Free energies of molecular clusters determined by guided mechanical disassembly

Hoi Yu Tang and Ian J. Ford

Department of Physics and Astronomy and London Centre for Nanotechnology,
University College London, Gower Street, London WC1E 6BT, United Kingdom

The excess free energy of a molecular cluster is a key quantity in models of the nucleation of droplets from a metastable vapour phase; it is often viewed as the free energy arising from the presence of an interface between the two phases. We show how this quantity can be extracted from simulations of the mechanical disassembly of a cluster using guide particles in molecular dynamics. We disassemble clusters ranging in size from 5 to 27 argon-like Lennard-Jones atoms, thermalised at 60 K, and obtain excess free energies, by means of the Jarzynski equality, that are consistent with previous studies. We only simulate the cluster of interest, in contrast to approaches that require a series of comparisons to be made between clusters differing in size by one molecule. We discuss the advantages and disadvantages of the scheme and how it might be applied to more complex systems.

I. INTRODUCTION

The formation of droplets from a metastable vapour phase is a commonplace event in nature, but so far it has resisted quantitative analysis, despite repeated attention [1–4]. The phenomenon plays a role in atmospheric aerosol and cloud formation [5, 6], as well as in industrial processes [7, 8]. Theoretical analysis often begins with the Becker-Döring equations [9] that describe changes in the populations n_i of clusters of i molecules brought about by the processes of gain and loss of single molecules, or monomers. They take the form

$$dn_i/dt = \beta_{i-1}n_{i-1} + \alpha_{i+1}n_{i+1} - (\beta_i + \alpha_i)\,n_i, \quad (1)$$

where β_i and α_i are growth and evaporation rates, respectively. The rate of nucleation J of droplets from a metastable vapour phase may then be expressed as [10]

$$J = n_1\beta_{i^*}Z\exp\left[-\left(\phi(i^*) - \phi(1)\right)/kT\right], \quad (2)$$

where k is the Boltzmann constant, T is the temperature, i^* is the size of the critical cluster, defined to have equal probabilities, per unit time, of molecular gain or loss, and Z is the Zeldovich factor that accounts for the nonequilibrium nature of the kinetics [11]. We shall refer to $\phi(i)$ as the thermodynamic work of formation of a cluster of i particles (or i-cluster) starting from the metastable vapour phase. A range of nomenclature is used for this quantity in the nucleation theory literature: the work of formation was denoted by $\epsilon(i)$ in [10], and elsewhere the same, or a very similar quantity has been labelled as ΔF, ΔG or ΔW, for example.

We note that ϕ in the nucleation rate expression has both a kinetic and a thermodynamic interpretation [12]. The quantity $\phi(i) - \phi(1)$ can be expressed in terms of ratios of cluster growth and evaporation rates:

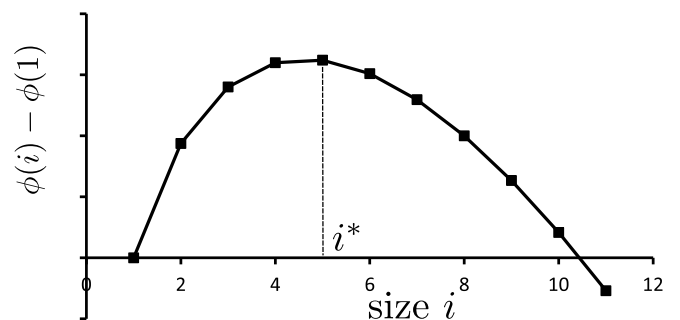$$\phi(i) - \phi(1) = -kT\sum_{j=2}^{i}\ln\frac{\beta_{j-1}}{\alpha_j}, \quad (3)$$

Figure 1. Typical work of formation of a cluster of i particles with a maximum at the critical cluster size i^*.

but ϕ is also related to the grand potential $\Omega_s(i) = F(i) - i\mu_s$ of an i-cluster at the chemical potential μ_s of the saturated vapour [10]:

$$\phi(i) = \Omega_s(i) - ikT\ln S, \quad (4)$$

where $F(i)$ is the Helmholtz free energy of the cluster, $S = p_v/p_{vs}$ is the vapour supersaturation, and p_v and p_{vs} are the vapour pressure and saturated vapour pressure, respectively. The role of the grand potential in this context is to specify the equilibrium population of clusters of size i in a saturated vapour, namely $n_i^s = \exp(-\Omega_s(i)/kT)$. The nucleation model is completed by representing the population of monomers as $n_1 = Sp_{vs}V/kT$, where V is the system volume, by assuming that the vapour pressure is dominated by the ideal partial pressure of single molecules.

In classical nucleation theory (CNT), clusters are viewed as scaled down versions of macroscopic droplets. According to this approach, the difference $\phi(i) - \phi(1)$ is replaced by $\phi(i)$ alone with

$$\phi(i) \approx \phi_{\mathrm{cl}}(i) = \gamma A(i) - ikT\ln S, \quad (5)$$

where γ is the surface tension of a planar interface between vapour and condensate, and $A(i)$ is the surface area of a cluster represented as a sphere with a density

equal to that of the bulk condensed phase. The work of formation is a combination of a free energy cost of forming the interface, and a free energy return proportional to the number of molecules in the cluster (or proportional to its volume since the condensed phase density is taken to be a constant). The neglected $\phi(1)$ term might be represented by $\gamma A(1) - kT \ln S$, which leads to the internally consistent classical theory [13].

The cluster size dependence of the CNT work of formation is illustrated in Figure 1. It represents a thermodynamic barrier, with a maximum at the critical size, that limits the natural tendency for small molecular clusters to grow into large droplets when exposed to a supersaturated vapour. CNT has been modified in several ways, for example by introducing a size-dependent surface tension [14] or by introducing compatibility with nonideal vapour properties [15, 16].

More fundamentally, the ratio of kinetic coefficients β_{j-1}/α_j might be evaluated using an underlying microscopic model for all clusters up to the critical size and beyond [12], and the work of formation determined through Eq. (3). It may be shown that

$$\beta_{j-1}/\alpha_j = S \exp\left[-\left[\Omega_s(j) - \Omega_s(j-1)\right]/kT\right], \quad (6)$$

which shifts attention to the free energy difference $F(j) - F(j-1)$ associated with the addition of a molecule to a $(j-1)$-cluster. Computing these differences is the basis of an approach has been used extensively in calculations of cluster free energies [17–23]. But nucleation is actually controlled by the properties of clusters near the critical size, and one drawback of computing the differences $F(j) - F(j-1)$ is that the predicted nucleation rate could be susceptible to the accumulation of errors in evaluating such a sequence.

In this paper, we describe a computational method for directly obtaining the cluster free energy without the need to perform calculations for a sequence of smaller clusters. We consider the following representation of the work of formation of a cluster minus that of a monomer:

$$\phi(i) - \phi(1) = F_s(i) - (i-1)kT \ln S. \quad (7)$$

We shall refer to $F_s(i)$ as the cluster excess free energy, though more accurately it is a *difference* between the excess free energies of an i-cluster and a monomer [10]. It is 'excess' in that it represents the free energy required to carve a cluster out of a bulk condensed phase, or equivalently to assemble it out of saturated vapour. It may be associated with the thermodynamic cost of creating an interface, which is why in CNT it is modelled by a surface term, and why we have given it a suffix s.

Our approach centres on *disassembling* a cluster into its component molecules using guided molecular dynamics in order to calculate the cluster excess free energy directly. The method employs the Jarzynski equality [24–26] and we provide details in Section II, including a comparison with the related method of thermodynamic integration. Tests of the method where we separate a
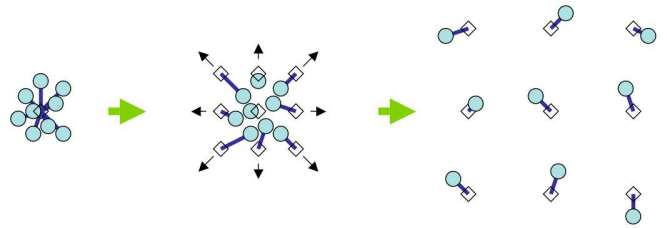

Figure 2. Guided disassembly process for an i-cluster. The real particles (circles) are initially weakly tethered to the guide particles (diamonds). The latter drift apart and the tethers gradually tighten leading to i independent, tethered particles upon completion of the process.

dimer according to a variety of protocols are described in Appendix A. The disassembly of argon-like Lennard-Jones clusters is presented in Section III and we compare our results with those obtained from Monte Carlo studies by Barrett and Knight [27] and Merikanto *et al.* [28, 29]. These studies gave consistent excess free energies, though they were not in agreement with experiments by Iland *et al.* [30]. We conclude with a discussion of the advantages and disadvantages of the approach compared with other treatments in Section IV.

II. GUIDED MOLECULAR DYNAMICS SIMULATIONS

A. Fundamentals of the method

We study the dynamical evolution of a cluster against a background of external manipulation. The cluster particles are harmonically tethered to a set of artificial 'guide particles', which lie initially at the origin but after a period of system equilibration are programmed to move apart, driving cluster disassembly. The strength of the tether forces is initially quite weak, in order to disturb the properties of the cluster as little as possible. Later, the tethers can be strengthened in order to guide the separation process more firmly, and to prevent the atoms from interacting with each other once the final guide particle positions have been reached. The mechanical work of the disassembly can then be related to the change in Helmholtz free energy.

The masses of the guide particles are taken to be very much greater than those of the cluster particles. This essentially fixes the trajectories of the guide particles in the molecular dynamics, in accordance with the velocities assigned to each at the beginning of the disassembly process. By choosing guide particle velocities, simulation times and a time-dependent tethering force, a range of cluster disassembly protocols can be explored. A simple illustration of the process is shown in Figure 2.

We shall consider clusters of argon-like atoms interacting through Lennard-Jones potentials, and so we shall refer to the cluster particles as atoms. We equilibrate this

system under the influence of the tethers for a suitable period, the duration of which will depend upon the cluster size and the desired temperature. A further molecular dynamics simulation is performed and from this trajectory we select initial configurations for cluster disassembly. In order that the configurations should represent a bound structure, we employ a Stillinger cluster condition [31] in the selection, allowing a separation of no more than $1.5\,\sigma_{\mathrm{ArAr}}$ between an atom and its nearest neighbour, where σ_{ArAr} is the usual Lennard-Jones range parameter. Such a Stillinger condition has been used in previous Monte Carlo approaches. The cluster definition is an important ingredient of a modelling strategy [2], and deserves careful consideration, but here we shall use this simple criterion for convenience.

The simulations were performed using the DL_POLY [32] molecular dynamics package, with modifications to the source code to implement the time-dependent harmonic tether potentials. We include a physical heat bath of helium-like Lennard-Jones atoms thermalised using a Nosé-Hoover thermostat [33]. We could instead have implemented a thermostat that acts on the cluster itself, but chose not to in order to achieve as natural a thermalisation as possible during the nonequilibrium processing.

B. Work performed on a system

Given an external control parameter λ in a Hamiltonian $H(\lambda)$, the work W done on a system due to the evolution of λ over a finite time period may be written

$$W = \int \frac{d\lambda}{dt}\frac{\partial H(\lambda)}{\partial \lambda}dt. \tag{8}$$

For example, consider the Hamiltonian H_1 of a single guided atom of mass m:

$$H_1 = \frac{\boldsymbol{p}^2}{2m} + \frac{1}{2}\kappa(t)\left[\boldsymbol{x}(t) - \mathbf{X}(t)\right]^2, \tag{9}$$

where $\boldsymbol{p}$ is the momentum, $\kappa(t)$ is the time-dependent tethering force or spring constant, $\boldsymbol{x}(t)$ is the atomic position and $\mathbf{X}(t)$ is the guide position. For a set of guided atoms, each controlled by a Hamiltonian H containing terms of the form given in Eq. (9) supplemented by interparticle interactions, $\kappa(t)$ and $\mathbf{X}(t)$ play the role of λ and the work W performed on the set is

$$\int_0^\tau \frac{d\kappa(t)}{dt}\frac{\partial H(\kappa, \{\mathbf{X}_k\})}{\partial \kappa}dt + \sum_j \int_0^\tau \frac{d\mathbf{X}_j(t)}{dt}\frac{\partial H(\kappa, \{\mathbf{X}_k\})}{\partial \mathbf{X}_j}dt$$
$$= \frac{1}{2}\int_0^\tau \frac{d\kappa(t)}{dt}\sum_{j=1}^i \left[\boldsymbol{x}_j(t) - \mathbf{X}_j(t)\right]^2 dt$$
$$- \int_0^\tau \kappa(t)\sum_{j=1}^i \left[\boldsymbol{x}_j(t) - \mathbf{X}_j(t)\right]\cdot\mathbf{V}_j(t)dt, \tag{10}$$

where τ is the length of the molecular dynamics simulation, and $\mathbf{V}_j(t)$ is the velocity of the guide particle associated with the jth atom, defined as $\mathbf{V}_j(t) = d\mathbf{X}_j(t)/dt$. The first term in Eq. (10) arises from the time dependence of the spring constant, and the second term is simply the conventional force times distance expression. It should be noted that all tethers within the system are characterised by the same spring constant, although more elaborate protocols could be imagined.

C. The Jarzynski equality

If we were able to perform an extremely slow, quasistatic process, then the mechanical work done would be equal to the difference in Helmholtz free energy between the initial and final equilibrium states. However, quasistatic processes are unfeasible in finite time molecular dynamics simulations and according to the second law [34], the average of the work done (as a result of a time-dependent change in the Hamiltonian of the system), performed over many realisations of a nonquasistatic process (indicated by angled brackets), will always be an overestimate of the free energy change, $\langle W \rangle > \Delta F$, allowing us only to infer an upper limit to ΔF.

However, the Jarzynski equality [24, 25]

$$\langle \exp\left(-W/kT\right)\rangle = \exp\left(-\Delta F/kT\right) \tag{11}$$

allows us to do better. For this identity to hold, the system must begin in thermal equilibrium, but need not remain so as the Hamiltonian changes during the simulation. Exploiting the work done in a nonequilibrium process is a powerful strategy for calculating cluster surface free energies and numerous computational studies [35–41] as well as experiments [42–47] have achieved this with the help of the Jarzynski equality. Systems studied include argon-like Lennard-Jones fluids, ion-charging in water, ideal gases confined to a piston, and one-dimensional polymer chains. Nevertheless, there are distinct aspects of this strategy for analysing the controlled disassembly of a cluster that need to be explored.

The Jarzynski equality ought to recover the free energy difference regardless of the nature of the evolution between initial and final Hamiltonians, but computed results might still depend upon the rate of the process as a consequence of a limited sampling of system trajectories in finite simulations [42]. We might expect 'slow' processes that gently pull a cluster apart to generate a narrower distribution of work compared with 'fast' processes that are violent and highly dissipative. A balance must therefore be struck between the poorer convergence of fast simulations and the demand for computational resources required for slow simulations.

Furthermore, a consequence of the exponential averaging in the Jarzynski equality is that occasional values of work that are well below the average, arising from unusual trajectories, can sometimes distort the extracted free energy change. This is a consequence of insufficient

sampling of the system trajectories and so we need to give careful attention to the statistical errors.

We have explored the outcomes of various guiding protocols, and the robustness of the Jarzynski equality in the face of limited statistics, in a test case of the separation of a dimer, for which the free energy change is easily calculable. These studies are described in Appendix A. We have used similar protocols to study the disassembly of larger clusters, which is described in Section III.

D. Comparison with thermodynamic integration

The method bears some similarity to thermodynamic integration, where the strength of the interparticle interactions is evolved over a sequence of equilibrium calculations in order to compare the system in question with another that has a known free energy [48–51]. The basic relationship $\Delta F = \int \langle \partial H(\lambda)/\partial \lambda \rangle d\lambda$ is analogous to Eq. (8). The reference system for clusters might, for example, be a set of noninteracting particles held together through the retention of the constraining cluster definition. Or indeed the cluster definition could be changed progressively along with the interactions in order to reach a more convenient final state, perhaps noninteracting particles inside a sphere.

However, there are some important differences. In our approach it is the tether potentials that change with time, not the interparticle interactions, and our reference system is a set of independent harmonic oscillators, not an ideal gas. Furthermore, we conduct the evolution by nonequilibrium molecular dynamics rather than by moving through a sequence of equilibrium ensembles, and we only need to impose a cluster definition when selecting the initial configurations, not throughout the evolution. An abrupt removal of the cluster definition constraint is acceptable in a nonequilibrium evolution, when the results are processed using the Jarzynski equation, but it would not be appropriate during a sequence of equilibrium calculations.

III. ARGON CLUSTER DISASSEMBLY

A. Preliminaries

We have investigated the disassembly of clusters consisting of 5, 10, 15, 20 and 27 argon-like atoms in order to obtain their excess free energies. Scaling up the guided molecular dynamics simulations from the test case of dimer separation is fairly straightforward. We perform simulations in a cubic cell with edge lengths of 100 Å, so that the initial clusters and the final disassembled configurations may be easily accommodated. We employ Lennard-Jones interaction potentials for each species (see Table III) and the helium temperature is set at 60 K in order to facilitate a comparison with the Monte Carlo studies by Barrett and Knight [27] and Merikanto *et al.*
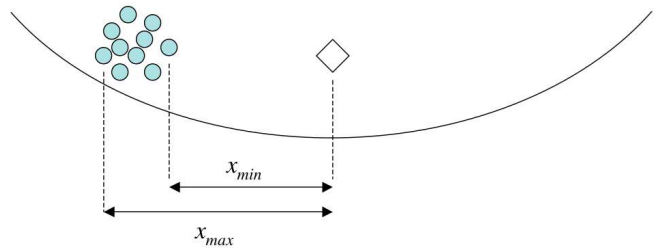

Figure 3. The difference in tether energy across a cluster configuration is given in terms of the maximum and minimum separations between an atom and its guide particle. The circles depict the argon atoms, while the diamond represents the position of all of the guides at the origin of the cell.

[28, 29], as well as the experimental studies of Iland *et al.* [30].

However, converting the free energy change associated with disassembly into an excess free energy requires some careful consideration of the statistical mechanics of tethered and free molecular clusters. We require the excess free energy of a cluster that is free to move anywhere inside a system volume, but our initial state is a cluster tethered to guide particles at the origin. The free energy change that emerges from our calculations will correspond to the disassembly of a cluster whose centre of mass explores a region around the origin, and furthermore, one that possesses energy due to the tethers in addition to that of the physical interactions between the atoms. These matters are discussed in detail in Appendix B.

The energetic perturbation of the cluster by the tethers can be reduced by choosing a small force constant. We take the view that the mean variation in tethering energy of an atom, as it explores different regions of the cluster during the equilibrated trajectory, should not exceed the thermal energy kT, or

$$\frac{1}{2}\kappa_i \langle x_{\max}^2 - x_{\min}^2 \rangle < kT, \qquad (12)$$

where $x_{\max}$ and $x_{\min}$ are, respectively, the maximum and minimum separations between an atom and its guide particle in a configuration (see Figure 3). This criterion may also be expressed as $\xi = \kappa_i \langle x_{\max}^2 - x_{\min}^2 \rangle/(2kT) < 1$.

From the equilibrated molecular dynamics trajectory, we select, for disassembly, a set of 'valid' cluster configurations that satisfy the Stillinger cluster definition [31], but this can be quite difficult for the smaller clusters at 60 K. Tethering the atoms keeps them closer together and more likely to form valid configurations. We therefore choose a tethering strength that satisfies the condition on ξ, but also helps to produce sufficient valid cluster configurations. The initial value of the tethering force constant was taken to be $\kappa_i = 0.01 \, \mathrm{kJ \, mol^{-1} Å^{-2}}$, which gives $\xi \sim 0.6 - 0.9$ for the five sizes of argon cluster studied. Table I shows the duration of the equilibrated cluster trajectory, the number of valid cluster configurations identified from candidates selected at intervals of

i	Duration/ns	Valid configurations	ξ
5	1000	152	0.604
10	250	411	0.745
15	225	1070	0.799
20	150	905	0.847
27	150	1020	0.922

Table I. The duration of equilibrated cluster trajectories at 60 K, as well as the number of valid cluster configurations selected at each size. The ratio ξ characterises the perturbation to the cluster energy due to the tether potentials.

100 ps from the equilibrated molecular dynamics trajectory, and the ratio ξ characterising the suitability of the tethering force constant.

Having obtained initial cluster configurations for the five sizes of cluster, the next stage is to disassemble them by a combination of guide particle motion and tether tightening. A range of separation times t_{sep} is explored, with the larger and more stable clusters expected to require longer disassembly processes in order to provide accurate estimates of the free energy change. As in the dimer calculations described in Appendix A, we use a tethering strength that strengthens in time according to Eqs. (A7), with a final value of $\kappa_f = 0.05\,\text{kJ}\,\text{mol}^{-1}\text{Å}^{-2}$.

The terminal positions for the guide particles are chosen from a $3 \times 3 \times 3$ grid with spacing of $33.33\,\text{Å}$. The largest cluster considered contains 27 argon atoms so after the process of disassembly, the tethered atoms move around each point on this grid. For smaller systems, the same grid of final guide positions is adopted, but employing only as many points as are necessary for the cluster in question. With initial guide positions at the origin and final positions defined in this way, it is straightforward to calculate the necessary drift velocities of the guide particles for a given separation time. Applying the Jarzynski procedure to the distribution of performed work then gives us the estimated free energy change ΔF associated with the disassembly of a cluster.

However, as mentioned previously, this free energy difference will only correspond to the disassembly of a tethered i-cluster, rather than of a freely translating, undistorted cluster. Furthermore, by necessity we obtain free energies of systems of *distinguishable* atoms in molecular dynamics, and we need to make an indistinguishability correction. An analysis of the thermodynamics is required in order to extract the excess free energy of an i-cluster from the free energy of disassembly, and the details are given in Appendix B. It turns out that we can write $F_s(i) = \sum_{k=1}^{5} f_s^k(i)$ with

$$f_s^1(i) = -\Delta F \tag{13}$$

$$f_s^2(i) = -ikT \ln\left(\rho_{vs} v_{\text{HO}}\right) \tag{14}$$

$$f_s^3(i) = kT \ln\left(\rho_{vs} v_c\right) \tag{15}$$

$$f_s^4(i) = -\frac{3i\kappa_i}{10}\left(\frac{3iv_l}{4\pi}\right)^{2/3} \tag{16}$$

$$f_s^5(i) = kT \ln i! \tag{17}$$

In the first term the free energy of disassembly ΔF appears with a negative sign because it refers to the process of taking a cluster apart while F_s is the free energy of interface formation. The f_s^2 term arises from relating the final state in the disassembly process, namely the separated harmonically bound particles, to the appropriate reference state of a saturated vapour. It represents the difference in free energy between the tethered particles, each effectively confined to a volume $v_{\text{HO}} = \left(2\pi kT/\kappa_f\right)^{3/2}$, and particles in the saturated vapour phase with density ρ_{vs} and volume per particle $1/\rho_{vs}$. The f_s^3 term is the entropy penalty associated with the initial tethering: the centre of mass of the cluster is effectively confined to a volume $v_c = \left(i\kappa_i/(2\pi kT)\right)^{-3/2}$ and needs to be referred to a situation where it is allowed, like a particle in saturated vapour, to explore a volume $1/\rho_{vs}$. The f_s^4 term is an approximate expression for the perturbation in the cluster energy due to the initial presence of the tethers, where $v_l = 1/\rho_l$ is the volume per particle in the condensed phase. Finally, f_s^5 converts calculations derived from molecular dynamics with distinguishable particles into results relevant to a system of indistinguishable particles.

B. Results and discussion

A typical example of the work $W(t)$ performed over a disassembly trajectory of duration 20 ns for a 27-atom cluster is shown in Figure 4. The gradual rise in the work performed prior to about 5 ns represents an accumulation of tethering energy as the guide particles move away from their initial positions at the origin. After this time, atoms begin to leave the cluster, and less work is needed to move the corresponding guides. After about 7 ns, the work rate reduces significantly as the cluster disintegrates and the guide particles move towards their final positions.

Visual representations of the disassembly process (see Figure 5) provide further insight into the manner in which the clusters are pulled apart. The onset of cluster disassembly is signalled by the loss of one or two atoms from the cluster, perhaps only temporarily. The cluster soon after breaks into several smaller clusters, which eventually disintegrate into fragments or single atoms. It is rare to see a complete and sudden disintegration of a cluster, where all the constituent atoms disassemble together within a short space of time.

Figures 6 and 7 show distributions of the work performed in disassembling the 5-cluster and the 27-cluster, along with estimates of the free energy change, for separation times between 0.5 ns and 20 ns. As expected, the work distributions are broader for the processes that are most rapid (smallest t_{sep}^{-1}) and hence least quasistatic in nature. Conversely, the work distributions become narrower, and lead to free energy changes that presumably provide the most accurate estimates of the true free energy change, as the rate of separation is reduced.

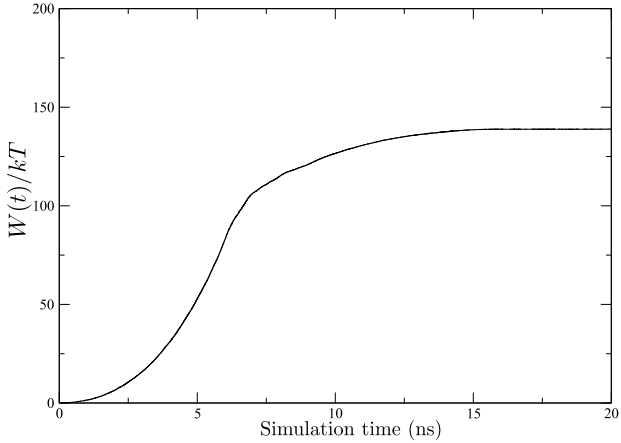The free energy change ΔF for the disassembly of each

Figure 4. A typical history of the work performed for one realisation of the disassembly of a 27-atom argon cluster with a separation time of 20 ns.

i	t_{sep}(ns)	$\langle W \rangle$	ΔF	$f_s^2(i)$	$f_s^3(i)$	$f_s^4(i)$	$f_s^5(i)$	$F_s(i)$
5	6	13.08	12.35	38.94	-7.79	-0.41	4.79	23.18
10	8	34.06	30.80	77.88	-8.83	-1.32	15.10	52.03
15	12	62.75	53.87	116.81	-9.44	-2.59	27.90	78.82
20	16	97.37	84.07	155.75	-9.87	-4.18	42.33	99.97
27	20	154.41	133.28	210.27	-10.32	-6.90	64.56	124.33

Table II. Results from the slowest set of disassembly simulations for each cluster size: the mean work $\langle W \rangle$, the free energy of disassembly ΔF and the other contributions to the excess free energy $F_s(i)$, all in units of kT.

size of cluster at the slowest rate studied is shown in Table II, along with the other contributions to the excess free energy F_s. We refer to a molecular dynamics study by Baidakov $et\,al.$ [53] to provide values of the saturated vapour density ρ_{vs} and liquid density $\rho_l = 1/v_l$ of the argon-like Lennard-Jones fluid at a temperature of 60.31 K.

Figure 8 shows our excess free energies $F_s(i)$ as a function of cluster size i. Statistical errors propagated from uncertainties in the free energy change ΔF are similar to the size of the symbols. We also include corresponding results from the Monte Carlo studies by Barrett and Knight [27] and Merikanto $et\,al.$ [28, 29]. Barrett and Knight employed a Lee-Barker-Abraham cluster definition [17] while Merikanto $et\,al.$ adopted a Stillinger cluster criterion similar to ours. The Barrett and Knight calculations are represented here by $F_s^{\text{BK}}(i)/kT = -\ln q_i - (i-1)\ln(\rho_{vs}\sigma_{\text{ArAr}}^3)$ with their fitting function $\ln q_i = 10.5 + 9.91(i-1) - 16.36(i^{2/3}-1)$, and the Merikanto $et\,al.$ values are derived from their Figure 1 in [28], which we interpret as a plot of $F_s^{\text{M}}(i)/kT - (i-1)\ln S$ with $S = 20$. The results of these earlier studies are consistent with one another, as well as with the excess free energy suggested by the internally consistent classical theory (ICCT) $F_s^{\text{ICCT}}(i) = \gamma \left(36\pi v_l^2\right)^{1/3}\left(i^{2/3}-1\right)$, con-
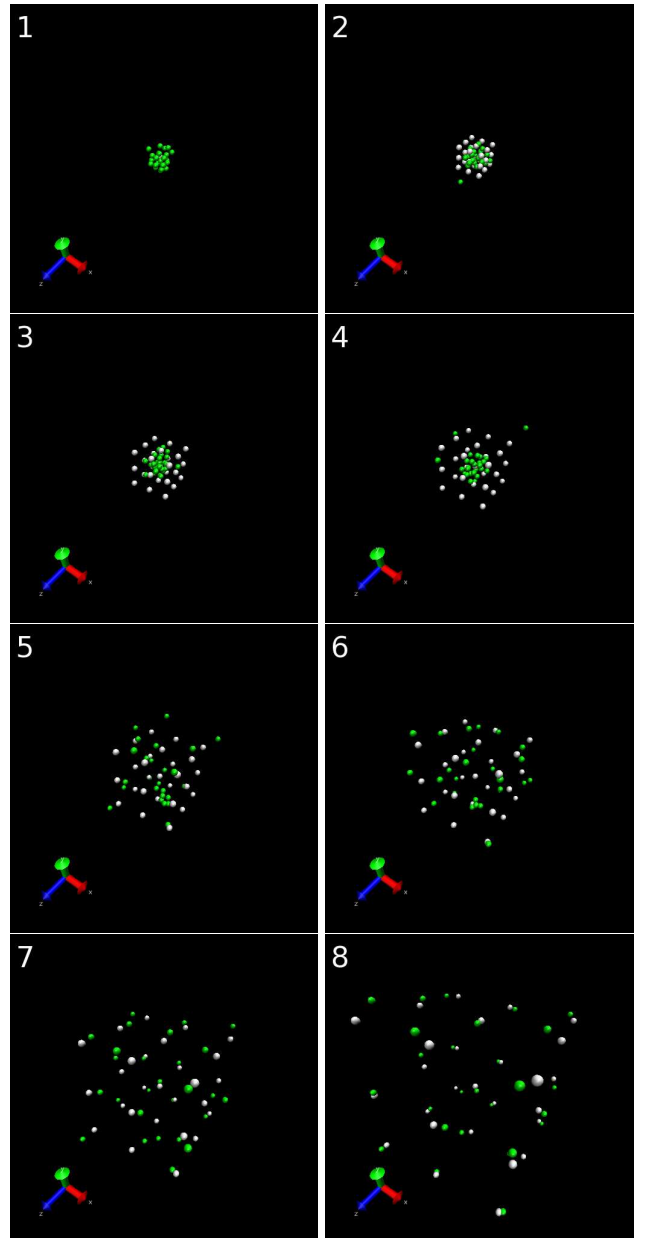

Figure 5. Illustration of the disassembly of a 27-atom argon cluster, with green spheres representing the argon atoms and lighter spheres the guide particles (helium atoms are not shown). In frame 1, all the guides lie at the origin of the cell. By frame 2, the guides have drifted far enough apart for a single argon atom to escape temporarily from the cluster before rejoining it in frame 3. In frame 4, several atoms have escaped, but remain in close proximity to the reduced cluster. A threshold is reached in frame 5, where many argon atoms break free to leave a fragment of about five atoms that also soon disintegrates as shown in frame 6. Shortly after, all of the atoms fall into motion about their partner guide particles which continue along steady paths away from one another (frames 7 and 8). The reader is encouraged to view movies of the disassembly provided in the Supplemental Material [52].
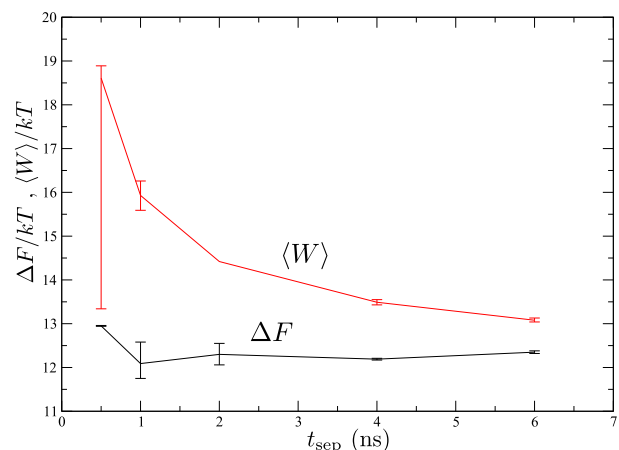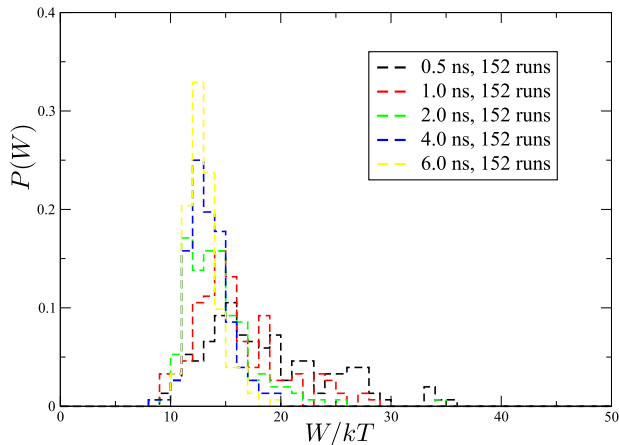
Figure 6. The distribution of work W (top) for sets of disassembly trajectories for the 5-atom argon cluster, for a range of separation times. The lower plot shows the mean of the work $\langle W \rangle$ and the corresponding free energy differences ΔF calculated via the Jarzynski equality for each t_{sep}.

Figure 7. Plots similar to those shown in Figure 6 but for the 27-atom argon cluster.

structed such that $F_s^{\text{ICCT}}(1) = 0$, where γ is the surface tension of the planar liquid-vapour interface, again taken from Baidakov *et al.* [53]. It is clear from Figure 8 that the calculations presented in this study are consistent with the previous Monte Carlo results. This is satisfactory support for the disassembly approach that we have developed. We note that all three are reasonably well represented by the ICCT model, which is somewhat surprising.

Note that the construction of a traditional plot of the nucleation barrier such as Figure 1 would require us to subtract a term $ikT \ln S$ from the excess free energies in Figure 8. Inserting a supersaturation of 30 would then yield a critical size of about 20, for example.
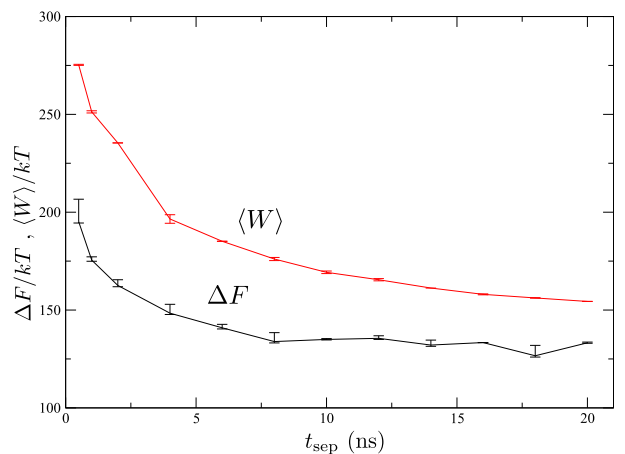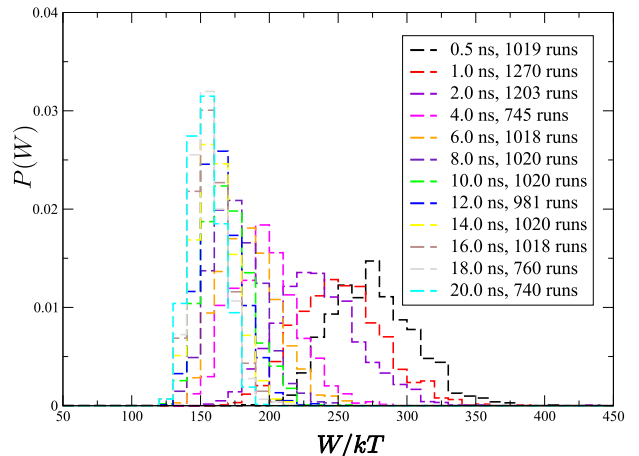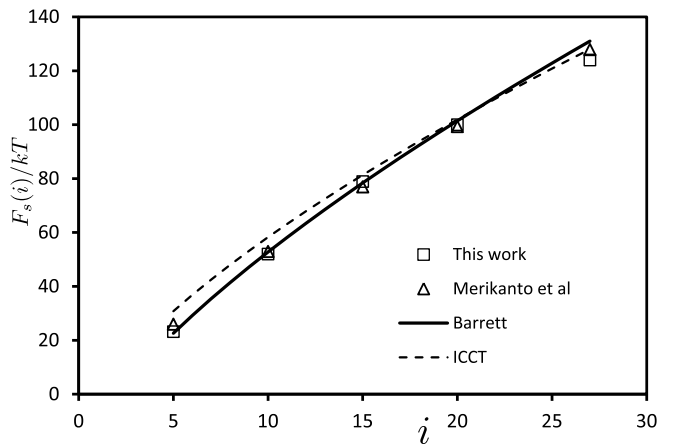

Figure 8. Excess free energies for argon-like Lennard-Jones clusters obtained from disassembly at 60 K are shown as squares and compared with values obtained in Monte Carlo studies by Barrett and Knight [27] at 59.88 K (solid line) and Merikanto *et al.* [28, 29] at 60.18 K (triangles). Also shown is the prediction from internally consistent classical nucleation theory for a temperature of 60.31 K (dashed line).

IV. CONCLUSIONS

We have developed a method of guided cluster disassembly in molecular dynamics, capable of extracting the excess free energy associated with the formation of a molecular cluster from the saturated vapour phase. This property is often regarded as a surface term and it plays a central role in kinetic and thermodynamic models of the process of droplet nucleation.

After exploring some aspects of the method by separating a dimer, the technique was applied to the controlled disassembly of Lennard-Jones argon clusters between 5 and 27 atoms in size. The extracted free energy of disassembly has been related to the excess free energy of the cluster through an analysis of the statistical mechanics of free and tethered clusters. Our calculations for clusters of various sizes are consistent with previous studies by Barrett and Knight [27] and Merikanto et al. [28, 29], both of which require the evaluation of a sequence of free energy differences between monomer and dimer, dimer and trimer, etc. A Lennard-Jones microscopic model of argon, within the standard kinetic and thermodynamic framework of nucleation theory, cannot account for the experimental argon nucleation data of Iland et al. [30], but we do not speculate here about this disparity.

The approach should be contrasted with methods of free energy estimation based on thermodynamic integration. In those methods, the strength of the interparticle interactions is evolved over a sequence of equilibrium calculations. Our approach also involves the evolution of a Hamiltonian, but it is the tether potentials that change with time, not the interparticle interactions. Furthermore, we evolve by nonequilibrium molecular dynamics rather than studying a sequence of equilibrium ensembles, and we are only required to apply a cluster definition when selecting the initial configurations, not during the evolution.

We believe that our process of mechanical disassembly offers an intuitive understanding of the meaning of the work of formation that plays such a central role in nucleation theory. We suggest that a direct evaluation of this quantity is preferable to an approach based on summing the free energy changes associated with the addition of single molecules to a cluster, on the grounds that we avoid the possible compounding of statistical errors. The computational costs of our current study of argon clusters have been higher than those of more traditional methods such as grand canonical Monte Carlo [29], for the same level of accuracy, largely because of our exploration of different protocols and our use of an explicit helium thermostat, but these can be reduced with further development. A particularly powerful variant of the disassembly scheme is to separate a cluster into two subclusters under similar mechanical guidance, in order to relate the distribution of work performed to a free energy of 'mitosis', essentially a difference in excess free energies between the initial cluster and the two final subclusters. Such comparisons would be unfeasible to perform in Monte Carlo. The calculations are not onerous and an evaluation of the excess free energy of clusters of up to 128 water molecules is to be reported [54]. Furthermore, the explicit thermostat can be replaced by an implicit scheme. With such tools, and guided by the experience developed in the current investigation of argon, we intend to carry out studies of clusters of water, acids and organic molecules, species that are particularly relevant to the process of aerosol nucleation in the atmosphere.

Appendix A: Argon dimer separation

We test the feasibility of the approach using two protocols of controlled dimer separation. First, the guide particles are made to drift apart with the tether strengths held constant, and then we allow the tethers to tighten over the course of the process. We determine the manner of dimer separation that leads to an accurate estimate of the free energy change.

We start by evaluating the free energy of a tethered dimer of argon-like atoms analytically. Particles are distinguishable in molecular dynamics simulations since they carry labels, so we take this into account in the analysis. The initial Hamiltonian of the dimer system is

$$H_i^{\text{dimer}} = \frac{\boldsymbol{p}_1^2}{2m} + \frac{\boldsymbol{p}_2^2}{2m} + \frac{1}{2}\kappa_i \left(\boldsymbol{x}_1 - \mathbf{X}_1\right)^2 \tag{A1}$$
$$+ \frac{1}{2}\kappa_i \left(\boldsymbol{x}_2 - \mathbf{X}_2\right)^2 + \Phi\left(|\boldsymbol{x}_1 - \boldsymbol{x}_2|\right),$$

where m is the argon mass, and $\Phi(|\boldsymbol{x}_1 - \boldsymbol{x}_2|)$ is a pairwise interaction potential. When the guide particles both lie at the origin ($\mathbf{X}_1 = \mathbf{X}_2 = 0$), the initial partition function is

$$Z_i^{\text{dimer}} = \frac{1}{h^6} \int \exp\left(-\frac{p_1^2 + p_2^2}{2mkT}\right) d\boldsymbol{p}_1 d\boldsymbol{p}_2 \tag{A2}$$
$$\times \int \exp\left(-\kappa_i \frac{x_1^2 + x_2^2}{2kT}\right) \exp\left(-\frac{\Phi\left(|\boldsymbol{x}_1 - \boldsymbol{x}_2|\right)}{kT}\right) d\boldsymbol{x}_1 d\boldsymbol{x}_2,$$

noting that there is no correction factor of one half since the atoms are distinguishable. Substituting $\boldsymbol{r} = \boldsymbol{x}_1 - \boldsymbol{x}_2$ and $\boldsymbol{R} = \boldsymbol{x}_1 + \boldsymbol{x}_2$, the partition function Z_i^{dimer} becomes

$$\frac{1}{\lambda_{\text{th}}^6} \int \frac{1}{8} \exp\left(-\kappa_i \frac{x_1^2 + x_2^2}{2kT}\right) \exp\left(-\frac{\Phi\left(|\boldsymbol{x}_1 - \boldsymbol{x}_2|\right)}{kT}\right) d\boldsymbol{r} d\boldsymbol{R}$$
$$= \frac{1}{\lambda_{\text{th}}^6} \frac{\pi}{2} \int \exp\left(-\kappa_i \frac{R^2 + r^2}{4kT}\right) \exp\left(-\frac{\Phi\left(r\right)}{kT}\right) r^2 d\boldsymbol{r} d\boldsymbol{R}$$
$$= \frac{1}{\lambda_{\text{th}}^6} \frac{\pi}{2} \left(\frac{4\pi kT}{\kappa_i}\right)^{\frac{3}{2}} \int_0^{r_c} r^2 \exp\left(-\frac{\kappa_i r^2 + 4\Phi(r)}{4kT}\right) dr, \tag{A3}$$

where $\lambda_{\text{th}} = h/(2\pi m k T)^{1/2}$ is the thermal de Broglie wavelength. We have imposed an upper limit r_c on the separation between the two atoms, corresponding to a definition of what we mean by a dimer.

For the final state in which the two argon atoms are tethered to respective guide particles that are far apart, the Hamiltonian is simply that in Eq. (A1) without the interaction term, and with a final tether strength κ_f. The corresponding final partition function is

$$Z_f^{\text{dimer}} = \frac{1}{h^6} \int \exp\left(-\frac{p_1^2 + p_2^2}{2mkT}\right) d\boldsymbol{p}_1 d\boldsymbol{p}_2 \qquad (A4)$$

$$\times \int \exp\left(-\kappa_f \frac{x_1^2 + x_2^2}{2kT}\right) d\boldsymbol{x}_1 d\boldsymbol{x}_2 = \frac{1}{\lambda_{\text{th}}^6}\left(\frac{2\pi kT}{\kappa_f}\right)^3 .$$

The free energy change in separating a dimer of tethered atoms can therefore be expressed as

$$\Delta F = kT \ln\left(Z_i^{\text{dimer}}/Z_f^{\text{dimer}}\right) \qquad (A5)$$

$$= kT \ln\left[\left(\frac{\kappa_f^2}{\kappa_i kT}\right)^{\frac{3}{2}} \int_0^{r_c} \frac{r^2}{2\sqrt{\pi}} \exp\left(-\frac{\kappa_i r^2 + 4\Phi(r)}{4kT}\right) dr\right],$$

which can be evaluated numerically. The parameter r_c is the Stillinger radius used to identify a dimer configuration in the equilibrated molecular dynamics simulation, to which we now turn.

We place two argon-like particles within a periodic cell with edge length 50 Å, each tethered to guide particles through a harmonic interaction $\frac{1}{2}\kappa(t)r^2$, where r is the separation between the argon atom and its guide, and $\kappa(t)$ is the tethering force constant. The argon atoms are thermalised through interaction with a gas of 100 helium-like atoms kept at constant temperature using a Nosé-Hoover thermostat. Conventional masses of 39.85 and 4.003 amu for the argon and helium-like particles are adopted, while the guide particles are assigned a vastly greater mass of 4×10^{12} amu. Interaction potentials are specified by

$$\Phi\left(r_{jk}\right) = 4\epsilon_{jk}\left[\left(\frac{\sigma_{jk}}{r_{jk}}\right)^{12} - \left(\frac{\sigma_{jk}}{r_{jk}}\right)^6\right], \qquad (A6)$$

with parameters shown in Table III, though it should be noted that only the repulsive part of the interaction between argon and helium is employed in order to prevent any binding between the two. Simulations are performed at a temperature of 15 K such that dimers are long-lived and a sufficient number of configurations satisfying the separation criterion $r \leq r_c = 1.5\sigma_{\text{ArAr}}$ can be obtained from the equilibrated trajectory. With a constant tethering force constant of $0.05\,\text{kJ mol}^{-1}\text{Å}^{-2}$, we generate an equilibrated molecular dynamics trajectory of duration 100 ns and choose 10^3 dimer configurations for use as starting points for the separation process.

j	k	ϵ_{jk} / kJ mol^{-1}	σ_{jk} / Å
Ar	Ar	0.995581	3.405
He	He	0.084311	2.600
Ar	He	0.289721	3.000

Table III. Parameters for the Lennard-Jones potentials, where j and k are the atomic labels, ϵ_{jk} is the depth of the potential well, and σ_{jk} is the range parameter [55].
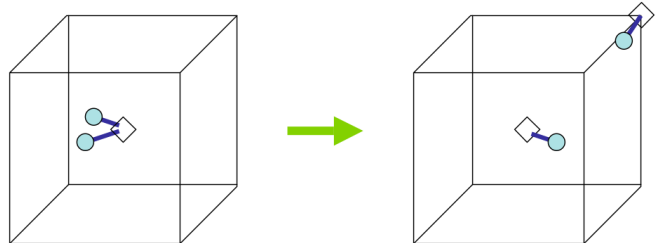

Figure 9. Illustration of the dimer separation process. Both guide particles (diamonds) are initially at the origin, but one is made to drift towards a corner of the simulation cell.

a. Guiding at constant tether strength

One of the guide particles drifts from the origin to a corner of the cubic simulation cell over a separation time t_{sep} while the other remains stationary (see Figure 9). We choose t_{sep} to be 1, 2 or 4 ns and the velocity of the moving guide particle (labelled 1) is given by $\mathbf{V}_1 = [\mathbf{X}_1(t = t_{\text{sep}}) - \mathbf{X}_1(t = 0)]/t_{\text{sep}}$.

For initial and final tethering force constants of $0.05\,\text{kJ mol}^{-1}\text{Å}^{-2}$, the expected free energy change in separating the dimer is $5.716\,kT$ according to Eq. (A5). Distributions of the work done for each rate of dimer separation are shown in Figure 10, and the corresponding estimates of the free energy change obtained from the Jarzynski equality are compared with the expected value in the lower part of Figure 11. A longer separation time leads to a better estimate of the free energy change since the process is then closer to being quasistatic.

b. Guiding with tether tightening

We now elaborate the process by tightening the tethers during guide drift according to

$$\kappa(t) = \kappa_i \quad \text{for } t \leq t_i$$

$$= \kappa_i + \frac{\kappa_f - \kappa_i}{2}\left[1 - \cos\left(\pi\frac{t - t_i}{t_s - t_i}\right)\right] \text{ for } t_i < t \leq t_s$$

$$= \kappa_f \quad \text{for } t > t_s, \qquad (A7)$$

where t_i is the time at which the force constant begins to change, and t_s is the time at which it reaches its final value. Once again starting with dimer configurations and an initial tethering force constant of $0.05\,\text{kJ mol}^{-1}\text{Å}^{-2}$
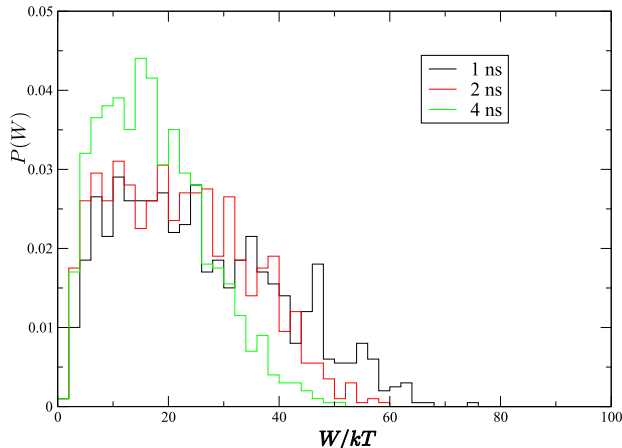
Figure 10. Distributions of the work done in the disassembly of a dimer for separation times t_{sep} of 1, 2 and 4 ns.
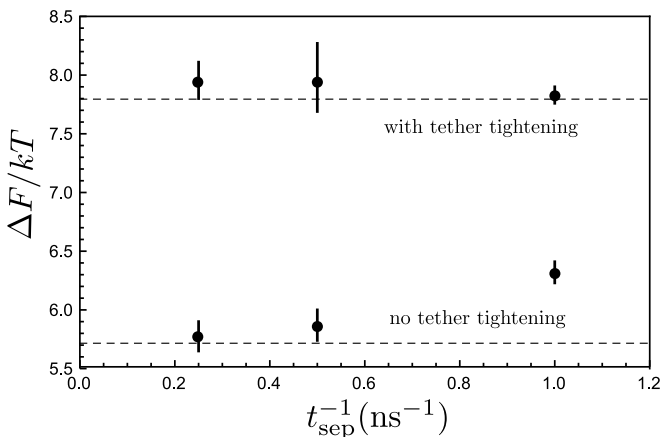

Figure 11. Convergence of the Jarzynski-estimated free energy change toward the expected value (dashed line) as the dimer separation rate is decreased, while keeping the tethering strength constant (lower set) and when the tethers are tightened (upper set).

at 15 K, three dimer separation times are investigated, during which the force constant rises by a factor of two. The times t_i and t_s are specified as 20% and 80% of the total separation time. The expected free energy change associated with dimer separation is $7.795\,kT$ according to Eq. (A5). It can be seen from the upper part of Figure 11 that all three separation rates give acceptable estimates of the free energy change. Furthermore, the greater compatibility between the distributions of the work performed at different separation rates shown in Figure 12, compared with those in the simulations with constant tether strength, suggests that a protocol where the tethers tighten while the guide particles drift apart is more effective. Intuitively, the separation is then conducted more firmly, and with less dissipation.
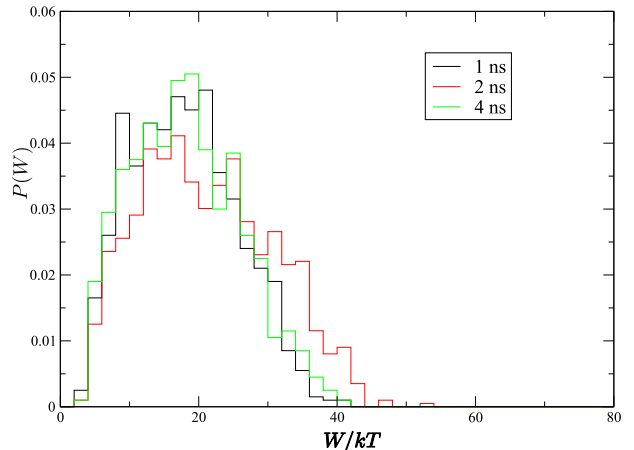

Figure 12. Distributions of the work of dimer disassembly where the atoms are guided apart and the tethers tightened for three different separation times.

Appendix B: Analysis of cluster free energies

1. Free and tethered clusters

The canonical partition function $Z_F = \exp\left(-F_F/kT\right)$ for an untethered, or 'free' cluster of i indistinguishable particles governed by a Hamiltonian H composed of kinetic energy terms and pairwise interactions is given by

$$Z_F = \frac{1}{i!h^{3i}} \int \prod_{j=1}^{i} d\boldsymbol{x}_j d\boldsymbol{p}_j \exp\left[-H\left(\{\boldsymbol{x}_k\}\right)/kT\right], \text{(B1)}$$

where F_F is the associated free energy. For a cluster tethered to the origin, the Hamiltonian will include an additional set of harmonic potentials, such that the partition function is

$$Z_T = \exp\left(-F_T/kT\right) = \frac{1}{i!h^{3i}} \int \prod_{j=1}^{i} d\boldsymbol{x}_j d\boldsymbol{p}_j$$

$$\times \exp\left[-\left(H\left(\{\boldsymbol{x}_k\}\right) + \sum_{j=1}^{i} \frac{1}{2}\kappa_i x_j^2\right)/kT\right], \text{(B2)}$$

where F_T is the free energy of the tethered cluster, and κ_i is the initial tethering force constant.

We insert a factor of unity in the form $1 = \int \delta\left(\frac{1}{i}\sum_{j=1}^{i} \boldsymbol{x}_j - \boldsymbol{x}_c\right) d\boldsymbol{x}_c$ into Eqs. (B1) and (B2), and transform to particle coordinates with respect to the cluster centre of mass $\boldsymbol{x}_c$, namely $\boldsymbol{x}'_j = \boldsymbol{x}_j - \boldsymbol{x}_c$. The partition

function for a free cluster becomes

$$Z_F = \frac{1}{i!h^{3i}} \int \prod_{j=1}^{i} d\boldsymbol{x}'_j d\boldsymbol{p}_j d\boldsymbol{x}_c \exp\left[-H\left(\{\boldsymbol{x}'_k\}\right)/kT\right]$$

$$\times \delta\left(\frac{1}{i}\sum_{j=1}^{i}\boldsymbol{x}'_j\right)$$

$$= \frac{V}{i!h^{3i}} \int \prod_{j=1}^{i} d\boldsymbol{x}'_j d\boldsymbol{p}_j \exp\left[-H\left(\{\boldsymbol{x}'_k\}\right)/kT\right]$$

$$\times \delta\left(\frac{1}{i}\sum_{j=1}^{i}\boldsymbol{x}'_j\right) = V Z_F^c, \tag{B3}$$

where V is the system volume and Z_F^c is the partition function for a cluster whose centre of mass is fixed at the origin. It should be noted that since the Hamiltonian contains pairwise interactions, it may be rewritten as $H(\{\boldsymbol{x}_k\}) = H(\{\boldsymbol{x}'_k\})$ after the change of variables.

Similarly, the partition function for a tethered cluster can be rewritten as

$$Z_T = \frac{1}{i!h^{3i}} \int \prod_{j=1}^{i} d\boldsymbol{x}'_j d\boldsymbol{p}_j d\boldsymbol{x}_c \delta\left(\frac{1}{i}\sum_{j=1}^{i}\boldsymbol{x}'_j\right) \tag{B4}$$

$$\times \exp\left[-\left(H\left(\{\boldsymbol{x}'_k\}\right)+\sum_{j=1}^{i}\frac{1}{2}\kappa_i x_j^2\right)/kT\right].$$

The second term in the exponent of Eq. (B4) may be simplified using the constraint $\sum_{j=1}^{i}\boldsymbol{x}'_j = 0$ and it follows that $\sum_{j=1}^{i} x_j^2 = \sum_{j=1}^{i} x_j'^2 + i x_c^2$, giving

$$Z_T = \frac{1}{i!h^{3i}} \int \prod_{j=1}^{i} d\boldsymbol{x}'_j d\boldsymbol{p}_j d\boldsymbol{x}_c \exp\left[-\frac{1}{2}\kappa_i i x_c^2/kT\right]$$

$$\times \exp\left[-\left(H\left(\{\boldsymbol{x}'_k\}\right)+\frac{1}{2}\kappa_i \sum_{j=1}^{i} x_j'^2\right)/kT\right]\delta\left(\frac{1}{i}\sum_{j=1}^{i}\boldsymbol{x}'_j\right)$$

$$= \left(\frac{2\pi kT}{i\kappa_i}\right)^{\frac{3}{2}} \frac{1}{i!h^{3i}} \int \prod_{j=1}^{i} d\boldsymbol{x}'_j d\boldsymbol{p}_j \exp\left[-H\left(\{\boldsymbol{x}'_k\}\right)/kT\right]$$

$$\times \exp\left[-\frac{1}{2}\kappa_i \sum_{j=1}^{i} x_j'^2/kT\right]\delta\left(\frac{1}{i}\sum_{j=1}^{i}\boldsymbol{x}'_j\right)$$

$$= \left(\frac{2\pi kT}{i\kappa_i}\right)^{\frac{3}{2}} Z_T^c, \tag{B5}$$

where Z_T^c is the partition function of a cluster constrained to have its centre of mass at the origin as well as having its constituent particles tethered to the origin by a harmonic potential.

Next, we employ the Gibbs-Bogoliubov approach [56, 57] to compare the free energies F_F^c and F_T^c of systems with Hamiltonians H_0 and Hamiltonian $H_0 + U$, defined by $\exp\left(-F_F^c/kT\right) = \int d\Gamma \exp\left[-H_0/kT\right]$ and

$\exp\left(-F_T^c/kT\right) = \int d\Gamma \exp\left[-\left(H_0+U\right)/kT\right]$, where Γ represents the configuration of a system, and $d\Gamma$ is proportional to the phase space volume element $\Pi_j d\boldsymbol{x}'_j d\boldsymbol{p}_j$. In the context of the tethered cluster described by Eq. (B5), U represents the term $\frac{1}{2}\kappa_i\sum_{j=1}^{i} x_j'^2$, while H_0 is the untethered Hamiltonian $H\left(\{\boldsymbol{x}'_k\}\right)$ modified by the delta function constraint. F_T^c is therefore the free energy of a tethered cluster with its centre of mass further constrained to lie at the origin, and is equal to $-kT\ln Z_T^c$. A similar relationship exists between F_F^c, the free energy of an untethered cluster with fixed centre of mass, and Z_F^c.

The free energies F_F^c and F_T^c may be related through

$$\exp\left(-F_T^c/kT\right) = \frac{\int d\Gamma \exp\left(-H_0/kT\right)\exp\left(-U/kT\right)}{\int d\Gamma \exp\left(-H_0/kT\right)}$$

$$\times \int d\Gamma \exp\left(-H_0/kT\right)$$

$$= \langle\exp\left(-U/kT\right)\rangle_0 \exp\left(-F_F^c/kT\right), \tag{B6}$$

where angle brackets represent an average in the statistical ensemble corresponding to H_0. For small $\langle U/kT\rangle_0$, we can write $\langle\exp\left(-U/kT\right)\rangle_0 \simeq \exp\left(-\langle U\rangle_0/kT\right)$, and hence

$$\exp\left(-F_T^c/kT\right) \simeq \exp\left[\left(-F_F^c - \langle U\rangle_0\right)/kT\right], \tag{B7}$$

with $\langle U\rangle_0$ given by

$$\langle U\rangle_0 = \frac{\int d\Gamma U\left(\{\boldsymbol{x}'_k\}\right)\exp\left(-H_0/kT\right)}{\int d\Gamma \exp\left(-H_0/kT\right)}. \tag{B8}$$

$U(\{\boldsymbol{x}'_k\})$ is a sum of single-particle harmonic potentials of the form $U_{\mathrm{HO}}(\boldsymbol{x}'_k) = \frac{1}{2}\kappa_i x_k'^2$, so Eq. (B8) can be written as

$$\langle U\rangle_0 = \frac{\sum_{k=1}^{i}\int d\Gamma U_{\mathrm{HO}}\left(\boldsymbol{x}'_k\right)\exp\left(-H_0/kT\right)}{\int d\Gamma \exp\left(-H_0/kT\right)}$$

$$= i\frac{\int d\Gamma U_{\mathrm{HO}}\left(\boldsymbol{x}'_k\right)\exp\left(-H_0/kT\right)}{\int d\Gamma \exp\left(-H_0/kT\right)} = i\langle U_{\mathrm{HO}}\rangle_0. \tag{B9}$$

We next introduce the spatial density profile of a single particle (labelled k without loss of generality) in a cluster constrained to have its centre of mass at the origin but not tethered, namely

$$\rho_0(\boldsymbol{y}) = \frac{\int d\Gamma \exp\left(-H_0/kT\right)\delta(\boldsymbol{x}'_k - \boldsymbol{y})}{\int d\Gamma \exp\left(-H_0/kT\right)}, \tag{B10}$$

with $\int \rho_0(\boldsymbol{y}) d\boldsymbol{y} = 1$. We can write

$$\langle U_{\mathrm{HO}}\rangle_0 = \int \rho_0(\boldsymbol{y}) U_{\mathrm{HO}}(\boldsymbol{y}) d\boldsymbol{y}, \tag{B11}$$

which represents the average tethering energy of a particle that is spatially distributed according to the density $\rho_0(\boldsymbol{y})$. The condition that the tether potential makes a relatively small contribution to the mean energy of the cluster is $\langle U_{\mathrm{HO}}\rangle_0 = \frac{1}{2}\kappa_i \int \rho_0(\boldsymbol{y}) y^2 d\boldsymbol{y} \ll kT$, in which case

the approximations involved in the Gibbs-Bogoliubov approach are acceptable and the initial tethering potential weak enough that the cluster is only slightly distorted in comparison with a free cluster. Thus we write

$$Z_T^c = \exp\left(-F_T^c/kT\right) \simeq \exp\left[\left(-F_F^c - i\left\langle U_{\mathrm{HO}}\right\rangle_0\right)/kT\right].$$
(B12)

Eq. (B5) can then be written as

$$Z_T = \left(\frac{2\pi kT}{i\kappa_i}\right)^{\frac{3}{2}} \frac{1}{i!h^{3i}} \int \prod_{j=1}^{i} d\boldsymbol{x}_j' d\boldsymbol{p}_j \exp\left[-H\left(\{\boldsymbol{x}_k'\}\right)/kT\right]$$

$$\times \exp\left[-i\int \rho_0\left(\boldsymbol{y}\right)\kappa_i y^2 d\boldsymbol{y}/2kT\right] \delta\left(\frac{1}{i}\sum_{j=1}^{i}\boldsymbol{x}_j'\right), \quad (B13)$$

such that the relationship between the partition function of a tethered cluster, and the partition function of a free cluster with a constrained centre of mass Z_F^c, is

$$Z_T = Z_F^c \left(\frac{2\pi kT}{i\kappa_i}\right)^{\frac{3}{2}} \exp\left[-i\int \rho_0\left(\boldsymbol{y}\right)\kappa_i y^2 d\boldsymbol{y}/2kT\right].$$
(B14)

Combining Eqs. (B3) and (B14) then gives

$$\ln Z_T = \ln\left[\frac{Z_F}{V}\left(\frac{2\pi kT}{i\kappa_i}\right)^{\frac{3}{2}}\right] - \frac{i\kappa_i}{2kT}\int \rho_0(\boldsymbol{y}) y^2 d\boldsymbol{y}, \text{ (B15)}$$

or

$$F_F - F_T = -kT\ln\left[\rho_c(0) V\right] - \frac{i\kappa_i}{2}\int \rho_0(\boldsymbol{y}) y^2 d\boldsymbol{y}, \text{ (B16)}$$

where $(i\kappa_i/2\pi kT)^{3/2}$ has been replaced by a function $\rho_c(0)$, representing the probability density that the centre of mass of the tethered cluster lies at the origin. This equivalence can be demonstrated by deriving the distribution of the cluster centre of mass, through considering a single particle with mass $M = im$ and coordinates $\boldsymbol{x}_c$ and $\boldsymbol{p}_c$ residing in a potential $i\kappa_i x_c^2/2$. The positional probability density at $\boldsymbol{z}$ is

$$\rho_c(\boldsymbol{z}) = \frac{\int d\boldsymbol{x}_c d\boldsymbol{p}_c \exp\left(-\frac{i\kappa_i x_c^2}{2kT} - \frac{p_c^2}{2MkT}\right)\delta\left(\boldsymbol{x}_c - \boldsymbol{z}\right)}{\int d\boldsymbol{x}_c d\boldsymbol{p}_c \exp\left(-\frac{i\kappa_i x_c^2}{2kT} - \frac{p_c^2}{2MkT}\right)}$$

$$= \left(\frac{i\kappa_i}{2\pi kT}\right)^{3/2} \exp\left(-\frac{i\kappa_i z^2}{2kT}\right), \quad (B17)$$

such that $\rho_c(0) = (i\kappa_i/(2\pi kT))^{3/2}$.

The purpose of the substitution is that the first term on the right hand side in Eq. (B16) may be interpreted as two competing contributions to the free energy difference $F_F - F_T$. We write

$$-kT\ln\left[\rho_c(0) V\right] = -T\left[-k\ln\left(\frac{1}{\rho_c(0)}\right) + k\ln V\right], \text{ (B18)}$$

such that the first term corresponds to the removal of the entropic contribution to free energy associated with the

freedom of motion of the cluster centre of mass within a constrained volume $1/\rho_c(0)$, brought about by the tethers, and the second term represents the addition of entropic free energy corresponding to the freedom of motion in volume V. Finally, the second term in Eq. (B16) is an estimate of the removal of tethering potential energy when relating a tethered to a free cluster.

2. Excess free energy from the free energy of disassembly

We now establish the relationship between the free energy of a free cluster to the cluster work of formation defined as $\phi(i) = \Omega_s(i) - ikT\ln S$, where $\Omega_s(i) = F_F(i) - i\mu_s$ is the grand potential of a free cluster of i particles in an environment at chemical potential μ_s for which the bulk condensed and vapour phases coexist. The excess free energy (difference) of the cluster is therefore

$$F_s(i) = \phi(i) - \phi(1) + (i-1) kT\ln S$$
$$= F_F(i) - F(1) - (i-1)\mu_s, \quad (B19)$$

having used Eq. (7).

Assuming the vapour is ideal, the coexistence chemical potential μ_s and the monomer Helmholtz free energy $F(1)$ are simply $\mu_s = kT\ln(\rho_{vs}\Lambda)$ and $F(1) = -kT\ln(V/\Lambda)$, respectively, where ρ_{vs} is the particle density in a saturated vapour and $\Lambda = \lambda_{\mathrm{th}}^3$ with $\lambda_{\mathrm{th}} = h/(2\pi mkT)^{1/2}$. The excess free energy $F_s(i)$ can now be expressed as

$$F_s(i) = F_F + kT\ln(V/\Lambda) - (i-1) kT\ln(\rho_{vs}\Lambda)$$
$$= F_T - kT\ln\left[\rho_c(0) V\right] - \frac{i\kappa_i}{2}\int \rho_0(\boldsymbol{y}) y^2 d\boldsymbol{y}$$
$$+ kT\ln(V/\Lambda) - (i-1) kT\ln(\rho_{vs}\Lambda). \quad (B20)$$

Now we consider the free energy change associated with the process of cluster disassembly. The difference in free energy between separated constituent particles each tethered to a guide particle, and a tethered cluster, is $\delta F = F_f - F_T$, where $F_f = -3ikT\ln(kT/\hbar\omega_f)$ is the free energy of i harmonic oscillators in three dimensions, where the angular frequency $\omega_f = (\kappa_f/m)^{1/2}$ of the oscillators is related to the final value of the tethering force constant κ_f.

It should be recognised, however, that the quantity δF is *not* the free energy difference extracted from the molecular dynamics simulations of cluster disassembly. Molecular dynamics simulations always involve distinguishable particles, since they are assigned labels, and δF is a difference between the free energy of i indistinguishable particles in a cluster, and i particles that are distinguishable through having been physically separated to regions around their final tether points.

The free energy difference that is extracted in our procedure is actually $\Delta F = F_f - F_T^{\mathrm{dist}}$, where the superscript

in F_T^{dist} reminds us that it is the free energy of a tethered cluster of distinguishable particles. But we can relate the partition function of such a cluster to the partition function Z_T for indistinguishable particles by the usual classical procedure, namely $Z_T^{\text{dist}} = i! Z_T$, and since $F_T^{\text{dist}} = -kT \ln Z_T^{\text{dist}} = -kT \ln Z_T - kT \ln i! = F_T - kT \ln i!$ we have

$$\Delta F = F_f - F_T + kT \ln i! = \delta F + kT \ln i!, \qquad (\text{B21})$$

such that $F_T = F_f - \delta F = F_f - \Delta F + kT \ln i!$. Substituting into Eq. (B20) then gives

$$F_s(i) = -\Delta F - i kT \ln (\rho_{vs} v_{\text{HO}}) + kT \ln i!$$
$$- kT \ln \left(\frac{\rho_c(0)}{\rho_{vs}} \right) - \frac{i \kappa_i}{2} \int \rho_0 (\boldsymbol{y}) y^2 d\boldsymbol{y}, \qquad (\text{B22})$$

where $v_{\text{HO}} = (2\pi kT/\kappa_f)^{3/2}$ is a volume scale associated with the confinement of particles within the final harmonic tether potentials. It should be noted that the excess free energy F_s does not depend upon the Planck constant h, nor on the system volume V, as is to be expected.

In order to complete our specification of $F_s(i)$ in terms of ΔF and material properties, we need to estimate the final term in Eq. (B22). We write $\int \rho_0(\boldsymbol{y}) y^2 d\boldsymbol{y} = \int_0^\infty \rho_0(r) 4\pi r^4 dr$, where r is the distance from the cluster centre of mass, and recall that $\rho_0(r)$ is the single-particle density profile in an untethered cluster with fixed centre of mass. As an approximation, we imagine the cluster to be spherical with a constant particle density, such that $\rho_0(r) \simeq \rho_l/i$ for $0 < r < r_{\max}$, where ρ_l is the particle density in the condensed phase, and $r_{\max}$ is the radius of the cluster. Since the probability density $\rho_0(r)$ is normalised, we have $\int_0^{r_{\max}} (\rho_l/i) 4\pi r^2 dr = 1$, such that $r_{\max} = (3i/4\pi\rho_l)^{1/3}$ and so

$$\int_0^{r_{\max}} \frac{\rho_l}{i} 4\pi r^4 dr = \frac{4\pi \rho_l}{5i} r_{\max}^5 = \frac{3}{5} \left(\frac{3 i v_l}{4\pi} \right)^{2/3}, \quad (\text{B23})$$

where $v_l = 1/\rho_l$ is the volume per particle in the condensed phase. Substituting this into Eq. (B22) gives Eqs. (13-17) in the main text.

[1] D. Kashchiev, *Nucleation: Basic Theory with Applications* (Butterworth-Heinemann, 2000).
[2] I. J. Ford, Proc. Instn. Mech. Engrs. **218**, 883 (2004).
[3] H. Vehkamäki, *Classical Nucleation Theory in Multicomponent Systems* (Springer, 2006).
[4] V. I. Kalikmanov, *Nucleation Theory* (Springer, 2013).
[5] R. Zhang, A. Khalizov, L. Wang, M. Hu, and W. Xu, Chem. Rev. **112**, 1957 (2011).
[6] M. Kulmala, T. Petäjä, M. Ehn, J. Thornton, M. Sipalä, D. Worsnop, and V.-M. Kerminen, Annu. Rev. Phys. Chem. **65**, 21 (2014).
[7] I. Agranovski, *Aerosols: Science and Technology* (Wiley-VCH, 2010).
[8] M. Bakhtar, M. Ebrahimi, and R. A. Webb, Proc. Instn Mech. Engrs. **209(C2)**, 115 (1995).
[9] R. Becker and W. Döring, Ann. Physik. (Leipzig) **24**, 719 (1935).
[10] I. J. Ford, Phys. Rev. E **56**, 5615 (1997).
[11] J. Zeldovich, Soviet Physics JETP **12**, 525 (1942).
[12] I. J. Ford and S. A. Harris, J. Chem. Phys. **120**, 4428 (2004).
[13] S. L. Girshick and C.-P. Chiu, J. Chem. Phys. **93**, 1273 (1990).
[14] R. C. Tolman, J. Chem. Phys. **17**, 333 (1949).
[15] A. Laaksonen, I. J. Ford, and M. Kulmala, Phys. Rev. E **49**, 5517 (1994).
[16] V. I. Kalikmanov and M. E. H. van Dongen, J. Chem. Phys. **103**, 4250 (1995).
[17] J. K. Lee, J. A. Barker, and F. F. Abraham, J. Chem. Phys. **58**, 3166 (1973).
[18] B. N. Hale and R. C. Ward, J. Stat. Phys. **28**, 487 (1982).
[19] P. R. ten Wolde and D. Frenkel, J. Chem. Phys. **109**, 9901 (1998).
[20] K. J. Oh and X. C. Zeng, J. Chem. Phys. **110**, 4471 (1999).
[21] I. Kusaka and D. W. Oxtoby, J. Chem. Phys. **113**, 10100 (2000).
[22] B. Chen, J. I. Siepmann, K. J. Oh, and M. L. Klein, J. Chem. Phys. **115**, 10903 (2001).
[23] J. Merikanto, H. Vehkamäki, and E. Zapadinsky, J. Chem. Phys. **121**, 914 (2004).
[24] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[25] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).
[26] R. E. Spinney and I. J. Ford, "Nonequilibrium statistical physics of small systems: Fluctuation relations and beyond," (Wiley-VCH, eds. R. Klages, W. Just and C. Jarzynski, 2012) Chap. 1. Fluctuation relations: a pedagogical overview.
[27] J. C. Barrett and A. P. Knight, J. Chem. Phys. **128**, 086101 (2008).
[28] J. Merikanto, E. Zapadinsky, and H. Vehkamäki, J. Chem. Phys. **125**, 084503 (2006).
[29] J. Merikanto, E. Zapadinsky, A. Lauri, and H. Vehkamäki, Phys. Rev. Lett. **98**, 145702 (2007).
[30] K. Iland, J. Wölk, and R. Strey, J. Chem. Phys. **127**, 154506 (2007).
[31] F. H. Stillinger, J. Chem. Phys. **38**, 1486 (1963).
[32] W. Smith, T. R. Forester, and I. T. Todorov, DL-POLY molecular simulation program. (1996).
[33] H. Y. Tang and I. J. Ford, J. Chem. Phys. **125**, 144316 (2006).
[34] I. J. Ford, *Statistical Physics: an entropic approach* (Wiley, Chichester, UK, 2013).
[35] D. A. Hendrix and C. Jarzynski, J. Chem. Phys. **114**, 5974 (2001).
[36] G. Hummer, Molecular Simulation **28**, 81 (2002).
[37] H. Hu, R. H. Yun, and J. Hermans, Molecular Simulation **28**, 67 (2002).

[38] D. Rodriguez-Gomez, E. Darve, and A. Pohorille, J. Chem. Phys. **120**, 3563 (2004).

[39] R. C. Lua and A. Y. Grosberg, J. Phys. Chem. B **109**, 6805 (2005).

[40] A. Dhar, Phys. Rev. E **71**, 036126 (2005).

[41] B. Palmieri and D. Ronis, Phys. Rev. E **75**, 011133 (2007).

[42] G. Hummer, J. Chem. Phys. **114**, 7330 (2001).

[43] F. Ritort, C. Bustamante, and I. Tinoco, Proc. Natl. Acad. Sci. USA **99**, 13544 (2002).

[44] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).

[45] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, Nature **437**, 231 (2005).

[46] F. Douarche, S. Ciliberto, and A. Petrosyan, J. Stat. Mech. , P09011 (2005).

[47] S. Joubaud, N. B. Garnier, F. Douarche, A. Petrosyan, and S. Ciliberto, C. R. Physique **8**, 518 (2007).

[48] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).

[49] T. P. Straatsma, H. J. C. Berendsen, and J. P. M. Postma, J. Chem. Phys. **85**, 6720 (1986).

[50] M. A. Miller and W. P. Reinhardt, J. Chem. Phys. **113**, 7035 (2000).

[51] T. Schilling and F. Schmid, J. Chem. Phys. **131**, 231102 (2009).

[52] See Supplemental Material at [URL will be inserted by publisher] for movies of cluster disassembly.

[53] V. G. Baidakov, S. P. Protsenko, Z. R. Kozlova, and G. G. Chernykh, J. Chem. Phys. **126**, 214505 (2007).

[54] G.V. Lau, P.A. Hunt, E.A. Müller, G. Jackson and I.J. Ford, in preparation.

[55] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, 1964).

[56] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, 2nd ed. (Academic Press Limited, 1986).

[57] A. Ishihara, J. Phys. A **1**, 539 (1968).