ABSTRACT: The crystallization of methane hydrates via homogeneous nucleation under natural, moderate conditions is of both industrial and scientific relevance, yet still poorly understood. Predicting the nucleation rates at such conditions is notoriously difficult due to high nucleation barriers, and requires, besides an accurate molecular model, enhanced sampling. Here, we apply the transition interface sampling technique, which efficiently computes the exact rate of nucleation by generating ensembles of unbiased dynamical trajectories crossing predefined interfaces located between the stable states. Using an accurate atomistic force field and focusing on specific conditions of 280 K and 500 bar, we compute for nucleation directly into the sI crystal phase at a rate of $\sim 10^{-17}$ nuclei per nanosecond per simulation volume or $\sim 10^2$ nuclei per second per cm$^3$, in agreement with consensus estimates for nearby conditions. As this is most likely fortuitous, we discuss the causes of the large differences between our results and previous simulation studies. Our work shows that it is now possible to compute rates for methane hydrates at moderate supersaturation, without relying on any assumptions other than the force field.

INTRODUCTION

Mixtures of methane gas and water can spontaneously form a solid hydrate at low temperatures and/or high pressures due to hydrophobic interactions$^3$ in which the gas molecule stabilizes the encapsulating water cage.$^2$ Methane hydrates naturally occur in abundance at the ocean floors and in permafrost, exceeding the natural gas reserve substantially.$^3$ As such, methane hydrates are envisioned not only as a future energy resource$^2,5,7$ but also as very relevant for the stability of ocean floors$^8$ and global climate change. In addition, interest in methane hydrates follows from the possible design of inhibitors that prevent hydrate formation in industrial pipelines.$^8$ A better understanding of the molecular mechanism of methane hydrate formation will be of interest to a large scientific audience.

Under moderate conditions, hydrates form via a nucleation and growth mechanism.$^3$ Such a hydrate formation mechanism can be understood within the framework of the widely used classical nucleation theory (CNT).$^{10}$ This theory postulates a spherical solid nucleus growing in a metastable liquid phase. The creation of a solid–liquid interface is unfavorable due to surface energy until the growing nucleus reaches a certain critical size to overcome the activation free energy barrier. From then on, the driving force toward the solid is large enough to allow spontaneous growth into a bulk crystalline phase. The thermodynamically stable crystal phase for methane hydrate is called structure type I (sI) and is composed of two standard methane cages in a ratio of 3:1. These cages are characterized by how their hydration shell is geometrically organized. The majority cage type $5^{12}6^2$ is built from 12 pentagons and 2 hexagons, while the minority cage type $5^{12}$ consists of 12 pentagons only. While the sI form is the most abundant in nature, methane hydrates can also form via a nonclassical mechanism in which an amorphous metastable intermediate (composed of nonstandard cages and with many fewer $5^{12}6^2$ cages) precedes the formation of a crystalline state. Under natural conditions, hydrate formation will be mostly dominated by heterogeneous nucleation. Still, as a first step, it is essential to understand the homogeneous process.$^{11}$

CNT successfully models homogeneous nucleation for many systems, not only giving a general expression for the barrier height, but also for the nucleation rate, defined as the number of nucleation events per unit volume and time. This value scales exponentially with the free energy barrier. Notwithstanding the success of CNT, a detailed understanding of the molecular mechanism of methane hydrate nucleation is still lacking, as current experimental analysis techniques are limited in spatiotemporal resolution and, hence, cannot directly give atomistic insight (although we note that recently breakthrough experiments were able to glean several nucleation features in other systems such as nanoparticles$^{11,12,13}$).
Such insight into the molecular motions is fundamental in understanding and eventually controlling the kinetics and thermodynamics of the nucleation and growth process. Direct molecular dynamics (MD) simulations using atomistic force fields could in principle provide this knowledge and predict thermodynamic and kinetic properties. However, the current computer hardware can only access μs–ms timescales even for a relatively small simulation volume. At moderate undercooling, where an experiment would still be able to observe homogeneous nucleation, the free energy barrier for nucleation is already large, leading to very long induction times in the small simulation volume and, thus, rendering straightforward MD utterly unfeasible. For instance, overcoming a barrier of roughly 50 kBT would already take billions of years of (wall clock) computer time. For a higher degree of undercooling or supersaturation, the driving force toward nucleation becomes stronger. That is why many simulation studies are performed at lower temperatures or higher supersaturation so that spontaneously nucleation is observed within a realistic simulation time.214–22 At this high driving force, the nucleation process might be significantly different from the nucleation at higher temperatures, as recent experiments indicated.23 As an illustration, simulations often find methane hydrates nucleating into amorphous phases, instead of the more thermodynamically stable sI phase. The formation of an amorphous phase is rationalized by Oswald’s step rule,24 which postulates the formation of an intermediate metastable phase with a different geometry before the thermodynamically stable state is reached.

Fortunately, nucleation of methane hydrates at moderate supersaturation can also be studied using enhanced sampling techniques and/or coarse-grained water models. For instance, Bi et al.25 conducted forward flux sampling (FFS) simulations at a strong undercooling (T = 220 K), where they found the sII clathrate structure as the most dominant phase. Lauricella et al.26 also reported sII clathrates based on a metadynamics study at moderate undercooling (T = 273 K). Later, DeFever and Sarupria18 employed FFS to find (using a water-soluble guest at T = 230 K) the formation of an amorphous solid. In all of the above studies, the formation of the thermodynamically stable phase sI was never observed. Also, note that all of these studies used a coarse-grained model of water.27

Nucleation rate predictions for methane hydrates vary widely between 2 × 10−24 nuclei cm−3 s−1 (T = 220 K, P = 500 bar, using FFS)17 and 3 × 10−111 nuclei cm−3 s−1 (T = 273 K, P = 900 atm, using CNT with computed thermodynamic properties)28, more than 132 orders of magnitude difference. Experimentally established nucleation rates range between 3 × 105 and 3 × 10−7 nucleus cm−3 s−1 for various conditions.29–37 Clearly, predicting accurate nucleation rates is notoriously difficult. Amongst others, the origins of the discrepancies could be due to the force field (atomistic or coarse grained), the choice of the collective variable (e.g., nucleus size metric)38 in the enhanced sampling methods, system size, specific assumptions (e.g., which phase is formed), and of course thermodynamic conditions. Therefore, one would prefer to use an enhanced sampling method making the least amount of assumptions. Transition path sampling39 is among those methods.

In this work, we apply transition interface sampling (TIS)40 to compute the nucleation rate for realistic undercooling41 (T = 280 K) at a relevant pressure (P = 500 bar), employing an accurate atomistic force field (Tip4P/ICE).42 Recently, we performed extensive transition path sampling (TPS)13 simulations on hydrate nucleation in a methane/water mixture with an sI stoichiometry at moderate temperatures between 270 and 285 K at a pressure of 500 bar using the same system setup as in refs 2, 19, 21, 22. The melting point for the model used is 303 ± 2 K.43 In the metastable liquid state, the system is phase separated into a (super)saturated water–methane mixture in equilibrium with a (spherical) methane gas reservoir. The supersaturation is moderate, resulting in a large nucleation barrier. TPS circumvents the long induction times involved in the nucleation process while retaining the advantage of the unbiased dynamics by sampling an ensemble of MD trajectories connecting the liquid and (amorphous or crystalline) solid states. Analysis of this path ensemble showed that the nucleation pathways switch from forming only amorphous solid hydrates to a crystallization mechanism ending in the sI state. Interestingly, at 280 K, both mechanisms can coexist. Now, while TPS does not directly give access to the nucleation rate nor the free energy barrier, the efficient TIS extension of TPS is capable of doing so.40 TIS computes the flux through a series of interfaces between the liquid and solid states, yielding the total rate constant. The TIS algorithm has been applied before to crystallization in Lennard-Jones models45 and colloidal systems46 and was recently employed to study the solidification of Ni.47 In general, path sampling techniques are more robust to estimate rates, avoiding any approximations made by CNT. For instance, the rate of nucleation of ice from water (a process related to hydrate formation) was rigorously measured by sampling the pathways along the nucleation barrier.48

Here, we compute the nucleation rate of methane hydrate to be ~102 nuclei per second per cm3, for a moderate undercooling of 280 K and a pressure of 500 bar, in rough agreement with the known experimental values. Knott et al. estimated almost 120 orders of magnitude lower rate under nearby conditions,28 which essentially ruled out the presence of homogeneous nucleation due to the very high free energy barrier obtained. We argue that this discrepancy can be explained to a large extent by the difference in the system setup used in the simulations. Our findings suggest that homogeneous nucleation, while unlikely to be realistic under natural conditions, could still be relevant for moderate undercooling and high pressure. Most importantly, this study shows that it is now possible to obtain accurate crystal nucleation rates at moderate supersaturation and, thus, for a high free energy barrier, without influence due to the choice of the progress variable, and without any assumption, except for the force field.

The remainder of this article is organized as follows. In Methods section, we introduce the methods and system setup. In Results and Discussion section, we discuss the results, followed by conclusions in the section Conclusions.

# METHODS

Transition Interface Sampling. TIS belongs to the TPS family of Monte Carlo-based algorithms for creating ensembles in trajectory space.49 The TIS algorithm was specifically developed to compute the kinetic rate constants for rare events more efficiently than the original TPS algorithm, which required a slow transformation of an unrestricted path ensemble into the rare event path ensemble.50 To do so, TIS introduces a set of n nonintersecting interfaces, placed along a reasonable order parameter or collective variable capable of separating the stable states. In the TIS framework, the rate

https://dx.doi.org/10.1021/acs.jpcb.0c04582
constant is viewed as the flux of trajectories through the final interface 0 (equivalent to state B), which in turn can be written as the product of the flux through the first interface, and the so-called crossing probability

\[ k_{AB} = \Phi_{i,0} = \Phi_{1,0} P_i(n1) = \Phi_{1,0} \prod_{i=1}^{n-1} P_A(i + 1\mid i) \]  

Here, interface 0 defines state A and interface \( n \) delineates the boundary of state B. In the final equation, the rate is thus composed of two ingredients. The first factor \( \Phi_{i,0} \) is the flux of trajectories through the first interface, the one closest to A. The second factor \( P_A(i + 1\mid i) \) is the probability that trajectories coming from state A and crossing interface \( i \) are able to reach interface \( i + 1 \). For each interface \( i \), path sampling is used to estimate this crossing probability \( P_A(i + 1\mid i) \) under the conditions that all trajectories cross the interface \( i \) and come directly from the initial state A. To generate new pathways, a shooting algorithm similar to standard TPS creates forward and backward trial MD trajectories from a randomly selected frame, which can be accepted based on a Metropolis criterion. For efficiency, TIS allows flexible lengths of pathways, which can be halted as soon as the paths reach a stable state or cross the required interface. We apply the Gaussian biased shooting scheme, which selects shooting points close to the interface leading to higher acceptance of paths. In TIS, the location of the interfaces is an important aspect of sampling. Interfaces should neither be too close nor too far from each other. Hence, the location of the interfaces needs to be optimized, usually by trial and error. This makes it hard to parallelize the methods. Therefore, we use a slightly different TIS implementation, where all trajectories are continued until they reach stable state A or B. Thus, all paths are either leading from A to A or from A to B. Note that this is the standard implementation in the OpenPathSampling (OPS) software package. The general scheme of the TIS implementation is illustrated in Figure 1.

To enhance the decorrelation of pathways, the replica exchange TIS algorithm (RETIS) allows for the exchange of paths between neighboring interfaces. In combination with a move that samples the initial state, RETIS enhances the path sampling of the path space tremendously.

**System Setup, Force Field, and Simulation Details.** Our system consists of a cubic box with 2944 water and 512 methane molecules. Water was represented by the TIP4P/Ice model, and methane was modeled using united atom Lennard-Jones interactions (\( \epsilon = 1.22927 \text{ kJ mol}^{-1} \) and \( \sigma = 3.700 \text{ Å} \)). This combination has been shown previously to mimic experimentally determined properties very well. Most of the molecular dynamics (MD) simulations were performed using OpenMM 7.1.1. The Velocity Verlet with the velocity randomization (VVVR) integrator (from openmmtools) was used to integrate the equations of motion. The integration time step was set to 2 fs. The van der Waals cutoff distance was 1 nm. Long-range interactions were handled by the Particle Mesh Ewald technique. The MD
simulations were performed in the NPT ensemble using the VVVR thermostat (frequency of 1/ps) and a Monte Carlo barostat (frequency of 4 ps). TIS simulations were performed using the CUDA platform of OpenMM on NVIDIA GeForce GTX TITAN 1080Ti GPUs. TIS was executed using the OpenPathSampling package.

**Order Parameters and Collective Variables. Nucleus Size Parameter.** We employ the size of the largest solid cluster as the order parameter to place the interfaces. The size of the nucleus has been used previously in simulation studies\(^7,47,48\) to calculate the rate of nucleation. The mutually coordinated guest (MCG) order parameter\(^60\) counts the number of methane molecules involved in the largest solid nucleus in the nucleating hydrate system. Each methane molecule (guest molecule) is checked whether its neighboring (methane and water) molecules satisfy a set of geometric constraints. If so, then methane is an MCG monomer. Neighboring MCG monomers are part of the same cluster. The largest connected cluster in the system is then identified using a cluster algorithm. The MCG order parameter is defined as the size of this largest (solid) cluster. Here, we use MCG-1 (and refer it to as MCG) as it checks for any possible occurrence of nucleus formation compared to MCG-3, which only identifies the stable nucleus.\(^60\) We determined the MCG using a home-written analysis code.

**Cage Type Analysis.** The structure of the growing nucleus can be identified by cage types that form it. We analyze the cage type for each methane in the MCG-based cluster using an algorithm similar to the one employed in ref\(^61\), using a home-written code. We identified seven main types of cage structures, namely, 5\(^4\), 5\(^6\)\(^2\), 5\(^6\)\(^3\), 6\(^3\)\(^6\), 4\(^3\)\(^6\), 4\(^5\)\(^6\), and 4\(^7\)\(^6\), where the superscript indicates the number of polygons made by the hydrogen-bonded water molecules in the cage facet. The base number gives the type of polygon (4: square, 5: pentagon, 6: hexagon). The ratio of the number of cage facets to the number of polygons involved in the largest solid nucleus is then identified using an MCG monomer. Neighboring MCG monomers are part of the same cluster. The largest connected cluster in the system is then identified using a cluster algorithm. The MCG order parameter is defined as the size of this largest (solid) cluster. Here, we use MCG-1 (and refer it to as MCG) as it checks for any possible occurrence of nucleus formation compared to MCG-3, which only identifies the stable nucleus.\(^60\) We determined the MCG using a home-written analysis code.

Table 1. Sampling Details for 24 Interfaces\(^a\)

<table>
<thead>
<tr>
<th>MCG</th>
<th>&lt;path length&gt; (ns)</th>
<th>acceptance</th>
<th>decorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.35</td>
<td>2097/3174 (66.0)</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>0.39</td>
<td>2162/3047 (70.9)</td>
<td>94</td>
</tr>
<tr>
<td>5</td>
<td>0.42</td>
<td>2099/3105 (64.7)</td>
<td>62</td>
</tr>
<tr>
<td>6</td>
<td>0.43</td>
<td>2039/3071 (66.0)</td>
<td>26</td>
</tr>
<tr>
<td>7</td>
<td>0.46</td>
<td>1937/3032 (63.8)</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>0.76</td>
<td>1698/3029 (56.0)</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>0.85</td>
<td>1756/3207 (54.7)</td>
<td>105</td>
</tr>
<tr>
<td>10</td>
<td>0.87</td>
<td>1682/3111 (54.0)</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>0.86</td>
<td>1711/3120 (54.8)</td>
<td>101</td>
</tr>
<tr>
<td>12</td>
<td>1.41</td>
<td>1539/3091 (49.7)</td>
<td>95</td>
</tr>
<tr>
<td>13</td>
<td>2.15</td>
<td>1577/3134 (50.3)</td>
<td>128</td>
</tr>
<tr>
<td>14</td>
<td>3.62</td>
<td>2767/5000 (55.3)</td>
<td>221</td>
</tr>
<tr>
<td>20</td>
<td>7.12</td>
<td>2684/5000 (53.6)</td>
<td>218</td>
</tr>
<tr>
<td>25</td>
<td>10.62</td>
<td>2676/5000 (53.5)</td>
<td>258</td>
</tr>
<tr>
<td>30</td>
<td>13.4</td>
<td>2838/5000 (56.7)</td>
<td>265</td>
</tr>
<tr>
<td>35</td>
<td>19.79</td>
<td>1350/2399 (56.2)</td>
<td>161</td>
</tr>
<tr>
<td>40</td>
<td>22.8</td>
<td>1659/3060 (54.2)</td>
<td>102</td>
</tr>
<tr>
<td>45</td>
<td>29.9</td>
<td>2150/3625 (59.3)</td>
<td>176</td>
</tr>
<tr>
<td>50</td>
<td>42.2</td>
<td>2897/5000 (57.9)</td>
<td>407</td>
</tr>
<tr>
<td>60</td>
<td>71.2</td>
<td>1094/1830 (59.7)</td>
<td>195</td>
</tr>
<tr>
<td>70</td>
<td>105.8</td>
<td>721/1199 (64.3)</td>
<td>145</td>
</tr>
<tr>
<td>80</td>
<td>169.4</td>
<td>518/838 (61.8)</td>
<td>116</td>
</tr>
<tr>
<td>90</td>
<td>197.30</td>
<td>182/354 (51.4)</td>
<td>48</td>
</tr>
<tr>
<td>100</td>
<td>140.9</td>
<td>541/968 (55.8)</td>
<td>113</td>
</tr>
</tbody>
</table>

**TIS and RETIS Simulation Details.** The MCG order parameter acts as a progress variable to define the interfaces along the nucleation process. Even though the use of path sampling speeds up the sampling exponentially, the average duration (length) for a full transition path from liquid-to-solid at 280 K and 500 bar pressure is in the order of a few hundred nanoseconds. This makes the path sampling computationally expensive, requiring the use of GPUs. Due to I/O and storage size limitations, we use a saving frequency of 100 ps, i.e., during the MD simulation, the snapshots of the system are stored only once every 100 ps. This relatively low saving frequency should not affect the mechanism or the rate constant, as the average transition time is already hundreds of nanoseconds, and the residence time is many orders of magnitude longer. Indeed, the resulting rate from TIS is independent of saving frequency, as a lower flux, due to missed interface crossings for lower frequencies, is compensated by an increase in the crossing probability for the final interface. Note that the low saving frequency might lead to large jumps in the order parameter from one frame to the next. The effect of such jumpiness has been studied in the context of FFS in ref\(^64\).

The interfaces for regular TIS were set at 13 different locations along the MCG order parameter, as given in Table 1. The choice for these values was based on the requirement for a sufficient overlap of the crossing histograms from one interface to the next. That is, paths that are sampled in interface \(i\) should have a reasonable chance of reaching interface \(i + 1\). The rule of thumb for TIS is that this probability should be roughly
Information. Table 1 gives the TIS sampling details in terms of \( \alpha \) centered around the interface value, with a width parameter of \( \sigma \). The shooting points were selected from a Gaussian distribution in the MCG variable, one-way shooting move, where the shooting points were located in the nucleus. Blue and red cage represent 512\( \times 62 \) and 512\( \times 62 \), respectively. Methane gas trapped inside is also shown in the same colors. The green spheres indicate interface locations. The horizontal line (in gray) demarcates the 0.2 threshold for the next interface. Bottom: the total crossing probability (TCP) along the progress parameter MCG, joined using WHAM. The TCP curve plateaus at \( 2.7 \times 10^{-26} \). The snapshot insets depict the growth of the nucleus. Blue and red cage represent 512\( \times 62 \) and 512\( \times 62 \), respectively. Methane gas trapped inside is also shown in the same colors. The green spheres around the fully formed cages show the MCG molecules part of the growing cluster but yet to form cages.

0.2. However, due to the steepness of the barrier (especially in the beginning), this is not always the case.

As an initial path to kickstart the TIS sampling, we take a full TPS path that nucleated into an sI crystal.\(^{43} \) We applied the one-way shooting move, where the shooting points were selected from a Gaussian distribution in the MCG variable, centered around the interface value, with a width parameter of \( \alpha = 0.03 \). More information is provided in the Supporting Information. Table 1 gives the TIS sampling details in terms of the number of trial shots, the acceptance ratios, the average path length, and the number of decorrelated paths. The total aggregate simulation time was 1.2 ms, which took a total of 15 months of wall clock time on our GPU setup.

At interfaces with low MCG values, paths tend to become very short, often below a nanometer, resulting in a reduced path decorrelation. Replica exchange TIS (RETIS) can significantly enhance this decorrelation. Therefore, we performed an independent RETIS simulation run for the interfaces MCG = 3–13. In total, 50,000 MC moves were performed with an average acceptance of 60%. The path exchange move was attempted 2266 times and the minus move 1188 times (with 100% acceptance\(^{54} \)).

**RESULTS AND DISCUSSION**

Sampling Interfaces and Extracting the Crossing Probability. We performed transition interface sampling (TIS) of methane hydrate nucleation at a realistic temperature of 280 K and a relevant pressure of 500 bar, using the same system setup as in ref 43. The acceptance ratio and the number of decorrelated paths are given in Table 1. Selected path trees are shown in the SI.

The crossing probabilities as a function of MCG were recorded for each of the TIS and RETIS ensembles and are depicted in Figure 2. Clearly, the crossing probabilities steeply decrease with MCG, certainly at the lower interfaces, but level off for higher interface values, where the barrier becomes flatter, as is expected when the critical nucleus is reached. The steepness at the interfaces with lower MCG values is also the reason that the interfaces are placed closer together there. Although not every TIS run showed the optimal 20% crossing with the next interface, there is a significant overlap between each successive interfaces so that we can apply the Weighted Histogram Analysis Method (WHAM)\(^{65} \) to join the histograms. Figure 2 shows the resulting total crossing probability. Here, the log probability indeed exhibits an initial steep decrease that slowly becomes flatter for higher interfaces. We estimated the error bar in the final crossing probability using the bootstrapping method to be 1–2 orders of magnitude (see the SI).

Flux Calculation. The flux out of the liquid state (defined by MCG \( \leq 2 \)) through the first interface at MCG = 3 is computed from the RETIS simulation by summing the average path lengths of the minus and the first interface,\(^{66} \) as implemented in OPS\(^{52,53} \). This average path length is around five frames, which, with a saving frequency of 100 ps, translates into 500 ps on average. The flux through MCG = 3 is thus \( \Phi_{3,2} = 0.002 \pm 0.00008 \) ps\(^{-1} \). Note that this flux will be dependent on the saving frequency and should be matched with a crossing probability obtained with the same saving frequency.

Nucleation Rate Calculation. The product of the flux and the crossing probability leads directly to the nucleation rate. As nearly all paths for the interface MCG = 110 end in the solid state (shown in the Supporting Information), the crossing probability will not change anymore beyond that interface. Therefore, \( \Phi_{3,2} = 0.002 \times 2.7 \times 10^{-26} \) nuclei per picosecond per simulation volume. The error on this number is mostly caused by the uncertainty in the crossing probability. Indeed, the bootstrapping results indicate that the error in the rate constant is roughly 1–2 orders of magnitude (calculation shown in the SI).

---

Figure 2. Top: the individual crossing probabilities for each sampled interface plotted along the progress order parameter MCG. Vertical lines indicate interface locations. The horizontal line (in gray) demarcates the 0.2 threshold for the next interface. Bottom: the total crossing probability (TCP) along the progress parameter MCG, joined using WHAM. The TCP curve plateaus at \( 2.7 \times 10^{-26} \). The snapshot insets depict the growth of the nucleus. Blue and red cage represent 512\( \times 62 \) and 512\( \times 62 \), respectively. Methane gas trapped inside is also shown in the same colors. The green spheres indicate interface locations. The horizontal line (in gray) demarcates the 0.2 threshold for the next interface.
As this rate is the expected homogeneous nucleation events per unit time in the simulation volume, we can easily estimate the rate in terms of the number of nuclei per second per cubic cm. Assuming that each nucleation event is independent of each other, this is simply determined by the size of the used simulation box. For our simulation, the box volume \( V = 110.5 \text{ nm}^3 \), yielding an overall nucleation rate of \( J_{\text{TIS}} = k_{A}V = 5.08 \times 10^2 \text{ nuclei cm}^{-3} \text{ s}^{-1} \) (Table 2). The computed nucleation rate will be similar for both the amorphous and the crystalline channel of the hydrate nucleation at the imposed temperature, as, in our previous TPS study, we found an equal population of amorphous and crystalline pathways.

**Free Energy and Comparison with CNT.** As many simulation studies focus on the free energy barrier and employ CNT to estimate the rate, it is natural to compare our TIS rate prediction to an estimate based on the free energy barrier. The TIS simulation of the nucleation process does not give the free energy barrier directly because the pathways that come from a high MCG value (the B state) are missing from the sampling. To obtain an estimate of the free energy barrier, we therefore performed several TIS simulations for the reverse process, i.e., to the data (solid red curve) is also substantially different. We conclude that the CNT expression is only an approximate, qualitative description for the hydrate nucleation free energy at these conditions.

**Molecular Mechanism of Nucleation.** All pathways in the TIS ensembles start in the metastable liquid phase and progress up the nucleation barrier until they cross the imposed interface and relax back to the liquid phase or continue until full solidification. Previously, we found that the nucleation mechanism shifts from forming amorphous phases at lower temperatures to sI crystal formation at higher temperatures. At 280 K, both mechanisms can coexist. The two mechanisms can be identified by the cage ratio CR, the ratio of large and small cages, which is higher than unity for crystalline structures and lower than 1 for amorphous structures. While the cage ratio is not the only indicator of crystallinity, e.g., the connectivity of the \( S^{126} \) cages is also indicative of an sI crystal. However, we observe in all trajectories that the large \( S^{126} \) cages always form into a connected, spherical grid, that is part of a single growing cluster (see also movies provided in the SI).

We analyzed close to 2.5 million frames from the TIS simulated pathways and plotted histograms of the cage ratio for several path ensembles in Figure 4. At the lower interfaces, paths are only sampling amorphous structures, but when reaching interfaces closer to the critical nucleus, the histograms clearly show signatures of crystallinity. Note that this observed behavior might change for a different progress variable. We, however, stress that the final result of TIS is independent of the choice of the order parameter.

**Comparison to Previous Rate Predictions.** Naturally, our rate predictions should be compared to the experimental

---

**Table 2. Nucleation Data from TIS Simulations**

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flux through MCG = 3 (nuclei ps(^{-1}))</td>
<td>0.002</td>
</tr>
<tr>
<td>Crossing probability</td>
<td>2.77 \times 10^{-26}</td>
</tr>
<tr>
<td>Average liquid phase volume, ( V_{liq} ) (nm(^3))</td>
<td>110</td>
</tr>
<tr>
<td>Nuc. rate from TIS, ( J ) (nuc. cm(^{-3}) s(^{-1}))</td>
<td>5.08 \times 10^{2}</td>
</tr>
<tr>
<td>Monomer number density, ( \rho ) (cm(^{-3}))</td>
<td>5 \times 10^{9}</td>
</tr>
<tr>
<td>Zeldovich factor, ( Z )</td>
<td>0.025</td>
</tr>
<tr>
<td>Diffusivity, ( D ) (MCG(^2)/ns)</td>
<td>7.8</td>
</tr>
<tr>
<td>Nucleation barrier, ( G(a)/k_B T )</td>
<td>56.9</td>
</tr>
<tr>
<td>Nuc. rate from CNT, ( J_{\text{CNT}} ) (nuc. cm(^{-3}) s(^{-1}))</td>
<td>2 \times 10^{4}</td>
</tr>
</tbody>
</table>

---

**Figure 3.** Free energy from projecting the reweighted path ensemble (blue circles). Error bars are estimated from the error in the crossing probability. Dashed blue curves are the partial densities from the forward and reverse path ensembles. Red curves are plots from standard CNT expressions: the solid curve is the best fit to the simulation data; the dotted red curve is a fit to the maximum of the barrier.
As in our previous work, we can estimate the driving force as $\Delta \mu = 2\gamma/(\rho_r r^*$), from the observed radius of the crucial nucleus $r^*$, the surface tension of the liquid–crystal interface $\gamma$, and the density of the solid nucleus $\rho_r$. Setting $\gamma \approx 32 \text{ mJ m}^{-2}$, and $\rho_r = 4.57$ (in units of cages per nm$^3$), and using the relation between size and number of methane molecules $(4/3)\pi r^3 = n^*$, the observed critical nucleus $n = 90$ yields a system driving force $\Delta \mu^{\text{sys}} \approx 5 \text{ kJ mol}^{-1} \approx 2k_BT$. The presence of the spherical bubble induces a Laplace pressure of roughly 750 bar, which in turn gives an excess chemical potential $\Delta \Delta \mu = k_BT \ln(P_{\text{Bubble}}/P) \approx 0.92$ k_BT. The driving force in the absence of the bubble is thus $\Delta \mu^{\text{flat}} = \Delta \mu^{\text{sys}} + \Delta \Delta \mu \approx 1.1$ k_BT, in line with the previous estimates. Employing the CNT expression $\Delta G^* = (16/3)\pi r^3/\rho (\Delta \mu)^2$, the barrier due to the Laplace pressure of the spherical bubble is a factor of 3–4 lower compared to a reservoir with a flat interface. With a predicted CNT barrier of 90 k_BT for the spherical reservoir, this would mean around 300 k_BT for the flat gas–liquid interface. For our TIS computed barrier of $\approx 60$ k_BT, this would still mean around 200 k_BT. Adjusting the TIS-based barrier height for the flat interface thus brings the free energy much more in line with that of ref 28. We note that these barrier estimates are very rough and should not be taken as accurate quantitative predictions. However, in any case, with such high barriers, we expect homogenous nucleation not to play a significant role in the imposed moderate supersaturation conditions with a flat gas–liquid interface.

Besides the extreme difference in rates based on the used models, system setup, and conditions, there are also more subtle explanations for differences in results. For instance, there is a difference between the FFS and the TIS approach. Previously, FFS has been used for methane nucleation (using CG models). However, while TIS creates full trajectories that can be influenced by the choice of the interfaces, TIS simulations also showed that the nucleation barrier reduces with the decreasing temperature.

On the other hand, Knott et al. predicted for 273 K and 900 atm, an exceedingly small rate that is 100 orders of magnitude lower than our prediction and the values reported in ref 29. While the conditions differ slightly from the pressure and temperature used here, this difference is not sufficiently large to explain the enormous discrepancy with our prediction.

Instead of the nucleation rate, we can also compare the corresponding free energy barriers. The predicted free energy barrier based on the TIS simulations was $\approx 56.9$ k_BT. The two previous estimates for the nucleation free energy barrier close to these conditions are 173 and 300 k_BT. Several arguments can be put forward to explain these large discrepancies in the rate and free energy barrier. First, we notice that the simulation setup was different. Due to the necessity for a limited system size, the choice of stoichiometric conditions, and the need for a complete solidification in our path sampling simulations, the methane reservoir in the initial state is a bubble with a radius on the order of a few nanometers. This leads to an additional Laplace pressure that induces a higher methane solubility and, hence, a higher driving force. This higher driving force naturally leads to a lower barrier and a higher rate and contributes to the difference with the work of Knott et al. and of Lauricella et al.

Figure 4. Cage ratio histogram for selected interfaces. A total of 2.5 million frames were analyzed, which included the data from all of the interfaces (shown in the SI). The amorphous phase is identified by a cage ratio < 1 (corresponding to the first bar at any interface value).

There is a gradual shift in the population of the cage ratio from lower (MCG = 40) to higher (MCG = 100) interfaces. At MCG = 100, crystalline nuclei become the most abundant. (Also, see Figure S11 in the Supporting Information).
were done for the same system setup and pressures. Figure 5: TIS rate and the rate predictions from refs 17, 18, 20, which for further details). We can additional driving force due to the curved interface (see the SI (F) or bubble (B)), the water force (T4I:Tip4P/Ice). Note that ref 20 uses the multiple system sizes where the rate varies between $10^{24}$ and $10^{34}$. Ref 28 reports the total number of particles on the order of 70 000.

**Table 3. Comparison of the Predicted and Experimental Data**

<table>
<thead>
<tr>
<th>$T$ (K)</th>
<th>$P$ (bar)</th>
<th>system (W/G)</th>
<th>method</th>
<th>FF</th>
<th>rate $(\text{cm}^{-3} \text{s}^{-1})$</th>
<th>refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>220</td>
<td>500</td>
<td>6912/1280 (F)</td>
<td>FFS</td>
<td>mW</td>
<td>$6 \times 10^{11}$</td>
<td>Bi17</td>
</tr>
<tr>
<td>230</td>
<td>500</td>
<td>7555/465 (F)</td>
<td>FFS</td>
<td>mW</td>
<td>$1.3 \times 10^{12}$</td>
<td>DeFever18</td>
</tr>
<tr>
<td>250</td>
<td>500</td>
<td>2944/512 (B)</td>
<td>MD</td>
<td>T4I</td>
<td>$5 \times 10^{13}$</td>
<td>Zhang20</td>
</tr>
<tr>
<td>250</td>
<td>300</td>
<td>2944/512 (B &amp; F)</td>
<td>MD</td>
<td>T4I</td>
<td>$5 \times 10^{14}$</td>
<td>Walsh21</td>
</tr>
<tr>
<td>255</td>
<td>500</td>
<td>2944/512 (B)</td>
<td>MD</td>
<td>T4I</td>
<td>$9.43 \times 10^{23}$</td>
<td>Yuhara22</td>
</tr>
<tr>
<td>255</td>
<td>500</td>
<td>2944/512 (B)</td>
<td>MD</td>
<td>T4I</td>
<td>$9.07 \times 10^{23}$</td>
<td>Barnes22</td>
</tr>
<tr>
<td>273</td>
<td>911</td>
<td>70 000 (F)</td>
<td>MD/CNT</td>
<td>mW</td>
<td>$3 \times 10^{21}$</td>
<td>Knott23</td>
</tr>
<tr>
<td>273</td>
<td>0.98</td>
<td>exp</td>
<td></td>
<td></td>
<td>$10^{2} - 10^{7}$</td>
<td>refs 30–37</td>
</tr>
<tr>
<td>280</td>
<td>300</td>
<td>exp/model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>280</td>
<td>500</td>
<td>2944/512 (B)</td>
<td>TIS</td>
<td>T4I</td>
<td>$5 \times 10^{2}$</td>
<td>this work</td>
</tr>
</tbody>
</table>

“System details for computational methods include the number of water (W) and guest (G) molecules used, the type of liquid–gas interface (flat (F) or bubble (B)), the water force field (FF) (T4I:Tip4P/Ice). Note that ref 20 uses the multiple system sizes where the rate varies between $10^{24}$ and $10^{34}$. Ref 28 reports the total number of particles on the order of 70 000.

equal for both the crystal and amorphous phases at 280 K, we do not expect a difference in the rate.

Finally, we note that previous rate predictions for moderate undercooling are usually computed by indirect means. They are either based on CNT using several thermodynamic ingredients obtained from different sources or they employ a free energy calculation as a function of a fixed-order parameter. In contrast, our approach is an internally consistent approach that is not dependent on the choice of the order parameter. Instead, it just depends on the choice of the force field and the thermodynamic conditions imposed.

Table 3 summarizes several previous simulation-based predictions for homogeneous methane hydrate nucleation rates. The methodology and force fields used for each prediction was slightly different. Each study also used a different setting like undercooling (and pressure), but overall the nucleation rates are at least 20 orders away from the experimentally available results. Our results, performed at 280 K, are much more in line with the available experimental data.

Although this seems to suggest that at natural, moderate undercooling conditions, the homogenous nucleation of methane hydrate is feasible, the larger supersaturation due to the spherical gas bubble, together with the finite size scaling argument, leads to effectively much larger barriers for reservoirs with flat interfaces. The fact that experimental results are in line with our results must thus be interpreted as fortuitous. Indeed, ref 29 employed a slightly lower surface tension, which could have resulted in an overestimate of the experimental nucleation rate. Moreover, many experimental results refer to hydrates in general and/or to heterogeneous nucleation, and, thus, cannot be easily compared.

To estimate how the predicted rate will change with temperature, we can use a simple approximation based on CNT, which states that the log rate is $\ln j \sim -\Delta \mu^2 + \text{const}$, where the proportionality factor and constant involve the kinetic prefactors, and properties such as the surface tension, which we assume constant over the temperature region of interest. We further assume that the driving force is given by $\Delta \mu = \Delta S (T_m - T) + \Delta \Delta \mu$, where $\Delta S$ is the difference in the entropy of the liquid and the solid phases, and $\Delta \Delta \mu$ is the additional driving force due to the curved interface (see the SI for further details). We can fit this relation to match both the TIS rate and the rate predictions from refs 17, 18, 20, which were done for the same system setup and pressures. Figure 5 plots the resulting fit of the temperature, together with the TIS rate, and the rate predictions from refs 17, 18, 20 as well as the consensus estimate from ref 29. Naturally, the fitted (black) curve agrees well with the literature results. However, the slope for this curve is lower than the experimental data as due to the additional driving force of the curved interface the rate remains finite even at the melting temperature. Therefore, we also include the CNT-based green curve for a flat interface, which is estimated by subtracting the additional driving force due to the spherical bubble. The rate of nucleation for a flat interface is thus substantially smaller and behaves more similarly to the results from ref 29. However, due to the various assumptions in these CNT estimates and the strong sensitivity of the rate, it is not clear whether or not this is incompatible with the conclusion that nucleation is unfeasible.\cite{ref30} In the SI, we present several other CNT-based curves, which all show similar behavior. We stress once more that all of these CNT predictions should be taken as qualitative.

**CONCLUSIONS**

Understanding the nucleation kinetics of methane hydrates is of importance for both theoretical and practical reasons. We have presented extensive transition interface sampling simulations using an accurate atomistic force field. The TIS methodology enables an efficient evaluation of the exact rate of nucleation by generating ensembles of unbiased dynamical
trajectories, leading from the metastable liquid up to the nucleation barrier. For the imposed specific conditions of 280 K and 500 bar, we find a nucleation rate of $\sim 5 \times 10^2$ nuclei per cm$^3$ per second, which is, probably fortuitously, in agreement with a comparative analysis study.29

This study shows that it is now possible to obtain accurate molecular crystal nucleation rates at moderate undercooling and, thus, for a high free energy barrier, without influence due to the choice of the progress variable, and without relying on assumptions other than the force field.

A previous estimate found almost 120 orders of magnitude lower rate under nearby conditions,28 which essentially ruled out the presence of homogeneous nucleation due to the very high free energy barrier obtained. It is likely that most of this discrepancy can be explained by differences in the system setup. Especially, the higher observed solubility of methane under the increased Laplace pressure helps in lowering the barrier and increasing the rate. However, this correction might not be sufficient to explain the discrepancy entirely, and future research might look into examining the rate for a flat methane gas interface using TIS. As this would require a much larger system size, this is beyond the scope of the current work.

Finally, we stress that heterogeneous nucleation can also reduce the free energy barrier of transformation, leading to a faster formation process. A natural extension of this project will be to estimate the rate in a system with heterogeneity.74

### ASSOCIATED CONTENT

#### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c04582.

Supporting information text contains information about the system setup, the Gaussian shooting algorithm, avoiding a cylindrical gas reservoir, the effect of small system size, details about RETIS, various plots showing sampling at each interface, the crystallinity of the nucleus, computation of the error in the rate and flux, computation of the free energy barrier and comparison with other rate calculations and CNT (PDF)

Reactive trajectories of transition from liquid phase to solid SI hydrate phase (MCG value of 2–300 in terms of TIS) sampled from the ensemble belonging to interface MCG = 100; $^1$H$^2$O and $^1$H$^2$O$^2$ are tracked using blue and red color, respectively. Video 1 (MP4)

Video 2 (MP4)

### ACKNOWLEDGMENTS

A.A. and P.G.B. acknowledge funding from the Industrial Partnership Program “Computational Sciences for Energy Research” (Grant no. 1SCE0501) of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organization for Scientific Research (NWO). This research program is co-financed by Shell Global Solutions International B.V. Both authors thank David W. H. Swenson and Gyorgy Hantal for critically reading the manuscript. The authors acknowledge support from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) for the use of supercomputer facilities.

### REFERENCES


(72) Juraszek, J.; Bolhuis, P. G. Rate constant and reaction coordinate of Trp-cage folding in explicit water. Biophys. J. 2008, 95, 4246−4257.