

Utilizing Averaged Configurations from Molecular Dynamics Simulation Trajectories

P.G. Kusalik*, K. Gillis and J. Vatamanu

Department of Chemistry, University of Calgary, Calgary, Alberta, T2N 1N4, Canada
peter.kusalik@ucalgary.ca

Abstract

One of the challenges in the large-scale simulations required for many molecular systems (such as those of biological interest) is the recording, monitoring and visualization of configurational information from molecular dynamics trajectories spanning millions, and sometimes billions, of timesteps. A detailed record of instantaneous configurations along the full trajectory can quickly become unmanageable. In this paper we will describe an alternative approach that utilizes time coarse-graining, where configurations averaged over trajectory segments are used to follow the detailed molecular behaviour of a system over multiple-nanosecond simulations. We will demonstrate that the sensitivity of structural measures (order parameters) can be dramatically enhanced when applied to averaged configurations. We then specifically discuss the successful application of this approach to molecular dynamics simulations of crystal growth.

1. Introduction

Analysis of system configurations can be a key component in the computer simulation of molecular systems. These configurations, which supply full molecular details of the (relative) positions of the particles that make up the system, are frequently used to provide a means of exploring the microscopic behaviour captured by a molecular simulation. It is typical for these configurations to consist simply of the positions of all particles at a particular instant in time along the dynamical trajectory of the system. A set of such instantaneous configurations from a simulation can be saved and later analyzed further or visualized.

Complications arise in simulations, such as those required for biomolecular systems, where large length and time scales are necessary to capture the behaviour

of interest [1-4]. In these cases, the storage required to maintain this detailed record of the system's evolution (for example, as instantaneous configurations recorded every 50 fs) can be prohibitive (possibly requiring 10^6 configurations, or more). Moreover, the visualization of a system at this resolution over a multiple nanosecond trajectory becomes quite impractical, and may well contain detailed information (for example, due to the thermal motion of the atoms or molecules of interest) that tends to obscure the most relevant behaviour. One remedy to these problems is to select and record instantaneous configurations on a far coarser grid in time (perhaps every 50 ps, for example). However, while such an approach may retain some basic aspects of the dynamics exhibited by the system, considerable information is lost nonetheless. For instance, it may not be clear if the instantaneous configuration recorded at the end of a relatively long trajectory segment is representative of the behaviour exhibited by the system during that segment. Below we will demonstrate that an alternative approach, that employs averaged configurations to provide a true time coarse-graining of a simulation trajectory, captures considerably more detailed molecular information and allows larger scale behaviour in such systems to be tracked more readily.

2. Generation of Averaged Configurations

Spatial coarse-graining has become a widely used approach [3-6] in molecular simulations of large systems (such as membrane system) where by the detailed behaviour of groups of atoms are represented by (averaged into) single effective interaction sites. The present approach is similar in character in that to will rely on a coarse-graining, but now in time. Specifically, an averaged configuration can be produced for any particular trajectory segment by averaging molecular (or atomic) coordinates,

$$\bar{x} = \frac{1}{N_s} \sum_{i=1}^{N_s} x_i, \quad (1)$$

over the N_t time steps of the segment spanning a time τ . In Eq. (1), x can represent a positional coordinate; the case of orientational coordinates will be discussed below. The choice of segment length can be an important consideration; it typically is made as long as possible while still providing a reasonable vantage point (i.e., frequency of sampling) from which to observe the complex processes characterizing the molecular behaviour of interest. With an appropriate selection of trajectory segment length over which to average, extraneous particle motion (e.g. thermal motion) can be effectively removed so that any net (more gross) motion can be more easily observed.

For each trajectory segment one has in principle a distribution of values for each degree of freedom of the system. To provide additional information into the nature of these distributions (beyond their means, or first-moments), we also find it advantageous to monitor their second-moments, or widths. The root mean-squared (RMS) deviations,

$$\sigma(\Delta x) = (\overline{x^2} - \bar{x}^2)^{1/2}, \quad (2)$$

are also measures of the diffusive motion exhibited by a molecule during the trajectory segment and hence are clearly related to Debye-Waller factors.

The treatment of molecular orientations requires specific attention. At least for small molecules (like water), we find it advantageous to separate and average their degrees of freedom as positions and orientations. This approach allows for the conservation of molecular geometry during the averaging process. However, Eq. (1) cannot be applied directly to orientational coordinates due to the lack of commutativity of finite rotations. Fortunately, an averaging procedure for orientations has been developed recently [7] in terms of an average quaternion, or orientational centroid. In this procedure the orientational centroid, q_c , minimizes the function

$$G(q_c) = \sum_{i=1}^{N_t} \Gamma^2(q_c, q_i), \quad (3)$$

where $\Gamma(q_c, q_i) = 2\cos^{-1}(q_c \cdot q_i)$ is the arc length (or rotational angle) between the centroid (or average) orientation and the orientation q_i . A simple Monte Carlo search algorithm can be used [7] to determine a value of q_c for any particular set of N_t orientations.

3. Applications

3.1. Homogeneous Nucleation

In simulation studies of homogeneous crystal growth, where one attempts to observe the spontaneous formation of crystalline order in an otherwise bulk liquid, the identification and characterization of the critical nucleus is a crucial aspect. Local structural order parameters has been developed and used for this

purpose [8-10], although the challenge has been to identify order parameters that are both sensitive and generic. The local structural order parameters utilized by Frenkel and co-workers [9,10], based on spherical harmonics, are the most widely used. Here we will only briefly outline how these order parameters are constructed.

One starts by defining for each particle i

$$\bar{q}_{lm}(i) = \frac{1}{N_1(i)} \sum_{j=1}^{N_1(i)} Y_{lm}(\hat{r}_{ij}) \quad (4)$$

where $N_1(i)$ is the number of first neighbors of i , Y_{lm} is a spherical harmonic and \hat{r}_{ij} is the unit vector representing the direction of the separation vector joining particles i and j . It has been previously shown [9,10] that the choice of $l=6$ provides a rather robust measure of local order in most systems. One then generates the normalized 13-dimensional complex vector $q_6(i)$ from the components $\bar{q}_{6m}(i)$ for each particle i . The order around any particular particle i can then be quantified by the scalar measure

$$q_i(i) = \left[\frac{4\pi}{2l+1} \sum_{m=-l}^l |\bar{q}_{lm}(i)|^2 \right]^{1/2}. \quad (5)$$

We subsequently consider the coherence between the measures $q_6(i)$ and $q_6(j)$ for the neighboring particles i and j . Specifically, if $q_6(i) \cdot q_6(j) > a_{con}$, where a_{con} is some fixed value, then the pair is labeled as connected. The number of connected neighbors of each i , $N_{con}(i)$, is obtained and if $N_{con}(i)$ is greater than some threshold, i is labeled as "solid-like".

In utilizing this order parameter, previous workers [9-11] have employed the coordinates from instantaneous configurations from their simulations. Here we will test the impact of using (rolling) average coordinates as obtained from Eq. (1). Since the goal is to detect solid-like particles in systems undergoing homogenous nucleation, where the character of particles might be expected to change rapidly, relatively short trajectory segments, composed of $N_t=50, 100, 200$ and 400 timesteps, were examined. Molecular dynamics simulations of systems of 4000 Lennard-Jones (LJ) particles were carried out at a reduced density of 0.95. Particles were considered to be first neighbors if their separation was less than 1.5σ , where σ is the particle diameter.

To demonstrate the utility of averaged configurations in monitoring local structure simulations were performed at a reduced temperature of 0.65 for a LJ liquid and an FCC crystalline solid. Fig. 1 compares probability distributions functions for the local order parameter $q_6(i)$ recorded when instantaneous coordinates are used with those obtained from averaged coordinates. One immediately observes that while averaging has little effect on the distribution of values observed to a liquid system, there is considerable

sharpening of the distribution for the local order parameter in a solid system as the level of averaging of the coordinates is increased.

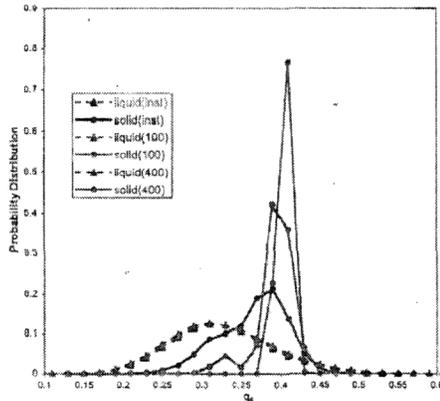


Figure 1. Probability distribution functions for values of $q_6(i)$ from simulations of liquid (dashed lines) and FCC solid (solid lines) LJ systems at a reduced temperature of 0.65. The blue lines represent data obtained from instantaneous configurations, while the red and green lines are results obtained from coordinates averaged over 100 and 400 timesteps, respectively.

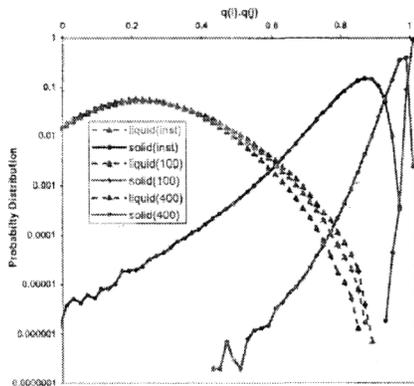


Figure 2. Probability distribution functions for values of $q_6(i) \cdot q_6(j)$ from simulations of liquid (dashed lines) and FCC solid (solid lines) LJ systems at a reduced temperature of 0.65. The symbols are defined as in Fig. 1. Note the log scale.

We also find that the use of averaged configurations has considerable impact on our ability to identify appropriate values for the parameters a_{con} and $N_{con}(i)$ (i.e. to distinguish solid-like from liquid-like

molecules). In Fig. 2 probability distributions functions for $q_6(i) \cdot q_6(j)$ obtained from instantaneous and averaged coordinates are compared. It can be seen that while the distributions from the liquid systems are relatively unaffected by local time averaging (coarse-graining), there is considerable sharpening and shifting of the distributions towards their ideal value (1.0) for crystalline systems. Consequently, the overlap between liquid and solid distributions is significantly reduced (by at least two orders of magnitude) even when configurations averaged over as few as 100 timesteps are employed. The crossover points in Fig. 2 were utilized to provide appropriate values for the connection threshold, a_{con} , in each case.

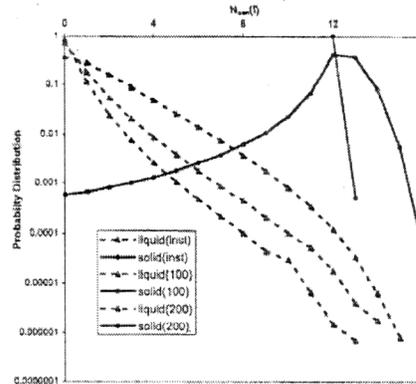


Figure 3. Probability distribution functions for values of $N_{con}(i)$ from simulations of liquid and FCC solid LJ systems at a reduced temperature of 0.65. The blue lines represent data obtained from instantaneous configurations, while the red and black lines are results obtained from coordinates averaged over 100 and 200 timesteps, respectively. It should be noted that the lines from averaged configurations for the FCC solid are superimposed on this plot.

Fig. 3 shows probability distributions functions for $N_{con}(i)$ obtained when coordinates from instantaneous and averaged configurations are used. We see a rather dramatic improvement in the resolution of the distributions from crystalline and liquid systems as local time averaging (coarse-graining) is enabled. The distributions from the liquid simulations consistently shift to the left (to lower values) as more averaging is performed. While the solid distribution appears somewhat broad and spans the full range of values when instantaneous coordinates are utilized, all the levels of averaging shown in Fig. 3 produce a distribution that has essentially become a delta function, centered at its ideal value of 12. The obvious conclusion is that the sensitivity of this order-

parameter, and hence ones ability to distinguish between solid-like and liquid-like particles in these simulations, has been greatly enhanced. Even with only modest amounts of averaging (i.e. 100 timesteps) the overlap between liquid and solid distributions is significantly reduced, i.e. by two orders of magnitude.

3.2. Heterogeneous Crystal Growth

We have employed averaged configurations extensively in our molecular simulation studies of heterogeneous crystal growth [12-14], where the detailed analysis of multiple nanosecond trajectories is required to uncover the underlying processes associated with crystal growth. The details of the simulation methodology we have employed can be found elsewhere [12,15].

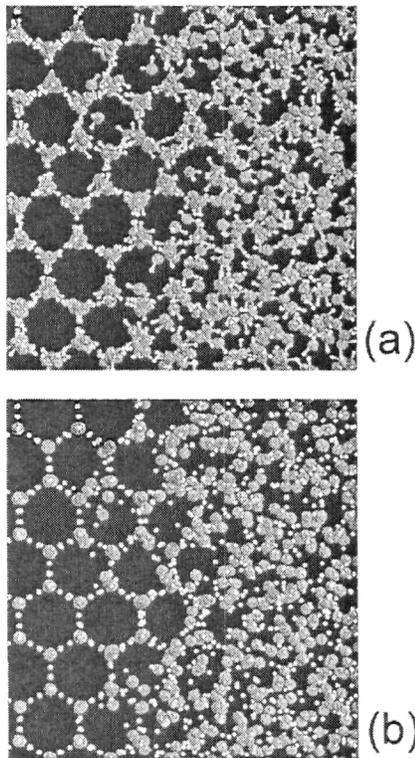


Figure 4. (a) Instantaneous and (b) averaged configurations of the same interfacial region of an ice/water system during crystal growth. The averaging in (b) is over 20 ps.

Fig. 4 compares an averaged configuration from a 20 ps trajectory segment with the instantaneous configuration from the end of this trajectory segment. The ice/water system pictured in Fig. 4 is looking down the c-axis of hexagonal ice (I) during its crystal growth. Perhaps the most striking aspect of the averaged configuration is the clarity of its crystalline structure and the distinctiveness of the interfacial layer. Clearly even at this level, the averaged configuration is providing a superior view of the systems behaviour. We point that while some molecular overlaps can occur in an averaged configuration within the liquid region of a system (due to the diffusive motion of these particles), these are not problematic in our analysis.

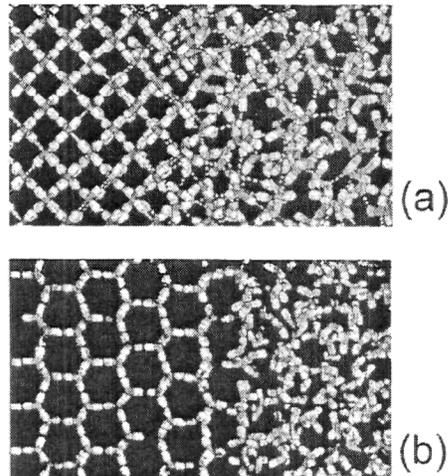


Figure 5. (a) Averaged configuration of the interfacial region of the [001] face of a growing cubic ice (I) crystal. A trajectory segment of 25 ps was used to produce the averaged positions and orientations of the molecules. The blue dashed lines represent hydrogen-bonded molecules. (b) Averaged configuration of the interfacial region of the [0001] face of a growing hexagonal ice (I) crystal. The length of averaging trajectory segment was 75 ps. The molecules are colored as discussed in the text.

To enhance further our ability to extract visual information from averaged configurations, we have labeled water molecules within each averaged configuration as being translationally solid-like or liquid-like, and similarly as being rotationally solid-like or liquid-like. This was done by identifying appropriate thresholds for the RMS deviations (i.e., diffusive behaviour) for both positions and orientations that are consistent with values found in the bulk crystal; molecules with values above these thresholds are identified as being liquid-like. The molecules could

be then colored according to their solid-like/liquid-like labels. Specifically, if a molecule was labeled translationally solid-like, its oxygen was colored red, otherwise the oxygen was colored magenta. If a molecule was labeled rotationally solid-like, its hydrogens were colored white, otherwise the hydrogens were colored yellow. We can see from Fig. 5, where the interfacial region of an averaged configuration for two ice/water systems are shown, that these labels provide considerable insight into the molecular behaviour at the interfaces. It should also be noted that the qualitative characteristics observed in Fig. 5 do not apparently change in any significant way if the length of the trajectory segment is changed somewhat (e.g. increased from 25 to 75 ps. as in Figs. 5(a) and 5(b), respectively). This indicates that as long as a reasonable choice for the length of time coarse-graining is utilized, the results obtained are rather insensitive to this value.

4. Conclusions

We have shown that averaged configurations, representing a time coarse-graining over the trajectory of a system, can be very useful for following the molecular mechanisms of crystal growth. The inclusion of second-moment information was also observed to add significant depth to the information contained in these averaged configurations. We have used this approach extensively [12-16] to studying crystal growth of pure and mixed crystals of both atomic and molecular systems. We would expect that it would prove similarly useful in simulations of other systems characterized by rather slow processes, for example in the folding of a protein or in the transport of an ion across a membrane.

References

[1] Bergethon, P.R.: *The Physical Basis of Biochemistry: The Foundations of Molecular Biophysics* (Springer-Verlag, NY, 2000)

[2] Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M.: *Computational Biochemistry and Biophysics* (Marcel Dekker, USA, 2001)

[3] Lyubartsev, A.P., Karttunen, M., Vattulainen, I., Laaksonen, A.: *Soft Mater.* **1** (2003) 121-137

[4] Nielsen, S. O., Lopez, C. F., Srinivas, G., Klein, M. L.: *J. of Phys.: Cond. Matter* **16** (2004), R481-R512

[5] Ortiz, V., Nielsen, S. O., Klein, M. L., Discher, D. E.: *J. of Polymer Sc., B: Polymer Phys.* **44** (2006) 1907-1918

[6] Srinivas, G., Discher, D. E., Klein, M. L.: *Nature Materials* **3** (2004), 638-644

[7] Hernandez de la Peña, L., Kusalik, P. G.: *Mol. Phys.* **102** (2004) 927-937

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

[9] ten Wolde, P. R., Ruiz-Montero, M. J., Frenkel, D.: *Phys. Rev. Lett.* **75** (1995) 2714

[10] ten Wolde, P. R., Ruiz-Montero, M. J., Frenkel, D.: *J. Chem. Phys.* **104** (1996) 9932

[11] Desgranges, C., Delhommelle, J.: *J. Am. Chem. Soc.* **128** (2006) 10368-10369

[12] Razul, M. S. G., Tam, E. V., Lam, M. E., Kusalik, P. G.: *Mol. Phys.* **103** (2005) 1929

[13] Razul, M. S. G., Hendry, J. G., Kusalik, P. G.: *J. Chem. Phys.* **123** (2005) 204722

[14] Vatamanu, J., Kusalik, P. G.: *J. Phys. Chem. B.* **110** (2006) 15896

[15] Vatamanu, J., Kusalik, P. G.: accepted *J. Chem. Phys.*

[16] Vatamanu, J., Kusalik, P. G.: *J. Am. Chem. Soc.* **128** (2006) 15588