

# Chapter 3

## Metadynamics Methods

One of the major difficulties with studying nucleation and crystal growth with computer simulations is the temporal limitations of molecular dynamics and the lack of dynamic information in Monte Carlo simulations. Metadynamics has been proposed as a alternative computer simulation method that overcomes the deficiencies of molecular dynamics and Monte Carlo simulations. Metadynamics is based on the energy landscape framework, which has had great success in describing microscopic behavior in a range of materials. In this chapter, we will first present the energy landscape framework and the multitude of information contained in the landscape. Then, we will present our improved metadynamics method for directly sampling the energy landscape. Lastly, we will discuss the implementation of the method into GROMACS, and how to analyze the metadynamics simulation output.

### 3.1 Energy Landscape

The potential energy landscape formulation is a framework developed originally by Goldstein in 1969 [21] for the study of viscous liquids and the glass transition. This framework has been extended by many since the original formulation to explain slow phenomena and out of equilibrium behavior including protein folding, supercooled liquids, phase behavior, and more [22][23][24][25]. A two dimensional schematic of an potential energy landscape is shown in Figure 3.1 [25].

Figure 3.1 shows a schematic of the potential energy landscape and the diverse features typical of the potential energy landscape [25]. It is important to note that this schematic drawing is reduced to two dimensions (the potential energy and a single configurational coordinates), a typical potential energy landscape is hyper dimensional ranging from one dimension to  $3N$  dimensions, where  $N$  is the number of atoms in the system. Examples of configurational coordinates are the bond angles in protein (several configurational coordinates) [26], the system bond orientational order parameter for ordered structures (one to a couple of configurational coordinates) [27], and the XYZ position coordinates of each atom ( $3N$  configurational coordinates) [28, 29]. The more degrees of freedom the higher the dimensionality, and the more complex

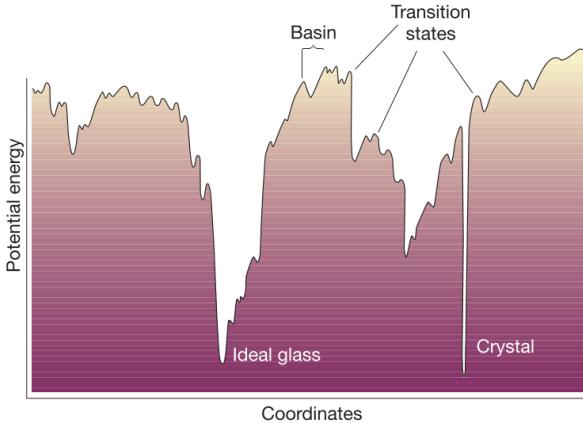


Figure 3.1: A schematic plot of the potential energy landscape in 2-D (potential energy versus 1 configurational coordinate). The true potential energy landscape is multi-dimensional depending on the number of configurational coordinates; the dimensionality can range from 1 to  $3N$  if all atoms coordinates are used. [25]

the potential energy landscape becomes. As a result of the hyper dimensionality, methods to calculate the landscape are still limited. Generally, the landscape is used as a theoretical framework to explain atomic scale phenomena, both structural and dynamical.

While the potential energy landscape is typically higher dimensional than Figure 3.1, much can be learned from the 2-D dimensional potential energy landscape. Figure 3.1 shows the potential energy landscape is rich in energy minimums, with variable potential energy. The probability of occupying a particular potential energy minimum with energy  $G$  is proportional to  $\exp(-G/k_B T)$ . Thus, at high temperatures, the system can occupy both high energy and low energy minimums with high likelihood [30]. As the temperature decreases the chance of occupying high energy (disordered states) becomes less likely, and the system tends to occupy the lower energy minimums (ordered states) [30]. Because the potential energy landscape is richer in high energy states than low energy states, the occupation of lower energy states with decreasing temperature leads to a decrease in configuration entropy [30]. At sufficiently low temperatures, any medium to high energy basins become inaccessible leading to dynamic arrest, and the formation of metastable materials [30].

In this way, the potential energy landscape is used to explain the structural state of a system, the minimums represent different structural states [25]. In order for the system to evolve structurally, the system must overcome energy barriers separating the minimums [25]. This dynamical process is related to the height of the energy barrier and the temperature of the system similar to the occupation probability of different energy minimum [25]. The larger the barrier the less likely the system will relax from one minimum to another [25]. Further the likelihood of crossing a barrier is related to the thermal energy in the system. In this way, as a system is cooled, the system can become stuck in deep energy minimums that are not the

lowest energy state. Because the energy barriers are too high to overcome the system is trapped in this metastable state, or glass state.

In MD simulations, we can study many equilibrium features of the potential energy landscape. In the crystal state, MD will capture the vibrations in the deep crystal minimum, or phonons [25]. In the disordered supercooled liquid or glassy state, MD will capture the relaxation times as the system evolves over energy barriers between energy minimum of similar potential energy [25]. However, as Figure 3.1 shows, the energy barrier between the glassy basin and the crystal basin is significantly large, such that the probability of the system to evolve from the ideal glass state to the crystal state is improbable, and likely to occur over only long timescales inaccessible by MD. Similarly, MC is used to sample the energy minimums of the potential energy landscape with likelihood  $\exp(-\frac{G}{k_B T})$ . However, MC does not provide energy barrier information (dynamical information), which is essential to understanding the temporal element of a kinetic process, such as nucleation. For this reason, in order to study kinetic events at the atomic scale, advanced sampling methods capably of sampling macroscopically long time behavior is required. Further, the method needs to sample the energy barrier height and the energy minimums separating the two deep basins.

## 3.2 Metadynamics Method

Currently, the prevalent method for simulating atomic scale systems with computers is with molecular dynamics and Monte Carlo simulations. Molecular dynamics provides a great deal of information about the structure and dynamics of a system. However, molecular dynamics is heavily limited by its accessible timescale [31]. In order to accurately integrate Newton's equations (the foundation of molecular dynamics) a time step on the scale of molecular vibrations is required, typically on the scale of femtoseconds [31]. As a result, molecular dynamics at best can access time scales of up to microseconds. While microseconds can be sufficient for equilibrium studies, transient events such as nucleation requires orders of magnitude larger in timescale. Recent work on metadynamics methods show promise for overcoming this temporal restraint while simultaneously providing basin and barrier information. Metadynamics use a potential bias to drive the system from energy minimums, preventing arrest in anyone minimum [31]. The method overcomes large barriers allowing for direct sampling of a larger region of the potential energy landscape [31].

Metadynamics simulations begin with an arbitrary configuration, typically this will be a metastable configuration, as shown in Figure 3.2A. This configuration is represented by  $\mathbf{r}_{\min}^\alpha$ , where  $\alpha$  represents the number of penalties applied, in this case 0 [28, 29]. The system configuration during the simulation is represented by  $\mathbf{r}$ . From this point we add a penalty function  $\phi$ , which can take any shape, but we choose to use Gaussian

functions

$$\phi(\mathbf{r}, \mathbf{r}_{\min}^\alpha) = H \exp\left(\frac{-(\mathbf{r} - \mathbf{r}_{\min}^\alpha)^2}{2\sigma^2}\right) \quad (3.1)$$

where  $H$  is the height of the penalty function, and  $\sigma$  is the width of the penalty function [28, 29]. Appropriate choices for the penalty function variables  $H$  and  $\sigma$  are discussed in the following sections. We choose to use a Gaussian function because it is symmetric in all directions and quickly decays, thus it is local on the landscape [28, 29]. If we choose a function that does not locally decay, the function may also affect surrounding basins and saddle points. This added potential results in a force on each atom of

$$\mathbf{F}^\alpha(\mathbf{r}, \mathbf{r}_{\min}^\alpha) = -\nabla\phi(\mathbf{r}, \mathbf{r}_{\min}^\alpha) \quad (3.2)$$

$$\mathbf{F}^\alpha(\mathbf{r}, \mathbf{r}_{\min}^\alpha) = \frac{H}{\sigma^2} (\mathbf{r} - \mathbf{r}_{\min}^\alpha) \exp\left(\frac{-(\mathbf{r} - \mathbf{r}_{\min}^\alpha)^2}{2\sigma^2}\right) \quad (3.3)$$

This force is added to the force created by the pair potential of the system. As more penalties are applied the total potential and force on the system becomes

$$\begin{aligned} \Phi(\mathbf{r}) &= U(\mathbf{r}) + \sum_{\alpha} \phi(\mathbf{r}, \mathbf{r}_{\min}^\alpha) \\ \mathbf{F}(\mathbf{r}) &= \mathbf{F}_{pp}(\mathbf{r}) + \sum_{\alpha} \mathbf{F}^\alpha(\mathbf{r}, \mathbf{r}_{\min}^\alpha) \end{aligned} \quad (3.4)$$

where  $\mathbf{F}_{pp}(\mathbf{r})$  is the force generated from the pair potential,  $\Phi(\mathbf{r})$  is the potential energy of the system post penalty function,  $U(\mathbf{r})$  is the potential energy of the system without the penalty functions, and  $\mathbf{F}(\mathbf{r})$  is the force applied to the system with the penalties applied. Again,  $\mathbf{r}_{\min}^\alpha$  represents the different locations that penalties have been applied, and  $\alpha$  is the number of penalties applied [28, 29]. By applying these penalty functions the potential energy landscape has now been changed, as shown by Figure 3.2B. The figure shows the added penalty energy in blue, thus the landscape has changed from figure panel A to B. We then apply Newton's steepest descent algorithm to search for the most local minimum on the new landscape [28, 29].

$$\begin{aligned} \mathbf{r}_{i+1} &= \mathbf{r}_i - \delta \nabla \Phi(\mathbf{r}) \\ &= \mathbf{r}_i + \delta \mathbf{F}(\mathbf{r}) \end{aligned} \quad (3.5)$$

where  $i$  is the iteration of Newton's steepest descent, and  $\delta$  is the step size of each iteration in units of distance. The force in the previous equation is normalized by force of the atom with the maximum force, and is as a result unitless.  $\delta$  determines the stability of the algorithm. Once Newton's steepest descent has converged, the configuration is considered to be in a new local minimum of  $\Phi$ -space, we denote this

configuration as  $\mathbf{r}_{\min}^{\alpha+1}$  [28, 29]. At this point, we begin the algorithm again by applying another penalty of same size and shape [28, 29]. Once a penalty has been applied, the penalty is permanent in order to prevent atoms from back tracking into minimums we have already sampled. Original metadynamics methods did not use this history dependent penalty function application resulting in reduced simulation efficiency [28, 29]. As Figure 3.2C shows, once enough penalties have been applied, the penalty energy builds up allowing the system to overcome very large energy barriers. At this point, the system is now sampling high energy configurations. As more penalties are applied, the system is driven over the saddle points and enters a new large basin of the landscape, as shown by Figure 3.2D. At this point, the system is now sampling a new region of the landscape typically inaccessible to the system within the timescale of molecular dynamics [28, 29]. Thus, the result of the simulation is such that the system has undergone macroscopically long time dynamics, especially compared to the time scale accessible to molecular dynamics [28, 29].

At this point, it is important to distinguish  $\mathbf{r}_{\min}^\alpha$ , which is a minimum on the  $\Phi$ -space potential energy landscape, and the  $\mathbf{r}_{\min}$ , which is a minimum on the inherent potential energy landscape  $U(\mathbf{r})$  [28, 29]. These two minimums can be very different. During a simulation, we sample far more minimums on the  $\Phi$ -space landscape, but it may require multiple penalties to overcome a barrier on the inherent landscape and sample a new minimum on the inherent landscape [28, 29]. Typically,  $\phi(\mathbf{r}, \mathbf{r}_{\min}^\alpha)$  is zero at a new minimum of the inherent landscape because as discussed above, the penalty functions are heavily localized. The metadynamics simulation tracks the potential energy of the system as the system is biased over the landscape [28, 29]. Because the potential energy trajectory is known post computation, the minimums (basins) and saddles (barriers) can be calculated from this trajectory.

### 3.3 Implementation and Benchmarking

The metadynamics method described above was implemented into the open-source molecular dynamics package GROMACS [32]. Details on the code’s implementation and structure are discussed in Appendix A. Briefly, the method was implemented into the code such that the implementation does not affect GROMACS’ other capabilities, thus the package is still a fully functional edition. The code was implemented with OpenMP support to help reduce computational loads [33]. The parallelization is performed mostly on loops over atoms. One issue noticed with parallelization is that it causes the code to no longer create reproducible results. If the code is run serially, the results are always consistent for the same inputs. However, because metadynamics has summations over thousands to millions of small numbers once several penalties have been applied, the parallelization of these summations leads to finite precision errors. As a result, the first tens of

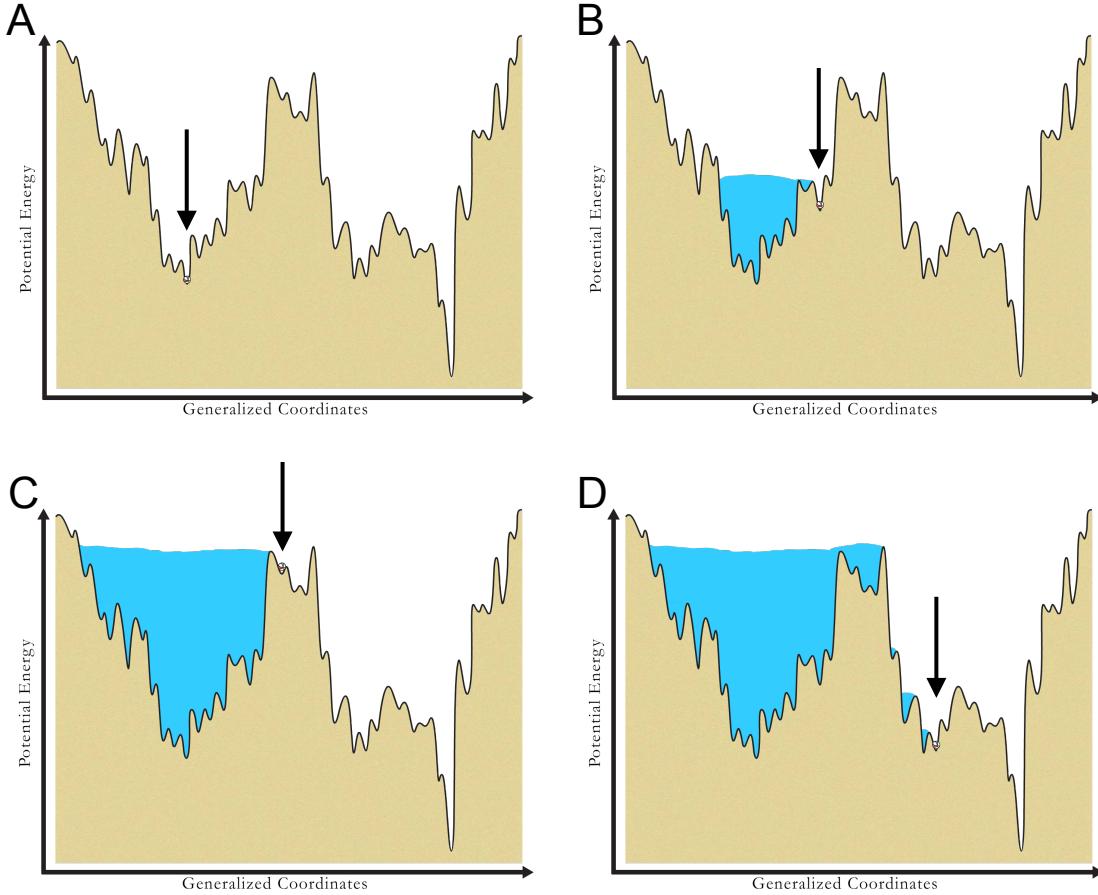


Figure 3.2: These four figures depict the metadynamics method with application to the potential energy landscape framework. The vertical axis is the potential energy of the system as a function of generalized coordinate. The ball, pointed to by the arrow, indicates the current configuration of the system. The tan region is inaccessible by the system. Figure A shows the system starting in an energy minimum on the landscape. From A to B, potential energy penalty functions are added in the form of Gaussians, shown in the figure by the blue region, which forces the system to evolve away from the original configuration and sample new energy minimums. From B to C, application of many more penalty functions has resulted in the left side of the landscape filling allowing the system to now sample high energy configurations. Lastly, from C to D, the system is driven over the large energy barrier separating two deep minima allowing the system to now sample a very different area of the landscape, previously inaccessible in a molecular dynamics simulation.

penalties applied are consistent for the same input values, but once hundreds and thousands of penalties are applied finite precision arithmetic results in different trajectories along the potential energy landscape. This is unavoidable, but also, presents no error in the final results as the finite precision errors merely leads to sampling the landscape in a different direction. Due to the high dimensional nature of the landscape, finite precision errors leads to sampling along a different path of the landscape, but we assume that as long as a sufficiently long trajectory is sampled, the trajectory is representative of the landscape as a whole.

While algorithmically metadynamics is simple, implementation is computationally intensive. In order to perform metadynamics, first the code must compute the pair potential forces in the system after every step, which alone is the most computationally demanding portion of a molecular dynamics simulation. Then, we add computation of the force on the system imposed by all the penalties added, which requires a loop over all atoms for every penalty function applied. Thus, after  $\alpha$  penalties, metadynamics adds  $\mathcal{O}(N\alpha)$  computations per iteration of Newton's steepest descent, which can become expensive. The initial metadynamics simulations required a large amount of computer time, far more than we expected, so while debugging we did time diagnostics to possibly increase the computational speed. As a result, we calculated the time spent per penalty applied, and, because the number of penalties grows during the simulation, we noticed a linear increase in computational time per penalty as more penalties are applied.

The increase in time is because each new penalty leads to an additional loop over all the atoms for every iteration of Newton's steepest descent. This significant increase in computational cost per penalty motivated us to add OpenMP parallelization. Further, this motivated us to seek other forms of computational speed up, because the success of metadynamics is dependent on applying as many penalties as possible in an efficient manner. A larger portion of the computation is performing Newton's steepest descent, which is dependent on the search parameter,  $\alpha$ .  $\alpha$  can be defined by the user, or several methods to compute the most efficient search parameter,  $\alpha$ , exists as well. However, these methods require the calculation of the inverse of the Hessian matrix, which is extremely expensive of a computation for large matrices. These methods are beneficial for small linear systems, but typical metadynamics systems have Hessian matrices of size  $N \times N$  (typically  $N$  ranges from large hundred to thousands), which is too expensive to compute. On the other hand, user fixed values can often be unstable or inefficient. Thus, we decided that a dynamically determined search parameter is the optimal method. Initially, the search parameter is set to an user estimated value. If Newton's steepest descent yields a step towards a minimum (energy was decreased), then the search parameter is increased by 20%. If the search yields an increase in energy, then we reject the configuration and decrease the search parameter by 50%. This resulted in significant speed-up from the fixed user value method. Further, this provided another check for convergence of the algorithm. We implemented that the

Newton's steepest descent was converged if, either, a successful step resulted a percent change of potential energy less than an user defined threshold percentage, or an unsuccessful step resulted in a minimization search parameter less than  $10e-5$  nm. While debugging, we noticed that the method typically converged by the energy criteria rather than the minimization step criteria.

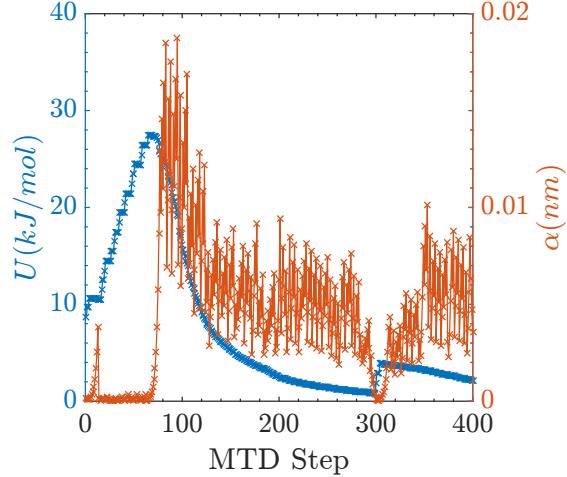


Figure 3.3: Added penalty energy and minimization step size,  $\alpha$ , as a function of metadynamics iteration. This figure shows the change in the added potential energy and minimization step size during convergence of one Newton's steepest descent. The figure shows that as the energy is decreased the minimization step is dynamically increased and decreased.

Figure 3.3 shows the results once the dynamically determined search parameter was implemented. The Figure shows the Newton's steepest descent for one convergence of a penalty function. The figure shows that, at the start, the search parameter is small and, as successful steps are made, the search parameter increases as the potential energy is decreased. Then, once the step size has become large, the steepest descent iteration results in an increase of energy (loss of stability), so the step size is decreased to retain stability. Clearly, from this figure, the original value is significantly smaller than the average value used for  $\alpha$ , thus, showing the success of the method. Note, this method resulted in a significant increase in computational speed with significantly less overhead than typical methods of determining  $\alpha$ , such as using the inverse of the Hessian matrix.

As explained in the previous section, the configurational coordinates used in our metadynamics method is the spatial coordinate of all atoms in the system, resulting in  $3N$  degrees of freedom, where  $N$  is the number of atoms. We first implemented this as the configurational coordinates of choice, and performed a

simulation of a binary Lennard Jones system. The simulation showed that after each penalty is applied, newton's steepest descent always resulted in the full minimization of the penalty energy. This was universal across penalty heights and widths. Once we visualized the simulation, the system was shown to have no internal relaxation of the atoms, instead the entire system drifted spatially. From this, we learned that  $3N$  degrees of freedom allows the system to minimize by drifting rather than internal reorganization. Then, we removed the net force of the system, or removed three degrees of freedom. As a result the force calculation for the metadynamics simulation becomes

$$\mathbf{F}_i = \mathbf{F}_i - m_i \mathbf{A} \quad (3.6)$$

where  $\mathbf{A}$  is a three dimensional vector defined as

$$\mathbf{A} = \frac{1}{N} \sum_i^N \frac{\mathbf{F}_i}{m_i} \quad (3.7)$$

where  $m_i$  is the mass of atom  $i$  and  $\mathbf{F}_i$  is the force vector of atom  $i$ . By removing this net force from the system, the system internally reorganized to minimize the penalty energy, as we should expect. Thus, our metadynamics method, and resulting landscape configurational coordinates space, has  $3N-3$  degrees of freedom.

### 3.4 Appropriate Simulation Parameters

The metadynamics method has several user definable parameters (energy convergence percent change, initial energy minimization step size  $\alpha$ , maximum number of minimization steps, penalty function height  $H$ , penalty function width  $\sigma$ , and maximum penalties). Most of these options have little impact on the computation and can therefore be chosen with ease. Percent change for energy convergence affects the margin of error of the minimums sampled. Initial energy minimization, step size can virtually be any number since the dynamic step size algorithm will correct for poor initial values, but generally a value on the scale of picometers is appropriate. The maximum number of minimization steps can be on the order of a few thousand, as can the maximum number of penalties. However, the penalty function width and height heavily influence the computation of the potential energy landscape.

The penalty function width and height impact both the results of the potential energy landscape sampled and the computational speeds of the simulation. As discussed in the previous sections, the potential energy landscape is rich with large and small energy barriers. The simplified picture of the potential energy land-

scape presented previously gives the illusion that a single path along the landscape exists. Contrary to this, a system has a plethora of paths to escape an energy minimum, each unique. The potential energy landscape contains more energy minimums and energy barriers separating these minimums than is calculable in any reasonable compute time. However, depending on the penalty energy function used in the metadynamics method, different paths and barriers will be preferentially sampled.

To study the height and width dependence of the potential energy penalty functions, we performed several simulations on the Kob-Andersen model (the parameters for this model are provided in Section 5.2). The penalty height was varied while maintaining a constant width squared of  $.1 \text{ nm}^2$ , and the penalty width was varied while maintaining a penalty height of  $1 \text{ kJ/mol}$ . The results are summarized in Figure 3.4. Both the height and the width affects the path along the potential energy landscape taken by the metadynamics method, as shown in Figure 3.4.

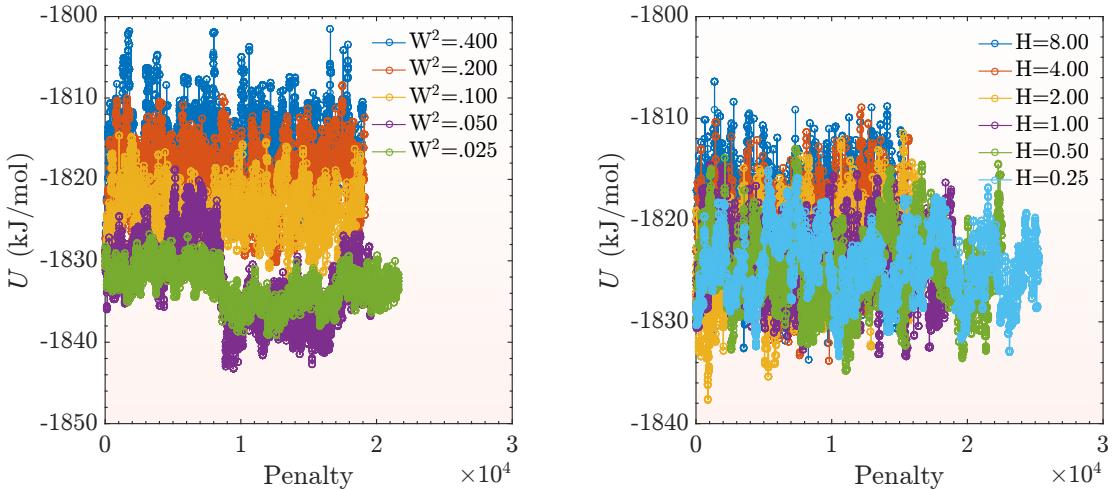


Figure 3.4: Potential energy versus penalty number for various penalty function heights and widths. The left figure shows the sampled potential energy landscape for varying penalty widths squared. The right figure shows the sampled potential energy landscape for varying the penalty heights. For the width variation, the height of the barrier was fixed to  $1 \text{ kJ/mol}$ . For the height variation, the width squared was fixed to  $.1 \text{ nm}^2$ .

Figure 3.4 shows the effect of the two penalty function parameters, the height and width of the penalty. The left figure shows the width dependence, and the right figure shows the height dependence. The width should in theory affect how well a single penalty fills the energy basin currently occupied by the system. If the width is too small, the penalty will under fill the broadness of the basin, and the system will inefficiently be biased out of the basin. However, a penalty with a small width will provide high resolution landscape sampling. The simulation will sample the small energy barriers of the landscape effectively, but will inefficiently sample the large barriers. On the other hand, if the width is too large, the penalty will no longer obey

the locality requirement of the penalty function. The penalty function will affect the system in nearby basins as well as the current basin. A large width will result in a penalty filling multiple basins. As a result, the simulation provides a sampled landscape with poor resolution but efficiently samples large barriers. From the figure, the small widths leads to a landscape sampled with low energy, with many low energy barriers sampled. The simulation also leads to more penalties being applied. The large widths resulted in much higher potential energy landscape sampling. Thus, we can conclude that wide penalties leads to a more effective filling of basins. A good estimate of an approximate width is from the approximate cage size of the system. The cage size can be estimated from the mean squared displacement of the system,  $\langle (r^2(t)) \rangle$ , and the cage is where the ballistic regime ends. For this system, this approximates to  $\langle (r^2(t)) \rangle = 0.1 \text{ nm}^2$ , which from the figure we can see leads to a landscape that appears to not under fill the basins like the lower widths.

Figure 3.4 also shows the height dependence of the penalty function. The height affects how accurately the potential energy landscape is sampled. If the height is small, the estimated values of the potential energy landscape will be very accurate, because the error of the values on the landscape are accurate to within the height of the penalty. However, with a small height the basins are filled slowly, thus requiring vastly more penalties to overcome energy barriers. This leads to a reduced sampling of the landscape. If the height is large, the accuracy of the simulation is low, but the barriers are sampled very efficiently. Thus there is an accuracy-efficiency trade-off for the height. We see this in Figure 3.4. The small heights lead to more penalties applied, because small penalties are easier for steepest descent to converge on, however, less barriers are sampled overall. On the other hand, the figure shows the largest penalty leads to low accuracy and sampling of only large barriers in the landscape. Overall, the best estimation for a strong height is to use the pair potential's value for  $\epsilon$ , which is 1 kJ/mol for this system. This is a good estimate, because  $\epsilon$  is the depth of the basin on the two atom potential energy landscape for Lennard-Jones. For non Lennard-Jones systems, the minimum value of the pair potential is the best estimate for the penalty height. This value will sample the potential energy landscape efficiently while minimizing the error induced by larger penalty function heights. Overall, the choice of the width and height is dependent on the desired resolution and efficiency desired. If high resolution is desired, small widths and heights are needed. If sampling a large portion of the landscape is desired, then large widths and heights are needed, but at the cost of resolution of the landscape.

### 3.5 Mapping the Landscape to Temperature

One aspect about metadynamics that has been overlooked is the concept of temperature. Metadynamics simulations are performed without a temperature component, because the goal of the simulation is to sample the potential energy landscape. Thus, in order to compare the simulation to experimental or MD results, the metadynamics results must be mapped to temperature in some manner. The method to determine the mapping from temperature to potential energy is as follows [28, 29]. At various temperatures and the same density as the metadynamics simulation, molecular dynamics simulations are performed. After set time intervals,  $\Delta t$ , the current configuration of the MD simulation is used as the initial configuration for energy minimization [28, 29]. The potential energy of the system after energy minimization is recorded, and then the MD simulation is continued from where it was prior to energy minimization. This method was also employed by a research group to sample the potential energy landscape [34]. The issue with this method for that purpose is that, like MC simulations, the energy barriers are not determined and there is no guarantee that subsequently sampled basins are adjacent on the landscape. Also for temperatures near  $T_g$ , the system struggles to escape deep minimums, thus the method provides an unrepresentative sample of the potential energy landscape. However, this method still serves as a method to map the inherent energy to the temperature. An example results of this method are shown in Figure 3.5.

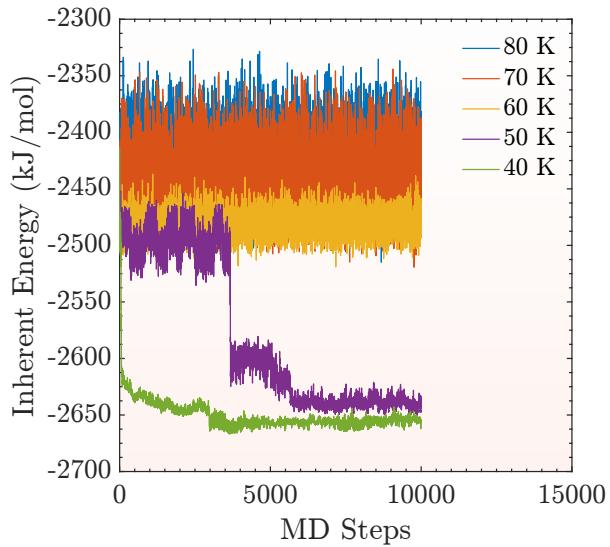


Figure 3.5: Inherent energy plotted as a function of MD steps for various temperatures. This figure shows the sampled inherent energies for calculating the mapping from inherent energy to temperature. Clearly, there is a transitional temperature around 50 K to 60 K. Below this temperature the sample values are concentrated around a particular value, and above this temperature the sample values have a large standard deviation.

Figure 3.5 shows the potential energy determined by energy minimization periodically during an MD simulation. As the figure shows, at high temperatures, the simulation samples a wide range of energy values because at high temperature, the system has the thermal energy to overcome high energy barriers and freely sample the entire landscape. As the temperature is decreased, the landscape begins to dominate the dynamics, and the energies sampled begin to decrease in value and in variance, until only discrete values are sampled. If sufficiently large numbers of samples are taken along the MD simulation, we can plot the average inherent energy versus temperature, shown in Figure 3.6 [28, 29].

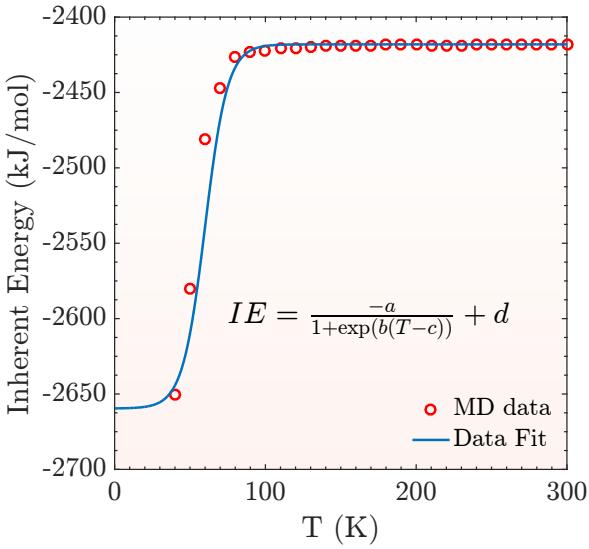


Figure 3.6: Inherent energy plotted as a function of temperature. This figure shows the data values determined from MD simulations in red open circles, and the data fit used to determine the mapping from inherent energy to temperature as a solid blue line. The figure shows the equation used to fit the data.

Figure 3.6 shows the average inherent energy (IE) as a function of temperature (T), the red circles show the values computed by the MD simulations. This figure shows that at high temperatures higher inherent energies may be sampled by the system. In order to create a relationship between the inherent energy and the temperature, the MD data is fit with a line. The MD data is fit to the line with the following equation

$$IE = \frac{-a}{1 + \exp(b(T - c))} + d \quad (3.8)$$

where  $IE$  is the inherent energy of the system,  $T$  is the temperature in Kelvin, and  $a$  is a fitting parameter calculated as  $a = \max(IE) - \min(IE)$ ,  $b$  is a free fitting parameter calculated by Matlab to best fit the data,  $c$  is a fitting parameter calculated as the temperature value located at  $\frac{a}{2}$ , and  $d$  is a fitting parameter

calculated as  $d = \max(IE)$ . This equation form was chosen because the shape of the data is assumed to be a logistic function. At high temperatures, the system will sample a large range of configurations with a large range of inherent energies. At temperatures below, but near the melting temperature, the system will no longer have the thermal energy to sample high inherent energy configurations with high frequency. The frequency with which inherent energies are sampled will decay exponentially, until a cross over temperature is reached. Below this temperature, the system will get stuck in local minimums and not have the thermal energy to escape, leading to sampling of only low inherent energy values. Thus, we assume a constant value at high temperature, a constant value at low temperature, and a logistic regression over all temperature.

### 3.6 Measure of Atomic Structure

To study phase transitions, we need to know the structural changes that occur in the system during the simulation. Many different order parameters exist to measure atomic structure, for example, the structure factor, the pair distribution function, or Voronoi polygons, to name a few. While each of these order parameters have use and merit in certain situations, they do not fully capture the information we seek in identifying a nucleation site.

Liquid to solid phase transitions are characterized by unique structural changes. The liquid state is isotropic (independent of direction) while the crystal state is anisotropic experiencing order in preferential crystallographic directions. The crystal state has both translational and rotational order, which does not exist in the liquid. Further, the crystal state shows long range repetition of atom positions in certain crystallographic directions. Typical measures of atomic structure such as the static structure factor or pair distribution function are useful for showing the difference in the structure between the liquid state and crystal state. However, they are sensitive to the orientation of the crystal structure. Further, while many studies have used the first peak in the structure factor to measure the structure during a phase transition, the parameter fails to provide local information of the structure. The structure factor and pair distribution function are system averaged quantities that do not provide information about where areas of high and low structure exists during the phase transition. This makes these parameters unsuitable for studying the structure of the system during nucleation.

As the name implies, one such order parameter that is orientationally invariant and provides information on the local degree of structure on a per atom basis is the local bond orientational order parameter (BOP). The local bond orientational order parameter is a quantity for measuring medium and short range order in a system.  $q_l$  is the bond order parameter for individual atoms in the system,  $Q_l$  is the system wide bond order

parameter,  $l$  is the order of the bond order parameter (typical values are 4 and 6),  $i$  is the atom number,  $t$  is the time.  $Q_l$  is equivalent to  $q_l$  averaged over all the atoms in the system [35].  $\bar{q}_l$  is equivalent to  $q_l$  averaged over an atom's nearest neighbors, this has the benefit of reducing the overlap of liquid and crystal values [36]. This quantity cannot be compared to experiments but can be very useful for determining structural order and phases in a system. The quantities are defined as below. The bond order parameter per atom is defined as

$$q_l(i, t) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^l |q_{lm}(i, t)^2|} \quad (3.9)$$

where

$$q_{lm}(i, t) = \frac{1}{N(i)} \sum_{j=1}^{N(i)} Y_{lm}(\hat{r}_{ij}(t)) \quad (3.10)$$

$m$  ranges from  $-l$  to  $l$ ,  $N(i)$  is the number of nearest neighbors around atom  $i$ , and  $Y_{lm}$  is the spherical harmonic of order  $l$  and  $m$ , and  $\hat{r}_{ij}$  is the unit vector between atom  $i$  and  $j$ . The unit vector can be converted to spherical coordinates which are required for the spherical harmonics. The full system bond order parameter of the system is defined as

$$Q_l(t) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^l |Q_{lm}(t)^2|} \quad (3.11)$$

where

$$Q_{lm}(t) = \frac{\sum_{i=1}^N N(i) q_{lm}(i, t)}{\sum_{i=1}^N N(i)} \quad (3.12)$$

and  $N$  is the number of atoms in the system. The nearest neighbor averaged bond order parameter is defined as

$$\bar{q}_l(i, t) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^l |\bar{q}_{lm}(i, t)^2|} \quad (3.13)$$

where

$$\bar{q}_{lm}(i, t) = \frac{1}{N(i)} \sum_{j=0}^{N(i)} q_{lm}(j, t) \quad (3.14)$$

Lastly,

$$\bar{Q}_l(t) = Q_l(t) \quad (3.15)$$

there is no difference between the system bond order parameter and the nearest neighbor averaged system bond order parameter. The nearest neighbors are defined as the atoms within a cutoff distance of a particular atom. The cutoff distance is typically determined to be the distance of the first peak of the pair distribution

Table 3.1: Mean values for the bond orientational order parameters for several crystal structures for a monoatomic Lennard-Jones liquid [35, 37].

	$Q_4$	$Q_6$	$q_4$	$\bar{q}_4$	$q_6$	$\bar{q}_6$
BCC	.0820	.5008	.0899	.0334	.4405	.4080
FCC	.1909	.5745	.1709	.1582	.5073	.4914
HCP	.0972	.4848	.1079	.0841	.4454	.4218
liq	.0000	.0000	.1090	.0312	.3601	.1620

function. Using a smaller value leads to smaller bond orientational order parameter values.

# References

- [1] Jean-Pierre Hansen and Ian R. McDonald. *Theory of Simple Liquids (Third Edition)*. Academic Press, Burlington, third edition edition, 2006.
- [2] P. Egelstaff. *An Introduction to the liquid state*. Elsevier Science, 2012.
- [3] D.R. Askeland, P.P. Fulay, and W.J. Wright. *The Science and Engineering of Materials, SI Edition*. Cengage Learning, 2011.
- [4] Richard Sear. What do crystals nucleate on? What is the microscopic mechanism? How can we model nucleation? *MRS Bulletin*, 41(05):363–368, 2016.
- [5] J.W.P. Schmelzer. *Nucleation Theory and Applications*. Wiley, 2006.
- [6] C.R. Barrett, W.D. Nix, and A.S. Tetelman. *The Principles of Engineering Materials*. Prentice-Hall, 1973.
- [7] Andrea Cavagna. Supercooled liquids for pedestrians. *Physics Reports*, 476(4-6):51–124, 2009.
- [8] Pieter Rein ten Wolde, Maria J. Ruiz-Montero, and Daan Frenkel. Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling. *The Journal of Chemical Physics*, 104(24):9932, 1996.
- [9] E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L F Abascal, and C. Valeriani. Homogeneous ice nucleation at moderate supercooling from molecular simulation. *Journal of the American Chemical Society*, 135(40):15008–15017, 2013.
- [10] David Reguera. How far is classical nucleation theory from predicting nucleation rates accurately? *AIP Conference Proceedings*, 1527(2013):9–14, 2013.
- [11] Guram Chkonia, Judith Wölk, Reinhard Strey, Jan Wedekind, and David Reguera. Evaluating nucleation rates in direct simulations. *Journal of Chemical Physics*, 130(6), 2009.
- [12] P.G. Debenedetti. *Metastable Liquids: Concepts and Principles*. Physical chemistry : science and engineering. Princeton University Press, 1996.
- [13] B. Mutaftschiev. *The Atomistic Nature of Crystal Growth*. Springer Series in Materials Science. Springer Berlin Heidelberg, 2001.
- [14] H. Vehkämäki. *Classical Nucleation Theory in Multicomponent Systems*. Springer Berlin Heidelberg, 2006.
- [15] David W Oxtoby. Homogeneous nucleation: theory and experiment. *Journal of Physics: Condensed Matter*, 4(38):7627–7650, 1992.
- [16] Jorge R. Espinosa, Carlos Vega, Chantal Valeriani, and Eduardo Sanz. Seeding approach to crystal nucleation. *Journal of Chemical Physics*, 144(3), 2016.

- [17] Stefan Auer and Daan Frenkel. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature*, 409(6823):1020–1023, February 2001.
- [18] Jim De Yoreo and Stephen Whitelam. Nucleation in atomic, molecular, and colloidal systems. *MRS Bulletin*, 41(05):357–360, 2016.
- [19] Takeshi Kawasaki and Hajime Tanaka. Formation of a crystal nucleus from liquid. *Proceedings of the National Academy of Sciences*, 107(32):14036–14041, 2010.
- [20] Peter G. Vekilov, Sungwook Chung, and Katy N. Olafson. Shape change in crystallization of biological macromolecules. *MRS Bulletin*, 41(May):375, 2016.
- [21] M Goldstein. Viscous Liquids and the Glass Transition: A Potential Energy Barrier Picture. *The Journal of Chemical Physics*, 51(9):3728–3739, 1969.
- [22] M. Scott Shell, Pablo G. Debenedetti, Emilia La Nave, and Francesco Sciortino. Energy landscapes, ideal glasses, and their equation of state. *The Journal of Chemical Physics*, 118(19):8821, 2003.
- [23] Thomas Middleton and David Wales. Energy landscapes of some model glass formers. *Physical Review B*, 64(2):1–20, 2001.
- [24] C.a. Angell. Landscapes with megabasins: Polyamorphism in liquids and biopolymers and the role of nucleation in folding and unfolding diseases. *Physica D: Nonlinear Phenomena*, 107(2-4):122–142, September 1997.
- [25] Pablo G Debenedetti, Thomas M Truskett, Catherine P Lewis, and Frank H Stillinger. Theory of supercooled liquids and glasses: Energy landscape and statistical geometry perspectives. *Advances in Chemical Engineering*, 28:21–79, 2001.
- [26] David J Wales. Decoding the energy landscape: extracting structure, dynamics and thermodynamics. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 370(1969):2877–99, June 2012.
- [27] D. Quigley and P. M. Rodger. Metadynamics simulations of ice nucleation and growth. *Journal of Chemical Physics*, 128(15), 2008.
- [28] Akihiro Kushima, Xi Lin, Ju Li, Jacob Eapen, John C Mauro, Xiaofeng Qian, Phong Diep, and Sidney Yip. Computing the viscosity of supercooled liquids. *The Journal of chemical physics*, 130(22):224504, July 2009.
- [29] Akihiro Kushima, Xi Lin, Ju Li, Xiaofeng Qian, Jacob Eapen, John C Mauro, Phong Diep, and Sidney Yip. Computing the viscosity of supercooled liquids. II. Silica and strong-fragile crossover behavior. *The Journal of chemical physics*, 131(16):164505, October 2009.
- [30] C. Patrick Royall and Stephen R. Williams. The role of local structure in dynamical arrest. *Physics Reports*, 560:1–75, 2015.
- [31] M. Ferrario, G. Ciccotti, and K. Binder. *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*. Number v. 2 in Lecture Notes in Physics. Springer Berlin Heidelberg, 2010.
- [32] Erik Lindahl, Berk Hess, and David van der Spoel. Gromacs 3.0: a package for molecular simulation and trajectory analysis. *Molecular modeling annual*, 7(8):306–317.
- [33] OpenMP Architecture Review Board. OpenMP application program interface version 3.0, May 2008.
- [34] Thomas B. Schrder, Srikanth Sastry, Jeppe C. Dyre, and Sharon C. Glotzer. Crossover to potential energy landscape dominated dynamics in a model glass-forming liquid. *The Journal of Chemical Physics*, 112(22), 2000.

- [35] Paul J. Steinhardt, David R. Nelson, and Marco Ronchetti. Bond-orientational order in liquids and glasses. *Physical Review B*, 28:784–805, Jul 1983.
- [36] Wolfgang Lechner and Christoph Dellago. Accurate determination of crystal structures based on averaged local bond order parameters. *The Journal of Chemical Physics*, 129(11), 2008.
- [37] Yanting Wang, S. Teitel, and Christoph Dellago. Melting of icosahedral gold nanoclusters from molecular dynamics simulations. *The Journal of Chemical Physics*, 122(21), 2005.
- [38] J. E. Jones. On the determination of molecular fields. ii. from the equation of state of a gas. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 106(738):463–477, 1924.
- [39] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, Jul 1967.
- [40] Jean-Pierre Hansen and Loup Verlet. Phase transitions of the lennard-jones system. *Physical Review*, 184:151–161, Aug 1969.
- [41] A. R. Dexter and A. J. Matheson. Elastic moduli and stress relaxation times in liquid argon. *The Journal of Chemical Physics*, 54(1):203–208, 1971.
- [42] Seungho Park, Sung San Cho, Joon Sik Lee, Young Ki Choi, and Ohmyoung Kwon. Molecular dynamics study on external field induced crystallization of amorphous argon structure. *KSME International Journal*, 18(11):2042–2048, 2004.
- [43] Sandip De, Bastian Schaefer, Ali Sadeghi, Michael Sicher, D. G. Kanhere, and Stefan Goedecker. Relation between the dynamics of glassy clusters and characteristic features of their energy landscape. *Physical Review Letters*, 112:083401, Feb 2014.
- [44] Walter Kob and Hans C. Andersen. Testing mode-coupling theory for a supercooled binary lennard-jones mixture i: The van hove correlation function. *Physical Review E*, 51:4626–4641, May 1995.
- [45] Amir Haji-Akbari and Pablo G Debenedetti. Direct calculation of ice homogeneous nucleation rate for a molecular model of water. *Proceedings of the National Academy of Sciences*, 112(34):10582–8, 2015.
- [46] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117(1):1 – 19, 1995.
- [47] Zhikun Cai and Yang Zhang. RMA: a relaxation excitation mode analysis. *To be submitted*.
- [48] Zhikun Cai and Yang Zhang. RMA: a relaxation mode analysis. *To be submitted*.
- [49] Martin Brehm and Barbara Kirchner. Travis - a free analyzer and visualizer for monte carlo and molecular dynamics trajectories. *Journal of Chemical Information and Modeling*, 51(8):2007–2023, 2011. PMID: 21761915.
- [50] T. Rg, K. Murzyn, K. Hinsen, and G. R. Kneller. nmoldyn: A program package for a neutron scattering oriented analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 24(5):657–667, 2003.
- [51] Nathan Walter, Abhishek Jaiswal, Zhikun Cai, and Yang Zhang. LiquidLib: A comprehensive tool for post processing of classical and *ab initio* molecular dynamics simulations of liquids with application to neutron scattering experiments. *To be submitted*.
- [52] G. Kresse and J. Hafner. *Ab initio* molecular dynamics for liquid metals. *Physical Review B*, 47:558–561, Jan 1993.
- [53] P. Charbonneau, a. Ikeda, G. Parisi, and F. Zamponi. Dimensional study of the caging order parameter at the glass transition. *Proceedings of the National Academy of Sciences*, 109(35):13939–13943, 2012.

- [54] Benoit Charbonneau, Patrick Charbonneau, Yuliang Jin, Giorgio Parisi, and Francesco Zamponi. Dimensional dependence of the Stokes-Einstein relation and its violation. *Journal of Chemical Physics*, 139(16), 2013.
- [55] Frank H. Stillinger and Aneesur Rahman. Improved simulation of liquid water by molecular dynamics. *The Journal of Chemical Physics*, 60(4), 1974.