Analysis of nucleation using mean first-passage time data from molecular dynamics simulation

David A. Nicholson and Gregory C. Rutledge

Citation: The Journal of Chemical Physics **144**, 134105 (2016); doi: 10.1063/1.4945256

View online: http://dx.doi.org/10.1063/1.4945256

View Table of Contents: http://aip.scitation.org/toc/jcp/144/13

Published by the American Institute of Physics

Analysis of nucleation using mean first-passage time data from molecular dynamics simulation

David A. Nicholson and Gregory C. Rutledge
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 16 February 2016; accepted 16 March 2016; published online 7 April 2016)

We introduce a method for the analysis of nucleation using mean first-passage time (MFPT) statistics obtained by molecular dynamics simulation. The method is based on the Becker-Döring model for the dynamics of a nucleation-mediated phase change and rigorously accounts for the system size dependence of first-passage statistics. It is thus suitable for the analysis of systems in which the separation between time scales for nucleation and growth is small, due to either a small free energy barrier or a large system size. The method is made computationally practical by an approximation of the first-passage time distribution based on its cumulant expansion. Using this approximation, the MFPT of the model can be fit to data from molecular dynamics simulation in order to estimate valuable kinetic parameters, including the free energy barrier, critical nucleus size, and monomer attachment pre-factor, as well as the steady-state rates of nucleation and growth. The method is demonstrated using a case study on nucleation of *n*-eicosane crystals from the melt. For this system, we found that the observed distribution of first-passage times do not follow an exponential distribution at short times, rendering it incompatible with the assumptions made by some other methods. Using our method, the observed distribution of first-passage times was accurately described, and reasonable estimates for the kinetic parameters and steady-state rates of nucleation and growth were obtained. © *2016 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4945256]

I. INTRODUCTION

Most first order phase transitions are characterized by a kinetic barrier that is crossed upon conversion from old phase to new phase. In this case, the transition is a nucleation-mediated process that occurs through stochastic fluctuations giving rise to formation of clusters of the new phase. The conventional reaction coordinate in such nucleation-mediated processes is the size of the cluster. The kinetic barrier arises from the competition between the driving force for formation of the new phase and an interfacial free energy penalty, the balance of which inhibits the growth of small, subcritical clusters due to their large specific surface areas, but allows large, supercritical clusters to grow. The critical nucleus size n^* marks the transition state at the top of the barrier. A nucleation event occurs when a cluster of size greater than n^* arises, surpassing the kinetic barrier and precipitating a period of rapid cluster growth.

Despite its importance to the overall kinetics of the phase change, nucleation is difficult to study experimentally due to the small spatiotemporal scale at which it occurs. Molecular-level simulation methods, on the other hand, are readily applied to phenomena that occur at small time- and length-scales, rendering them effective for nucleation studies.[1,2] In particular, molecular dynamics (MD) has proven useful for kinetic studies; the method has been applied to transitions ranging from the condensation of liquid from the vapor of a Lennard-Jones fluid,[3] to more complex transitions such as the crystallization of ice,[4] *n*-alkanes,[5] and NaCl.[6] In an MD simulation, time advances linearly, making the dynamics of a simulated nucleation event analogous to dynamics in an experimental phase change. Due to the computational expense of MD, studies are typically conducted at deep supercooling or high superheating in order to observe a nucleation event within a reasonable time frame.[1] Alternative methods that do not preserve the linear advancement in time, including linear response,[7,8] metadynamics,[9] and transition-path sampling,[10] have been developed in order to facilitate kinetic studies of slower nucleation events.[1,2] These methods are promising, but they come with assumptions regarding underlying time scales[7,8] and other theoretical challenges, particularly with regard to the identification of suitable order parameters.[9,10] Where applicable, MD remains a method of choice for the simulation of phase change kinetics at the molecular level and serves as an important point of comparison for emerging methods.

Despite the simplicity of MD, a fundamental challenge in the analysis of MD data arises from the fact that the equations that govern nucleation kinetics, i.e., the master equation[11] or Fokker-Planck equation,[12] do not permit closed-form solutions for the time-evolution of a reaction coordinate. As a result, simplifying assumptions are often invoked, such as a large separation between the time scales of nucleation and growth. The methods proposed by Bartell and Wu[13] and Wedekind *et al.*[14] invoke this assumption in order to estimate the nucleation rate, critical nucleus size, and Zeldovich factor, which is related to the curvature of the free energy surface around the critical nucleus size, from MD simulations. The method described by Wedekind *et al.*[14] is based on the mean first-passage time (MFPT) of the largest cluster, obtained from

a large set of independent nucleation runs, and is applicable when the characteristic time scale for nucleation is much larger than the time scale for growth in the system under study. The relationship between time scales is dependent on both the size of the kinetic barrier to nucleation and the system size. The system size effect arises from the fact that a successful nucleation event is the eventual result of many trials to form a supercritical cluster, where each trial initiates from a single monomer. As a system becomes larger it contains more monomers, thus allowing for more trials in a given unit of time. As a result, the time scale for nucleation decreases with increasing system size. Once a supercritical cluster is formed, however, it grows at a rate that is independent of system size, leaving the time scale for growth unchanged. Therefore, as the system size is increased, the separation of time scales is diminished. When the separation becomes too small, due to a combination of the free energy barrier and system size, the assumptions made in the MFPT-based method of Wedekind *et al.*[14] can no longer be expected to hold.[15] Alternative methods have been employed to study systems with very rapid nucleation, including the methods of Yasuoka and Matsumoto,[3] and Mokshin and Galimzyanov.[16] These methods are applicable when many nucleation events can be observed in a single system, allowing for the evaluation of the time between subsequent events.

In this paper, we introduce an alternative method for analyzing nucleation data on the MFPT that can be applied to systems where the separation of time scales for nucleation and growth is small, either due to a small free energy barrier or large system size. We invoke a stochastic model for the nucleation process governed by a birth-death master equation and make use of Gillespie's model for the system size dependence of first-passage statistics.[17] Parameterization of this model to MFPT data from MD simulations yields estimates for valuable kinetic quantities, including the free energy barrier, critical nucleus size, nucleation rate, and a characteristic growth rate. In order to make this procedure computationally practical, we report an efficient method to evaluate the first-passage time distribution (FPTD) based on its cumulant expansion. This method is not specifically restricted to nucleation and may prove useful for computing the FPTD for other physical processes that follow a birth-death master equation, including phenomena in chemical kinetics[18] and biology.[19]

II. THEORY

A. Stochastic model

The stochastic nucleation model employed here describes the time-evolution of cluster formation from an initial state with N monomers. A cluster may consist of one or more monomers, and monomers are equivalent to clusters of size 1. The model depends explicitly on N and contains three time-independent parameters: the monomer attachment pre-factor f_1, the critical nucleus size n^*, and the critical free energy barrier $\beta\Delta G^*$. The model follows closely that presented by Shugard and Reiss[20] and Gillespie.[17] We first review the model as presented by them, along with simplifications that follow from the literature on the topic of birth-death processes. Per Gillespie,[17] the following assumptions are made at the outset:

i. Clusters evolve in time through the attachment or detachment of a single monomer. The rates of these events are time-independent and follow the law of mass action.
ii. The concentration of monomers C_1 is approximately constant over time.
iii. The total number of clusters in the system is approximately constant over time, and approximately equal to the total number of monomers in the system, N.

The first assumption is known as the Szilard model and is frequently employed in kinetic nucleation theory.[11] Following this assumption, clustering occurs through bimolecular attachment $X_n + X_1 \rightarrow X_{n+1}$, and unimolecular detachment, $X_n \rightarrow X_1 + X_{n-1}$, steps, leading to the modified Becker-Döring equations of Penrose and Oliver.[21] These equations present a good model for the physical system, but they are nonlinear and stiff[22] and thus difficult to solve.

The second assumption follows from the notion that in the nucleation stage the system is primarily composed of monomers and results in a drastic simplification of the equations. Under this condition, any single cluster is nearly completely surrounded by monomers, and the rate of attachment to the cluster can therefore be approximated by unimolecular steps, including attachment $X_n \rightarrow X_{n+1}$ with rate $f(n)$ and detachment $X_n \rightarrow X_{n-1}$ with rate $g(n)$. This assumption yields the standard Becker-Döring equations,[23] which have the form of the birth-death master equation for the transient cluster size distribution $C(n,t)$,

$$\frac{\partial C(1,t)}{\partial t} = 0, \tag{1a}$$

$$\frac{\partial C(n,t)}{\partial t} = f(n-1)C(n-1,t) + g(n+1) \times C(n+1,t) - (f(n)+g(n))C(n,t). \tag{1b}$$

Under assumption (ii) the total mass in the system is not conserved; however, if the system is composed predominantly of monomers, the additional mass associated with clusters of size $n > 1$ is taken to be negligible.

The third assumption is a necessary consequence of assumptions (i) and (ii) under an initial condition where there are initially N monomers in the system.[17] It asserts that nucleation in a system which initially contains N monomers behaves as a system with N clusters. This number is an overestimate of the number of clusters, but the degree of overestimation is small in a system that is composed primarily of monomers.

The rates of attachment to and detachment from a cluster of size n are related to each other by the stationary solution to the birth-death master equation. The flux for the transition $n \rightarrow n+1$ is

$$I(n,t) = f(n)C(n,t) - g(n+1)C(n+1,t). \tag{2}$$

The stationary, or equilibrium, solution $C_{eq}(n)$ is obtained under zero flux, $I(n,t) = 0$, and is equivalent to the detailed

balance condition for a step-wise process,

$$f(n)C_{eq}(n) = g(n+1)C_{eq}(n+1), \tag{3a}$$

$$f(n)e^{-\beta\Delta G(n)} = g(n+1)e^{-\beta\Delta G(n+1)}. \tag{3b}$$

The expression in Eq. (3b) is obtained from an equilibrium cluster distribution that follows a Boltzmann distribution, e.g., $C_{eq}(n) = C_1 e^{-\beta\Delta G(n)}$ where $\beta\Delta G(n)$ is the free energy of forming a cluster of size n with $\beta\Delta G(1) = 0$, and the value C_1 is the time-independent concentration of monomers. The free energy can be expressed using the capillary approximation for a 3 dimensional cluster with a barrier height ΔG^* and a critical nucleus size n^*[24]

$$\beta\Delta G(n) = \beta\Delta G^* \left[3\left(\frac{n^{2/3}-1}{(n^*)^{2/3}} \right) - 2\left(\frac{n-1}{n^*} \right) \right]. \tag{4}$$

The rate of attachment to a cluster that is surrounded by a bath of monomers is expected to scale as the surface area of the cluster; assuming the cluster shape to be static, the size dependence is

$$f(n) = f_1 n^{2/3}. \tag{5}$$

The value $f(1) = f_1$ is the monomer attachment pre-factor, which reflects the frequency at which a monomer is added to a labeled monomer. With these relationships, the evolution of a single cluster under Eq. (1) is completely specified by the choice for the parameters $\beta\Delta G^*$, n^* and f_1, along with a reflecting boundary condition, $g(1) = 0$, that prevents removal of monomers from the system. None of these three parameters depend on system size, due to assumptions (i) and (ii).

For a single cluster evolving according to Eq. (1) the first-passage time is defined as the first time $T(n_0 \to n)$ that a cluster reaches size n given that it started at size n_0 at $t = 0$

$$T(n_0 \to n) = \min\{ t | X(t) = n, X(0) = n_0 \}. \tag{6}$$

For a given set of conditions $\{n_0, n\}$, the first-passage time $T(n_0 \to n)$ is a random variable governed by a probability density function $f_T(\mathrm{t};\ n_0 \to n)$, referred to as the first-passage time distribution (FPTD). The general form of $f_T(\mathrm{t};\ n_0 \to n)$ for arbitrary n_0 is known,[17,20] but expensive to compute, and numerically unstable.[25] For the special case of $n_0 = 1$, which is of primary interest for the model described here, the FPTD takes on a comparatively simple form. $T(1 \to n)$ has the same distribution as a sum of $n-1$ exponential random variables with rate parameters: $\{\lambda_1^{(n)}, \lambda_2^{(n)}, \ldots, \lambda_{n-1}^{(n)}\}$.[26,27] These rates are the eigenvalues of $-P^{(n)}$, where $P^{(n)}$ is the tridiagonal generator matrix on $\{1, 2, \ldots, n-1\}$

$$P^{(n)} = \begin{bmatrix} -f(1) & g(2) & & & \\ f(1) & -(f(2)+g(2)) & g(3) & & \\ & f(2) & \ddots & \ddots & \\ & & \ddots & \ddots & g(n-1) \\ & & & f(n-2) & -(f(n-1)+g(n-1)) \end{bmatrix}. \tag{7}$$

The eigenvalues are distinct and positive[27] and may thus be ordered as $0 < \lambda_1^{(n)} < \lambda_2^{(n)} < \cdots < \lambda_{n-1}^{(n)}$. The distribution of the sum of exponential random variables is the hypoexponential distribution, yielding the following form for the FPTD:[28]

$$f_T(t; 1 \to n) = \sum_{k=1}^{n-1} \left(\lambda_k^{(n)} e^{-\lambda_k^{(n)} t} \prod_{j=1, j\neq k}^{n-1} \frac{\lambda_j^{(n)}}{\lambda_j^{(n)} - \lambda_k^{(n)}} \right). \tag{8}$$

The cumulative distribution function $F_T(1; 1 \to n) = \int_0^t f_T(t'; 1 \to n) dt'$ denotes the probability that a single cluster starting at the monomer state at time zero has reached size n at time t; the complementary cumulative distribution function, $\bar{F}_T(t; 1 \to n) = 1 - F_T(t; 1 \to n)$, is the probability the cluster has not yet reached size n. The form of the complementary cumulative distribution function follows from Eq. (8),

$$\bar{F}_T(t; 1 \to n) = \sum_{k=1}^{n-1} \left(e^{-\lambda_k^{(n)} t} \prod_{j=1, j\neq k}^{n-1} \frac{\lambda_j^{(n)}}{\lambda_j^{(n)} - \lambda_k^{(n)}} \right). \tag{9}$$

Up to this point, the distributions have described the first-passage time for a single cluster. In order to find the FPTD of the largest cluster, consider a system of N independently evolving clusters each of which is initially of size 1. Following Gillespie's assmuptions,[17] the largest of N independent clusters corresponds to the largest cluster in a system that initially contains N monomers, which is the reaction coordinate used for the analysis of MD simulation data. The probability $\bar{F}_{T,\mathrm{largest}}(t; 1 \to n)$ that the largest of all N clusters has not yet reached size n at time t is

$$\bar{F}_{T,\mathrm{largest}}(t; 1 \to n) = (\bar{F}_T(t; 1 \to n))^N. \tag{10}$$

Eq. (10) is the complementary cumulative distribution function for the first-passage time of the largest cluster, and its time derivative yields the negative of $f_{T,\mathrm{largest}}(t;\ 1 \to n)$, the FPTD for the largest cluster[17,20]

$$f_{T,\mathrm{largest}}(t; 1 \to n) = N f_T(t; 1 \to n)(\bar{F}_T(t; 1 \to n))^{N-1}. \tag{11}$$

The MFPT for the largest cluster, $\tau_{\mathrm{largest}}(1 \to n)$, is the mean of this distribution and takes on the following form after

integration by parts:[17]

$$\tau_{\text{largest}}(1 \to n) = \int_0^\infty t f_{T,\text{largest}}(t; 1 \to n) dt$$
$$= \int_0^\infty [\bar{F}_T(t; 1 \to n)]^N dt. \quad (12)$$

For the case of a single cluster ($N = 1$), the MFPT has a simple, closed form expression[24,25]

$$\tau(1 \to n) = \sum_{k=1}^{n-1} \frac{e^{\beta\Delta G(k)}}{f(k)} \sum_{j=1}^{k} e^{-\beta\Delta G(j)}. \quad (13)$$

No such simplification has been reported for $N \neq 1$; however, the integral in Eq. (12) can be evaluated numerically, using standard quadrature techniques. Thus, following Gillespie's formulation, we have a rigorous, yet computationally expensive, procedure to obtain the first-passage time curve $\tau_{\text{largest}}(1 \to n)$ for a system that obeys the assumptions (i-iii). To be clear, the procedure is as follows:

1. Given values for the parameters f_1, $\beta\Delta G^*$, and n^*, calculate the monomer attachment and detachment rates from Eqs. (3)–(5).
2. For each size n, find the $n-1$ eigenvalues $-P^{(n)}$, defined in Eq. (7).
3. For each size n, compute the $\tau_{\text{largest}}(1 \to n)$ using Eqs. (12) and (9).

In principle, given the mean first-passage time data for a nucleation process from molecular simulation, one can estimate f_1, $\beta\Delta G^*$, and n^* through this procedure by adjusting their values until Eq. (12) provides a suitable fit to the data. In practice, the cost associated with the eigenvalue computation and the numerical instability of the resulting FPTD[25] make this a very difficult task. In Section III, we present a method based on cumulants to compute an approximation to the FPTD that avoids matrix diagonalization, so that the analysis of nucleation according to Gillespie's model, from mean first-passage time data of a molecular simulation, is made significantly more practical.

B. Kinetic rates

The parameters f_1, $\beta\Delta G^*$, and n^* can be used to compute steady-state kinetic rates for a phase transition, including the nucleation rate and a characteristic growth rate. The steady-state nucleation rate is the constant flux solution to Eq. (1) and has a closed-form solution based on the equilibrium cluster distribution function $C_{eq}(n)$ and monomer attachment rate $f(n)$ (or equivalently a set of ΔG^*, n^*, and f_1)[29]

$$I_S = \left[\sum_{k=1}^{M} \frac{1}{C_{eq}(k) f(k)}\right]^{-1} = C_1 \left[\sum_{k=1}^{M} \frac{e^{\Delta G(k)}}{f(k)}\right]^{-1}. \quad (14)$$

The value M should be chosen so that the expected rate of decay $g(M+1)$ is negligibly small and has no discernible effect on the computed rate of nucleation as long as its value is sufficiently larger than the critical nucleus size. Gillespie[17] measured no difference between nucleation rates computed using $M = 2n^*$ and $M = 4n^*$. The concentration of monomers C_1 can be defined under the assumption that the concentration of clusters is equal to N/V and can be derived from the steady-state distribution of clusters[29]

$$C_1 = \frac{N}{V}\left[\frac{\sum_{k=1}^{M} e^{\Delta G(k)}/f(k)}{\sum_{n=1}^{M} e^{-\Delta G(n)} \sum_{k=n}^{M} e^{\Delta G(k)}/f(k)}\right]$$
$$= \frac{N}{V}\left[\frac{\sum_{k=1}^{M} e^{\Delta G(k)}/f(k)}{\tau(1 \to M+1)}\right]. \quad (15)$$

This expression is in apparent contradiction with the assumption that the concentration of monomers is constant and initially equal to N/V. This contradiction is an inevitable consequence of the application of the assumptions listed in Section II A to steady-state nucleation. However, resolution of this contradiction can be found in the limit that the number of clusters with size $n > 1$ is negligibly small, such that Eq. (15) approaches N/V. Combining Eqs. (14) and (15) yields an expression for the nucleation rate, which is similar to a commonly held expression for the steady-state flux in barrier crossing problems[14,30]

$$I_S = \frac{N}{V\tau(1 \to M+1)}. \quad (16)$$

With a few additional assumptions, an expression for growth rate can be derived from the parametric description of the attachment and detachment processes used in the nucleation model. The derivation relies on the translation of the attachment and detachment rates for a growing nucleus to a situation where the phase transition occurs over an ideal, flat interface. The full derivation can be found in Appendix A, but the final result is a familiar expression for the continuous growth rate[31]

$$G_S = \frac{f_1}{\rho A_1}\left(1 - \exp[-\beta\mu]\right). \quad (17)$$

The value μ is the driving force for the phase change, ρ is the crystalline number density, and A_1 is the shape factor for a cluster with a surface area that follows $A = A_1 n^{2/3}$. According to the capillary approximation, the driving force is $\mu = 2\Delta G^*/n^*$; assuming a spherical cluster, the shape factor is $A_1 = (6/\rho)^{2/3}\pi^{1/3}$. Together, these relationships lead to the following steady-state growth rate:

$$G_S = \frac{f_1}{(36\pi\rho)^{1/3}}\left(1 - \exp[-2\beta\Delta G^*/n^*]\right). \quad (18)$$

Here we have assumed continuous, or liquid-like, growth. The calculation can be applied readily to vapor-liquid systems, but not as readily to crystal growth, where the growth rate changes depending on the crystallographic direction. In this case, G_S should be regarded as a characteristic growth rate rather than that corresponding to any particular growth plane. Despite this caveat, we expect that the greatest contribution to the characteristic growth rate comes from the fastest growth direction.

III. APPROXIMATIONS FOR THE FIRST-PASSAGE TIME DISTRIBUTION

A. The single exponential approximation

A drastic simplification to the first-passage time distribution can be made in the event that one eigenvalue of $-P^{(n)}$ is much smaller than all of the remaining ones $\lambda_1^{(n)} \ll \lambda_{j\geq 2}^{(n)}$.[17,20] In this case the first-passage time for a single cluster is well approximated by a single exponential random variable with time constant $\tau(1 \to n)$. For a system of size N, this assumption yields a FPTD that is a decaying exponential with a rate that is larger by a factor of N,

$$f_{T,\text{largest}}(t; 1 \to n) \cong N\lambda_1^{(n)} e^{-N\lambda_1^{(n)} t} \cong \frac{N e^{-tN/\tau(1\to n)}}{\tau(1 \to n)}. \tag{19}$$

In this case, the solution to Eq. (12) reduces to a simple relationship for the system-size dependent first-passage time[17,20]

$$\tau_{\text{largest}}(1 \to n) \cong \tau(1 \to n)/N. \tag{20}$$

If this approximation holds, then the first-passage time curve should scale exactly with the inverse of system size. In general, the true FPTD is well-approximated by a single decaying exponential at long times, but at short times it has low probability density. When N is large, the small time region becomes especially important, because the tail associated with the longest relaxation time is diminished. This effect is demonstrated in Figure 1, which shows that the deviation between the approximation in Eq. (19) and the full expression in Eq. (11) becomes larger as N gets larger.

In order to determine whether the single exponential approximation holds for a given set of nucleation data, we consider the implications of Eq. (19) on the standard deviation in the first-passage time of the largest cluster, $\sigma_{\text{largest}}(1 \to n)$. Shneidman[32] demonstrated that $\sigma_{\text{largest}}(1 \to n)$ is nearly n-independent for large post-critical nucleus sizes. Furthermore, if the single exponential approximation is valid, $\sigma_{\text{largest}}(1 \to n)$ is equal to $\tau_{\text{largest}}(1 \to n)$, and therefore $\tau_{\text{largest}}(1 \to n)$ should also be nearly independent of n for $n \gg n^*$. This implies

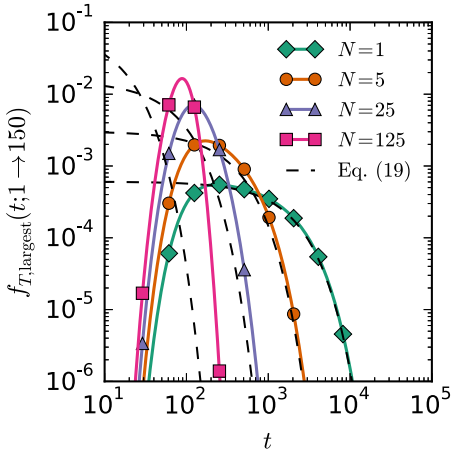

FIG. 1. The first-passage time distribution $f_T(t;\ 1 \to n)$ at $n = 150$ for a system with $n^* = 40$, $\beta\Delta G^* = 5$, and $f_1 = 1$ for various system sizes, N. The solid lines are for the true distribution from Eq. (11) and the dashed lines are the approximate relationship from Eq. (19).

that we should expect Eqs. (19) and (20) to be valid only if the standard deviation in the first-passage time and the MFPT for the largest cluster are equal, and level off at large n. In general, this only occurs when the first-passage time curve for the largest cluster takes on a shape that closely matches the error function suggested by Wedekind *et al.*[14] The form of the error function implies a large separation in time scale between nucleation and growth in this class of problems.

B. The cumulant expansion

Here, we introduce a computationally efficient method to obtain an approximation to the FPTD in Eq. (8) that may be used even when conditions necessary for the single exponential approximation do not hold. Our method makes use of cumulants of the first-passage time, which can be computed without diagonalization, in order to estimate the eigenvalues of $-P^{(n)}$. These estimated eigenvalues, along with the known structure of the FPTD,[26,28] can be used to efficiently approximate the true FPTD in order to evaluate $\tau_{\text{largest}}(1 \to n)$.

The cumulants of the first-passage time are transformations of the moments of the first-passage time, which can be found efficiently using the procedure outlined in Section 6.6B of Gillespie's text on Markov processes.[25] From these moments, the cumulants $\kappa_i(1 \to n)$ follow from a recursive relationship.[33] Recalling from Section II A, $T(1 \to n)$ is distributed as a sum of $n-1$ independent exponential random variables, and therefore $\kappa_i(1 \to n)$ are the sums of cumulants of the $n-1$ exponential random variables. A single exponential random variable with rate λ has cumulants $\kappa_i = \lambda^{-i}(i-1)!$, and therefore the cumulants for $T(1 \to n)$ are

$$\kappa_i(1 \to n) = (i-1)! \sum_{k=1}^{n-1} (\lambda_k^{(n)})^{-i}. \tag{21}$$

The dominant contribution to the ith cumulant comes from the smallest rate, allowing us to write the following expression, which includes an upper bound for the contributions from the remaining rates:

$$\frac{\kappa_i(1 \to n)}{(i-1)!} = \frac{1}{(\lambda_1^{(n)})^i} + O\left(\frac{n-2}{(\lambda_2^{(n)})^i}\right). \tag{22}$$

In the limit of large i we get a limiting expression for smallest rate based on the ith cumulant,

$$\lim_{i\to\infty} \left[\frac{\kappa_i(1 \to n)}{(i-1)!}\right]^{1/i} = \frac{1}{\lambda_1^{(n)}}. \tag{23}$$

Even for nucleation processes with small barriers $\beta\Delta G^*$, the cumulants were observed to converge to this limiting expression for modest values of i. This behavior is demonstrated in Figure 2, in which the limiting expressions for $i = 2$ and $i = 3$ are nearly indistinguishable for barrier heights in the range of $1 \leq \beta\Delta G^* \leq 10$; convergence improves with increasing $\beta\Delta G^*$.

The convergence shows that we can get a good estimate of the smallest rate $\hat{\lambda}_1^{(n)} \approx \lambda_1^{(n)}$ from $\kappa_i(1 \to n)$ using Eq. (23) with a modest value of i. We can extend this approach by assuming that the dominant contributions to $\kappa_{i-1}(1 \to n)$

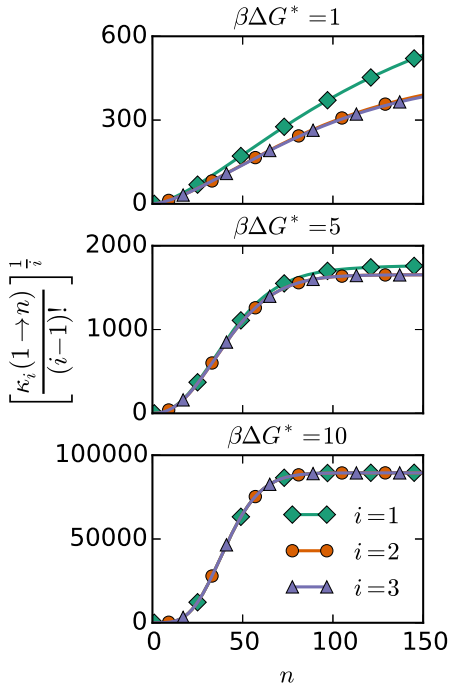

FIG. 2. Convergence of cumulants to the limiting expression for $n^* = 40$, $f_1 = 1$ with varying barrier heights. The variable i denotes the order of the cumulant used in the limiting expression.

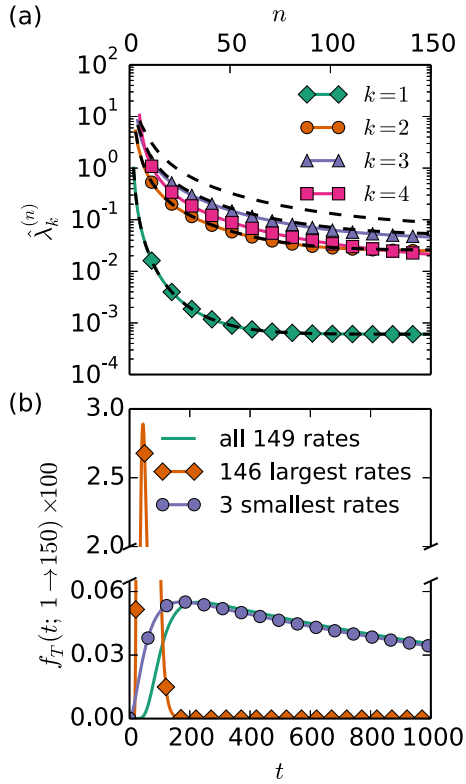

FIG. 3. For a system with $\beta\Delta G^* = 5$, $n^* = 40$, and $f_1 = 1$: (a) the approximate eigenvalues computed using Eq. (24) with $i = 4$ (solid lines) and the smallest four eigenvalues of $-P^{(n)}$ (dashed lines), (b) the FPTD for $T(1 \to 150)$ using all 149 rates as well as the individual contributions based on the 3 smallest rates and the $149 - 3 = 146$ largest rates.

are from $\lambda_1^{(n)}$ and $\lambda_2^{(n)}$, to get an estimate of the second smallest rate $\hat{\lambda}_2^{(n)} \approx \lambda_2^{(n)}$, and so on. In general, one may compute the estimated values of i rates, under the assumption that $\kappa_{i-k}(1 \to n)$ is well approximated by $\{\lambda_1^{(n)}, \ldots, \lambda_{k+1}^{(n)}\}$ in Eq. (21). This reduces to the following recursive relation to compute the estimated rates $\{\hat{\lambda}_1^{(n)}, \ldots, \hat{\lambda}_i^{(n)}\}$:

$$\hat{\lambda}_k^{(n)} = \left[\frac{\kappa_{i-k+1}(1 \to n)}{(i-k)!} - \sum_{j=1}^{k-1} (\hat{\lambda}_j^{(n)})^{k-i-1}\right]^{\frac{1}{k-i-1}}, \quad k = 1, 2, \ldots, i. \tag{24}$$

For example, if we choose $i = 2$, the estimated rates are: $\hat{\lambda}_1^{(n)} = \kappa_2(1 \to n)^{-1/2}$ and $\hat{\lambda}_2^{(n)} = [\kappa_1(1 \to n) - \kappa_2(1 \to n)]^{-1/2}$.

In Figure 3(a) we compare the $\hat{\lambda}_k^{(n)}$ values from Eq. (24) with $i = 4$ to the true 4 smallest rates, obtained from the diagonalization of $-P^{(n)}$, for $\beta\Delta G^* = 5$, $n^* = 40$, and $f_1 = 1$. The first 3 estimates were found to be good approximations to the true rates, which are represented by the dotted lines. The largest of the 4 rate estimates, however, was found to be consistently smaller than the true rate. In general, Eq. (24) was found to give accurate estimations for the first $i - 1$ rates, but not for the ith. Recalling that $f_T(t;\ 1 \to n)$ is the distribution for a sum of $n - 1$ independent exponential random variables, we can determine the origin for this discrepancy by considering the separate contributions to the full FPTD from the first $i - 1$ rates and the remaining $n - i$ rates. Each of these contributions follows the hypoexponential distribution, given by Eq. (8), and is drawn in Figure 3(b) for the transition $1 \to 150$. We see that, relative to the full FPTD, the contribution from the $n - i = 146$ largest rates is a sharply peaked curve, and the contribution from the $i - 1 = 3$ rates looks similar to the full FPTD, but shifted. This behavior is consistent with the $n - i = 146$ rates acting as a shift with mean

$$\mu_{\text{peak}} \cong \kappa_1(1 \to n) - \sum_{k=1}^{i-1} \left(\hat{\lambda}_k^{(n)}\right)^{-1} = \left(\hat{\lambda}_i^{(n)}\right)^{-1}. \tag{25}$$

Consequentially, $\hat{\lambda}_i^{(n)}$ reflects the inverse of the shift caused by the $n - i$ largest rates, rather than an approximation for the ith smallest rate.

Based on this observation we propose that the FPTD is well-approximated by the following hypoexponential distribution, in which $s = t - 1/\hat{\lambda}_i^{(n)} \geq 0$ is a shifted time:

$$f_T(s; 1 \to n) = \sum_{k=1}^{i-1} \left(\hat{\lambda}_k^{(n)} e^{-\lambda_k^{(n)} s} \prod_{j=1, j\neq k}^{i-1} \frac{\hat{\lambda}_j^{(n)}}{\hat{\lambda}_j^{(n)} - \hat{\lambda}_k^{(n)}}\right), \quad i > 1. \tag{26}$$

We designate that this expression is for $i > 1$ since it would not make sense to use a shift if there is only one approximate rate. For $i = 1$, we use the exponential distribution with rate $\hat{\lambda}_1^{(n)} = \kappa(1 \to n)^{-1} = \tau(1 \to n)^{-1}$, which is equivalent to the approximation discussed in Section III A. In Figure 4, we show the FPTD based on Eq. (26) for the same system used to produce Figure 3. As i is increased from 1 to 4, the approximation improves dramatically.

Using this approximate form for the FPTD in the expression for the MFPT of the largest cluster, given in

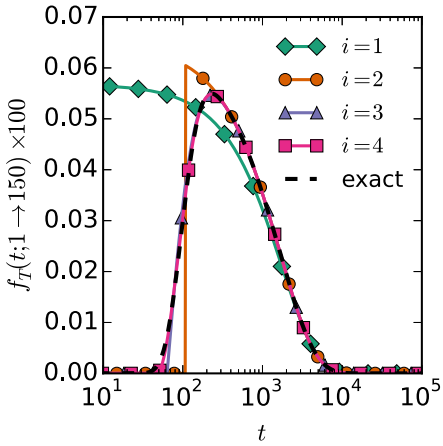

FIG. 4. The first-passage time distribution for a system with $\beta\Delta G^* = 5$, $n^* = 40$, and $f_1 = 1$. Approximations are plotted with solid lines and symbols, and the value of i denotes the number of approximate rates that were used to compute the FPTD according to Eq. (26). The $i = 1$ case corresponds to the single exponential approximation discussed in Section III A. The dashed line is the exact expression according to Eq. (8).

Eq. (12), yields

$$\tau_{\text{largest}}(1\to n) = \left(\hat{\lambda}_i^{(n)}\right)^{-1} + \int_0^\infty \left[\sum_{k=1}^{i-1}\left(e^{-\lambda_k^{(n)}s}\prod_{j=1,j\neq k}^{i-1}\frac{\hat{\lambda}_j^{(n)}}{\hat{\lambda}_j^{(n)}-\hat{\lambda}_k^{(n)}}\right)\right]^N ds. \tag{27}$$

In general, the integral in Eq. (27) must be carried out numerically. However, for the $i = 2$ approximation this integral can be found analytically, making the procedure for computing the $\tau_{\text{largest}}(1 \to n)$ curve significantly simpler. The full details of this procedure are shown in Appendix B. In general, the calculation of $\tau_{\text{largest}}(1 \to n)$ becomes increasingly accurate as i increases, although the specific number of terms required depends on the values of the parameters ΔG^*, n^*, and f_1, and the system size N. Despite the fact that i is adjustable, it is always possible to start with a fit using a small value of i, and subsequently increase i, performing fits until convergence upon a set of values for ΔG^*, n^*, and f_1 is achieved.

IV. CASE STUDY: N-EICOSANE NUCLEATION

We have applied the method described here to data obtained by molecular dynamics (MD) simulations for crystal nucleation of n-eicosane (C20) from the melt. The simulated system and force field are the same as those used by Yi and Rutledge.[34,35] The system consisted of 336 C20 molecules, with a united atom (UA) representation in which each UA represents a CH_3 or CH_2 group. The force field was initially proposed by Paul, Yoon, and Smith[36] with subsequent modifications by Waheed *et al.*,[37,38] and includes bond stretching, bond angle bending, torsion, and non-bonded Lennard-Jones forces. We used a rRESPA integration scheme[39] with a time step of 2 fs for bonded interactions and 4 fs for intermolecular interactions. All simulations were carried out using the LAMMPS software package.[40] We observed 60 independent nucleation events by first equilibrating the melt at 310 K and 1 atm, which is above the melting point, and then quenching 60 independent configurations, taken from this trajectory at intervals of 6 ns, to 260 K (17% supercooling).

To identify crystalline clusters, we first invoked the local nematic order parameter, as defined by Yi and Rutledge,[34] as a criteria for identifying crystalline UAs

$$\langle P_2\rangle_{i,local} = \frac{1}{2}(3\cos^2\theta_{ij}-1)_{j\in\Gamma_i},$$
$$\Gamma_i = \{k : |\mathbf{r}_k - \mathbf{r}_i| < r_{P2}, k \neq i\}. \tag{28}$$

In this expression θ_{ij} is the angle between chord vectors located on atoms i and j, and the average for the ith UA is taken over all atoms within a distance r_{P2}. UAs that exceeded a threshold value $P_{2,\text{th}}$ for the local order parameter were eligible to form clusters of size greater than one. Eligible UAs were determined to be part the same cluster if they were within a threshold distance r_{th} of another eligible UA. All clusters of size one were considered to be in the monomer state regardless of whether they exceeded the threshold value of $P_{2,\text{th}}$. In keeping with Yi and Rutledge,[34] we used the values $P_{2,\text{th}} = 0.4$, $r_{P2} = 2.5\sigma$, and $r_{\text{th}} = 1.3\sigma$, where $\sigma = 4.01$ Å is the van der Waals radius of a CH_3 united atom.

For each run, we monitored the size of the largest cluster in the system and computed the first-passage time curves $T_{\text{largest,sim}}(1 \to n)$ based on the earliest time that a cluster of size $\geq n$ was observed. The MFPT curve $\tau_{\text{largest,sim}}(1 \to n)$ was found by averaging the first-passage times at each value of n for all 60 runs. We also computed the standard deviation in the first-passage time $\sigma_{\text{largest,sim}}(1 \to n)$ in a similar manner and plotted both quantities as solid and dashed lines in Figure 5. We find that there is a significant discrepancy between the mean and standard deviation obtained by MD, which indicates that the FPTD for the largest cluster in this system is not well described by a single exponential, as was discussed in Section III A.

We performed a least-squares fit of the simulated MFPT curve $\tau_{\text{largest,sim}}(1 \to n)$ to $\tau_{\text{largest}}(1 \to n)$ from Eq. (27), which is the expression from the stochastic nucleation model

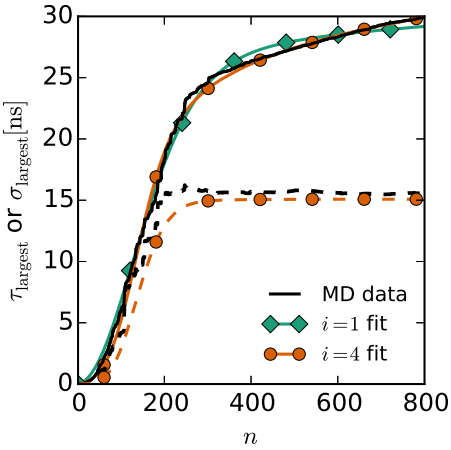

FIG. 5. The mean (solid lines) and standard deviation (dashed lines) in the first-passage time computed using MD for n-eicosane nucleation at 260 K and 1 atm, and the fitting results to Eq. (27) using $i = 1$ and $i = 4$ terms. The fits using $i = 2$ and $i = 3$ terms are nearly identical to the $i = 4$ case but have been omitted for clarity. We have also plotted the standard deviation in the first-passage time of the largest cluster for the $i = 4$ fit. In the $i = 1$ case, the standard deviation is equal to the mean first-passage time.

TABLE I. Computed kinetic parameters and rates for n-eicosane nucleation at 260 K and 1 atm. Errors in fitting parameters are given in parentheses.

i[a]	$\beta\Delta G^*$	n^*	f_1 (ns^{-1})	I_S (10^{30} m^{-3} s^{-1})	G_S (cm s^{-1})
1	4.05(0.07)	152.8(1.38)	0.029(0.001)	181	0.009 05
2	12.60(0.03)	142.7(0.73)	25.01(0.42)	367	24.5
3	12.59(0.03)	142.9(0.76)	24.48(0.45)	362	23.9
4	12.59(0.01)	142.9(0.03)	24.49(0.004)	362	23.9

[a] i is the number of rates used to approximate the FPTD.

using the approximate FPTD. For this system, the number of monomers from which clusters can form is the total number of UAs, and therefore the value $N = 6720$ was used in Eq. (27). The fitting parameters were $\beta\Delta G^*$, n^*, and f_1, and we used a number of terms ranging from $i = 1$ (the single exponential approximation) to $i = 4$. Additionally we computed the nucleation rate from Eq. (16) and the characteristic growth rate, under the assumption of a spherical nucleus, from Eq. (18) using the parameters obtained at each value of i. The value $M = 800$ was used for both rate calculations. The results are displayed in Table I, and the fits for $i = 1$ and $i = 4$ are shown in Figure 5. For the rate calculations, a density of 0.83 g cm^{-3} was used for the melt, and 0.936 g cm^{-3} for the crystal; these values were obtained from linear interpolations of the data from Yi and Rutledge.[34]

For i ranging from 2 to 4, we observed a good quality of fit to the MFPT, with little variation in the parameters. For $i = 1$, however, the best-fit parameterization was not consistent with the others, and the quality of fit was not that good. This result corroborates our earlier conclusion that the single exponential approximation is not sufficient to describe nucleation in this system. We find that in this case, the $i = 2$ approximation of the FPTD is sufficient, with slight improvement shown upon increasing the number of terms to 3 or 4. Additionally, we find that the standard deviation computed from the FPTD used to fit the MFPT corresponds well to the standard deviation in the first-passage time observed in simulations for i values greater than 1. This implies that the higher moments of the FPTD are well described by the fit, despite the fact that they did not explicitly enter into the objective function.

The non-exponential nature of the FPTD for the largest cluster is further illustrated in Figure 6, which shows a histogram of first-passage times for the largest cluster from MD for the transition n = 1 → 800, along with the approximate $f_{T,\text{largest}}(t;\ 1 \to n)$ curves resulting from the parameterizations in Table I for $i = 1$ and $i = 4$. The MD data show clear non-exponential behavior and therefore deviate strongly from the $i = 1$ parameterization. The deviation is greatest at short time, where the divergence between the single exponential distribution and the FPTD is expected to be large, as is shown in Figure 1. On the other hand, the agreement between the $i = 4$ parameterization and the MD data is very good, especially considering the fact that the $i = 4$ curve was the result of a three-parameter fit to the MFPT curve from MD, and included no higher moments of the first-passage time.

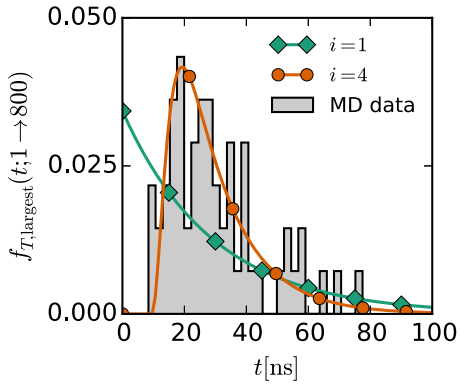

FIG. 6. The histogram of first-passage times to the critical nucleus size of n = 800 from 60 nucleation runs is plotted alongside the approximate FPTD for the largest cluster resulting from the $i = 1$ and $i = 4$ fits given in Table I.

From Figure 6 we may obtain an estimate of the time lag τ', which is defined as the time elapsed before an initially quenched system reaches a steady-state nucleation rate.[1] Once the steady-state rate is reached, nucleation events are expected to occur with exponentially distributed waiting times,[34] where the time constant is $1/I_S V$. From Figure 6, it is apparent that after waiting a period of roughly $\tau' \approx 20$ ns, corresponding to the peak in the FPTD, any subsequent nucleation events would be roughly exponentially distributed. This means that 20 ns is a reasonable estimate of the time lag. The value of the time constant $1/I_S V = 14.6$ ns, computed using the $i = 4$ approximation, is the same order of magnitude as the time lag, implying that transient nucleation is important for a system of this size. As the system size becomes larger, $1/I_S V$ decreases and the role of transient nucleation becomes more important.

The critical nucleus size of ~140 UAs at 260 K obtained by our fitting procedure was found to agree well with the value reported at 265 K by Yi and Rutledge using umbrella sampling Monte Carlo simulation.[34] The free-energy barrier of 12.59 ± 0.01 kT, on the other hand, is higher than the 9.5 ± 1 kT obtained previously. Using the relationship from classical nucleation theory, $\beta\Delta G_v/2 = \beta\Delta G^*/n^*$, we estimated from nucleation parameters the experimentally measureable quantity ΔG_v, the free energy difference per UA between melt and crystal at the crystallization temperature T. Using a relationship for the driving force for crystallization at deep supercooling[35,41] and the measured heat of fusion for C20 $\Delta H_{f,\text{C20}} = 46.17$ kJ/mol at the melting temperature $T_m = 309.2$ K,[42] we obtain

$$\frac{\beta\Delta G_v}{2} = \frac{\Delta H_{f,\text{C20}}(T_m - T)}{2mT_m{}^2R} = 0.0714 \qquad (29)$$

at $T = 260$ K, where $m = 20$ is the number of CH_2 or CH_3 groups per chain. The value obtained for $\beta\Delta G^*/n^*$ from our fitting procedure is 0.088, which deviates from this estimate by 23%. Under the conditions studied by Yi and Rutledge,[34] Eq. (29) gives a value of 0.0642, while their umbrella sampling Monte Carlo method yielded $\beta\Delta G^*/n^* = 0.036$, a discrepancy of 44%. Although classical nucleation theory and the extrapolation using Eq. (29) are not without their own uncertainties, this analysis suggests that the results of our MD study and analysis method are at least as reasonable as those reported previously, based on umbrella sampling Monte Carlo results, with regard to agreement with this experimentally measurable quantity.

The nucleation rate of 362×10^{30} m^{-3} s^{-1} at 260 K is higher than the value of 65.9×10^{30} m^{-3} s^{-1} at 250 K observed in the study by Yi and Rutledge.[34] This discrepancy cannot be due to the difference in temperature alone. We have traced this deviation in part to the overly large time step of 5 fs used in that work; such a large time step was reported by Waheed *et al.*[37] to increase the chain relaxation time by only 10%, but its effect in suppressing nucleation is apparently more dramatic, compared to the more accurate rRESPA scheme used in this work. Our estimate for nucleation rate is also higher than the rate of 105×10^{30} m^{-3} s^{-1} reported by Anwar *et al.* at 250 K.[43] In this case, we suspect that the discrepancy is due to the application of the MFPT analysis method proposed by Wedekind *et al.*[14] which employs the single-exponential approximation; as shown for our system in Table I, the single-exponential approximation results in an underestimation of the true nucleation rate. From the analysis herein, we conclude that the single exponential approximation should not hold for their conditions, given that their system is larger and the supercooling is deeper, both of which are expected to give rise to a greater degree of non-exponential behavior in the FPTD. The characteristic growth rate of 23.9 cm s^{-1} cannot be directly compared to more detailed studies of growth rates since it is an average of growth in all crystalline directions. However, our result is consistent with the study performed by Waheed and Rutledge for crystal growth of the (110) plane,[37] in which rates in the range of 4-28 cm s^{-1} were reported.

V. CONCLUSIONS

We have presented a new method to analyze nucleation from mean first-passage time data obtained by molecular dynamics simulations. We explicitly invoke a stochastic model for the nucleation process and, in contrast to prior approaches, account for the system size in a manner that does not rely on the assumption that the FPTD is exponential. Although it has been widely used in the literature, this assumption is not always justified for a given set of simulation conditions. In order for this assumption to hold, the standard deviation in the first-passage time should be approximately equal to the MFPT. This constitutes a simple test that can be applied to first-passage time data in order to determine an appropriate method for further analysis. Our method is of greatest utility where the described test fails, which in turn can be traced to cases where the free energy barrier is small and/or the system size is large.

The rigorous solution for the system size dependence of the MFPT for the stochastic model proved to be of little practical utility due to the large cost associated with eigenvalue computations. This led us to develop a diagonalization-free approximate method based on the cumulant expansion of the FPTD for a birth-death process. Using this method, $\tau_{\text{largest}}(1 \to n)$ can be determined efficiently, with little deviation from the rigorous solution. Although the approximation was developed with nucleation in mind, it could be used to evaluate the FPTD for other birth-death processes, in particular, those that include a diffusive barrier-crossing event.

By fitting $\tau_{\text{largest}}(1 \to n)$ curves from simulation to the model expression, we estimate the critical barrier height, critical nucleus size, and attachment pre-factor for a given set of nucleation conditions. These quantities can be further used to estimate the nucleation rate and a characteristic growth rate. We demonstrate this process for the nucleation of C20 crystals from the quiescent melt at 260 K, a system for which the single-exponential approximation fails. The critical barrier height does not agree particularly well with previous simulations; however, the critical nucleus size was similar, and the ratio of the critical barrier height to nucleus size proved to be more consistent with the experimental heat of fusion for *n*-eicosane than previous results. The characteristic growth rate under the assumption of a spherical nucleus showed good agreement with the previously calculated fastest growth rate in C20.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the ExxonMobil Research and Engineering Co.

APPENDIX A: DERIVATION OF THE STEADY-STATE GROWTH RATE

The growth rate is derived by translating the attachment and detachment rates for a cluster to an ideal system where the phase change occurs normal to a flat surface that separates the old phase from the new phase. In this case, the volume of the new phase can be written in terms of the cross-sectional area of the dividing surface A_c and its width l, $V = A_c l = n/\rho$ where ρ is the number density of the new phase. The growth rate G_S is the time derivative of the width

$$G_S = \frac{dl}{dt} = \frac{1}{\rho A_c}\frac{dn}{dt} = \frac{1}{\rho A_c}\left[\frac{dt}{dn}\right]^{-1}. \tag{A1}$$

In the discrete limit, the quantity dt/dn is the average time it takes to add a single monomer across a flat interface. This quantity can be expressed as $\tau_g(n \to n+1)$, the mean first-passage time from cluster size n to $n+1$ for the idealized growth process. With this substitution, our expression for the growth rate is

$$G_S \cong \frac{1}{\rho A_c}\left[\tau_g(n \to n+1)\right]^{-1}. \tag{A2}$$

Since $\tau_g(n \to n+1)$ is based on attachment and detachment processes for a flat surface, it is different from the MFPT for cluster formation. For a cluster, the monomer attachment rate scales with the surface area of the cluster. Assuming that the cluster shape is constant, the surface area is related to the cluster size by some factor A_1 where $A(n) = A_1 n^{2/3}$. Therefore, the rate of addition to a cluster per unit area is $f(n)/A(n)$, and the attachment rate for a flat surface of size A_c is

$$f_g(n) = \frac{f_1 A_c}{A_1}. \tag{A3}$$

Since the surface area is constant, there is no variable interfacial dependence to the free energy. The interfacial energy term in the capillary approximation may be eliminated, leaving only the bulk driving force μ

$$\beta \Delta G_g(n) = -\beta \mu n. \tag{A4}$$

Based on the free energy and attachment rate for growth, $\tau_g(n \to n+1)$ follows from the standard expression for a birth-death process in Eq. (13)

$$\tau_g(n \to n+1) = \tau_g(1 \to n+1) - \tau_g(1 \to n) = \frac{e^{\Delta G_g(n)}}{f_g(n)} \sum_{k=1}^{n} e^{-\Delta G_g(n)}. \quad \text{(A5)}$$

This mean first-passage time contains a reflecting boundary condition at $n = 1$; however, we are actually interested in the limiting growth rate that is achieved infinitely far away from this boundary. Taking the limit as $n \to \infty$, we find that $\tau_g(n \to n+1)$ can be expressed as a convergent sum that does not depend on n

$$\tau_g(n \to n+1) = \frac{A_1}{f_1 A_c} \sum_{k=0}^{\infty} e^{\Delta G_g(k)} = \frac{A_1}{f_1 A_c} \left(\frac{1}{1 - \exp[-\beta\mu]} \right). \quad \text{(A6)}$$

Combining Eqs. (A6) and (A2) yields the growth rate given in Eq. (17)

$$G_S = \frac{f_1}{\rho A_1} \left(1 - \exp[-\beta\mu]\right). \quad \text{(A7)}$$

APPENDIX B: MFPT OF THE LARGEST CLUSTER USING THE i = 2 APPROXIMATION FOR THE FPTD

A two term approximation to the FPTD yields a significant improvement over the single exponential approximation without adding a great deal of complexity to the computation of the MFPT of the largest cluster. Following the procedure in Section III B, we first compute the first two cumulants of $T(1 \to n)$. The first cumulant is $\kappa_1(1 \to n) = \tau(1 \to n)$ and is given in Eq. (13). The second cumulant is the variance and can be computed using the iterative method from moment calculations outlined by Gillespie,[25] or alternatively using the following form from Jouini and Dallery:[44]

$$\kappa_2(1 \to n) = \sum_{l=2}^{n} \left(\frac{2e^{\beta\Delta G(l-1)}}{f(l-1)} \sum_{m=2}^{l-1} \frac{e^{\beta\Delta G(m-1)}}{f(m-1)} \left(\sum_{p=1}^{m-1} e^{-\beta\Delta G(p)} \right)^2 + \frac{e^{2\beta\Delta G(l-1)}}{f(l-1)^2} \left(\sum_{p=1}^{l-1} e^{-\beta\Delta G(p)} \right)^2 \right). \quad \text{(B1)}$$

From these two cumulants, the estimates of the two smallest rates in the FPTD follow from Eq. (24)

$$\hat{\lambda}_1^{(n)} = \kappa_2(1 \to n)^{-1/2},$$
$$\hat{\lambda}_2^{(n)} = \left[\kappa_1(1 \to n) - \kappa_2(1 \to n)^{-1/2}\right]^{-1}. \quad \text{(B2)}$$

The second rate $\hat{\lambda}_2^{(n)}$ represents a shift, leaving the FPTD with the form of a shifted exponential distribution ($s = t - 1/\hat{\lambda}_2^{(n)}$)

$$f_T(s; 1 \to n) = \hat{\lambda}_1^{(n)} e^{-\lambda_1^{(n)} s}. \quad \text{(B3)}$$

Since the shift is not affected by the size of the system N, the FPTD for the largest cluster is also a shifted exponential distribution

$$f_{T,\text{largest}}(s; 1 \to n) = N\hat{\lambda}_1^{(n)} e^{-N\lambda_1^{(n)} s}. \quad \text{(B4)}$$

This form allows the MFPT of the largest cluster to be calculated without numerical integration of Eq. (27) as

$$\tau_{\text{largest}}(1 \to n) = 1/\lambda_2^{(n)} + 1/(N\lambda_1^{(n)}). \quad \text{(B5)}$$

Better accuracy can be obtained by using a FPTD approximation with $i > 2$. However, for some nucleation conditions, including the n-eicosane case study discussed in Section IV, there is only a slight improvement in the convergence of the estimated parameters upon using the higher order approximation.

[1] P. Yi and G. C. Rutledge, Annu. Rev. Chem. Biomol. Eng. **3**, 157 (2012).
[2] J. Anwar and D. Zahn, Angew. Chem., Int. Ed. **50**, 1996 (2011).
[3] K. Yasuoka and M. Matsumoto, J. Chem. Phys. **109**, 8451 (1998).
[4] M. Matsumoto, S. Saito, and I. Ohmine, Nature **416**, 409 (2002).
[5] K. Esselink, P. A. J. Hilbers, and B. W. H. van Beest, J. Chem. Phys. **101**, 9033 (1994).
[6] T. Koishi, K. Yasuoka, and T. Ebisuzaki, J. Chem. Phys. **119**, 11298 (2003).
[7] C. H. Bennett, *Algorithms for Chemical Computations* (American Chemical Society, 1977), pp. 63–97.
[8] D. Chandler, J. Chem. Phys. **68**, 2959 (1978).
[9] A. Laio and F. L. Gervasio, Rep. Prog. Phys. **71**, 126601 (2008).
[10] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Annu. Rev. Phys. Chem. **53**, 291 (2002).
[11] D. Kashchiev, Cryst. Res. Technol. **19**, 1413 (1984).
[12] D. Reguera, J. Rubí, and A. Pérez-Madrid, Physica A **259**, 10 (1998).
[13] L. S. Bartell and D. T. Wu, J. Chem. Phys. **125**, 194503 (2006).
[14] J. Wedekind, R. Strey, and D. Reguera, J. Chem. Phys. **126**, 134103 (2007).
[15] G. Chkonia, J. Wölk, R. Strey, J. Wedekind, and D. Reguera, J. Chem. Phys. **130**, 064505 (2009).
[16] A. V. Mokshin and B. N. Galimzyanov, J. Chem. Phys. **140**, 024104 (2014).
[17] D. T. Gillespie, J. Chem. Phys. **74**, 661 (1981).
[18] C. R. Doering, K. V. Sargsyan, L. M. Sander, and E. Vanden-Eijnden, J. Phys.: Condens. Matter **19**, 065145 (2007).
[19] A. S. Novozhilov, Briefings Bioinf. **7**, 70 (2006).
[20] W. J. Shugard and H. Reiss, J. Chem. Phys. **65**, 2827 (1976).
[21] O. Penrose, Commun. Math. Phys. **124**, 515 (1989).
[22] D. B. Duncan and A. R. Soheili, Appl. Numer. Math. **37**, 1 (2001).
[23] R. Becker and W. Döring, Ann. Phys. **416**, 719 (1935).
[24] D. Kashchiev, J. Chem. Phys. **127**, 064505 (2007).
[25] D. T. Gillespie, *Markov Processes: An Introduction for Physical Scientists* (Academic, 1992).
[26] S. Karlin and J. McGregor, Pac. J. Math. **9**, 1109 (1959).
[27] Y. Gong, Y.-H. Mao, and C. Zhang, J. Theor. Probab. **25**, 950 (2012).
[28] G. Bolch, S. Greiner, H. de Meer, and K. S. Trivedi, *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications* (Wiley, 2006).
[29] D. Kashchiev, *Nucleation* (Elsevier, 2000).
[30] P. Hänggi and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[31] D. Kashchiev, *Nucleation: Basic Theory with Applications* (Butterworth-Heinemann, Oxford, 2000).
[32] V. A. Shneidman, J. Chem. Phys. **141**, 051101 (2014).
[33] P. J. Smith, Am. Stat. **49**, 217 (1995).
[34] P. Yi and G. C. Rutledge, J. Chem. Phys. **135**, 024903 (2011).
[35] P. Yi, C. R. Locker, and G. C. Rutledge, Macromolecules **46**, 4723 (2013).
[36] W. Paul, D. Y. Yoon, and G. D. Smith, J. Chem. Phys. **103**, 1702 (1995).
[37] N. Waheed, M. S. Lavine, and G. C. Rutledge, J. Chem. Phys. **116**, 2301 (2002).
[38] N. Waheed, M. J. Ko, and G. C. Rutledge, Polymer **46**, 8689 (2005).
[39] M. E. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **97**, 1990 (1992).
[40] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).
[41] J. D. Hoffman and J. J. Weeks, J. Chem. Phys. **37**, 1723 (1962).
[42] P. Claudy and J. M. Létoffe, Calorim. Anal. Therm. **22**, 281 (1991) (in French).
[43] M. Anwar, J. T. Berryman, and T. Schilling, J. Chem. Phys. **141**, 124910 (2014).
[44] O. Jouini and Y. Dallery, Math. Methods Oper. Res. **68**, 49 (2007).