

Supporting Information to: Rate Prediction for Homogeneous Nucleation of Methane Hydrate at Moderate Supersaturation Using Transition Interface Sampling

A. Arjun and P. G. Bolhuis*

*van 't Hoff Institute for Molecular Sciences, University of Amsterdam, PO Box 94157,
1090 GD Amsterdam, The Netherlands*

E-mail: p.g.bolhuis@uva.nl

Methods

TIS and RETIS simulation details

We performed transition interface sampling (TIS) simulations using OpenPathSampling^{1,2} to accurately predict the homogenous nucleation rate for methane hydrate at moderate undercooling of 280K and a pressure of 500 bar. The mutually coordinated Guest (MCG) order parameter was used to define interfaces along the nucleation process. The duration of a single dynamical transition path exhibiting the formation of the solid phase from the liquid at 280K and 500 bar pressure is in the order of 0.5 μ s.³ To increase the efficacy of the dynamics engine and limit the required storage space, we use a frame saving rate of 100ps for each interface. This means that during the molecular simulation, the snapshots of the system will be stored after every 100 ps of simulation. Since the average time of transition

from state A(liquid) to state B(solid) is hundreds of nanoseconds, there are still many frames (like 1000 frames for interface at MCG= 100) for each path. Note that the frame saving frequency does not affect the rate of transformation or the rate coefficient calculation.

The interfaces are defined by a discrete variable, MCG. Due to the nature of the implementation in OPS, the crossing criterion imposes that paths passing the interface, in fact, need to reach interface MCG+1, in order to be accepted. While slightly inconvenient, this discreteness does not influence the results. As an example, if a trajectory belongs to interface MCG = 60 then as per our implementation, that trajectory should have crossed MCG= 61 to be considered as an accepted path. In one sense we can view this that any interface's actual location is MCG+1 instead of the mentioned values in table I of the Main Text (MT).

Biased Shooting

For the TIS sampling we employed biased shooting, rather than the spring shooting used in Ref.³ Biased shooting is a algorithm that selects the shooting points close to the set interface.⁴ This bias yields a higher acceptance of the pathways, as the shot is likely to cross the interface. A higher acceptance is beneficial for the sampling and improves the decorrelation of the paths. We apply a gaussian bias function to the shooting point selection, which is centered around the interfaces, with a width determined by the acceptance. The selection probability is $P_{\text{sel}}(x) \propto \exp(-\alpha(\lambda(x) - l_0)^2)$, where $l_0 = \lambda_i$ refers to the position of

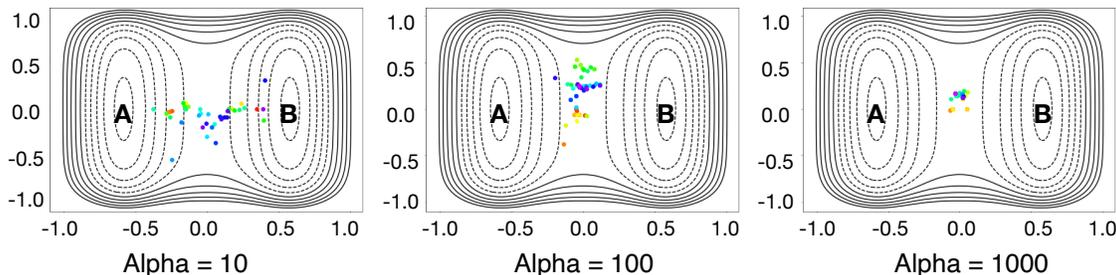


Figure S1: 2D toy model showing stable states A & B and the shooting points selected (different colors) with different values of α . The location of the interface is in the middle between A and B vertical to x -axis.

the interface i . The value of α determines the width of the bias. For a simple 2D potential shown in Fig. S1, the colored points show the spread of shooting points for different values of α . The applied bias is corrected for in the acceptance criterion.⁴ Note that this procedure has been implemented in OPS. In this work we set to $\alpha = 0.003$.

Preventing formation of a cylindrical gas reservoir

To make the path sampling simulation feasible, the system needs to be as small as possible. We chose the same system size as in Ref.³ For this system size there is a small chance that the spherical gas reservoir in the liquid state changes into a cylindrical shape across the periodic boundaries, which does alter the conditions dramatically. As this is an unphysical situation, we need to take care that this does not occur in the simulation. Note that we did not have this problem in Ref.,³ as there we halted the trajectory when reaching the liquid. In this work we do sample the liquid state substantially, especially in the minus ensemble(which is a part of the RETIS). To solve this problem, we reject each trial shooting move that leads to a trajectory with a final frame exhibiting a cylindrical gas reservoir. To identify the cylinder we project the water coordinates on a two-dimensional plane. In this two-dimensional projection of the cubic system, a cylindrical reservoir will show up as a (roughly) circular shape containing only methane gas and no water in the projected plane. In contrast, a spherical bubble will always have water molecules occurring throughout the plane. This is shown in Fig. S2 where the cylindrical bubble has a empty region in y-z plane. Indeed, a spherical bubble always has water molecules surrounding in all three dimensions. The presence of a empty cavity in the projected densities, indicative of the cylindrical gas reservoir, was tracked using a home-written script and implemented in OPS. Note that the rejection of the unphysical trajectory will not alter the validity of the conclusions.

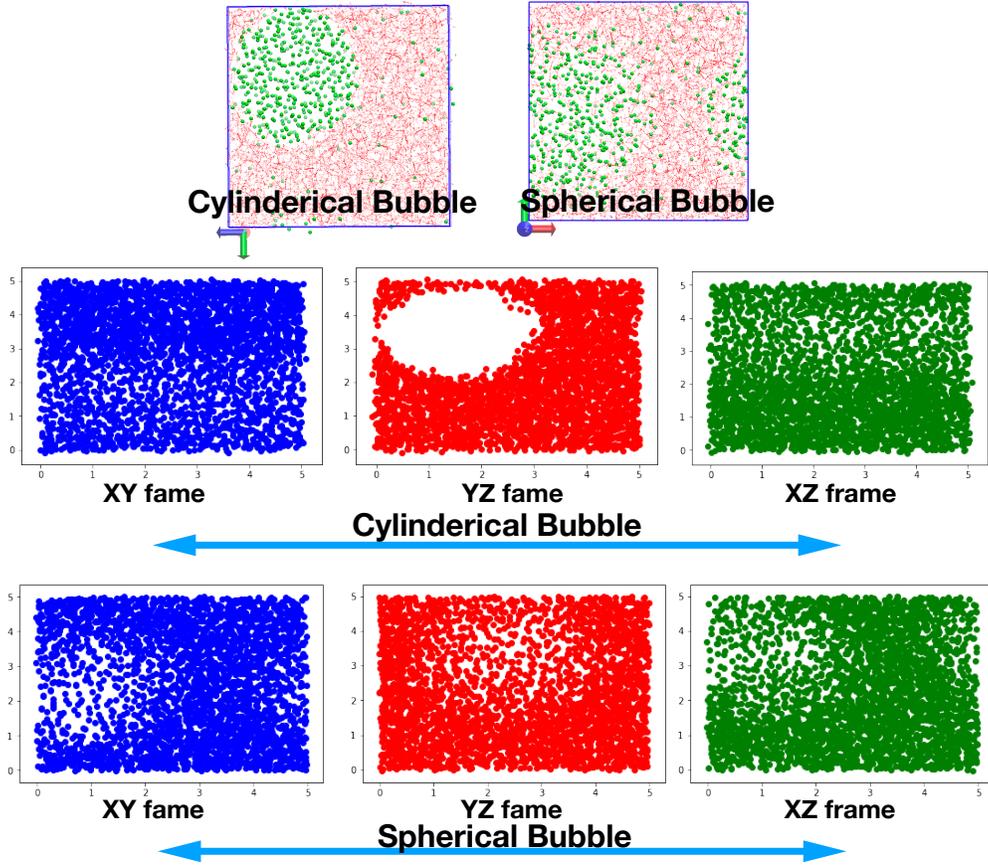


Figure S2: Top row: system snapshots of the initial liquid state in equilibrium with the gas reservoir. Cylindrical reservoir (left). Spherical reservoir (right). Middle and bottom row: The projection of water molecules on x-y, y-x and x-z planes (from left to right), are shown in blue, red and green respectively. The middle row shows the cylindrical reservoir projection and the bottom row the spherical reservoir projections.

Effects of the small system size

The system studied in this work is the smallest possible system able to contain three phases: a gas reservoir, a liquid phase and a growing solid nucleus. This small size was required to make the TIS rate computation feasible. However, an important check is whether the growing nucleus does not directly touch the spherical gas bubble, and that no significant size effect from the formation of a system spanning cluster exists, before the post-critical region is reached. Therefore, we provide in this section an analysis of the geometry of the growing cluster in the liquid phase, and of the simultaneously shrinking gas reservoir, in order to see whether there are significant size effects to be expected. To do so, we compute the following

average distances and sizes in the simulation box: 1) the radius of the spherical gas reservoir 2) the radius of the nucleus as determined by the MCG, 3) the distance between the surface of the bubble and the surface of the nucleus. 4) the distance between the surfaces of the nucleus and its periodic image. We present this data as a function of the nucleus size, as a progress variable for the nucleation.

The average nucleus radius is computed from the following relation

$$r = \left(\frac{MCG}{\frac{4}{3}\pi\rho_s} \right)^{1/3}, \tag{1}$$

where we assumed a spherical nucleus, and took the density $\rho_s = 4.57$ cages per nm^3 . Note that this is an approximation that is giving the average of the radius. Next, we compute the average distance between the cluster and its nearest periodic image as $d_{nuc-ni} = L - 2r$, where L is the box length. Next, we compute the bubble radius r_{bub} from the number of methanes in the gas reservoir and the methane density (based on the equation of state). This gives for the radius (in nm)

$$r_{bub} = 1.98 - 0.0012MCG. \tag{2}$$

Finally, the distance between the surface of the nucleus and the bubble is given by

$$d_{nuc-bub} = \sqrt{3}L - 2r_{bub} - 2r, \tag{3}$$

where we assumed that the center of mass of the bubble and nucleus are aligned along the diagonal of the box.

In Fig. S3 we plot these four properties as function of MCG. Several important observations can be made. First, the radius of the bubble hardly changes up to far beyond the critical nucleus size. This implies that also the Laplace pressure stays roughly constant, as we assumed during the analysis in the Results section. Further, the two distances $d_{nuc-bub}$

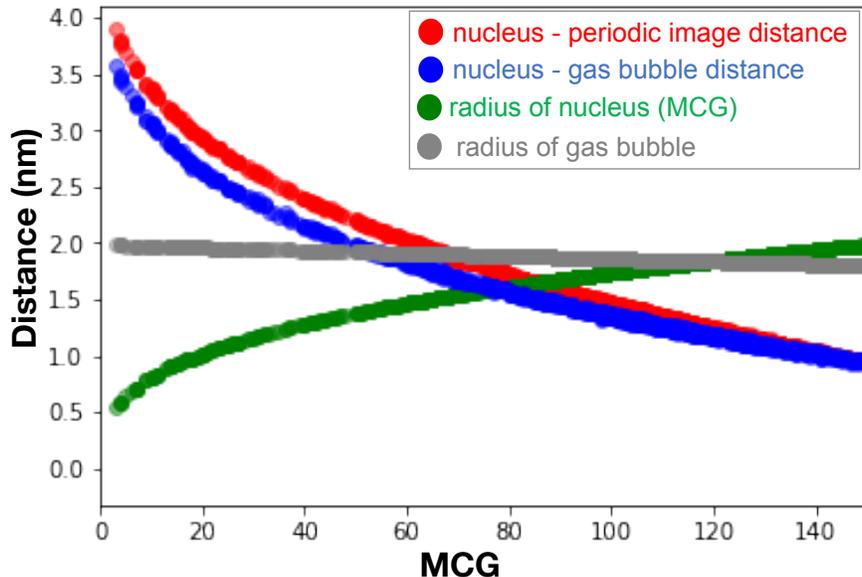


Figure S3: Average distance and radii as function of the MCG in the simulation box.

and d_{nuc-ni} drop quickly from roughly 4 nm in for MCG= 2 to around 1 nm for post-critical values. However, for critical values these distances are still higher than 1.2 nm, implying there is always a layer of liquid water between the two phases.

The conclusion is thus that the growing nucleus is not in contact with the gas reservoir, and there is always a layer of water wetting the solid surface up to post-critical nucleus sizes. Of course, the layer is not very thick, and we can expect fluctuations.

Driving force during nucleation

In the previous section, we showed that the radius of the spherical gas reservoir hardly changes along the nucleation process (till the critical nucleus), implying that the driving force remains more or less constant. Another way of determining the driving force along the nucleation process, is to measure the mole fraction / density of the solvated methane in the simulation box. We calculate this by dividing the solvated methanes in water (not part of MCG cluster) with the number of liquid water molecules (not engaged in the solid hydrate). This mole fraction is plotted as a function of growing cluster size (MCG) and time in Fig. S4. Although there is a slight downward trend, the mole fraction remains largely

constant up to the barrier region (MCG 90) and even beyond that. Plot D shows this as a function of time

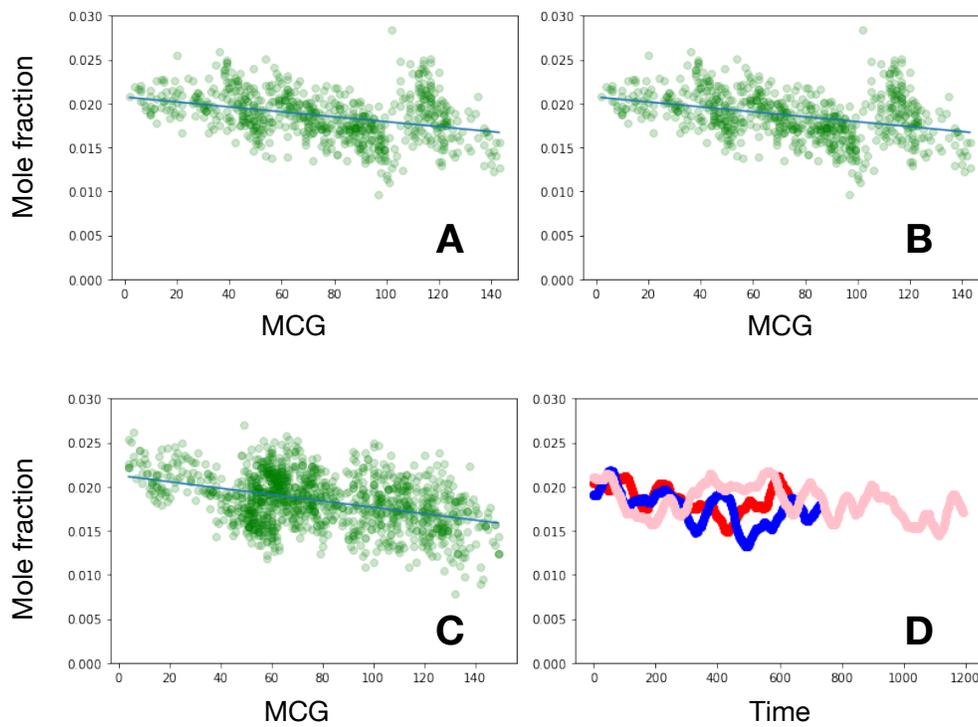


Figure S4: Mole fraction of the solvated methane during the nucleation process for three TIS trajectory (shown in A, B & C). The line shows the linear fit to the data. The running average (over 50 time steps) of the density for these three trajectory is shown in subplot D.

Results

Replica Exchange TIS

Replica exchange TIS (RETIS) was used to sample lower interfaces close to state A. RETIS has the benefit of the minus move, which helps in decorrelation of the pathways, and the replica exchange move, which helps in propagating decorrelated paths to higher interfaces. Since the paths in these initial ensembles are rather short compared to pathways belonging to higher interfaces, a sufficient degree of decorrelation is paramount. The path length histogram for the RETIS ensemble is shown in Fig. S5. Note that the path lengths are

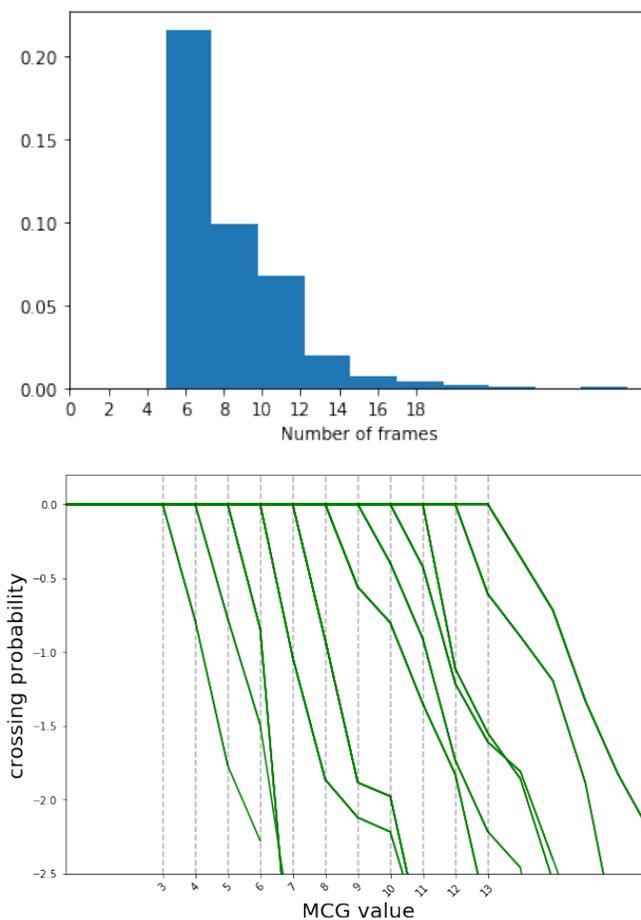


Figure S5: Replica exchange TIS results. Top: The distribution of path lengths in the minus ensemble. Bottom: Logarithmic plot of the crossing probabilities of the interfaces involved in the RETIS simulation.

much shorter than the ones for the higher interfaces. In the same figure we plot the crossing probabilities for each of the interfaces in the RETIS simulations. These crossing probabilities are only known at discrete MCG values, and are in fact very steep. Note that the results are clearly not perfectly converged, but are sufficiently accurate to be used in the WHAM procedure.

To check that RETIS can sufficiently decorrelate paths, we tracked the identity (i.e. the particle index) of the methane molecules belonging to the largest cluster in the MCG=13 ensemble (the last interface for the RETIS calculation). This is illustrated in Fig. S6. Clearly the identity of the cluster changes completely over the course of the sampling. This shows

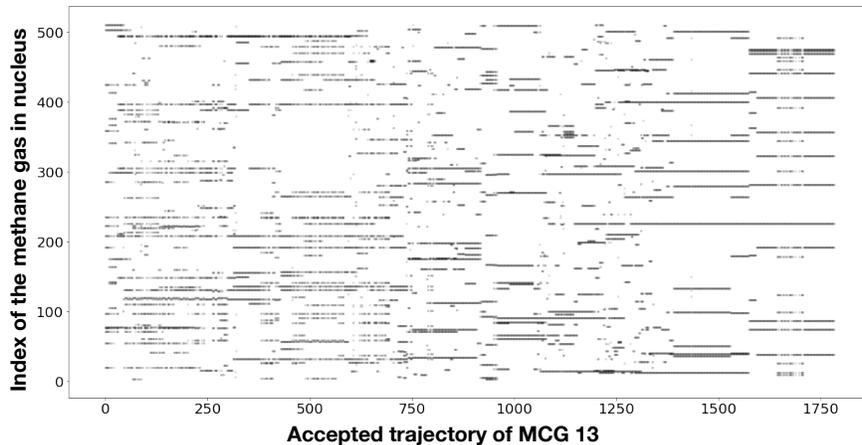


Figure S6: Decorelation of pathways belonging to Interface 13

that indeed the replica exchange move is enhancing decorrelation of clusters and thus the trajectories. As stated previously, this is an important aspect of the RETIS scheme and leads to sampling of a larger trajectory space.

Statistics for the TIS interfaces

Sampling interfaces

The acceptance, average path length, and number of decorrelated paths of the sampling across an interface in TIS are shown in Table I of the Main Text. For completeness, we also show partial path trees, shooting point locations, and path length distribution for selected interfaces in Fig. S7 and Fig. S8. The average path length increases with increasing interface value as shown in Fig 1 of the Main Text. The exception is interface MCG=100. Since MCG=100 is already beyond the maximum of the barrier, pathways can easily end in state B. Clearly this would mean that the AB paths are shorter than the AA pathways. This in turn would mean that the rate of growth in the post critical region is faster compared to precritical region, as also suggested in plot S7 in the supplementary information of Ref.³

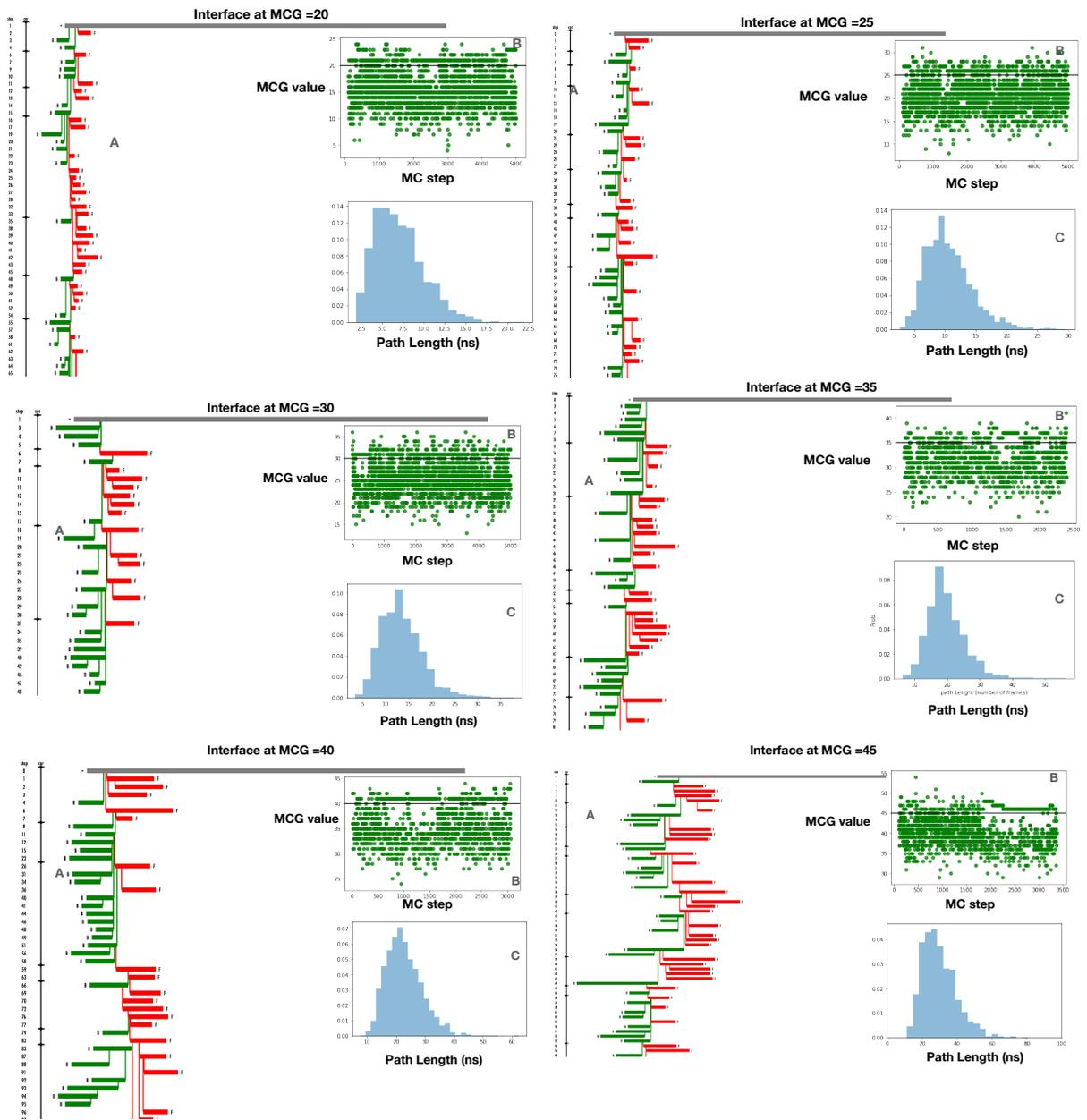


Figure S7: Details pertaining TIS simulation for Interface 20-45. A: Part of the path tree that shows successful paths. B: Selection of the shooting points based on gaussian shooting. The black horizontal line shows the position of the interface. x-axis and y-axis show MC trial number and MCG value, respectively. C: Path length histogram of the accepted paths.

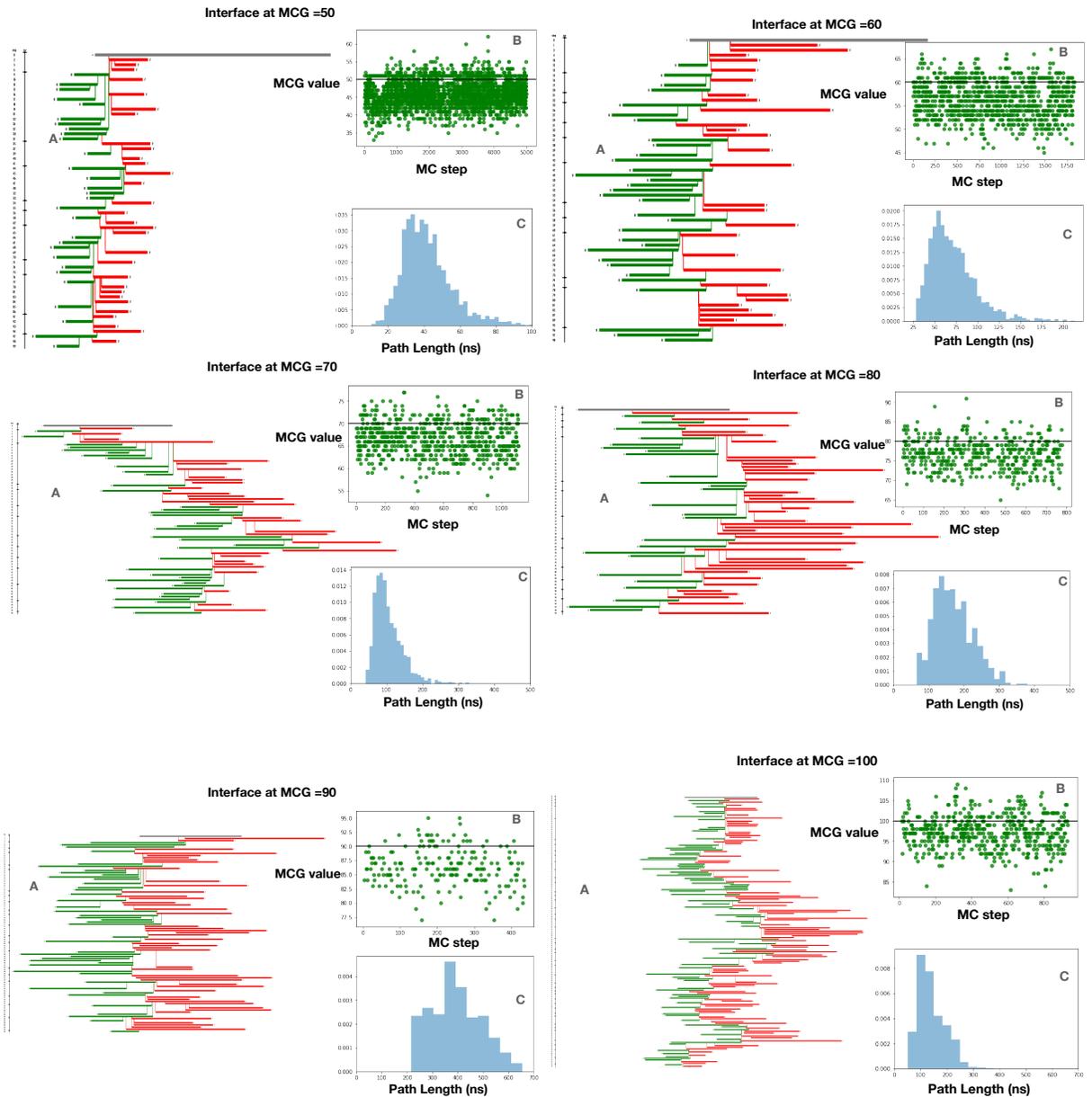


Figure S8: Details pertaining TIS simulation for Interface 50-100. A: Part of the path tree that shows successful paths. B: Selection of the shooting points based on gaussian shooting. The black horizontal line shows the position of the interface. x-axis and y-axis show MC trial number and MCG value, respectively. C: Path length histogram of the accepted paths.

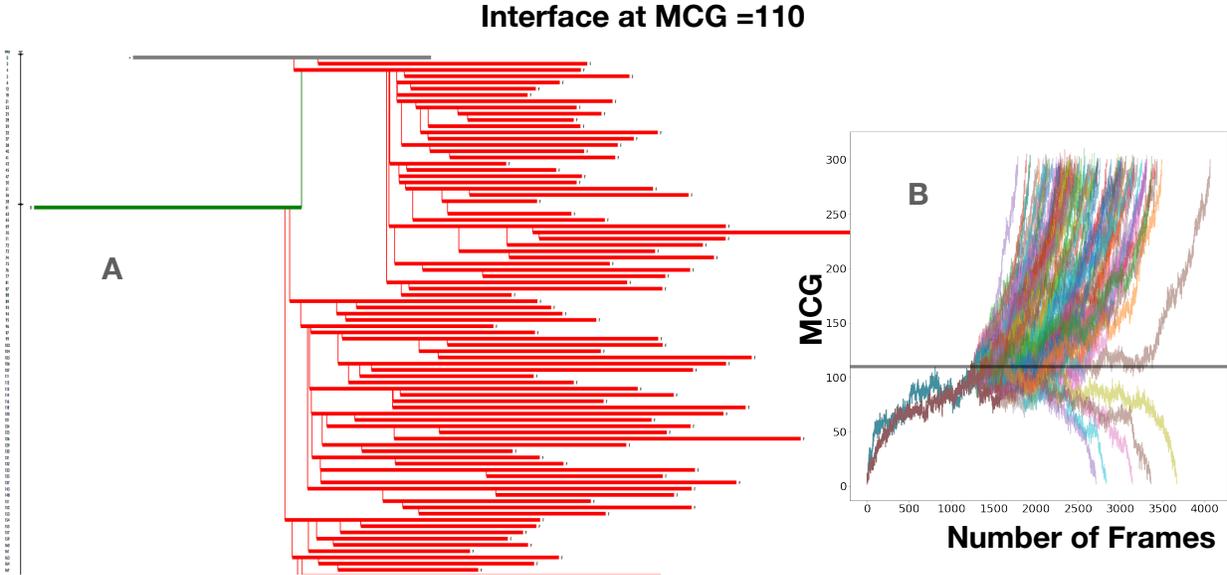


Figure S9: Sampling indicators for Interface 110. A: Path tree showing mostly forward paths (red lines). B: Accepted trajectories plotted along the order parameter showing most trajectories end up in state B.

Interface MCG= 110

As stated in the Main Text, the last interface for the TIS simulation was MCG=100. We tried to place an additional interface at MCG =110, but for this interface most trial pathways ended up in B, and backward moves were hardly accepted, see Fig. S9. This behavior indicates that the interface is located at post-critical cluster sizes. As the acceptance for valid $A \rightarrow A$ paths was extremely small, we did not include these interface results in the calculation of crossing probability. In any case the crossing probability is near unity for these MCG values.

Sampling the reverse process

To estimate the free energy profile for the nucleation, we can employ the total path ensemble, which can be obtained from the reweighted path ensemble.⁵ For that we require, besides the forward crossing probability discussed above, also the reverse crossing probability, which can be constructed from sampling the post-critical side of the nucleation barrier. To generate this data, we placed the interfaces at MCG= 105, 110, 120, 130 and 140, which are all post-

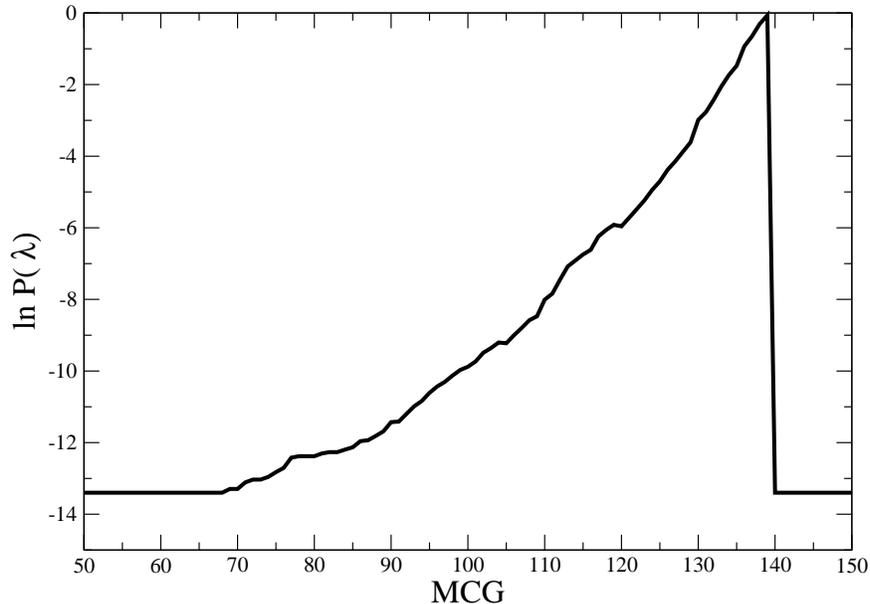


Figure S10: The reverse crossing probability to reach state A from state B. The black line shows the overall crossing probability joined using WHAM.

critical values. The resulting crossing probability data is shown in Fig. S10. Although this data is not as converged as the forward crossing probability used to calculate the nucleation rate (as shown in the Main Text) it gives a substantially good estimate to compute the reweighted path ensemble, and hence the free energy.

Crystallinity along the trajectories

As discussed in the Main Text, the cage ratio (CR) describes the extent of crystallinity or amorphousness of a frame or a pathway. We compute the cage ratio for all the frames on the accepted paths for the most important interfaces $MCG > 30$. Note that below $MCG=30$ there is no fully formed cage. We observe a gradual shift towards the crystalline domain, which is illustrated in figure 4 of the Main Text. The pathways showing liquid to solid transitions sampled at interface $MCG=100$ are plotted in Fig. S11. Clearly, pathways end up in the crystalline domain. We stress that our much longer TPS simulations showed transition between the crystalline and amorphous pathways in both directions.³

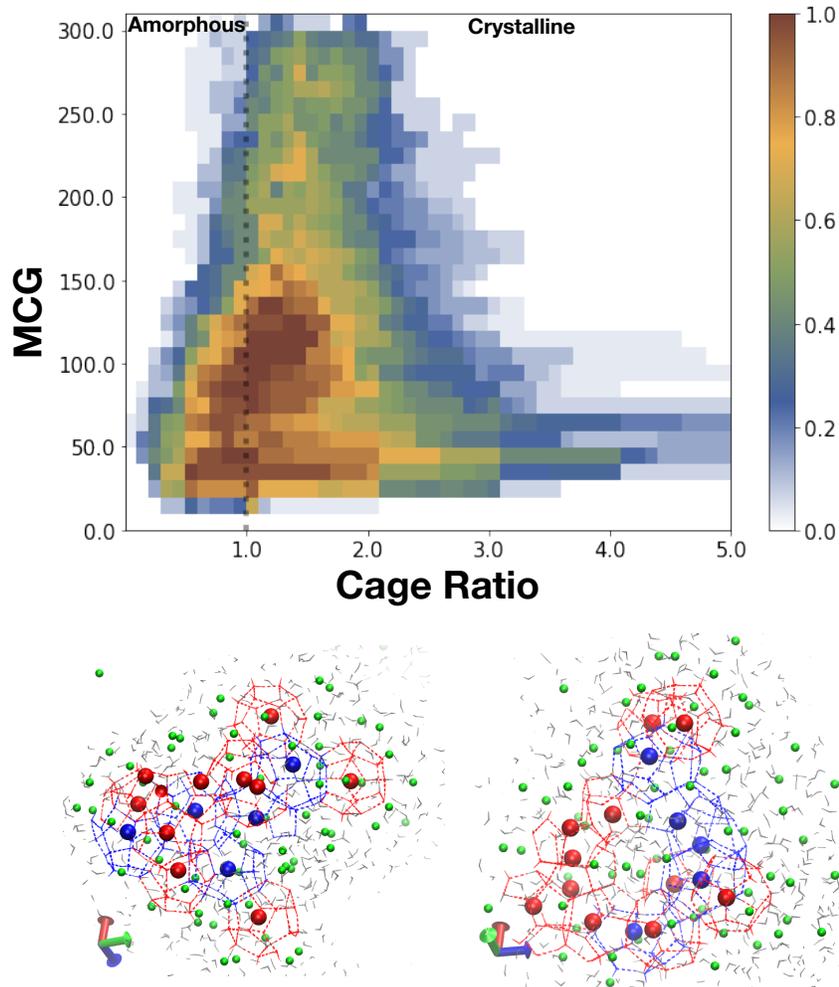


Figure S11: Above: Path density plot showing the liquid to solid transition paths sampled from interface $MCG=100$. 30 accepted pathways are plotted here. Almost all pathways end up in crystalline domain domain i.e $CR > 1$. Below: A typical snapshot of critical nucleus close to $MCG = 90$. 5^{12} and $5^{12}6^2$ cages are shown red and in blue respectively while methane in the MCG cluster with incomplete cages as green spheres. Waters around (5.5 \AA) the MCG is shown in grey.

Computation of rate and flux error

To compute the overall crossing probability curve, OPS applies the WHAM algorithm with a low cutoff value of 0.001 (which was 0.01 for the calculation in the main text). The statistical error in the rate constant is not immediately accessible as there is not enough data for block averaging, Nevertheless we can estimate the error using the bootstrapping approach⁶ for the crossing probability, and block averaging for the flux. This flux through the first interface

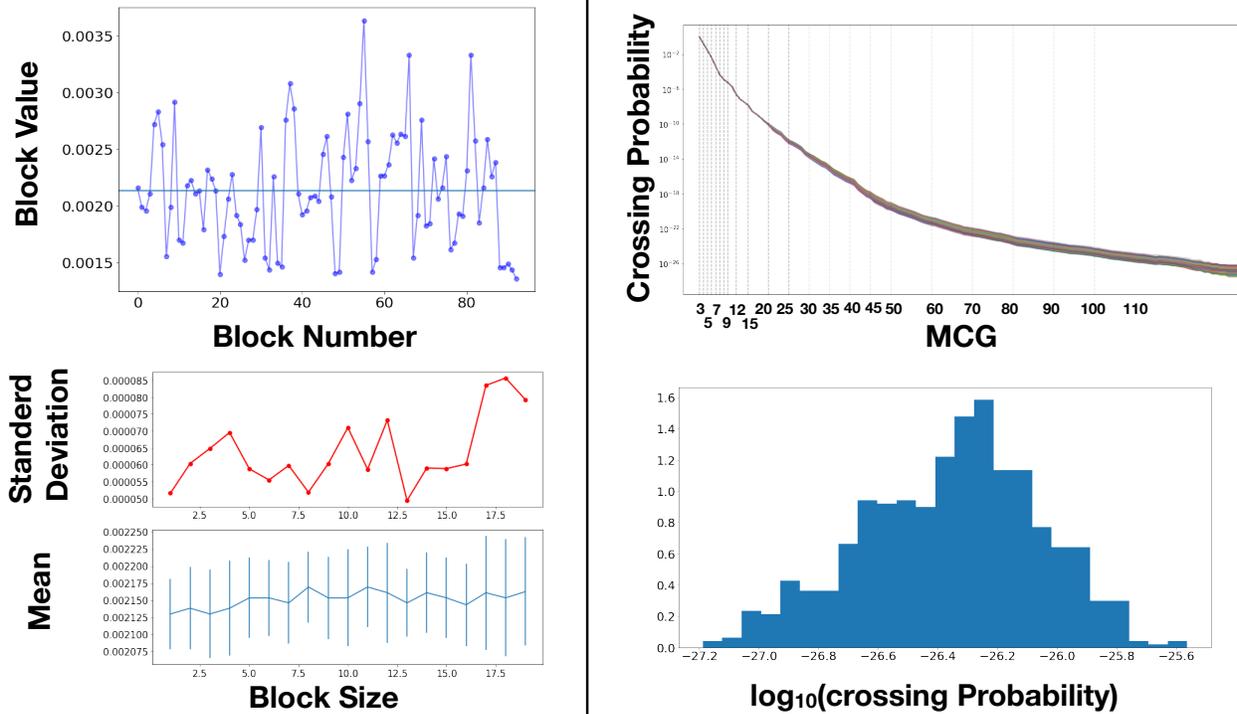


Figure S12: Error estimation in flux and rate calculation. Left column: error estimation in the flux using the block averaging method. RETIS simulation data is divided into 94 blocks and the flux is calculated in each block independently. This data is block averaged using a maximum block-size of 20, giving the mean and standard deviation in the blocks. Right column: the bootstrapping method applied to the crossing probability. The histogram shows the distribution in the recrossing probability.

was calculated using the 1182 minus moves (100% acceptance) in the RETIS (see Fig. S12). The flux was calculated in 2 separate RETIS simulations and the resulting value was also the same: 0.002 per picoseconds. In case of bootstrapping, each interface was resampled 1000 times and the resulting crossing probability was computed by WHAM method. The total crossing probability was between 10^{-27} to 10^{-25} as shown in the right column of Fig. S12. This is probably an underestimation of the error. We therefore take one order of magnitude as a lower boundary of the statistical error.

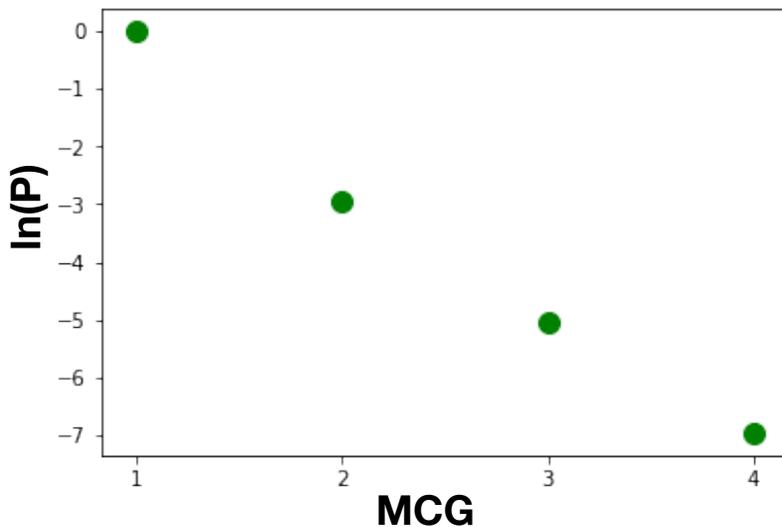


Figure S13: Logarithm of the population of MCG-based clusters of different sizes obtained from the MD simulation in the liquid phase.

Computation of the free energy profile

The free energy barrier profile for the hydrate nucleation as shown in Fig. 3 of the main text is computed from reweighting the path ensemble obtained by TIS using the WHAM weights for the crossing probabilities,⁵ and projecting the paths to the MCG variable. In this procedure the reverse path ensemble also was taken into account. To obtain the zero of the free energy curve, we perform histogramming of the cluster sizes from a straightforward MD simulation in the liquid phase. Analysing 40 ns of MD simulation gives for the first four cluster sizes (MCG=1-4) counts of 21289.0, 1106.0, 139.0, 20.0, respectively. The logarithm of the populations are plotted in Fig. S13. Note that the cluster population falls off exponentially, reflecting the steep barrier. Applying $G = k_B T \ln P$, and matching the free energy from reweighting leads then to the total free energy profile as shown in Fig. 3 of the MT.

The rate calculation for a different interface

The overall rate calculation involved estimation of the flux, which was done across the MCG = 3 interface. To check the consistency of our approach we can also compute the rate for

the next MCG interface at $MCG = 4$. From the RETIS simulation results, the flux through this interface is ($\Phi_{4,2} = 0.00049ps^{-1}$, one order of magnitude smaller than flux through $MCG = 3$, which is reasonable considering the steepness of the barrier in the initial part of the nucleation (see Fig. S5).

For the same interface $MCG = 4$ we estimate the crossing probability as $P(138, 4) = 4.74 \times 10^{-25}$. Using the same reasoning as in the Main Text, this translates into a nucleation rate of $2.091 \times 10^3 cm^{-1}s^{-1}$, which is within the error bar of the rate computation in the MT.

Calculation of diffusivity on top of the barrier

In the Main Text we compare the TIS based rate with the CNT expression for the nucleation rate based on the free energy. For this expression we need the diffusivity on the top of the barrier. The value of diffusivity follows from the mean squared displacement (MSD) along the MCG coordinate. Several trajectories from interface $MCG=100$ were selected, and truncated to focus on the barrier (i.e. in the range $80 < MCG < 100$). The MSD is shown in Fig. S14

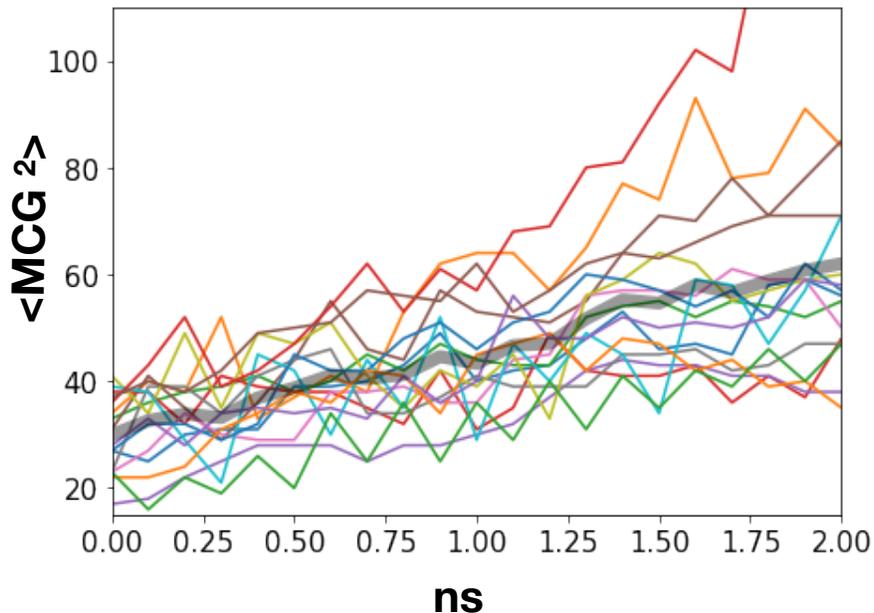


Figure S14: Different trajectories from interface $MCG=100$ used to compute the average MSD (shown in thick black line)

for each of these selected trajectories. We also plot the averaged MSD (thick black line). From the usual expression of the MSD, $\langle \Delta MCG(t)^2 \rangle = 2Dt$, with t the time and D , the diffusion constant, it follows that the diffusivity is $D = 7.8MCG^2/ns$.

Comparison to other rate calculations and CNT

In the Main Text we provide a discussion on the comparison of our TIS-based rate predictions with the different rate predictions for homogeneous methane hydrate nucleation from the literatures, and with experimental data. We also provided CNT predictions based on fitting to all the known rates done for the same system and pressure. Here we provide another CNT estimate, which is based on our previous TPS results.³

In Fig. S15 we show (again) the different rate predictions for homogeneous methane hydrate nucleation from the literature, and compare them to our results, and experimental data (light blue curve). We also include several hypothetical curves based on CNT scaling

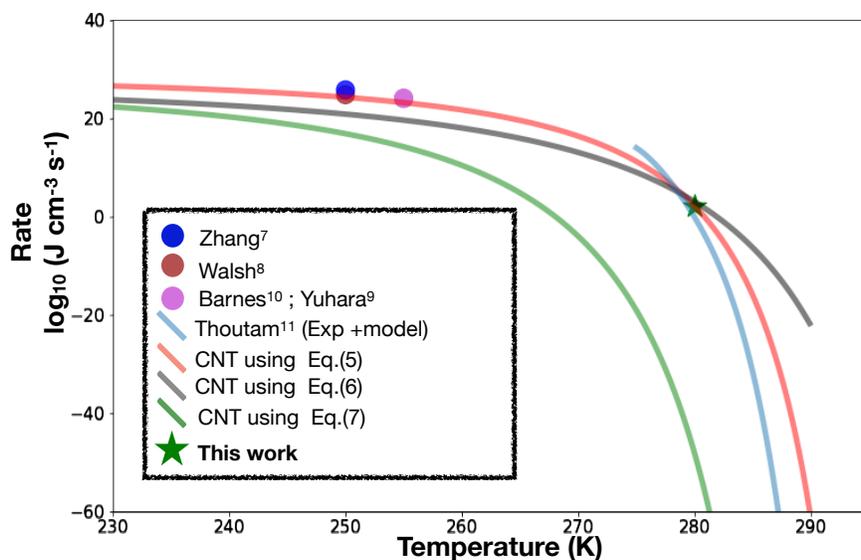


Figure S15: Graphical comparison of the logarithmic nucleation rate from literature^{7–11} and estimation from CNT. The circles corresponds to different rate calculation from simulation (for methane hydrate with same force field and pressure as discussed in the main text). The light blue curve is the experimental result from Ref.¹¹ The other solid curves present the CNT rate predictions using linear driving force (red curve), a linear fit to the TPS data of Ref³ (black curve) and the driving force corrected for the spherical interface (green curve)

for the log rate, given by

$$\ln J \sim (\Delta\mu)^{-2} + \text{const.} \quad (4)$$

The red curve is the CNT rate for the simple relation

$$\Delta\mu \sim T - T_m, \quad (5)$$

which is true only for a flat interface. The proportionality factor is obtained by matching this curve to the TIS predicted free energy and rate, hence this curve matches by definition our results (star symbol). This factor is dependent on both the entropy difference between the solid and the liquid and the corresponding surface tension. This curve matches the other rate predictions remarkably well, considering that in the TIS simulations and in most predictions there is not a flat gas-liquid interface.

We also include the black curve, which is based on a linear fit to the actual driving force in the simulation as giving by the critical nucleus size

$$\Delta\mu = \frac{2\gamma}{\rho_s r^*}. \quad (6)$$

with r^* , the radius of the crucial nucleus the γ surface tension of the liquid-crystal interface and ρ_s the density of the solid nucleus. We perform this fit by using the data from the TPS simulation.³ The CNT rate is again matched to match the TIS predicted free energy and rate at 280 K. Note that now this curve underestimates the literature rates at lower temperature.

Of course, a flat gas-liquid interface is much more realistic. Therefore, we also included the green curve, which corrects driving force for the spherical gas-liquid interface

$$\Delta\mu^{flat} = \Delta\mu - \Delta\Delta\mu. \quad (7)$$

This line is indeed substantially lowering the rate, as expected. However, it also shows that this estimated rate for a flat gas-liquid interface at 273K is still 10^{-12} nuclei $cm^{-3}s^{-1}$, or

10^{-6} nuclei $m^{-3}s^{-1}$, or 1 nucleus per second per million cubic meters.

This leads to the conclusion that homogenous nucleation is unlikely to play a significant role for moderate conditions with a flat interface, but that it cannot be ruled out. More measurements need to be done, with high precision, to make that distinction.

Supplementary videos

We provide supplementary videos that illustrate the reactive trajectories of transition from liquid phase to solid sI hydrate phase (MCG value of 2 to 300 in terms of TIS) sampled from the ensemble belonging to interface MCG =100. $5^{12}6^2$ and 5^{12} are tracked using blue and red color respectively . Water molecules and methane gas reservoir are not shown for clarity. The video is smoothed using time averaging over 10 frames(for water molecules) and 2 frames (for methane molecules) for visualization purposes.

References

- (1) Swenson, D. W. H.; Prinz, J.-H.; Noe, F.; Chodera, J. D.; Bolhuis, P. G. OpenPathSampling: A Python Framework for Path Sampling Simulations. 1. Basics. *Journal of Chemical Theory and Computation* **2018**, *15*, 813–836.
- (2) Swenson, D. W.; Prinz, J.-H.; Noe, F.; Chodera, J. D.; Bolhuis, P. G. OpenPathSampling: A Python Framework for Path Sampling Simulations. 2. Building and Customizing Path Ensembles and Sample Schemes. *Journal of chemical theory and computation* **2018**, *15*, 837–856.
- (3) Arjun,.; Berendsen, T. A.; Bolhuis, P. G. Unbiased atomistic insight in the competing nucleation mechanisms of methane hydrates. *Proceedings of the National Academy of Sciences* **2019**, *116*, 19305–19310.
- (4) Bolhuis, P. G.; Dellago, C. *Reviews in Computational Chemistry*; John Wiley & Sons, Inc., 2010; pp 111–210.
- (5) Rogal, J.; Lechner, W.; Juraszek, J.; Ensing, B.; Bolhuis, P. G. The reweighted path ensemble. *The Journal of chemical physics* **2010**, *133*, 174109.
- (6) Friedman, J.; Hastie, T.; Tibshirani, R. *The elements of statistical learning*; Springer series in statistics New York, 2001; Vol. 1.
- (7) Zhang, Z.; Liu, C.-J.; Walsh, M. R.; Guo, G.-J. Effects of ensembles on methane hydrate nucleation kinetics. *Physical Chemistry Chemical Physics* **2016**, *18*, 15602–15608.
- (8) Walsh, M. R.; Beckham, G. T.; Koh, C. A.; Sloan, E. D.; Wu, D. T.; Sum, A. K. Methane hydrate nucleation rates from molecular dynamics simulations: effects of aqueous methane concentration, interfacial curvature, and system size. *The Journal of Physical Chemistry C* **2011**, *115*, 21241–21248.

- (9) Yuhara, D.; Barnes, B. C.; Suh, D.; Knott, B. C.; Beckham, G. T.; Yasuoka, K.; Wu, D. T.; Sum, A. K. Nucleation rate analysis of methane hydrate from molecular dynamics simulations. *Faraday discussions* **2015**, *179*, 463–474.
- (10) Barnes, B. C.; Knott, B. C.; Beckham, G. T.; Wu, D. T.; Sum, A. K. Reaction coordinate of incipient methane clathrate hydrate nucleation. *The Journal of Physical Chemistry B* **2014**, *118*, 13236–13243.
- (11) Thoutam, P.; Rezaei Gomari, S.; Ahmad, F.; Islam, M. Comparative Analysis of Hydrate Nucleation for Methane and Carbon Dioxide. *Molecules* **2019**, *24*, 1055.