droplets, the true surface tension is only 5% larger than the bulk value.

Our study shows an interface thickness of 1.1 to 1.3 nm using the radial density as an indicator. For even smaller droplets with a diameter less than 4 nm, it could be argued that the surface fluctuations are on the same length scale as the whole droplet, there would be no bulk liquid inside the droplet. The very concept of a spherical droplet with a bulk-like interior may start to break down. This may have some implications when the classical nucleation theory is used to model very small droplets.

1.9. Summary

In this work, the internal pressure of liquid water was measured through a proxy, similar to experimental approaches for inferring internal pressures with probes. To the best of our knowledge, the use of proxy to measure pressure has not been done in a simulation. The best proxy will be a property of water that shows a good dynamic range and high degree of correlation with the internal pressure. Although it is still yet to be determined what the best order-parameter is to use as a proxy for pressure determination, a simple density-pressure isotherm seems to be adequate for liquid water. The density shows a good quadratic correlation with pressure in the range from 1 bar to 1000 bar at ambient temperature.

The surface tension deduced from the internal pressure using the Young-Laplace equation shows good agreement to the Tolman equation for all droplets above 4 nm in diameter with a Tolman length of -0.48 Å. Since the water-water distance in liquid water is approximately 3 Å, the negative Tolman length of around half molecule size is consistent with previous predictions from classical density functional theory for liquid water[31] and studies with the mitosis method for TIP4P/2005 and mW water models.[35]

Since the radial density profile shows an interface thickness of 1.1 to 1.3 nm, for even smaller droplets, we anticipate the very concept of a spherical water droplet with bulk-like environment in the center may cease to be valid.

1.10. Acknowledgement

This research was supported by the National Science Foundation (NSF), Grant No. DMR-1609650 and the Arkansas Bioscience Institute. The computer resources for this study were provided by the Arkansas High Performance Computational Center through Grant No. MRI-R2 0959124 provided by the NSF.

1.11. References:

1. E. J. Murray, B. J. Murray and V. Sivakumar, Chemical Society Reviews **42** (24), 9571-9572 (2013).

2. S. M. Kathmann, G. K. Schenter, B. C. Garrett, B. Chen and J. I. Siepmann, Journal of Physical Chemistry C **113** (24), 10354-10370 (2009).

3. S. H. Petrosko, R. Johnson, H. White and C. A. Mirkin, Journal of the American Chemical Society **138** (24), 7443-7445 (2016).

4. S. Boutet, *et al*, Science **337** (6092), 362-364 (2012).

5. R. Zhang, A. Khalizov, L. Wang, M. Hu and W. Xu, Chemical Reviews **112** (3), 1957-2011 (2011).

6. D. A. Hegg and M. B. Baker, Reports on Progress in Physics **72** (5), 056801 (2009).

7. M. P. Anisimov, Russian Chemical Reviews **72** (7), 664-705 (2003).

8. M. Ivan-V, *Crystal Growth for Beginners, Fundamentals of Nucleation, Crystal Growth and Epitaxy*, 2 ed. (Bulgarian Academy of Sciences, Bulgaria, 1995).

9. V. B. Warshavsky and X. Y. Song, Journal of Physics-Condensed Matter **22** (36), 364112 (2010).

10. S. M. Thompson, K. E. Gubbins, J. Walton, R. A. R. Chantry and J. S. Rowlinson, Journal of Chemical Physics **81** (1), 530-542 (1984).

11. M. H. Factorovich, V. Molinero and D. A. Scherlis, Journal of the American Chemical Society **136** (12), 4508-4514 (2014).

12. L. M. Skinner and J. R. Sambles, Journal of Aerosol Science **3** (3), 199-210 (1972).

13. T. Young, Philos. Trans. R. Soc. of London **95**, 65-87 (1805).

14. P. S. Laplace, Traite de Mecanique Celeste **4** (*Supplement au dixieme livre du Traite de Mecanique Celeste*), 1-79 (1805).

15. C. Vega and E. de Miguel, Journal of Chemical Physics **126** (15), 154707 (2007).

16. A. Ghoufi, P. Malfreyt and D. J. Tildesley, Chemical Society Reviews **45** (5), 1387-1409 (2016).

17. A. Malijevsky and G. Jackson, Journal of Physics-Condensed Matter **24** (46), 464121 (2012).

18. T. P. Bennett and J. C. Barrett, Journal of Chemical Physics **137** (12), 124702 (2012).

19. V. Holten, D. G. Labetski and M. E. H. van Dongen, Journal of Chemical Physics **123** (10), 104505 (2005).

20. J. H. Irving and J. G. Kirkwood, Journal of Chemical Physics **18** (6), 817-829 (1950).

21. A. Harasima, Journal of the Physical Society of Japan **8** (3), 343-347 (1953).

22. A. E. van Giessen and E. M. Blokhuis, Journal of Chemical Physics **131** (16), 164705 (2009).

23. E. M. Delrio, E. Demiguel and L. F. Rull, Physica A **213** (1-2), 138-147 (1995).

24. S. H. Lee, Bulletin of the Korean Chemical Society **33** (11), 3805-3809 (2012).

25. T. Nakamura, W. Shinoda and T. Ikeshoji, Journal of Chemical Physics **135** (9), 094106 (2011).

26. G. V. Lau, I. J. Ford, P. A. Hunt, E. A. Muller and G. Jackson, Journal of Chemical Physics **142** (11), 114701 (2015).

27. A. A. Homman, E. Bourasseau, G. Stoltz, P. Malfreyt, L. Strafella and A. Ghoufi, Journal of Chemical Physics **140** (3), 034110 (2014).

28. G. J. Gloor, G. Jackson, F. J. Blas and E. de Miguel, Journal of Chemical Physics **123** (13), 134703 (2005).

29. R. C. Tolman, Journal of Chemical Physics **17** (3), 333-337 (1949).

30. B. J. Block, S. K. Das, M. Oettel, P. Virnau and K. Binder, Journal of Chemical Physics **133** (15), 154702 (2010).

31. O. Wilhelmsen, D. Bedeaux and D. Reguera, Journal of Chemical Physics **142** (17), 171103 (2015).

32. Y. A. Lei, T. Bykov, S. Yoo and X. C. Zeng, Journal of the American Chemical Society **127** (44), 15346-15347 (2005).

33. T. V. Bykov and X. C. Zeng, Journal of Chemical Physics **111** (23), 10602-10610 (1999).

34. M. N. Joswiak, N. Duff, M. F. Doherty and B. Peters, Journal of Physical Chemistry Letters **4** (24), 4267-4272 (2013).

35. M. N. Joswiak, R. Do, M. F. Doherty and B. Peters, Journal of Chemical Physics **145** (20), 204703 (2016).

36. J. L. F. Abascal and C. Vega, Journal of Chemical Physics **123** (23), 234505 (2005).

37. J. J. Thomson and G. P. Thomson, *Conduction of Electricity through Gases*, 3 ed. (2013).

38. M. P. A. Fisher and M. Wortis, Physical Review B **29** (11), 6252-6260 (1984).

39. A. P. Thompson, S. J. Plimpton and W. Mattson, Journal of Chemical Physics **131** (15), 154107 (2009).

40. E. M. Blokhuis and D. Bedeaux, Journal of Chemical Physics **97** (5), 3576-3586 (1992).

41. A. J. Sodt and R. W. Pastor, Journal of Chemical Physics **137** (23), 234101 (2012).

42. V. V. Zakharov, E. N. Brodskaya and A. Laaksonen, Journal of Chemical Physics **107** (24), 10675-10683 (1997).

43. F. S. Shahrazad M. A. Malek, Peter H. Poole, and Ivan Saika-Voivod, (**arXiv:1711.03994**).

44. N. Fujiwara, K. Koyama-Nakazawa, T. Matsumoto, A. Hisada and Y. Uwatoko, Journal of Physics-Condensed Matter **19** (42), 425203 (2007).

45. F. Wang, O. Akin-Ojo, E. Pinnick and Y. Song, Molecular Simulation **37** (7), 591-605 (2011).

46. A. R. Leach, *Molecular Modelling: Principles and Applications .* (New York, Prentice Hall, 2001).

47. C. H. Bennett, Journal of Computational Physics **19** (3), 267-279 (1975).

48. H. Hu, Z. Ma and F. Wang, Annual Reports in Computational Chemistry **10**, 25–43 (2014).

49. T. Darden, D. York and L. Pedersen, Journal of Chemical Physics **98** (12), 10089-10092 (1993).

50. M. P. Allen, Tildesley, D. J., *Computer Simulations of Liquids.* (Oxford: Oxford Science Publications., 1987).

51. H. Y. Hu and F. Wang, Journal of Chemical Physics **142** (21), 214507 (2015).

52. M. Parrinello and A. Rahman, Journal of Applied Physics **52** (12), 7182-7190 (1981).

53. S. Nose and M. L. Klein, Molecular Physics **50** (5), 1055-1076 (1983).

54. Y. Liu and J. Z. Wu, Journal of Chemical Physics **139** (4), 041103 (2013).

55. A. E. Ismail, G. S. Grest and M. J. Stevens, Journal of Chemical Physics **125** (1), 014702 (2006).

56. Y. Mu, A. Houk and X. Y. Song, Journal of Physical Chemistry B **109** (14), 6500-6504 (2005).

57. J. Vrabec, G. K. Kedia, G. Fuchs and H. Hasse, Molecular Physics **104** (9), 1509-1527 (2006).

1.12. Appendix

Tables:

Table I: The target pressure of the barostat, the measured Virial pressure for the 12,000 saved configurations, and the fitted pressure according to Eq. 3 are reported. The associated error bar of the measured pressure is also shown.

Target Pressure (bar)	Fitted Pressure (bar)	Reference Virial Pressure
1	4.5	4.4 ± 4.0
150	151.9	150.9 ± 4.0
300	299.2	299.4 ± 4.1
450	449.4	453.7 ± 4.1
600	598.3	593.4 ± 4.1
750	748.7	749.3 ± 4.2
1000	1003.5	1004.0 ± 4.2

Table II: The number of water molecules in each droplet, the length of the production simulation, and the radius of the Gibbs surface for all the droplets investigated in this study.

Number of	Length of production simulation (ns)	Radius of Gibbs Surface (nm)
1116	72	1.973
3769	45	2.971
8934	45	3.970
17450	24	4.968
29869	11	5.947
70802	10	7.940

Table III: The surface tension at the Gibbs dividing surface calculated using the Young-Laplace Equation.

Gibbs Radius (nm)	Internal Pressure (bar)	Surface Tension (mN/m)
1.973	707.2	69.8(1)
2.971	463.7	68.9(1)
3.970	340.7	67.6(1)
4.968	273.4	67.9(1)
5.947	226.2	67.3(1)
7.940	170.1	67.5(1)

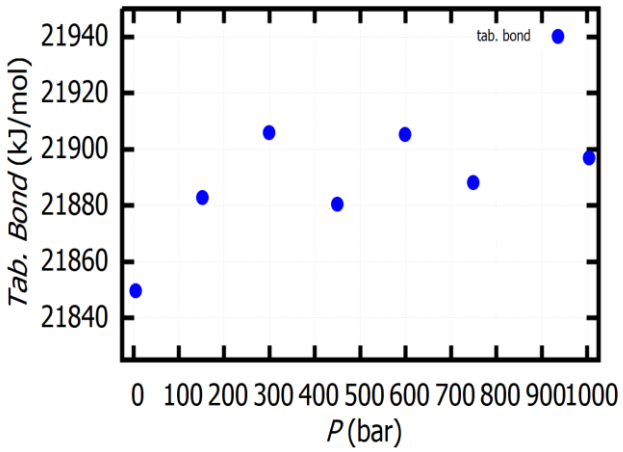

Figure 1: The tabulated bond energy-pressure correlation for BLYPSP-4F water at 298 K.

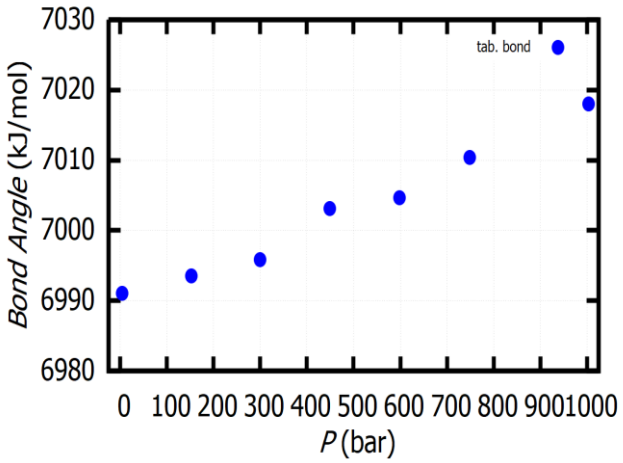

Figure 2: The tabulated bond angle-pressure correlation for BLYPSP-4F water at 298 K.

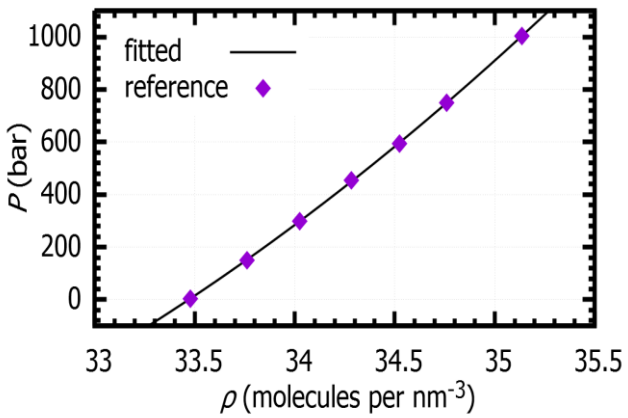

Figure 3: Density-pressure correlation for BLYPSP-4F water at 298 K.

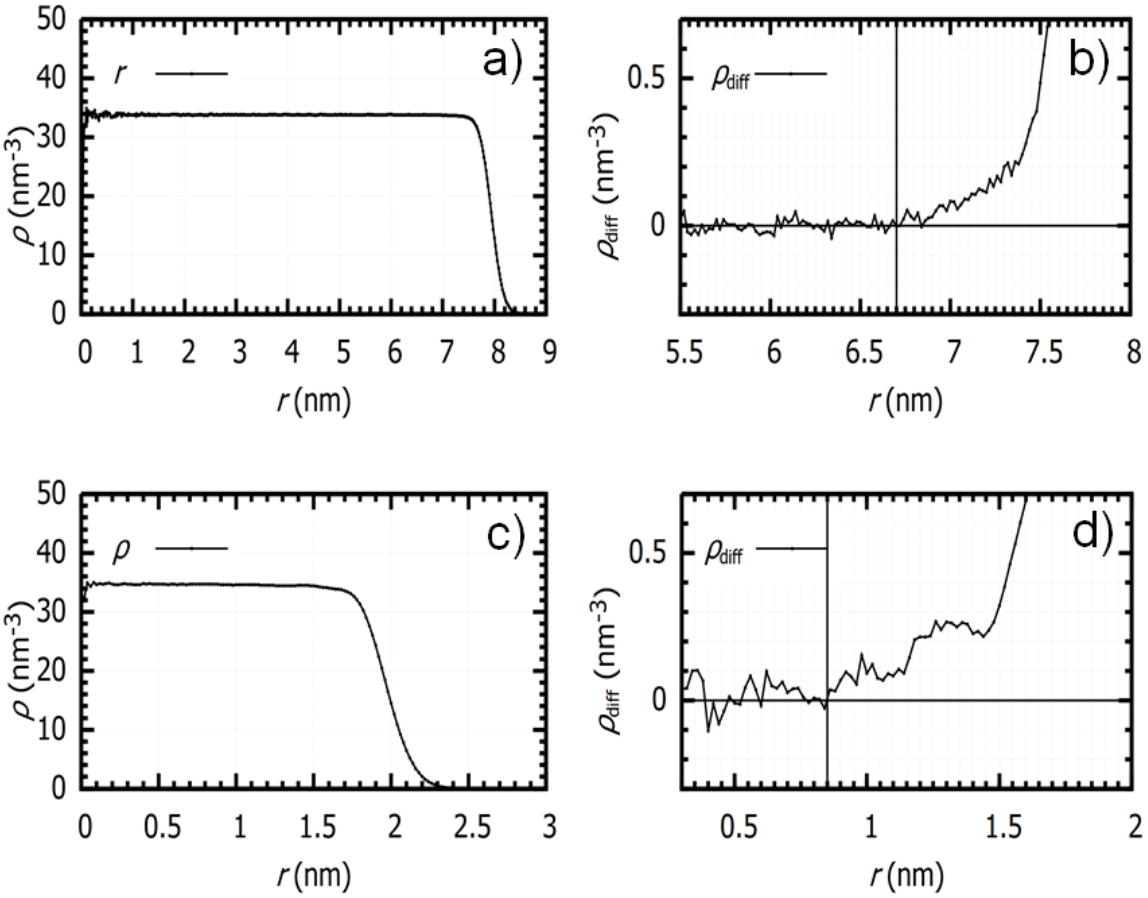

Figure 4: The radial density for the 16 nm diameter (a) and 4 nm diameter (c) droplets as a function of distance to the center. The difference between the radial density and the average density in the center of the droplet is also shown for the 16 nm (b) and 4 nm droplets (d).

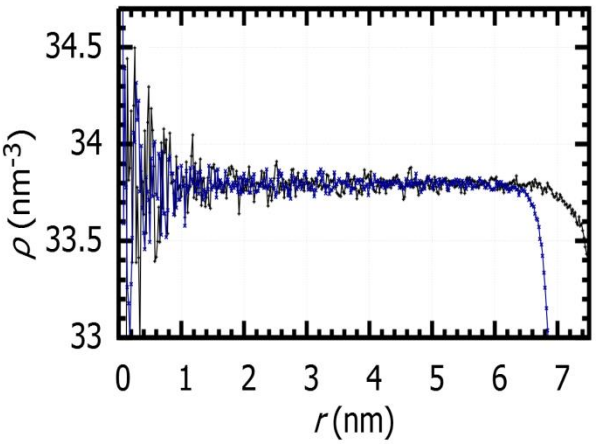

Figure 5: The radial density of water in the 16 nm droplet calculated using the center of mass (black) and an arbitrary point 1 nm in the positive z direction (blue) as the origin.

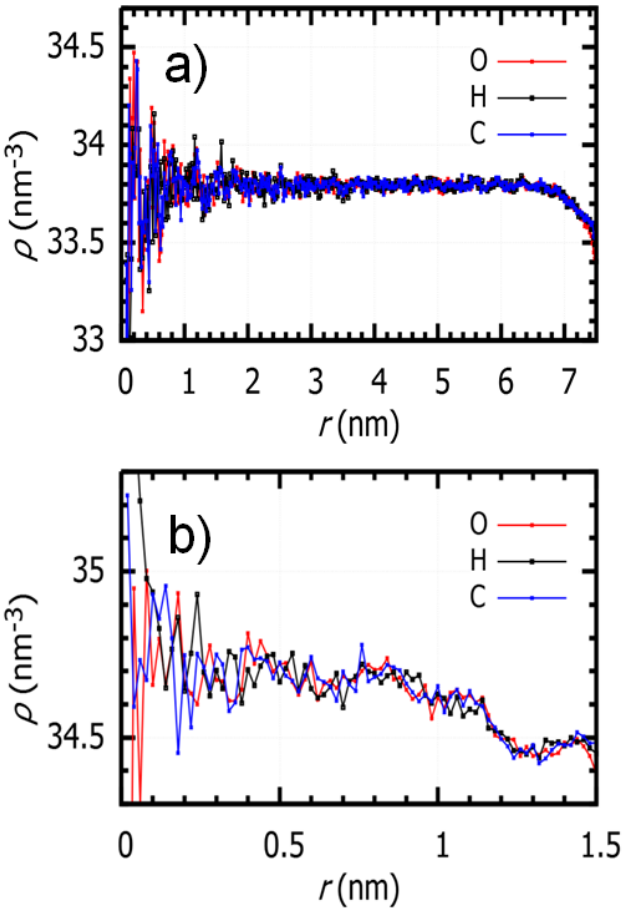

Figure 6: The radial density of water determined using the location of oxygen atoms (O), one of the two hydrogen atoms in each water (H), and the center of geometry (C) as reference for the 16 nm (a) and 4 nm diameter (b) droplets.

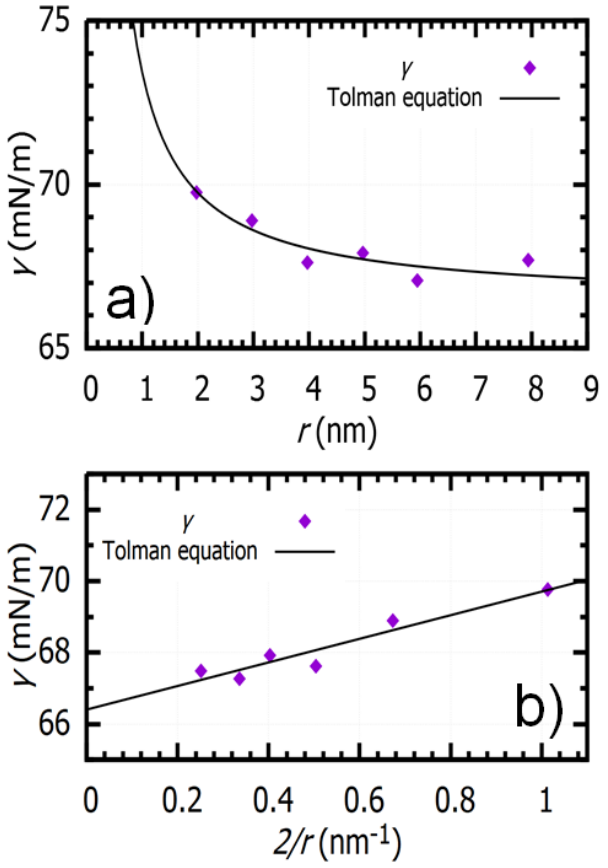

Figure 7: The surface tension of nanodroplets as a function of the Gibbs radius, *r*, (a) and as function of 2/*r* (b). The solid line is the fit to the Tolman equation.

On approximating a weak Markovian process as Markovian: are we justified when discarding longtime correlations

2.1. Introduction

Many real life processes may be influenced by longtime correlations. For example, it has been argued that protein folding occurs along a single funnel in the free energy landscape,[1, 2] thus there is a possibility that the entire folding process is weakly correlated over microsecond timescale or longer. It is probably reasonable to assume that correlations over extended periods of time will be weak. Under this assumption, it is a common practice to ignore longtime correlations. For example, when a long trajectory is simulated by a series of shorter ones, occasionally new initial velocities are generated at restarts. Certain stochastic thermostats will cause loss of longtime correlations in velocity.[3] Many enhanced sampling methods, especially those based on piecing together shorter trajectories, ignore correlations over an extended period of time.[4-17] For example, the Markov State Model (MSM) assume all correlations are lost over a lag time.[18-25] It is anticipated that, by choosing sufficiently long lag times, the evolution is approximately Markovian between updates. Several methods exist for verification of Markovianity after a lag time.[26-30] In practice, however, the detection of weak correlations is hampered by statistical noises that are omnipresent for finite temperature simulations in the condensed phase. The most challenging problems are those where the correlation is orders of magnitudes smaller than thermal fluctuations.

Protein folding is strongly influenced by random fluctuations. Longtime correlations, if present, are anticipated to be much smaller than such fluctuations. The objective of this work is to study the effect of very weak longtime correlations. We investigate to what degree ignoring such weak correlations will affect the overall kinetics. Our model system will be indefinitely correlated by design. However, longtime correlation is much smaller than fluctuations. Particular attention will be paid to the popular MSM method, although the insights from this study are

applicable to many other enhanced sampling algorithms,[4-17] where longtime kinetics are derived from shorter trajectories while ignoring longtime correlations.

In this work, longtime correlations are removed in one case by randomly resetting the system velocities and in another case by construction of MSM. We investigate whether such a treatment of a weakly non-Markovian process as Markovian will introduce an appreciable error. The paper is organized in four sections. After the introduction, the model system will be described in section II. The results will be presented in section III with a summary and conclusion in section IV.

2.2. Model system

In this study, we will investigate the first passage time (FPT) of a driven atom in super-critical Ar at a relatively high density. The simulation box contains 328 Ar atoms at a density of 34.04 mol/L and a temperature of 200 K.

The Ar atoms are modeled with a Lennard Jones potential V_{LJ} ,

$$V_{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad (1)$$

with an atomic size parameter, σ, of 0.3405 nm and a well depth, ε, of 0.9980 kJ/mol.[31] The Ar model has a critical point of 162.92 K and 15.14 mol/L.[32] The 200 K simulation temperature is thus above the critical point and the box density corresponds to the density of a liquid below the critical temperature.[33] Unlike systems with slow hydrogen bond relaxations, a rare gas system intrinsically decorrelates much faster; thus we are able to study the effect of applied correlated force with relatively short trajectories.

At these simulation conditions, the root mean square force is measured to be 784.8 kJ/(mol·Å). For one of the 328 Ar atoms, we apply a constant biasing force of 9.57 kJ/(mol·Å), which is slightly more than 1% of the root mean square force resulting from random collisions at

this temperature. The special particle with the constant driving force will be referred to as the tagged particle. The simulation is performed in an orthorhombic box with dimensions of 2 nm × 2 nm × 4 nm. The weak biasing force is applied along the longer dimension of the box, *z*. The FPT is defined as the time it takes for the tagged particle to travel the full length of 4 nm along *z* for the first time.

Although a small work is being done to the system as a result of the external driving force, this work is negligibly small when compared to the total energy. In addition, the system temperature is maintained with the Nosé-Hoover thermostat[34, 35] with a 2 ps relaxation time. The equation of motion is integrated with a 5 fs timestep.

The mass of the tagged particle should play an important role since it determines the inertia. A heavier particle is anticipated to amplify the effect of the constant driving force. A protein may have a relatively large mass although the driving force for any dynamical events, such as folding, might be small at any instant. In order to study the effect of inertia, the tagged particle is allowed to assume masses other than the atomic mass of Ar. In this study, we investigate tagged particles with masses of 5 g/mol, 40 g/mol, 320 g/mol and 3200 g/mol.

Figure 1 shows the force autocorrelation function,

$$C_{\mathrm{ff}}(t) = \frac{\langle \mathbf{f}(t)\cdot\mathbf{f}(0)\rangle}{\langle \mathbf{f}(0)\cdot\mathbf{f}(0)\rangle}, \qquad (2)$$

for the tagged particle. In Eq. 2, the angle brackets indicate ensemble averages and bold letters indicate vectors. If the one assumes the force, **f**, is a simple summation of a correlated component $\mathbf{f}_c$ being the driving force and a random component $\mathbf{f}_r$ being the environmental force, one would anticipate that $C_{\mathrm{ff}}(t)$ approaches $\langle \mathbf{f}_c / \mathbf{f}\rangle^2$ at large *t*. If $\mathbf{f}_r$ is assumed to be a white noise that averages to zero, it can be shown that the longtime asymptote would be

$$\langle \mathbf{f}_c / \mathbf{f}\rangle^2 = 0.012^2 = 1.44\times10^{-4}, \qquad (3)$$

which is marked as a horizontal line in Figure 1. It can be seen from Figure 1 that the fluctuation of $C_{\mathrm{ff}}(t)$ is orders of magnitude larger than the average at longtime, t, making the detection of any asymptotic behavior challenging. In order to remove the fluctuations and obtain the long time asymptote, 10 ps running averages of $C_{\mathrm{ff}}(t)$ are calculated and plotted in Figure 1c. It is clear that the true asymptote is not strongly affected by particle mass and is orders of magnitude smaller than that predicted by Eq. 3.

Such a small asymptotic average is actually not surprising. At the high simulation density, each particle has a free mean path of 0.335 Å, which leads to constant collisions. The collective effect of the collisions acts as a frictional resistance and almost completely cancels the driving force, thus the force from the environment could not be considered as a white noise. The white noise assumption that leads to Eq. 3 is not valid, so we observe a longtime average of $C_{\mathrm{ff}}(t)$ much smaller than the naïve anticipation. The vanishing asymptote is analogous to an object falling through the atmosphere under the force of gravity is expected to reach a terminal velocity without net acceleration. The vanishing asymptote along with the large fluctuations make the force correlation an ineffective metric for detecting longtime correlation. It is also worth noting that the longtime average of $C_{\mathrm{ff}}(t)$ approaches zero similarly for particles with different masses, which would seem to suggest that particles with different masses would behave similarly when correlations are ignored. However, it is clear from our subsequent calculations that removing longtime correlations from particles with larger inertia will have a much larger effect on transport kinetics, further suggesting that focusing only on the tail of $C_{\mathrm{ff}}(t)$ is likely to lead to improper conclusions.

Although the force autocorrelation functions average to virtually zero at large *t,* the longtime correlation is visible with the velocity autocorrelation function,

$$C_{\mathrm{vv}}(t) = \frac{\langle \mathbf{v}(t)\cdot\mathbf{v}(0)\rangle}{\langle \mathbf{v}(0)\cdot\mathbf{v}(0)\rangle}. \qquad (3)$$

Figure 2 reports the C_{vv} for both the tagged particle and a non-tagged particle. It can be seen that the effect of the biasing force is almost negligible for lighter tagged particles. The C_{vv} for the 40 g/mol tagged particle is indistinguishable from that of non-biased particles for the first 5 ps. After 5 ps, a faint tail in C_{vv} can be observed. The lighter 5 g/mol particle decorrelates much faster showing a residual correlation comparable to simulation noises in our study. The 320 g/mol tagged particle retains a stronger correlation in C_{vv} as expected. However, even for this mass, the correlation becomes less than 0.1% after 15 ps. The C_{vv} for the 3200 g/mol tagged particle drops to about 0.3 % after 60 ps, clearly demonstrating the effect of the increased inertia. It is worth emphasizing that even for this massive particle, the longtime asymptote of C_{vv} averages to about 2.5×10^{-3}. Although such a weak correlation of C_{vv} is detectable in our study, for a complex system, such as protein folding, correlation in the velocity of a collective variable, that may not be well-defined in itself, will be much harder to detect.

2.3. Results and discussion

Table I reports the mean FPT for the tagged particle to travel 4 nm. All the numbers were averaged over 10,000 repetitions from random initial conformations obtained from an equilibrium NVT ensemble without biasing. The initial conformations are taken from an equilibrium molecular dynamics (MD) trajectory at 100 ps intervals to eliminate any possible correlation between the conformations. The error bars reported in Table I are the standard error of the mean. The reference FPT was obtained by continuous simulations fully considering longtime correlations. It can be seen that heavier particles have longer FPTs. While the 3200 g/mol particle has a 2320 ps FPT, the lightest 5 g/mol particle has a FPT of 1045ps. This is not

surprising since the heavier particles are moving at a lower average velocity at the same temperature; also the heavier particles have a smaller acceleration with the same strength of biasing force.

In order to understand the effect of ignoring correlations, we measured the FPT by running trajectories where the correlations were removed at constant intervals. This was accomplished by restarting the trajectories with the velocity of each particle randomized. A series of restarting intervals were studied ranging from 12.5 ps to 100 ps. For the 5 g/mol and 40 g/mol particles, the C_{vv} have reached their respective asymptotes even for the shortest 12.5 ps intervals studied. For the 320 g/mol tagged particle, the C_{vv} is already very small at 0.17% after 12.5 ps and reaches the longtime plateau at 25 ps.

With the lighter 5 g/mol and 40 g/mol tagged particles, our results indicate that discarding the longtime correlation has negligible effect on the FPT. Considering that the one sigma error bar reported corresponds to a 68% confidence interval, all the estimated mean FPTs agree with each other within the statistical noise. For the 320 g/mol tagged particle, while an excellent agreement is obtained with a 100 ps restart frequency, the measured FPT increases as the correlations are discarded more frequently. With a restart interval of 25 ps, the FPT is overestimated by 56 ps, which is significantly larger than the statistical noise. Although 56 ps is a fairly small deviation, the residual of C_{vv} is only 0.057% after 25 ps. The test case with the heavy 3200 g/mol mass shows a quite significant 200 ps error, almost 10% over-estimation, even with the 100 ps restart time. This is fairly large considering the residual velocity correlation is only 0.25%. With a 12.5 ps restart time, the error grows to almost 100% despite an undetectable $C_{\mathrm{ff}}\left(t\right)$ and C_{vv} being about 0.2. We note the Ar test system was simulated far above the critical point, where strongly oscillatory forces from the environment lead to a very fast decorrelation. For a more complex system with hydrogen bonds and their associated slower

dynamics, a much slower decorrelation is anticipated. Thus the timescale used for our model system should not be taken literally. This study suggests that neglecting longtime correlation could, in certain cases, lead to fairly appreciable errors and should be done with care.

MSM is widely used to derive longtime evolution from a collection of short trajectories. With this approach, short-trajectories are simulated independently with no correlation between them. A transition matrix is constructed by studying transition probabilities between microstates after a lag time. The Chapman-Kolmogorov property is assumed for the application of the transition matrix. In other words, the system is assumed to be Markovian after each lag time.

Although the MSM assumes the system is not correlated after a lag time, it is not the only method with such an approximation. Any algorithms based on construction of overall kinetics from short-trajectories,[12-17, 36-39] potentially suffer the loss of longtime correlation. It is not within the scope of the present work to do a general survey of short-trajectory based methods. In this study, we will investigate the effect of forcing Markovianity on a system with weak longtime correlations with the MSM. During the application of a MSM, the validation of the Markovian approximation can be tested with several methods, such as the Swope-Pitera eigenvalues[30], information entropy[29], or Chapman-Kolmogorov analysis,[19, 26, 30, 40, 41] Such validation methods are much more practical for complex systems than force or velocity correlation functions. In this work, the lack of Markovianity will be checked with the Chapman-Kolmogorov analysis, which is one of the most straight-forward method for validating Markovianity.

For each tagged particle, a series of MSM matrices were constructed with different lag times from 5 ps to 50 ps. For our model system, the MSM matrices were constructed using 81 microstates. The relatively large number of microstates minimize discretization errors.[26] The microstates were defined using the location of the tagged particle along the z direction. We do

not lump microstates into macrostates to avoid lumping errors[30, 42, 43] that may lead to additional complications not directly related to the longtime correlation problem we are trying to investigate. We note all MSM transition matrices were constructed from the 10,000 simulations, where particles are allowed to travel a full box length, but with velocities reset once every 100 ps. In practical application of MSMs, the full evolution of interest is generally longer than what can be done in one continuous simulation. Thus the construction of the MSM is typically based on shorter trajectories. The use of 100 ps trajectories will thus better reflect real world application of MSMs. We note that for all test systems, the correlation at 100 ps is already sufficiently small. A deviation for the observed FPT can only be seen for the 3200 g/mol tagged particle. Using shorter trajectories to construct a MSM will only lead to larger deviations and are thus not tested.

Figure 3 compares the MD probability of the tagged particle in the initial states to the probability predicted with MSM. Such a test of Chapman-Kolmogorov property is frequently referred to as the Chapman-Kolmogorov analysis.[19, 26, 30, 40, 41] The reference curves were produced with the 100 ps restart simulations used to construct the MSM. It is clear that all MSM predictions agree within the error of the MD simulations for any tagged particle with masses up to 320 g/mol. For the 3200 g/mol mass, the Chapman Kolmogorov test indicates good agreement only for the 50 ps lag time, and the agreement with the 25 ps lag time can be considered fair. The agreement with the 12.5 ps lag time should probably be considered poor and would indicate a problem.

The FPT obtained with all MSM studies are summarized in Table II. It is not surprising that a MSM works well with the 5 g/mol and 40 g/mol particles without much inertia. All the FPT predictions agree with the reference calculations within statistical uncertainty. For the 320 g/mol tagged particle, the FPT predicted by the MSM with a 5 ps lag time is about 49 ps longer than the reference values. Although 49 ps is not significant, considering the Chapman Kolmogorov

test suggest excellent Markovanity for this test case. The fact that the error is larger than statistical uncertainty is still worth mentioning.

The MSM predicted FPTs for the 3200 g/mol tagged particle are most interesting. For this mass, the 100 ps trajectories on which the MSM were trained has a 2520 ps FPT. Thus, the 2520 ps FPT is the best the MSM could possibly achieve. The MSM predicts a FPT of 2495 ps and 2554 ps with a 50.0 ps and 25.0 ps lag time, respectively. This is very good agreement considering the Chapman-Kolmogorov test shows a minor deviation. With a lag time of 12.5 ps, the MSM only results in a 410 ps overestimation of the FPT. This is much better than the over-estimation of 1770 ps observed when velocity correlation is discarded every 12.5 ps, as shown in Table 1. Even the 5 ps lag time MSM simulation leads to a much smaller 658 ps overestimation when compared to simply discarding velocity correlation at a longer 12.5 ps interval.

The much better-than-expected performance of the MSMs can only be explained by considering that the 100 ps trajectories based on which the MSM transition matrices were constructed have their motions fully correlated during the entire simulation length. The tagged particles are continuously accelerated by the driving force during whole 100 ps. This led to accelerated transitions between microstates, which are built into the transition probability matrix elements. Subsequently, the use of MSMs led to better estimates of the FPT even when shorter tag times are used.

2.4. Summary and Conclusion

The effect of removal of weak longtime correlation was studied by simulating the motion of a tagged particle under a weak constant driving force. The particle is surrounded by super-critical Ar atoms at high density. The driving force is only about 1% of the random force experienced by each particle due to thermal fluctuations. The tagged particle is allowed to have

a range of masses from 1/8 to 800 times that of an Ar atom. The higher mass simulations will likely better mimic macromolecules with large inertia under a weak driving force.

Regardless of particle mass studied, the force auto-correlation function shows little evidence of longtime correlation as a result of the frictional force exerted by the dense environment. Particles with different inertia demonstrate similar longtime decay of the force auto-correlation functions. These observations indicate that the force autocorrelation function is a poor metric for detecting longtime correlation. A small tail can be observed for the velocity autocorrelation functions, which depends on the mass of the particle. However, for the model system studied, the longtime asymptote is still very small, being no more than 2.5×10^{-3}, even for the most massive 3200 g/mol tagged particle.

The FPT for the tagged particle to travel a distance of 4 nm along the direction of the driving force is measured. When the particles have little inertia, discarding the longtime correlation has little influence on its FPT. When the particle has a fairly large inertia, discarding longtime correlation starts to affect the observed FPT, slowing down the motion of the particle along the direction of the driving force. Appreciable change of the FPT can be observed even when the velocity correlation function shows negligible residual correlation. For example, for the 3200g/mol test case, while randomizing velocity once every 50 ps leads to a 16% overestimation of the FPT, the velocity correlation being discarded is only about 0.49% at the 50 ps mark.

It is interesting that simulation with MSMs shows very good agreement with the reference FPT even with very short lag times. For example, while enforcing Markovianity by randomizing velocities once every 12.5 ps will lead to a 100% overestimation of FPT, the MSM with a 12.5 lag time only lead to a 18% overestimation. Such a better-than-anticipated performance is a result of the transition matrices being constructed with the much longer 100 ps

trajectories. Although the way time-dependent probabilities are derived in the MSM assumes that the system completely loses memory after the lag time, in practice, a cancelation of errors could arise when the trajectory for transition matrix construction is correlated for a duration much longer than the lag time. This leads to a cautious optimism when using MSM with a high time resolution but based on longer trajectories.[44]

We note that the model system decorrelates quickly due to the lack of slow local relaxation. The specific restart interval and lag times used for this work should not be taken literally for more complex systems in a more viscous environment. We believe the simulations provide insight on the effect of longtime correlations in more complex systems with a much slower dynamical relaxation.

2.5. Acknowledgment:

The research was supported by National Institutes of Health (NIH), Grant No. NIGMS 1R01GM120578 and the Arkansas Bioscience Institute. The computer resources for this study were provided by the Arkansas High Performance Computational Center through Grant No. MRI-R2 0959124 provided by the NSF.

2.6. References:

1. K. A. Dill and H. S. Chan, Nature Structural Biology **4**, 10-19 (1997).

2. D. Bryngelson Joseph, N. Onuchic José, D. Socci Nicholas and G. Wolynes Peter, Proteins: Structure, Function, and Bioinformatics **21** (3), 167-195 (1995).

3. G. Bussi, D. Donadio and M. Parrinello, The Journal of Chemical Physics **126** (1), 014101 (2007).

4. D. Hamelberg, J. Mongan and J. A. McCammon, The Journal of Chemical Physics **120** (24), 11919-11929 (2004).

5. A. F. Voter, Physical Review Letters **78** (20), 3908-3911 (1997).

6. A. Laio and M. Parrinello, Proceedings of the National Academy of Sciences **99** (20), 12562-12566 (2002).

7. R. C. Bernardi, M. C. R. Melo and K. Schulten, Biochimica et Biophysica Acta (BBA) - General Subjects **1850** (5), 872-877 (2015).

8. S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, Science **220** (4598), 671 (1983).

9. C. Tsallis and D. A. Stariolo, Physica A: Statistical Mechanics and its Applications **233** (1), 395-406 (1996).

10. H. Szu and R. Hartley, Physics Letters A **122** (3), 157-162 (1987).

11. N. Metropolis and S. Ulam, Journal of the American Statistical Association **44** (247), 335-341 (1949).

12. A. K. Faradjian and R. Elber, The Journal of Chemical Physics **120** (23), 10880-10889 (2004).

13. D. Shalloway and A. K. Faradjian, The Journal of Chemical Physics **124** (5), 054112 (2006).

14. R. Elber, Biophysical Journal **92** (9), L85-L87 (2007).

15. A. M. A. West, R. Elber and D. Shalloway, The Journal of Chemical Physics **126** (14), 145104 (2007).

16. E. Vanden-Eijnden and M. Venturoli, The Journal of Chemical Physics **130** (19), 194101 (2009).

17. J. M. Bello-Rivas and R. Elber, The Journal of Chemical Physics **142** (9), 094102 (2015).

18. V. S. Pande, K. Beauchamp and G. R. Bowman, Methods **52** (1), 99-105 (2010).

19. J. Chodera, W. Swope, J. W. Pitera and K. Dill, Multiscale Model. Simul. **5** (4), 1214-1226 (2006).

20. W. Wang, S. Cao, L. Zhu and X. Huang, *WIREs Comput. Mol. Sci.* (8), e1343 (2018).

21. B. E. Husic and V. S. Pande, Journal of the American Chemical Society **140** (7), 2386-2396 (2018).

22. N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*. (North-Holland Physics Publishing, New York, 1992).

23. C. Schütte and M. Sarich, *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches.* (American Mathematical Society, 2013).

24. G. R. Bowman, V. S. Pande and F. Noé, *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. (Springer, The Netherlands, 2014).

25. A. A. Markov, *The theory of algorithms*. (Acad. Sci. USSR, Moscow–Leningrad, 1954).

26. J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, The Journal of Chemical Physics **134** (17), 174105 (2011).

27. F. Noe, Journal of Chemical Physics **128** (24), 13 (2008).

28. F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich and T. R. Weikl, Proceedings of the National Academy of Sciences **106** (45), 19011 (2009).

29. S. Park and V. S. Pande, The Journal of Chemical Physics **124** (5), 054118 (2006).

30. W. C. Swope, J. W. Pitera and F. Suits, The Journal of Physical Chemistry B **108** (21), 6571-6581 (2004).

31. N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark and F. v. G. Wilfred, Eur Biophys J **40** (7), 843-856 (2011).

32. L. Verlet and D. Levesque, Physica **36**, 254 (1967).

33. J.-P. Hansen and L. Verlet, Physical Review **184** (1), 151-161 (1969).

34. S. Nosé, Molecular Physics **52** (2), 255-268 (1984).

35. W. G. Hoover, Physical Review A **31** (3), 1695-1697 (1985).

36. T. S. van Erp, D. Moroni and P. G. Bolhuis, The Journal of Chemical Physics **118** (17), 7762-7774 (2003).

37. D. Moroni, P. G. Bolhuis and T. S. van Erp, The Journal of Chemical Physics **120** (9), 4055-4065 (2004).

38. G. A. Huber and S. Kim, Biophysical Journal **70** (1), 97-110 (1996).

39. B. W. Zhang, D. Jasnow and D. M. Zuckerman, The Journal of Chemical Physics **132** (5), 054107 (2010).

40. R. D. Malmstrom, C. T. Lee, A. T. Van Wart and R. E. Amaro, Journal of Chemical Theory and Computation **10** (7), 2648-2657 (2014).

41. J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill and W. C. Swope, The Journal of Chemical Physics **126** (15), 155101 (2007).

42. K. H. Hoffmann and P. Salamon, Applied Mathematics Letters **22** (9), 1471-1475 (2009).

43. J. D. Chodera and F. Noé, Current opinion in structural biology **25**, 135-144 (2014).

44. T. J. Lane, G. R. Bowman, K. Beauchamp, V. A. Voelz and V. S. Pande, Journal of the American Chemical Society **133** (45), 18413-18419 (2011).

2.7. Appendix

Tables:

Table I: The mean first passage time in ps for tagged particles of various masses. The reference values are obtained from continuous simulations. t_r indicates restart intervals at which a new random velocities are assigned to each particle in the system.

t_r	Mass of tagged particle (g/mol)			
(ps)	5	40	320	3200
Reference ($t_r = \infty$)	1045 ± 8	1210 ± 13	1478 ± 10	2320 ± 14
$t_r = 12.5$	1063 ± 6	1238 ± 6	1588 ± 13	4090 ± 42
$t_r = 25.0$	1037 ± 6	1220 ± 6	1534 ± 10	3147 ± 16
$t_r = 50.0$	1021 ± 7	1244 ± 12	1489 ± 17	2686 ± 10
$t_r = 100.0$	1027 ± 10	1212 ± 6	1478 ± 7	2520 ± 12

Table II: The mean FPT in ps for tagged particles of various masses predicted by MSMs. The transition matrices for MSM evolution were constructed using a series of trajectories with a 100 ps restart time. τ indicates the lag time chosen for the MSM. Both the true reference FPT and the FPT value based on a 100 ps restart are reported since the latter is the best the MSM is expected to be able to reproduce.

τ	Mass of tagged particle (g/mol)			
(ps)	5	40	320	3200
reference	1045 ± 8	1210 ± 13	1478 ± 10	2320 ± 14
t_r = 100.0	1027 ± 10	1212 ± 6	1478 ± 7	2520 ± 12
τ = 5.0	1029	1210	1527	2978
τ = 10.0	1054	1231	1513	2814
τ = 12.5	1063	1240	1512	2730
τ = 25.0	-	-	-	2554
τ = 50.0	-	-	-	2495

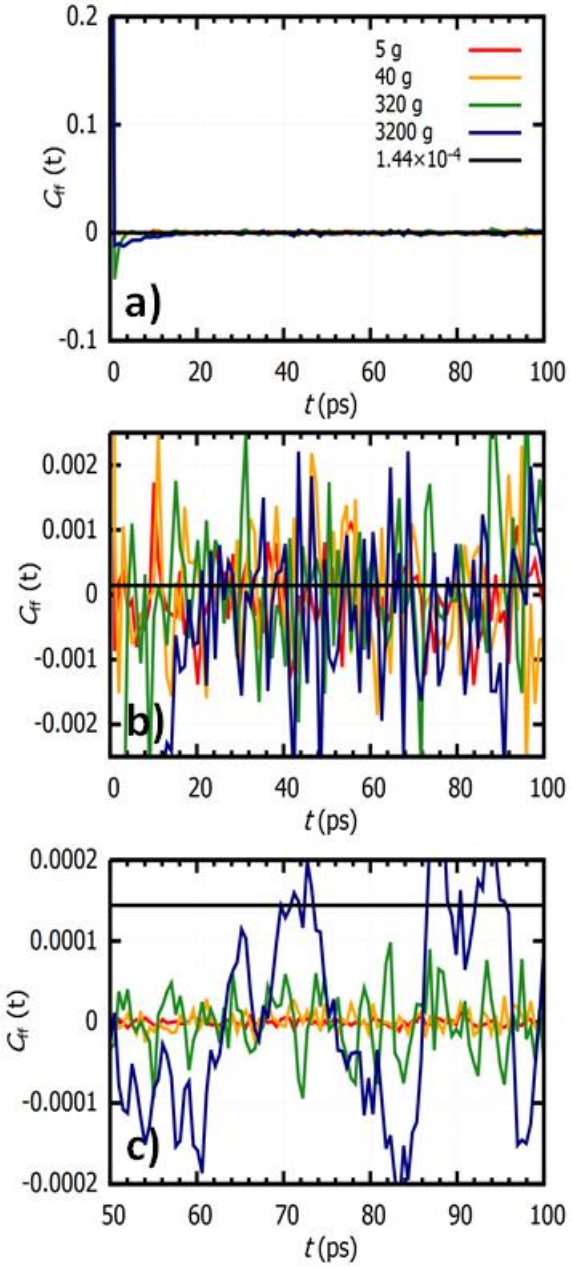

Figure 1: The force autocorrelation function, C_{ff}, of the tagged particle. Panel (b) presents a zoomed view to show the fluctuations of tail of the C_{ff}. Panel (c) shows a 10 ps running average of the C_{ff} to better reveal the asymptotic limit.

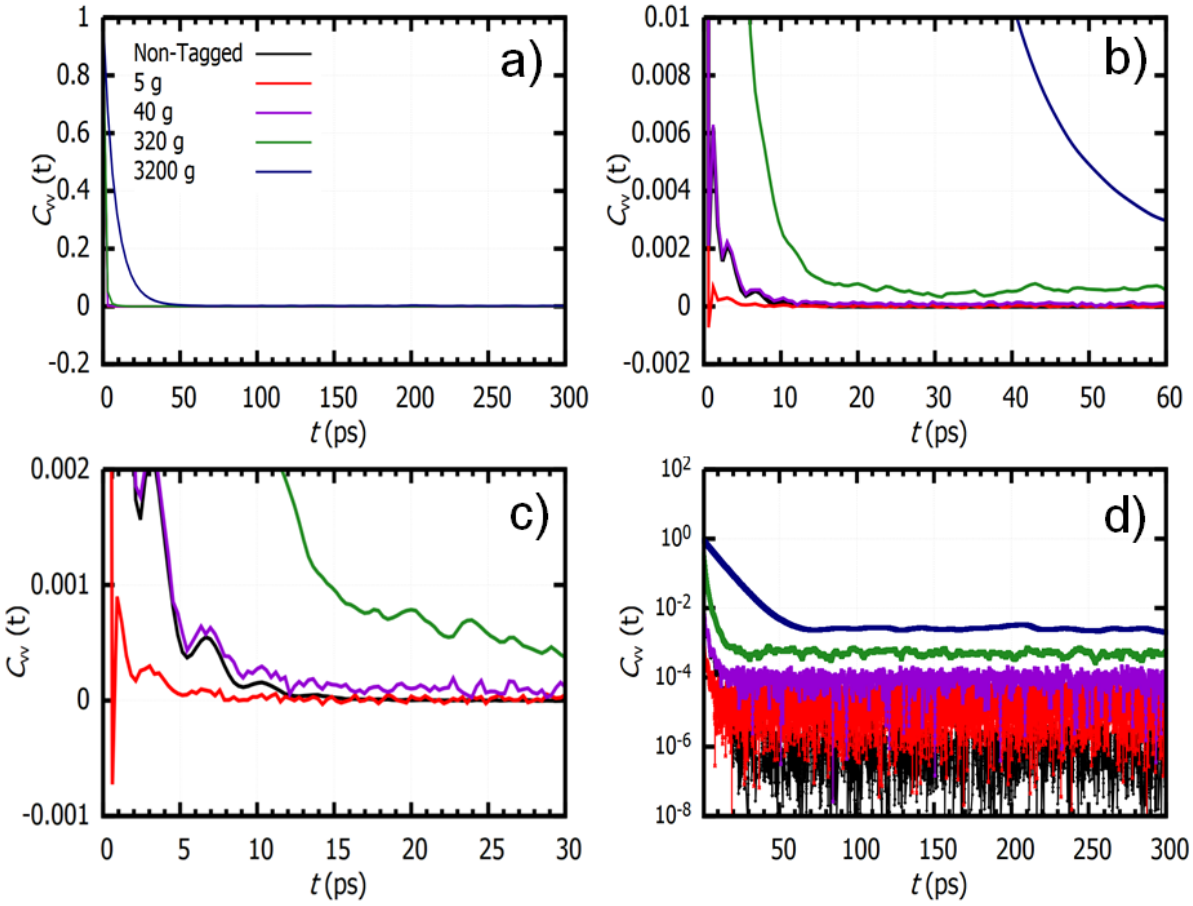

Figure 2: The velocity autocorrelation function, C_{vv}, of the tagged particle along with non-tagged Ar particles. Panels (b) and (c) are zoomed to better show fast decorrelation. Panel (d) shows the same functions with a semi-log scale.

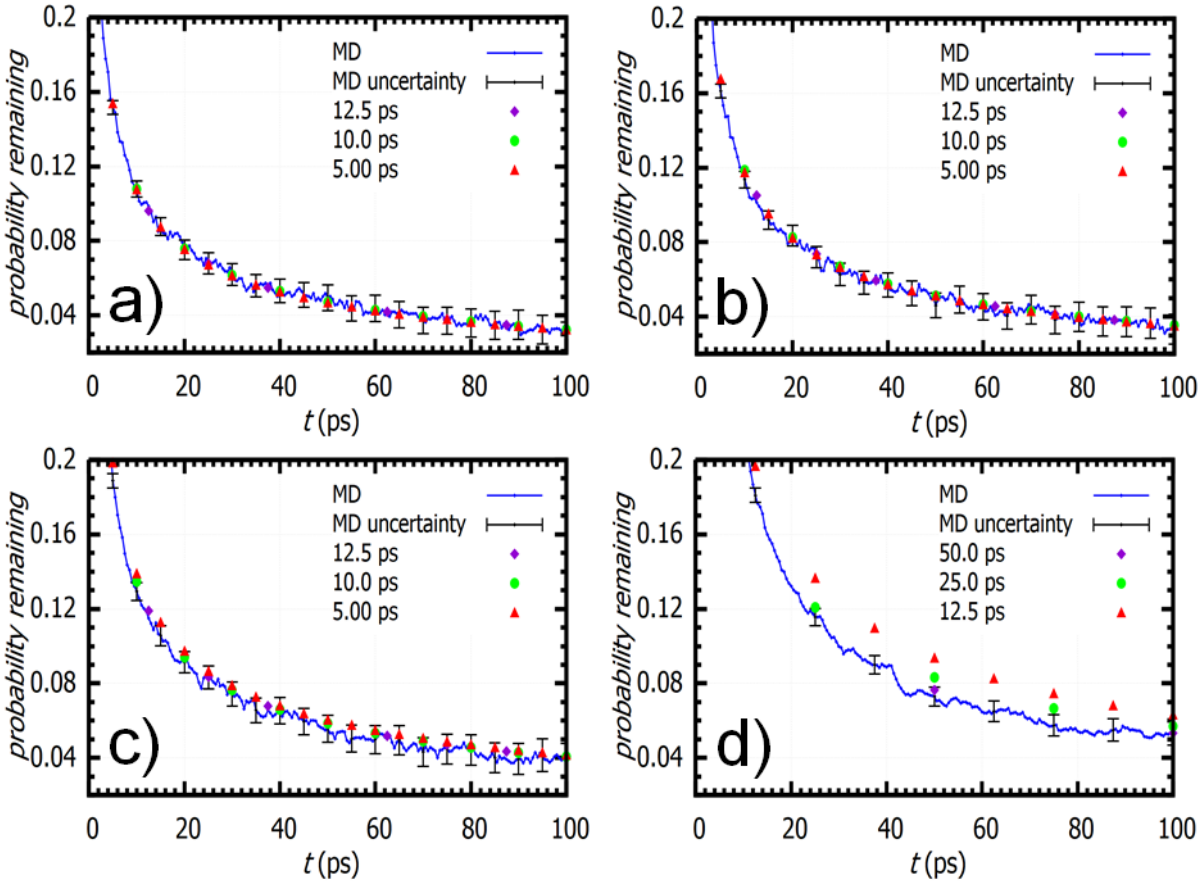

Figure 3: The Chapman-Kolmogorov test for tagged particles of various masses. The population of the initial microstate is shown for the simulation for the constructing the MSM matrices and those predicted with MSM with various lag times τ. Panels (a), (b), (c) and (d) correspond to particle masses of 5 g/mol, 40 g/mol, 320 g/mol and 3200 g/mol, respectively.

Thesis Conclusion

In this work, we establish a generalized empirical method for calculating the surface tension of nanoscale droplets using a density-pressure correlation. The surface tension deduced from the internal pressure using the Young-Laplace equation shows good agreement to the Tolman equation for all droplets above 4 nm in diameter with a Tolman length of -0.48 Å. Since the radial density profile shows an interface thickness of 1.1 to 1.3 nm, for even smaller droplets, we anticipate the very concept of a spherical water droplet with bulk-like environment in the center may cease to be valid.

A small tail can be observed for the velocity autocorrelation functions, which depends on the mass of the particle. For the model system studied, the longtime asymptote is still very small, being no more than 2.5×10^{-3}, even for the most massive 3200 g/mol tagged particle. The effect of removal of weak longtime correlation by a weak constant driving force of about 1% of the random force experienced by each particle due to thermal fluctuations can be significant. For example, for the 3200g/mol test case, while randomizing velocity once every 50 ps leads to a 16% overestimation of the FPT, the velocity correlation being discarded is only about 0.49% at the 50 ps mark.

It is interesting that simulation with MSMs shows very good agreement with the reference FPT even with very short lag times. Although the way time-dependent probabilities are derived in the MSM assumes that the system completely loses memory after the lag time, in practice, a cancelation of errors could arise when the trajectory for transition matrix construction is correlated for a duration much longer than the lag time. This leads to a cautious optimism when using MSM with a high time resolution but based on longer trajectories. We believe the simulations provide insight on the effect of longtime correlations in more complex systems with a much slower dynamical relaxation.