Supplemental Material: Nested transition path sampling

Peter G. Bolhuis[1] and Gábor Csányi[2]

[1] *Van 't Hoff Institute for Molecular Sciences, Universiteit van Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

[2] *Engineering Laboratory, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom*

In this supplemental material we present an additional illustrative example of the application of nested transition path sampling (NTPS) on a 3D Lennard Jones (LJ) gas-liquid transition. For simplicity we chose a small system of 16 LJ particles only, but scaling up to larger systems is straightforward. These particles interact via a truncated and shifted LJ potential[1]. As we investigate a gas system we impose a cubic box with periodic boundaries. We choose the box length to be $L = 12\sigma$ where σ is the LJ unit of length. This sets the density to $\rho \approx 0.0092$. Because of this low density condensation of the particles into a single droplet is a rare event on the molecular time scale. A NVT straightforward molecular dynamics (MD) simulation (using a time step $dt = 0.005$ and employing the Lowe thermostat[2]) shows that for $\beta \equiv 1/k_BT \approx 2.4$ the system transits between a gas phase and a liquid droplet, where the condensation rates from gas to liquid is about equal to the evaporation rate (see Fig.1 for an illustration of this behavior). For smaller β (higher temperature) evaporation is faster than condensation, and the system is mostly in the gas phase, while for higer β (lower temperature) the liquid droplet becomes more stable. We note that for this small system we can hardly speak about a phase transition, and that there are huge finite size effects. Indeed, the gas-liquid binodal is located at a higher temperature of approximately $\beta \approx 1.4$, even for the very low density we consider here[3]. Notwithstanding this observation, we will refer to the dilute phase as the 'gas' and the droplet as the 'liquid' for simplicity.

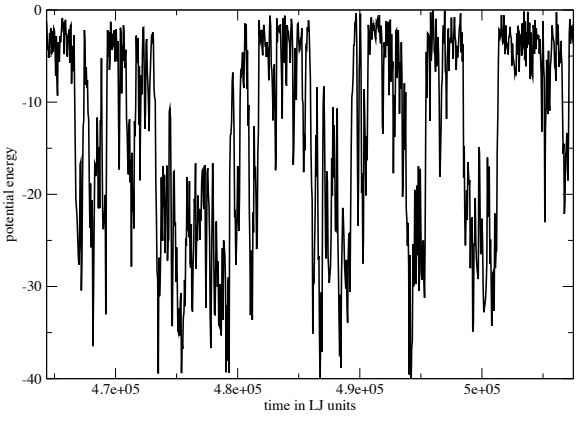

FIG. 1. A arbitrary zoom in of the time trace of the potential energy E_{pot} in a long NVT simulation at $\beta = 2.4$. The system traverses between a 'gas' phase $E_{pot} > -15k_BT$ and a 'liquid' droplet phase for $E_{pot} < -15k_BT$. Note the large fluctuations in energy. At this temperature the system still shows rare event behaviour.

We are interested in the transition rate from the gas to the liquid. The barrier toward formation of the liquid droplet is largely entropic, as all particles need to be confined to the same location in the box. Of course when particles come together they are stabilised by the attractive LJ potential. In large systems, in the region between the binodal and the spinodal where a phase transition occurs barrierlessly, the gas-liquid transition is dominated by homogenous nucleation. The formation of a new phase inside an other phase means the creation of a surface that is costly, due to unfavourable surface energy. While at the spinodal the barrier disappears, the nucleation barrier becomes higher at elevated temperatures, and becomes even infinite at the binodal (in the thermodynamic limit). Thus, when raising the temperature, the rate of liquid formation via homogenous nucleation goes decrease. This behaviour is the oppositie of that of the Arrhenius type transition discussed in the main text. This holds at constant pressure, the most familiar case, but is even true at constant volume, the case that we consider here.

To compute the rates for reciprocal temperature $\beta < 2.4$ we conduct NTPS using $K = 200$ path walkers. We initialise the path walkers using as an initial condition a liquid droplet at relatively low temperature $\beta \approx 3$. We define the stable state of the liquid droplet phase by an low potential energy $E_{pot} < -30k_BT$ and that of the gas phase by a high potential energy $E_{pot} > -5k_BT$. These boundaries follow from the straightforward NVT simulations at $\beta = 2.4$ (see Fig. 1). See for the rationale behind these choices e.g. Refs.[4, 5]). We bring this initial condition to a high total energy of $E = 30k_BT$, and run a NVE trajectory. At such high energy the droplet evaporates extremely fast. We take this trajectory as the initial path for NTPS. Each path sample in the $K = 200$ path ensemble was obtained from the previous one by copying and performing $l = 1000$ shooting moves on the new copy. In each shooting move the velocities were changed with a random vector distributed according to a gaussian with a width $dv_{max} = 2$. To conserve the linear (and angular) momentum we applied a relative velocity change along a particle pair vector[2] (Note that the angular momentum is not conserved anyway due to the periodic boundaries). Paths connecting the liquid $E_{pot} < -30k_BT$ and the gas phase as $E_{pot} > -5k_BT$ were accepted. Because the location of the barrier is always close to the liquid side of the

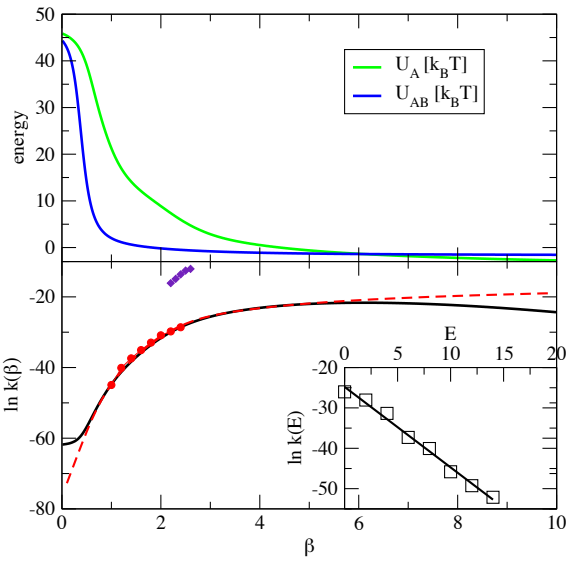

FIG. 2. (a) Path observables U_A and U_{AB} as function of β for a 3D LJ gas-liquid transition as obtained from two independent NTPS runs. (b) Logarithm of rate constant of 3D LJ gas-liquid condensation transitions obtained by direct NVT MD simulation (purple diamonds), and the NTPS rate (black curve). The red curve is the nested integral of Eq. 1, while the red circles indicate the canonically averaged $k(E)$ values using MD. The black NTPS is shifted to match the red curve at the reference temperature $\beta_0 = 2.4$. The inset gives the micro-canonical rate constant as obtained from SRTIS over the relevant energy range.

transition, whereas the gas side of the barrier creates a longer path, we employ the aimless shooting method[6], to avoid low acceptance. This method allows in each shooting move a shift of $-1, 0$, or 1 frame with respect to the previous shooting point. See Ref. [6] for more details.

We restrict the energy of the NVE paths to a maximum of $E = 48k_BT$ to avoid unphysically high energies that would result in integration errors. We obtained sufficiently decorrelated trajectories after $l = 1000$ shootings.

After the decorrelation phase, we apply the iterative nested sampling algorithm, using the same number of $l = 1000$ shooting trials. (We note that we probably overestimated the number of steps between samples in the calculation, since the gas phase space is quite easily traversed, and in fact requires fewer walkers. Using $l = 100$ we essentially obtain the same results). During the sampling we compute the cumulative density of states. We end when the total energy does not contribute any more to the temperatures of interest, which is the case around $E = -2k_BT$. This computation using NTPS took about 24 hours on a single Intel CPU core. As pointed out above, the paths to the gas phase can become relatively long, as the driving force to evaporation is not very high, especially not at low energy. From the cumulative density of states function we compute the average path (total) energy $U_{AB} \equiv \langle E \rangle_{AB}$ as a function of β via Eq. 5 in the Main Text, which is plotted in Fig. 2a,. The energy of the paths is high at low β but decreases quickly when $\beta > 0.5$, to reach around $E \approx -1$ for high β.

We then conduct a NTPS run in which only the initial state needs to be in the gas phase. Here we use uniform shooting with a fixed path length $\tau = 1000$ timesteps, as the end point is not defined, and all paths can be accepted as long as it starts in the gas phase. From the cumulative density of states we compute the energy $U_A \equiv \langle E \rangle_A$ via Eq. 5 in the Main Text, also shown in Fig. 2a. This computation took around 5 hours on a single Intel CPU core. As these paths are less restricted than the AB paths, the energy starts at higher E for low β, but of course stays below the upper limit of $E = 48k_BT$. Note that this is the opposite behaviour as for the examples in the Main Text, as there the unrestricted paths were drawn to lower energy by the potential energy wells. Returning to Fig. 2a, the total energy U_A then drops gradually, until at high β the energy drops even below the U_{AB} curve and ends at $E \approx -2.8k_BT$, which is still well above the gas stable state limit of $E_{pot} = -5k_BT$. So, while at high temperature the unrestricted paths are higher in energy than the restricted AB paths, at low temperature, the situation reverses, and becomes similar to the situation in the main text examples.

Next, we compute the derivative $d \ln k / d\beta$ from the path observables U_A and U_{AB}, using Eq. 12 in the Main Text. This derivative can be integrated to yield, up to a constant, the (logarithm of) the rate constant for condensation. This rate curve is shown in Fig. 2b, with the integration constant chosen such that the predicted rate constant curve matches with the rate constant computed via regular path sampling, as detailed below.

As the rate curve is only known up to an arbitrary (integration) constant, we need to compute a reference rate constant, just as in the Main Text for the 2D LJ cluster. We can perform straightforward MD at temperatures around $\beta = 2.4$, as there the transitions between the liquid and the gas occur spontaneously (See Fig. 1). We compute the mean first passage times from the dwell times for the gas to liquid transition, using the same stable state definition as for the path sampling, for temperatures $\beta = 2.1, 2.2, 2.3, 2.4, 2.5, 2.6$ and 2.7, by running straightforward NVT MD for 10^9 timesteps. For $\beta < 2.1$ we cannot get a direct estimate anymore, since the transition has become too unlikely, and the liquid droplet is no longer stable, or even metastable. These calculations took of the order of several hours on a single Intel CPU core, per temperature. The resulting rate constants are shown as purple diamonds in Fig. 2b.

However, the NTPS is done for constant energy path rather than constant temperature paths. As the paths can be relatively long, this does influence the rates significantly. A proper comparison with the NTPS results would require the computation of rates via direct MD with infinitely slow coupling to a heat bath. This turns out to be unfeasible. The direct rate at very low ther-

mostat coupling is simple too low to occur within the amount of reasonable wall clock time. (We estimate that direct computation of the rate constant at the required low temperature coupling would take on the order of a million CPU hours to see just a few events, even for such a small system.)

Therefore we compute the rates via regular TPS. To do so we employ the efficient single replica exchange transitin interface sampling (SRTIS) as detailed in Ref. [7]. We run SRTIS for several total energies between $E = 0$ and $E = 14k_BT$. The order parameter chosen is the potential energy E_{pot}. The interfaces were chosen between $E_{pot} = -5k_BT$ and $E_{pot} = -30k_BT$, with a regular spacing of $1k_BT$. This results in 25 interfaces and a minus interface for computing the flux (see [7] for more details). The move set for SRTIS consists of constrained shooting from the interface, single replica exchange moves, and time reversal moves. We used the standard Wang-Landau version of SRTIS. In total we performed 10^4 cycles of 100 moves of each type, which takes about a day on a single intel CPU. The SRTIS calculations result in an estimate of the micro-canonical rate $k(E)$, shown in the inset of Fig. 2b. However, to compare with the NTPS results we need the canonical rate $k(\beta)$. This can be obtained by taking the canonical average of the micro-canonical rate, using a Boltzmann weighted integral over configurations in stable state A,

$$k(\beta) = \langle k(E) \rangle_\beta = \frac{\int\ dE k(E) e^{-\beta E} \Omega_A(E)}{\int\ dE e^{-\beta E} \Omega_A(E)}. \qquad (1)$$

This average can be numerically computed as the time average of the predicted SRTIS $k(E)$ in a regular NVT MD in stable state A. The resulting $k(\beta)$ is plotted as red circles in Fig. 2b. The NTPS result are shifted to match these results. Alternatively, we could compute the canonical average using the NTPS results of state A, as the integral in Eq.1 is identical to a nested sampling integral. The resulting curve is plotted in Fig. 2b as a dashed red curve. This curve is an absolute rate prediction. The agreement between all three rate constant data sets is excellent, illustrating both the consistency and accuracy of the NTPS method. At very low β and very high β the agrement is less good, but this is to be expected. The low β limit does sample well above the energy scale that we sampled with SRTIS, while the high β limit suffers from extremely long NVE paths. Therefore, the fact that the predicted NTPS rate constant curve reaches a maximum around $\beta \approx 6$, where the two energy curves U_A and U_{AB} cross, might be a spurious result. The NTPS and SRTIS curves are probably also not trustworthy below $\beta \ll 1$ above due to the upper limit of $E = 48k_BT$ in the sampling. Nevertheless, the data shows that within relevant range of $1 < \beta < 5$ NTPS yield very accurate results.

Note also that the discrepancy between NVT and NVE rates is due to the small system size. In an NVT simulation, for $\beta \geq 2.5$, the gas-liquid transition ceases to be a rare event. The direct rate constants are then no longer determined by the dwell times in the states (gas or liquid) but are dominated by the transition times across the barrier. In other words, there is no separation of time scales, and in fact, no rare event dynamics. For the canonically averaged NVE simulation for $\beta \geq 2.5$ the behaviour is different. Here the gas-liquid transition is still a rare event. For large system sizes the difference between the ensembles will disappear, as then the internal degrees of freedom will play the role of termostat. One can thus gauge the importance of this difference by monitoring the instantaneous temperature along the NVE and NVT paths.

Note that for the formation of small liquid droplets in a dilute gas, as we study here, the NVE paths might actually be more accurate. Only when the liquid droplet can shed its heat somehow (e.g. by radiative cooling) that the temperature will be constant. However this process might be slow.

This example illustrates that NTPS is also capable of dealing with entropic transitions, where the rate constants are not determined by a potential energy barrier, but a free energy barrier. In that case the rate constant increases by lowering the temperature, as is also found in many nucleation transitions. Naturally, the high cost of the NTPS is only worth the investment if the rate of the process of interest is too slow to observe in direct MD. We therefore are confident that the method can be applied to larger systems, showing such transition, albeit at larger computational cost. We note that for this calculation we had to produce order 10^5 trajectories for the NTPS to converge. This is large, but not unfeasible for realistic systems, see e.g. Ref.[8, 9]. Moreover, the NTPS procedure can be sped up by a factor roughly of order K, by a parallelization scheme that evolves many walkers simultaneously. This is the subject of future research.

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science Series, Vol 1)* (Academic Press, 2001).

[2] C. P. Lowe, Europhys. Lett. **47**, 145 (1999).

[3] B. Smit, J. Chem. Phys. **96**, 8639 (1992).

[4] P. G. Bolhuis and C. Dellago, in *Reviews in Computational Chemistry*, Vol. 27 (John Wiley & Sons, Inc., 2010) pp. 111–210.

[5] P. G. Bolhuis and C. Dellago, Eur. Phys. J Special Topics **224**, 2409 (2015).

[6] B. Peters and B. L. Trout, J. Chem. Phys. **125**, 054108 (2006).

[7] W.-N. Du and P. G. Bolhuis, J. Chem. Phys. **139**, 044105 (2013).

[8] W. Du and P. G. Bolhuis, J. Chem. Phys. **140**, 195102 (2014).

[9] W. Du and P. G. Bolhuis, Biophys. J. **108**, 368 (2015).