Comparative study on methodology in molecular dynamics simulation of nucleation

Jan Julin, Ismo Napari, and Hanna Vehkamäki
Department of Physical Sciences, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland

(Received 24 January 2007; accepted 23 April 2007; published online 14 June 2007)

Gas-liquid nucleation of 1000 Lennard-Jones atoms is simulated to evaluate temperature regulation methods and methods to obtain nucleation rate. The Berendsen and the Andersen thermostats are compared. The Berendsen thermostat is unable to control the temperature of clusters larger than the critical size. Independent of the thermostating method the velocities of individual atoms and the translational velocities of clusters up to at least six atoms are accurately described by the Maxwell velocity distribution. Simulations with the Andersen thermostat yield about two times higher nucleation rates than those with the Berendsen thermostat. Nucleation rate is extracted from the simulations by direct observation of times of nucleation onset and by the method of Yasuoka and Matsumoto [J. Chem. Phys. **109**, 8451 (1998)]. Compared to the direct observation, the nucleation rates obtained from the method of Yasuoka and Matsumoto are higher by a factor of 3. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2740269]

I. INTRODUCTION

Nucleation in gaseous phase is usually described by classical nucleation theory (CNT), which subjects the inherently microscopic nature of nucleation process to macroscopic thermodynamic treatment. The advantage of this approach is simplicity and practicality: only measurable properties, such as liquid density, surface tension, vapor pressure, etc., are needed to obtain the rate of appearance of new particles, that is, nucleation rate, from a simple equation. The disadvantage is the unreliability of results. By assuming thermodynamics (macroscopic theory) to hold for clusters of few tens of molecules (or even less), CNT brings about an unpredictable and system-dependent error source. It is not uncommon to find a discrepancy of over ten orders of magnitude between the measured nucleation rate and the nucleation rate predicted by CNT.

To elucidate the problems of CNT and to explore and test more refined approaches to nucleation, a microscopic basis is required. The most straightforward molecular-level method to investigate gas-liquid nucleation is the molecular dynamics (MD) simulation.[1] In the so called direct MD method one places a large number of molecules in a simulation box, quenches the system to a supersaturated state, and follows the trajectories of the particles by integrating the equations of motion. The nucleation event and the subsequent growth of the molecular cluster can readily be observed during the simulation. This method contrasts with an indirect MD simulation, where a nucleated cluster is already present at the start of the simulation (see, for example, Ref. 2). Other methods, for example, Monte Carlo simulation and density functional theory, are also viable but MD has the advantage of closely emulating the actual dynamic nucleation process. Nevertheless, MD method has some intrinsic difficulties, especially related to thermostating and extracting nucleation rate from the simulations.

Molecular dynamics is essentially classical mechanics. From the statistical physics viewpoint the simulation is performed in constant energy ensemble (*NVE*), which is somewhat restricting because real nucleating systems are usually in contact with a heat bath, corresponding to the canonical (*NVT*) ensemble in the statistical description. In nucleation experiments the thermal coupling is achieved by adding carrier gas (usually an inert gas) to the nucleating vapor. In atmospheric nucleation, carrier gas is always present in the form of "air molecules" (mainly O_2 and N_2). The interaction with the carrier gas thermalizes the nucleating substance, and the nucleation process is then isothermal. If carrier gas is missing from a MD simulation, the nucleating clusters heat up during condensation, which increases the evaporation rate from clusters and, in consequence, lowers the nucleation rate.

Thermal contact in MD simulations via carrier gas can be a time-consuming method, because the carrier gas particles must outnumber the nucleating particles by several times for effective thermalization. Usually a more cost-effective method to imitate the presence of carrier gas is to couple the system to an artificial thermostat. A thermostat changes the velocities of the particles in order to achieve the desired temperature. However, the thermostat may disturb the dynamics in a nonrealistic manner, and generate a system which does not correspond to a canonical one. Previous direct MD studies on gas-liquid nucleation have used both the carrier-gas method[3–8] and thermostats[9,10] to thermalize the system. However, no studies exist where thermostating schemes are compared in a systematic manner.

In this work we perform gas-liquid nucleation simulations of simple Lennard-Jones particles. We compare two existing methods, the direct observation of nucleation events and the method by Yasuoka and Matsumoto,[4] to obtain the nucleation rate from the simulation and assess two popular thermostats, the Berendsen and the Andersen thermostats. The results are compared to simulations without thermostat-

ing. The paper is organized as follows. In Sec. II we present the system and describe the simulation method. Next, in Sec. III, we show our results for temperature control, velocity distributions, nucleation rate, and cluster distributions. The section ends with a comparison to CNT. Finally, in Sec. IV we summarize our main conclusions.

II. MODELS AND METHODS

We simulated a system of 1000 particles in a box of size $120\times120\times120$ Å^3 with periodic boundary conditions. As is shown by Wedekind *et al.*[3] this size is easily large enough to avoid possible finite-size effects. The interaction potential was a Lennard-Jones potential,

$$V(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6], \tag{1}$$

and the parameter values for argon, $\sigma=3.40$ Å and $\epsilon/k_B = 120$ K, were used. The potential cutoff was at 5σ.

The system starts at a temperature of 130 K, and is quenched after 1 ns to the target temperature of 85 K. A 6 fs timestep is used throughout the simulation. The simulation was terminated either after 10 ns of simulation time or after the size of the largest cluster exceeded 200 atoms, depending on which happened first.

Two well-known thermostats, the Berendsen[11] and Andersen[12] thermostats were used. When thermostating with the Berendsen thermostat the velocities of every particle is scaled by a factor

$$\lambda = \left[1 + \frac{\Delta t}{\tau}\left(\frac{T_0}{T} - 1\right)\right]^{1/2}, \tag{2}$$

where T_0 is the target temperature, T is the current kinetic temperature of the system, Δt is the timestep, and τ is a preset time parameter. For the time parameter the value $\tau = 400$ fs was used. This value was found appropriate by Berendsen *et al.*[11] (Further discussion on the effect of the strength of the coupling can be found in Sec. III A.)

When using the Andersen thermostat a number of particles are randomly selected to be given new velocities. These new velocities are drawn from a Maxwell-Boltzmann distribution corresponding to the target temperature. The probability for a particle to get a new velocity is $\nu\Delta t$, where Δt is again the timestep and ν is a parameter describing the collision frequency with an imaginary heat bath. In these simulations the value of the parameter was set to $\nu=5\times10^{-4}$ fs^{-1}. Should one consider this value to correspond to a collision frequency with carrier gas atoms of roughly the size of N_2 or O_2, there would be about seven times more carrier gas atoms than nucleating atoms present.

An apparent unnaturalness in both of these thermostats is the fact that they remove heat from particles inside large clusters as well. In order to test the effect this has on the nucleation process we also conducted simulations with both thermostats where only the free particles were thermostated. That is, those particles that do not belong to any clusters when a cluster is defined with the usual Stillinger criterion[13] with the cutoff at 1.5σ. In a more natural scenario one would of course wish to use the thermostat on all particles in the small clusters and the particles on the surfaces of larger clusters as well, but the downside is a considerable increase in simulation time as the list of neighbors for each atom should then be more frequently updated.

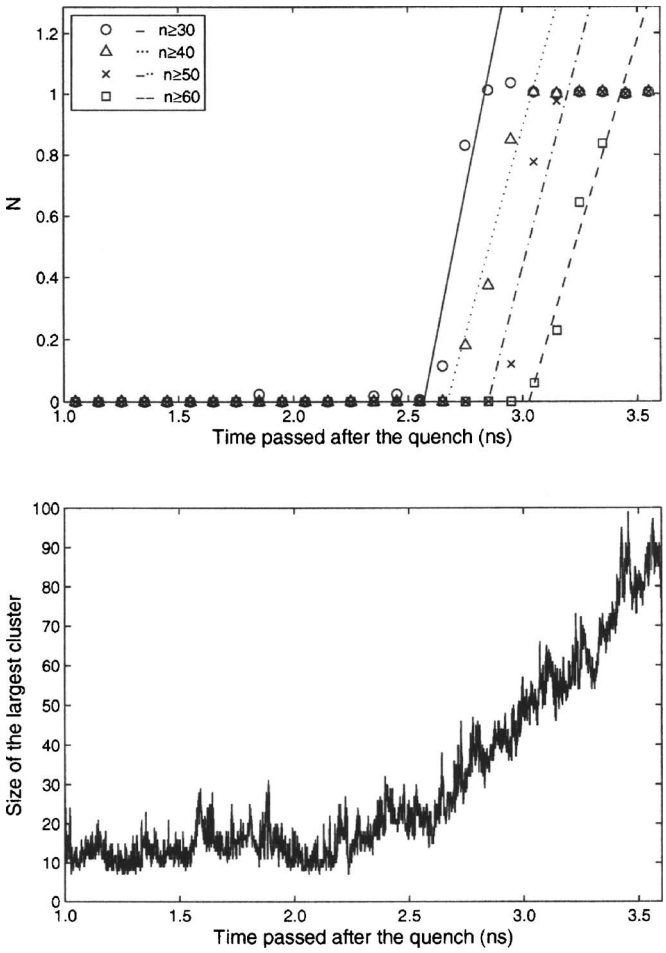

FIG. 1. Top: Time development of number of clusters larger than the threshold sizes 30, 40, 50, and 60. Bottom: Time development of the number of particles in the largest cluster for the same simulation run. The Berendsen thermostat was used for this run.

Two different methods to obtain the nucleation rate were used. First of these was the method of Yasuoka and Matsumoto.[4] In this method the number of clusters exceeding a certain threshold size is plotted as a function of simulation time. The slope of the linear dependence found is then divided with the volume of the simulation box to get the nucleation rate. In our simulation we used threshold cluster sizes of 30, 40, 50, and 60 particles. An example of a Yasuoka-Matsumoto graph used to determine the nucleation rate is shown in Fig. 1, and as can be seen the slopes are similar for the different threshold sizes. Also shown is the number of atoms in the largest cluster as a function of time from the same run.

The other method used was the so called direct observation method. In direct observation the nucleation rate is determined by observing the time of nucleation onset, the time when the first cluster that is able to grow appears. In the example run shown in Fig. 1, the onset is clearly seen happening at 2.2 ns after the quench. The nucleation rate is then estimated as (one cluster)/(time of onset $\times$ volume of simulation box). Naturally this method requires sufficient number

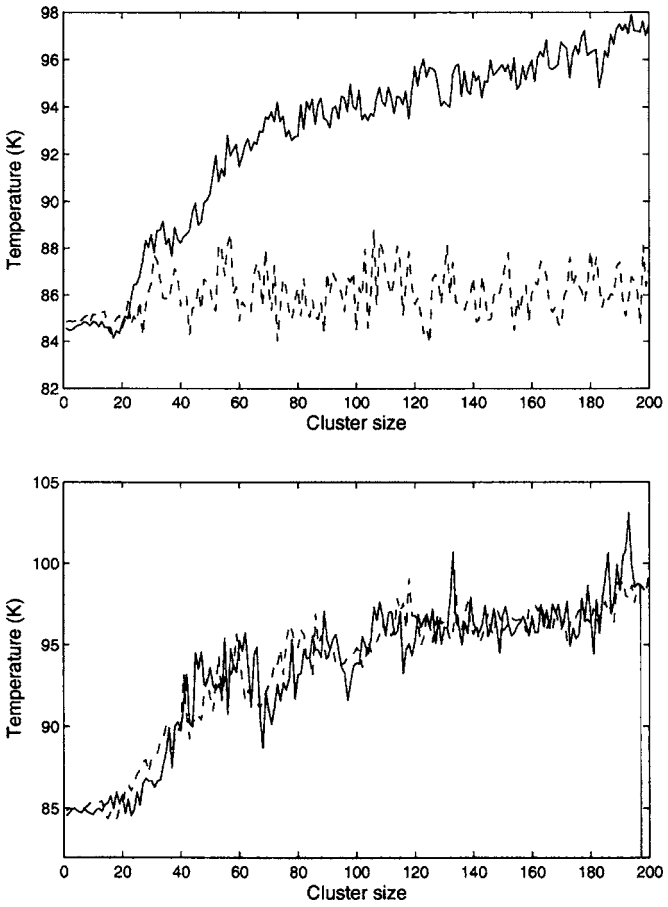

FIG. 2. Cluster temperatures averaged over several runs. The temperatures are collected starting from the quench until the end of the simulation. Top: Regular Berendsen (solid line) and Andersen (dashed line) thermostated runs. Bottom: Berendsen (solid line) and Andersen (dashed line) thermostats applied only to monomers.

of simulation runs to acquire reasonable statistics. This requirement is emphasized by the fact that the time of onset could vary greatly from one run to another. It is worth noting that the broad range of observed times of onset did not have an effect on nucleation rates obtained with the method of Yasuoka and Matsumoto.

III. RESULTS AND DISCUSSION

A. Temperature control

The two thermostats do their job, combating the temperature increase due to condensation heat, with varying success. As seen in Fig. 2, the Andersen thermostat manages to keep the temperature well in check for all cluster sizes, and the deviation from the desired temperature is at most about 3 K. The Berendsen thermostat manages to keep clusters up to size 25 at the desired temperature, which is near the critical cluster size of our simulations. However, for larger clusters than this the temperature rises rapidly and the largest clusters are over 10 K hotter than the desired temperature. Smaller values of the Berendsen time parameter τ were also tried, but it turned out that even these stricter couplings to the thermostat were unable to keep the temperatures of larger clusters at the desired level.

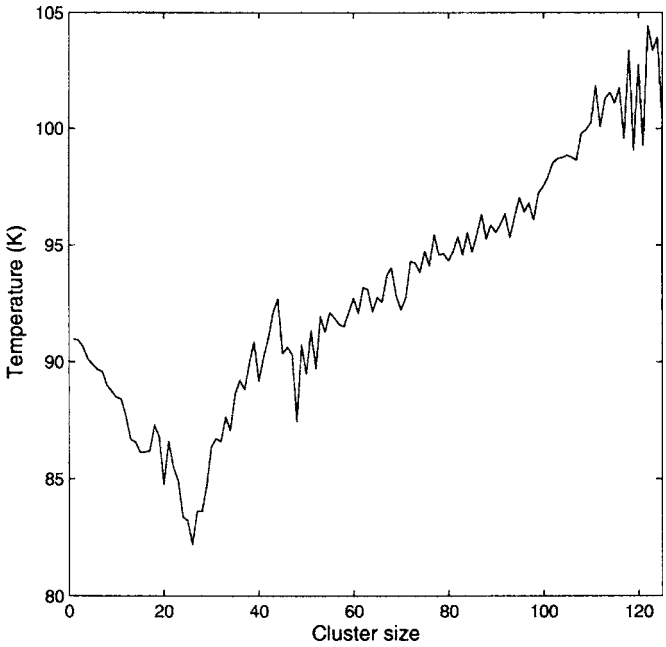

FIG. 3. Cluster temperatures for *NVE* runs.

When the thermostating is applied only to the free particles, in other words monomers, the temperature of clusters behaves in both cases almost similarly to the regular Berendsen thermostated runs. As one should expect, when the Berendsen thermostat is used only on monomers, the resulting cluster temperatures are slightly higher than in regular Berendsen thermostated runs. Since both thermostats managed to keep the temperature of the free particles where desired, it comes as no surprise that when the Andersen thermostat is used only on monomers, the cluster temperatures mimic the behavior seen in corresponding Berendsen thermostated runs.

With such temperature behavior it is clear that only the regular Andersen thermostated runs can be considered to produce a canonical ensemble. However, the gas phase does remain in constant temperature in all four thermostating methods considered, meaning that the situation is very different from constant energy (*NVE*) simulations, as Fig. 3 shows. In our *NVE* simulations the system is temperature controlled with the Berendsen thermostat until 500 ps after the quench, after which the thermostat is removed and the simulation runs as a *NVE* simulation. The cluster temperatures of Fig. 3 have been collected only from the *NVE* part of the simulations. When nucleation occurs in constant energy simulations, the smaller cluster sizes become over 5 K hotter than the target temperature. Inspection of the temperature of free particles as a function of simulation time reveals that their temperature begins to increase when the temperature of the clusters begins to increase after the nucleation onset. This heating of the vapor phase is due to heated atoms which evaporate from the nucleating clusters and slowly increase the vapor temperature. Similarly to the free particles, the temperature of clusters does not deviate much from the target temperature before nucleation has occurred, explaining why clusters with sizes near the critical size are cooler since these are only found in the simulation in this fairly isothermal period.

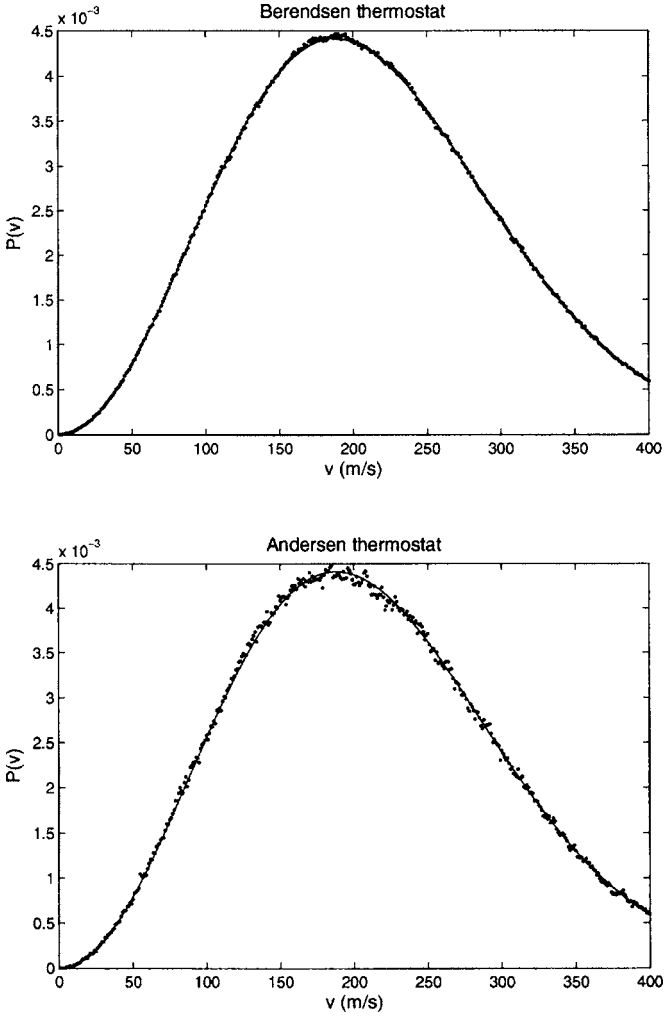

FIG. 4. Velocity distributions of Berendsen (top) and Andersen (bottom) thermostated runs. Solid line is the theoretical Maxwell-Boltzmann distribution at $T=85$ K. Velocities of all the atoms in the systems are included in the distribution.

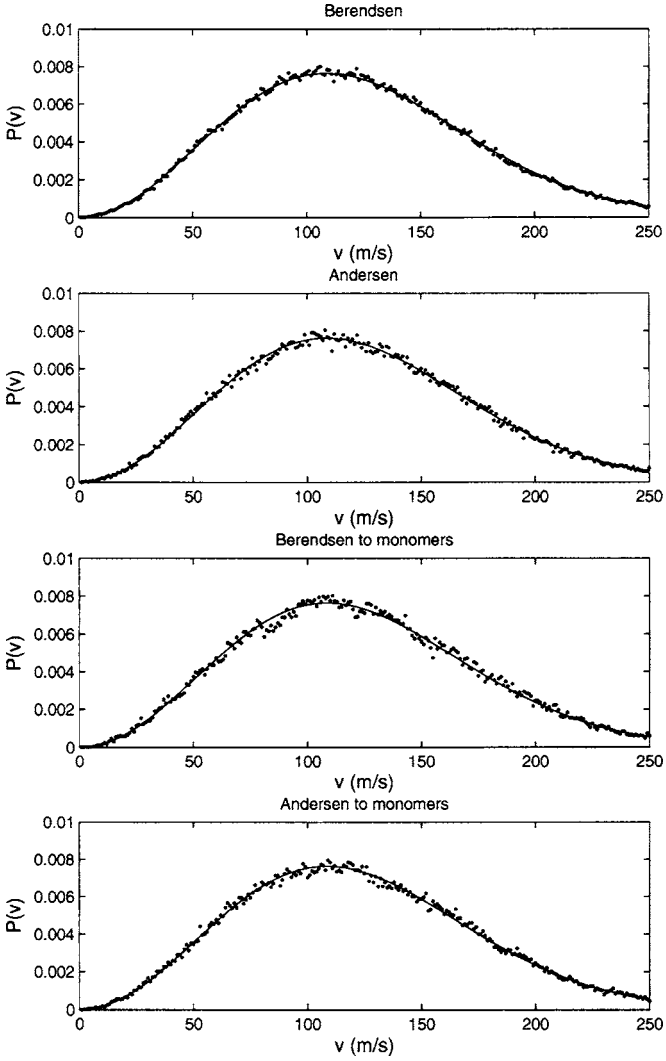

FIG. 5. Velocity distributions of the centers of mass of three-particle clusters for different thermostats. From top to bottom: Berendsen, Andersen, Berendsen only on monomers, and Andersen only on monomers. Solid lines represent the theoretical Maxwell-Boltzmann distribution at $T=85$ K.

B. Velocity distributions

As an initial configuration the velocities of the particles are drawn from a Maxwell-Boltzmann distribution. During the simulation the thermostats interfere with the velocities, but this does not change the shape of the distributions as can be seen in Fig. 4, where the dots represent the velocity distributions collected from the simulations and the solid line is the theoretical Maxwell-Boltzmann distribution corresponding to $T=85$ K. The time period from which the velocities are collected begins at 200 ps after the quench and continues until the nucleation onset. The smoother shape of the distribution from the Berendsen thermostated run is due to a late onset at about 8 ns after the quench compared to the onset in the Andersen thermostated run happening 3 ns after the quench, which results in more data for the Berendsen run.

During the simulation we kept track of which particles were in the same clusters, allowing us to find a velocity distribution for the centers of mass of clusters. This was done for clusters up to size six. It turns out that the centers of mass also have a Maxwell-Boltzmann velocity distribution, which is demonstrated in Fig. 5 for three-particle clusters. The distributions are not as smooth as the distributions of single particle velocities as we now have a lot less data points to draw the distributions from, even for cluster size as small as three. For example, while the Berendsen thermostated run provides 13×10^6 data points for the single particle velocity distribution, there is only around 4×10^5 velocities available for the three-particle case from the same run. Nevertheless, the Maxwell-Boltzmann distribution is clearly observed, as was the case for all the cluster sizes for which we collected the center of mass velocity distributions. In this sense clusters behave like large molecules with mass Nm, where N is the number of particles in the cluster and m is the mass of an individual atom.

C. Nucleation rate

Table I collects the nucleation rates obtained with the different methods and thermostats along with the number of simulations where a nucleation event was observed. The nucleation rates are averages over all the runs. They are given in the reduced units where we have taken the Lennard-Jones parameters σ and ϵ as the units of length and energy,

TABLE I. Nucleation rates obtained by different methods and thermostats. Also shown is the number of simulations where a nucleation event was observed.

	Yasuoka-Matsumoto	Direct observation	Number of simulations
Andersen	2.3×10^{-7}	1.0×10^{-7}	99
Berendsen	1.3×10^{-7}	0.5×10^{-7}	97
Andersen to free	1.5×10^{-7}	0.4×10^{-7}	87
Berendsen to free	1.3×10^{-7}	0.4×10^{-7}	84
NVE	0.8×10^{-7}	···	2

respectively, and the single particle mass m (6.63×10^{-26} kg for argon) as the unit of mass. The unit of nucleation rate is then $[J]=\sigma^{-3}\tau^{-1}$, where $\tau=\sqrt{m\sigma^2/\epsilon}=2.15$ ps is the unit time. In addition to the thermostated runs the table shows nucleation rates for *NVE* simulations. There were unfortunately only few *NVE* simulations where nucleation was observed. This was likely due to the rapid temperature rise of the clusters after their formation, which, combined with our rather high target temperature, causes the clusters to evaporate again. Inspecting Fig. 3 again shows that in our two *NVE* runs where nucleation did occur, the cluster temperatures near the critical size have actually been lower than 85 K, providing more favorable conditions for nucleation.

The nucleation rates for the different thermostats are fairly close to each other with the rates for regular Andersen thermostated runs a bit higher than the others, and the rates for *NVE* simulations are lower (as should be, see Barrett *et al.*[14]). The difference in nucleation rates between runs, where only the free particles are thermostated, and the regular thermostating is very small, for the Berendsen thermostat it is practically nonexistent. The Yasuoka-Matsumoto method yields two to three times higher nucleation rates than the direct observation for all thermostats. This is still a fairly good agreement.

We will next examine the distribution of nucleation onset times for the different thermostats. Getting nucleation rates as high as the Yasuoka-Matsumoto ones with direct observation would require the nucleation onsets to happen on average within about 0.5 ns after the quench or even sooner with the Andersen thermostat. Figure 6 shows that the distribution of onset times is quite wide, and in all cases most of the nucleation events start clearly later than 0.5 ns after the quench. Immediately after the quench to the lower temperature we see less nucleation events than expected if the times of onset followed an exponential distribution. While the vapor after the quench is in a supersaturated state, some period of time is always needed for the vapor atoms to agglomerate together to form a critical cluster. Some of the earliest events are probably facilitated by a subcritical cluster already present before the quench. The distribution for the Andersen thermostated runs seem to follow an exponential distribution, but closer inspection reveals that the onset happens considerably less frequently for the first 200 ps after the quench than in the following 200 ps (see Fig. 7). It should be noted that the large number of Andersen thermostated runs in which the onset happens during the first nanosecond after the quench indeed makes it statistically reasonable to use the narrower bin width of 200 ps for this time frame.

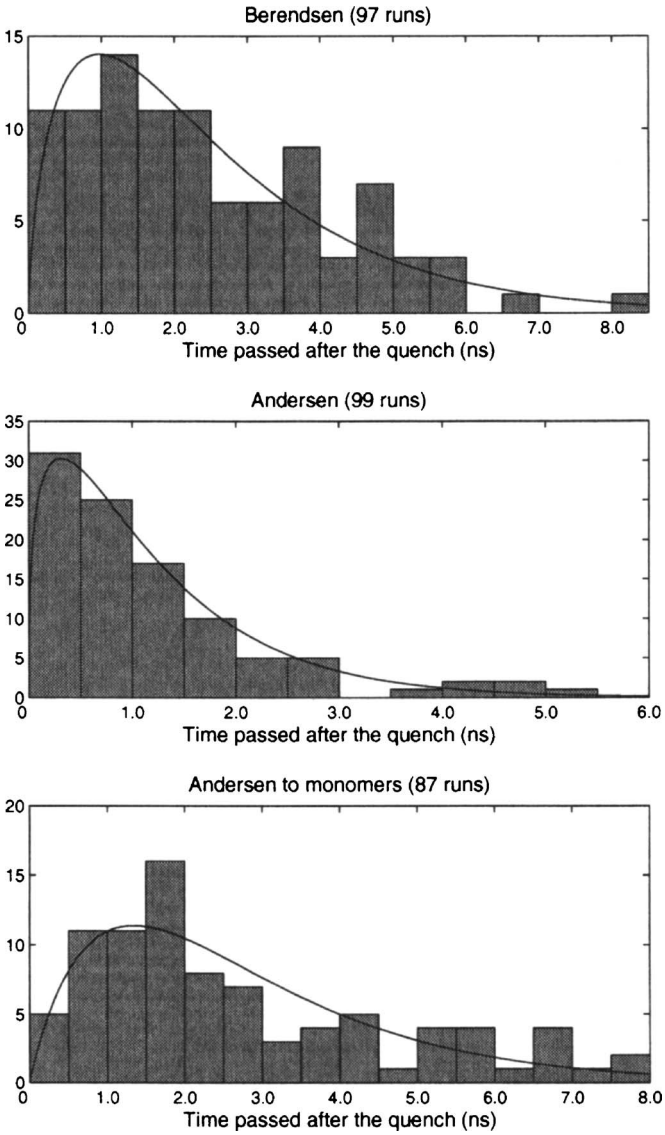

FIG. 6. Histograms for the times of onset with 500 ps bin width. From top to bottom: Berendsen, Andersen, and Andersen only applied to monomers. Total number of runs with observed nucleation events shown in the picture. The solid lines are the gamma distributions fitted to the data.

For the runs with other thermostats the histograms have peaks at 1–2 ps, and they have a tail of nucleation events spanning practically the whole range of our simulation time. Even though nucleation on average occurs soonest when using the Andersen thermostat, there still exists a clear tail in the histogram. An example of a general distribution that has this shape and can be derived from Poisson's assumptions is the gamma distribution,

$$f(x)=\frac{x^{\gamma-1}e^{-x/\beta}}{\beta^{\gamma}\Gamma(\gamma)}, \tag{3}$$

with the shape parameter γ in the range between 1 and 2. The gamma distributions fitted to the onset data are shown in Figs. 6 and 7.

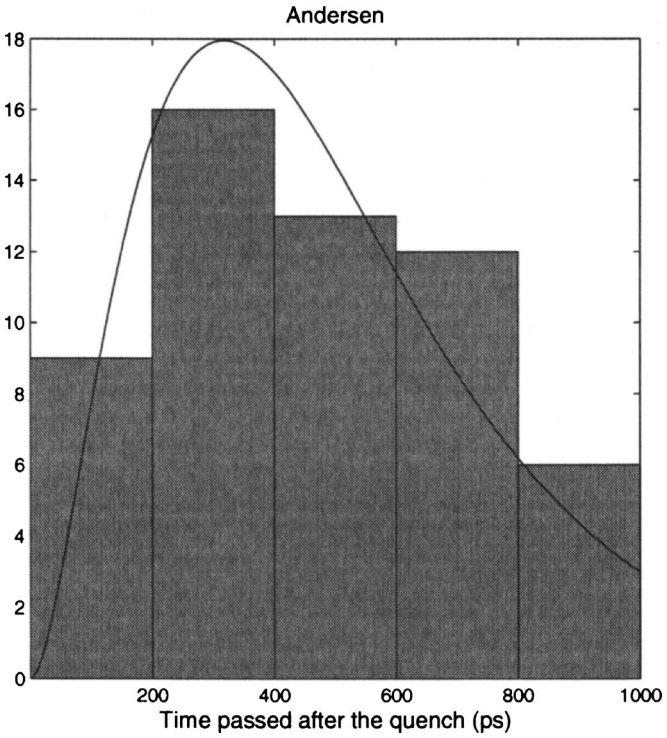

FIG. 7. Narrower histogram bins for the 1 ns of nucleation onsets for the Andersen thermostated runs, suggesting a gamma distribution shape. The solid line is again the fitted gamma distribution.

D. Cluster distributions and comparison with classical nucleation theory

In molecular dynamics the cluster distributions are directly available from the simulations. In a simulation like this, where the system is in a supersaturated state, we do not have equilibrium after the quench, but the system is in a metastable state prior to the nucleation onset. The vapor density remains practically unchanged when a constant nucleation rate is observed with the Yasuoka-Matsumoto graphs, so clusters smaller than the critical size can be considered to be in quasiequilibrium during this period. Cluster distributions shown in Fig. 8 are from the period after the onset where we have a constant nucleation rate. In the example run of Fig. 1 the period would be roughly from 2.5 up to 3 ns after the quench. These distributions are averaged over several runs with different times of nucleation onset. The time of onset did not affect the distributions, which is demonstrated for the Berendsen thermostat in the bottom part of Fig. 8. There appears to be some difference in the cluster distributions depending on the choice of thermostat. However, these differences clearly do not affect the nucleation rates for the different thermostats as was shown in Table I; additionally the formation free energy will not be significantly different for the different thermostats as we will see shortly.

Toxvaerd[5] found that the cluster size distributions follow a simple exponential form after a certain cluster size. Our critical size is about half of that of Toxvaerd, and as we have usually only one cluster exceeding the critical size, the statistics for the larger cluster sizes are rather poor. However, the cluster concentrations seem to depend linearly on cluster size at sizes above $n \approx 15$ in Fig. 8.

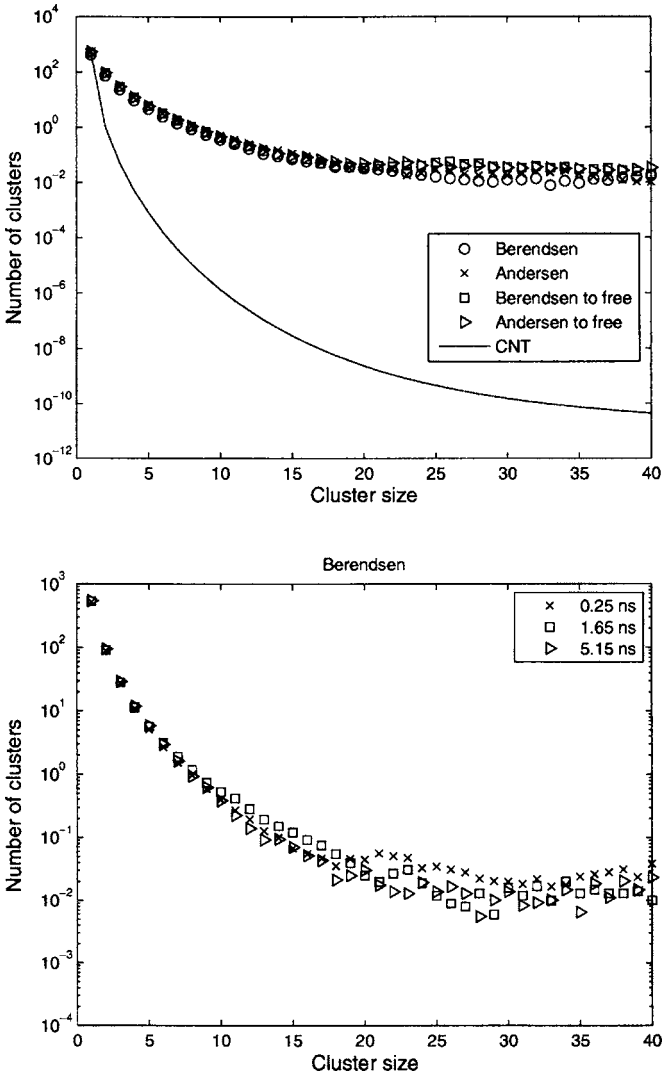

FIG. 8. Top: Averaged cluster distributions for the different thermostats taken from the period of constant nucleation rate after the onset. Bottom: Cluster distributions for Berendsen thermostated runs where the nucleation onset has occurred at different times, time passed between quench and the onset given in legend.

The equilibrium cluster distribution is expected to have an exponential form,

$$c_e(n) = \rho \exp\left(-\frac{\Delta G(n)}{k_B T}\right), \tag{4}$$

where ρ is the number density of monomers, k_B is the Boltzmann constant, T is the system temperature, and $\Delta G(n)$ is the formation free energy of an n cluster. In a vapor nucleating at a steady rate the cluster distribution is given by (see Ref. 15)

$$c(n) = c_e(n)\left(1 - J\sum_{i=1}^{n-1} \frac{1}{\beta(i)c_e(i)}\right), \tag{5}$$

where $\beta(n)$ is the rate at which monomers collide with an n cluster, and the nucleation rate J is given by

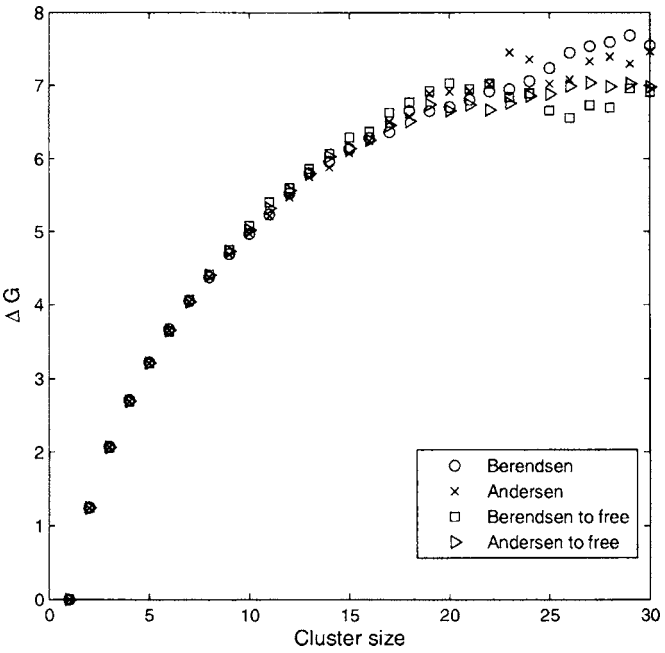

FIG. 9. Formation free energy approximated with Eq. (9) for the different thermostats given in the reduced Lennard-Jones units.

$$J = \left[\sum_{i=1}^{\infty} \left(\frac{1}{\beta(i) c_e(i)} \right) \right]^{-1}. \tag{6}$$

In the summation of Eq. (5) $1/c_e \propto \exp(\Delta G(n)/(k_B T))$ has a very sharp peak around the critical size n^* and approaches zero elsewhere, thus

$$c(n) \approx c_e(n) \quad \text{for } n < n^*. \tag{7}$$

If we let n go to infinity in the summation of Eq. (5), we see from Eq. (6) that the sum is equal to J^{-1}. From the near symmetry of the exponential in $1/c_e$ it is clear that adding the terms up to the critical size should yield about half of this. So we have

$$c(n^*) \approx c_e(n^*)/2 \tag{8}$$

for the critical cluster number density.

Solving Eq. (4) for the cluster formation free energy one obtains

$$\Delta G(n) = -k_B T \ln \frac{c_e(n)}{\rho}. \tag{9}$$

Even though the simulated cluster distributions of Fig. 8 are obtained from a vapor nucleating at a steady rate rather than an equilibrium situation, according to Eq. (7) a reasonable approximation for the formation free energy of cluster sizes clearly smaller than the critical size is given by Eq. (9). As the steady state cluster distribution starts to deviate from the equilibrium distribution at cluster sizes close to the critical size, we should take this into account when trying to plot the formation free energy. Doing this, however, would require us to make further approximations so we settle on plotting the formation free energy with the help of Eq. (9) and then correct for the critical value. Figure 9 shows the formation free energy obtained from our simulations for the different thermostats. The formation free energy levels off at the critical size, which in our system is about 25 for all thermostats. From Fig. 9 we see that the critical formation free energy would be ≈ 7 in the reduced units ($[\Delta G] = \epsilon$) should the equilibrium expression be valid at that size. But as we have $c(n^*) = c_e(n^*)/2$, we must correct this estimate by $-k_B T \ln 2$ finally arriving at $\Delta G^* \approx 6.5$ for a simulation-based estimate for the critical formation free energy of our system.

Assuming that the cluster has spherical shape, and the vapor is an ideal gas, the formation free energy is expressed in classical nucleation theory as

$$\Delta G(n) = \left(\frac{36\pi}{\rho_l^2} \right)^{1/3} \gamma n^{2/3} - n k_B T \ln\left(\frac{p}{p_e} \right). \tag{10}$$

Here γ is the surface tension of the planar liquid-vapor interface, ρ_l is the number density of the bulk liquid at liquid-vapor equilibrium, p is the pressure of supersaturated vapor, and p_e is the equilibrium pressure of the vapor. The maximum of the formation free energy ΔG^* corresponds to the critical cluster size n^*, and from Eq. (10) one gets

$$n^* = \frac{32\pi}{3} \frac{\gamma^3}{(k_B T \ln S)^3 \rho_l^2} \tag{11}$$

for the critical size and

$$\Delta G^* = \frac{16\pi}{3} \frac{\gamma^3}{(k_B T \rho_l \ln S)^2} \tag{12}$$

for the critical formation free energy, where $S = p/p_e$ is the saturation ratio.

From Eqs. (11) and (12) we get CNT estimates for the critical values, which turn out to be somewhat larger than the simulation results. Using values for ρ_l and γ from the results of Mecke *et al.*[16] and for the equilibrium vapor pressure p_e the result from Laasonen *et al.*[2] which gives us a saturation ratio $S = 4.7$, the CNT prediction for critical cluster size is $n^* = 38$ and critical formation free energy $\Delta G^* = 20.9$, again in the reduced units.

Finally, obtaining a value for the critical formation free energy allows for yet another way to determine the nucleation rate, as an expression for nucleation rate can be written as

$$J = \rho_l^{-1} \sqrt{\frac{2\gamma}{\pi m}} \left(\frac{p_e}{k_B T} \right) \rho \exp\left(-\frac{\Delta G^*}{k_B T} \right). \tag{13}$$

Plugging in the simulation-based value $\Delta G^* = 6.5$ results in a nucleation rate $J = 3.4 \times 10^{-9}$, about two orders of magnitude smaller than those in Table I. Using instead the CNT value for the formation free energy one gets a predicted nucleation rate $J_{\mathrm{CNT}} = 5.3 \times 10^{-18}$, which is over ten orders of magnitude lower than the simulated values. The two orders of magnitude difference between the nucleation rate obtained from Eq. (13) with the critical formation free energy from simulations and the nucleation rate obtained directly from the simulations is partly due to the uncertainties in the values of the surface tension, number density of the bulk liquid, and equilibrium vapor pressure as well as our estimate for the value

of the critical formation free energy, and partly due to the fact that the kinetic prefactor in Eq. (13) uses the bulk values for liquid density and surface tension, yet the critical cluster consists only of around 25 particles.

IV. CONCLUSIONS

We have performed MD simulation of 1000 simple Lennard-Jones atoms to assess the effects of simulation methods on nucleation. We have compared the Berendsen and the Andersen thermostats in regulation of vapor and cluster temperatures. Nucleation rate was derived from the simulations by direct observation of nucleation onset and by the method of Yasuoka and Matsumoto.[4]

Our simulations show that the Andersen thermostat regulates the temperature of clusters better than the Berendsen thermostat. This fact is reflected both in cluster distributions and in the nucleation rates. However, the differences are small compared to orders of magnitude difference to CNT results. The velocities of atoms and the center of mass velocities of small clusters conform to Maxwell distribution of velocities with all thermostats, which implies that a close approximation of true canonical ensemble is generated. If strict isothermality is required, the Andersen thermostat is preferred. However, it seems that thermostating method (Berendsen or Andersen) is not a critical issue in obtaining a reasonable approximation of nucleation rate. It would seem unnecessary to further refine which particles the thermostats are applied to, that is, using the thermostats on all particles in smaller clusters and on the surfaces of larger clusters in addition to the free particles.

There exist other thermostats which could possibly be used in nucleation simulations. In particular, the Nosé-Hoover thermostat[17] has been widely used and it is commonly available in MD software packages. Nevertheless, problems in thermal regulation of small clusters with Nosé-Hoover thermostat has been reported,[18,19] which is the reason why we have not considered it in this study.

In a practical sense, perhaps the most important difference between the two methods to obtain the nucleation rate is the fact that a relatively high number of simulation runs is needed for the direct observation method, while a single run could suffice for the method of Yasuoka and Matsumoto. This might rule out the option of using direct observation should the system size be larger, and consequently the simulation time longer. Determining the nucleation rate with the method of Yasuoka and Matsumoto results in two to three times higher nucleation rates than using the direct observation method. Acquiring nucleation rates similar to the Yasuoka-Matsumoto ones with direct observation would require that for the majority of the runs the nucleation onset would occur shortly after the quench, which seems to be an unlikely scenario according to the shape of the onset distributions acquired from our simulations. Observing the onset directly after the quench is not to be expected as some period of time is needed so that the vapor atoms can agglomerate together to form a critical cluster. Nevertheless, the agreement between the two methods is still fairly good.

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. (Academic, New York, 2002).
[2] K. Laasonen, S. Wonczak, R. Strey, and A. Laaksonen, J. Chem. Phys. **113**, 9741 (2000).
[3] J. Wedekind, D. Reguera, and R. Strey, J. Chem. Phys. **125**, 214505 (2006).
[4] K. Yasuoka and M. Matsumoto, J. Chem. Phys. **109**, 8451 (1998).
[5] S. Toxvaerd, J. Chem. Phys. **115**, 8913 (2001).
[6] K. Yasuoka and M. Matsumoto, J. Chem. Phys. **109**, 8463 (1998).
[7] S. Toxvaerd, J. Chem. Phys. **119**, 10764 (2003).
[8] P. Krasnochtchekov and R. S. Averback, J. Chem. Phys. **122**, 044319 (2005).
[9] K. Tanaka, K. Kawamura, H. Tanaka, and K. Nakazawa, J. Chem. Phys. **122**, 184514 (2005).
[10] N. Lümmen and T. Kraska, J. Aerosol Sci. **36**, 1409 (2005).
[11] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[12] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
[13] F. H. Stillinger, J. Chem. Phys. **38**, 1486 (1963).
[14] J. Barrett, C. Clement, and I. Ford, J. Phys. A **26**, 529 (1993).
[15] H. Vehkamäki, *Classical Nucleation Theory in Multicomponent Systems* (Springer, Berlin, 2006).
[16] M. Mecke and W. Winkelmann, J. Chem. Phys. **107**, 9264 (1997).
[17] M. Tuckerman, Y. Liu, G. Ciccotti, and G. Martyna, J. Chem. Phys. **115**, 1678 (2001).
[18] S. A. Harris and I. J. Ford, J. Chem. Phys. **118**, 9216 (2003).
[19] E. Kelly, M. Seth, and T. Ziegler, J. Phys. Chem. A **108**, 2167 (2004).