Stochastic self-assembly of incommensurate clusters

M. R. D'Orsogna,[1] G. Lakatos,[2] and T. Chou[3,a)]

[1]*Department of Mathematics, CSUN, Los Angeles, California 91330-8313, USA*
[2]*Department of Chemistry, The University of British Columbia, Vancouver, BC V6T-1Z1, Canada*
[3]*Departments of Biomathematics and Mathematics, UCLA, Los Angeles, California 90095-1766, USA*

(Received 11 January 2012; accepted 7 February 2012; published online 29 February 2012)

Nucleation and molecular aggregation are important processes in numerous physical and biological systems. In many applications, these processes often take place in confined spaces, involving a finite number of particles. Analogous to treatments of stochastic chemical reactions, we examine the classic problem of homogeneous nucleation and self-assembly by deriving and analyzing a fully discrete stochastic master equation. We enumerate the highest probability steady states, and derive exact analytical formulae for quenched and equilibrium mean cluster size distributions. Upon comparison with results obtained from the associated mass-action Becker-Döring equations, we find striking differences between the two corresponding equilibrium mean cluster concentrations. These differences depend primarily on the divisibility of the total available mass by the maximum allowed cluster size, and the remainder. When such mass "incommensurability" arises, a single remainder particle can "emulsify" the system by significantly broadening the equilibrium mean cluster size distribution. This discreteness-induced broadening effect is periodic in the total mass of the system but arises even when the system size is asymptotically large, provided the ratio of the total mass to the maximum cluster size is finite. Ironically, classic mass-action equations are fairly accurate in the coarsening regime, before equilibrium is reached, despite the presence of large stochastic fluctuations found via kinetic Monte-Carlo simulations. Our findings define a new scaling regime in which results from classic mass-action theories are qualitatively inaccurate, even in the limit of large total system size. [http://dx.doi.org/10.1063/1.3688231]

I. INTRODUCTION

Self-assembly arises in countless physical and biological processes, over many length and time scales.[1] Atoms and molecules can nucleate to form small multiphase structures that can influence overall bulk material properties. For example, adatoms adsorbed on growing surfaces aggregate to form islands whose shapes and characteristics control epitaxial synthesis of thin films and layered materials. In more recent years, advances in nanotechnology have opened the possibility of controlling the self-assembly of specifically designed mesoscopic[2] and macroscopic[3] parts into functional components. Because of the importance of nucleation and self-assembly in material science, these processes have been extensively studied.[4–6]

Nucleation and self-assembly are also ubiquitous in cellular biology. The polymerization of actin filaments[7–11] and amyloid fibrils,[12] the assembly of virus capsids[13–17] and of antimicrobial peptides into transmembrane pores,[18,19] the recruitment of transcription factors, and the self-assembly of clathrin-coated pits[20–22] are all important cell-level processes that can be cast as initial binding and self-assembly problems for which there is great interest in developing theoretical tools.

Because it is such a fundamental process across so many disciplines, there is a vast, longstanding literature on nucleation and self-assembly, from both the theoretical and experimental perspectives.[23] Theoretical models were initially developed using mass-action kinetics, as exemplified by the Becker-Döring (BD) equations describing the evolution of the *mean* concentrations of clusters of a given size.[24] These well-studied mass-action equations implicitly employ the mean-field assumption by neglecting correlations. Notwithstanding, solutions to the BD equations exhibit rich behavior, including metastable particle distributions,[25] multiple time scales,[26] and nontrivial convergence to equilibrium and coarsening.[25,27,28] Becker-Döring-type models implicitly assume infinite system sizes and/or the possibility of infinitely large cluster sizes, without addressing the importance of how these limits are taken. To date, all the biological and physical applications described above have been modeled almost exclusively using BD-type equations. As a result the discreteness and cluster "stoichiometry" in these problems have not been explored.

In this paper, we carefully investigated a simple homogeneous nucleation and growth process starting from the fundamental, multi-dimensional, fully stochastic master equation. In particular, we consider the *probability* of the system to be in a state with specific numbers of clusters of each size. The fully stochastic master equation governing the evolution of the state probabilities is derived, simulated, and solved analytically in different steady-state limits. Upon comparing the mean cluster concentrations found from the stochastic master equation with those calculated from numerical integration of the mean-field BD equations, we find qualitative differences, even in the large system size limit. Our results especially

[a)] Author to whom correspondence should be addressed. Electronic mail: tomchou@ucla.edu.

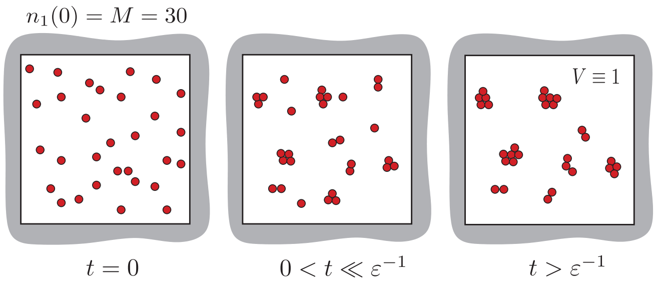

FIG. 1. Homogeneous nucleation and growth in a closed unit volume initiated with $M = 30$ monomers. If the constant monomer detachment rate q is small, monomers will be nearly exhausted in the long time limit. In this example, the final cluster size distribution consists of two dimers, one trimer, one 4-mer, one pentamer, and two hexamers. In this paper we will consider the limit of slow, but non-zero monomer detachment rates $q/p \equiv \varepsilon \ll 1$.

highlight the importance of system size in self-assembly, and how they lead to cluster concentrations dramatically different from those obtained by solving classical, mean-field BD equations.

We begin by considering the simple homogeneous nucleation and growth process in a closed system containing M total particles, as depicted in Fig. 1. Monomers first bind together to form dimers. Large clusters are formed by successive one-by-one binding of monomers. Individual monomers can also desorb shrinking clusters. Depending on the molecular details and application, different monomer binding and unbinding rate structures,[25,26,28,29] cluster fragmentation/coagulation rules,[24,30] and monomer sources[26,31,32] have been considered. For example, in epitaxial growth, atoms may be continuously deposited on the substrate, giving rise to a constant flux of adatoms (monomers) that contribute to island nucleation and growth. In other applications, self-assembly may occur in small volumes, such as during the assembly of biomolecular aggregates inside cells. Here, monomer production may be slow compared to the growth dynamics and the total number of monomers, bound and part of clusters, can be assumed constant.

Especially within biophysics, cluster sizes are usually limited by the finite total mass of the system, or by some intrinsic cluster stoichiometry. For example, virus capsids, clathrin coated pits, and antimicrobial peptide pores typically consist of $N \sim 100 - 1000$, $N \sim 10 - 20$, and $N \sim 5 - 8$ molecular subunits, respectively. By defining the self-assembly problem with fixed total mass and a maximum cluster size, we can now study the process under different limits of total mass M and maximum cluster size N, and clarify how different large system size limits can be properly taken, finding unexpected results. Our analysis highlights the importance of stochastics and particle discreteness in a wide class of applications of self-assembly models.

In general, the monomer attachment and detachment rates p_k and q_k depend on the cluster size k. For example, clusters that grow spherically can be modeled by attachment and detachment rates that are proportional to the cluster surface area, $p_k, q_k \sim k^{2/3}$. Other scalings for attachment and detachment rates have also been used to model specific physical limits of nucleation.[28] For the sake of simplicity we assume in this work that monomer attachment and detachment occur at constant, cluster size-independent rates p and q, respectively. This scaling would be appropriate for, say, the growth of a one-dimensional filament, where attachment and detachment always occur at only one or both ends. Henceforth, we will also focus on the case $p \gg q$, as this strong binding limit best reveals the importance of stochasticity and finite-size effects in this class of models.

II. MASS-ACTION TREATMENT

We first briefly review classical mass-action nucleation theory. Since the underlying assumption is that all clusters are well-mixed, we assume, without loss of generality, a closed system of fixed unit volume. The mass-action BD equations can be written in dimensionless form:

$$
\begin{aligned}
\dot{c}_1(t) &= -c_1^2 - c_1 \sum_{j=2}^{N-1} c_j + 2\varepsilon c_2 + \varepsilon \sum_{j=3}^{N} c_j,\\
\dot{c}_2(t) &= -c_1 c_2 + \frac{1}{2} c_1^2 - \varepsilon c_2 + \varepsilon c_3,\\
\dot{c}_k(t) &= -c_1 c_k + c_1 c_{k-1} - \varepsilon c_k + \varepsilon c_{k+1},\\
\dot{c}_N(t) &= c_1 c_{N-1} - \varepsilon c_N,
\end{aligned}
\tag{1}
$$

where $\varepsilon \equiv q/p$. For a system defined with fixed unit volume, the c_k in Eqs. (1) represent the mean number of clusters of size k. However, to better distinguish the solutions of BD equations from the mean cluster numbers derived from subsequent stochastic analyses, we shall continue to refer to $c_k(t)$ as "concentrations." Note that setting $\varepsilon = 0$ precludes monomer detachment, giving rise to irreversibility and ergodicity-breaking in the evolution to the final cluster size distribution.[33] When $0 < \varepsilon \ll 1$, detachment occurs, and a true equilibrium cluster size distribution will be slowly reached. Since we also assume mass conservation, the constraint $\sum_{k=1}^{N} k c_k = M$ is also imposed to fix the total number of monomer particles, whether free or part of clusters, to M. We also use the initial condition $c_k(t = 0) = M\delta_{k,1}$ (where the Kroenecker δ-function $\delta_{i,j} = 1$ if $i = j$, and zero otherwise), corresponding to all mass in the form of monomers at $t = 0$.

Each term in the BD equations represents a particular attachment or detachment event. For example, the $-c_1^2$ term on the RHS of the first equation arises from the $\sim c_1^2/2$ ways of dimerizing two monomers so that the time rate of change of $c_1(t)$ is proportional to $-2c_1^2/2 = -c_1^2$. Equations (1) have been solved numerically and are extensively analyzed in asymptotic limits under the total mass constraint.[25–28]

Figure 2 plots the numerical solution to Eqs. (1) as a function of time for $N = 4$ and $\varepsilon = 10^{-5}$. In general, $c_{k>1}(t)$ initially rise at the expense of $c_1(t)$. After monomers are significantly depleted, the mean cluster size distributions remain nearly constant at "quenched" or "metastable" values c_k^*. Since detachment is slow ($\varepsilon \ll 1$), the long-lived values c_k^* correspond to a quenched size distribution before detachment and redistribution of mass have had time to occur appreciably. After a long time $t_c \sim \varepsilon^{-1}$, this metastable size distribution begins to "coarsen," eventually reaching an equilibrium distribution $c_k^{\rm eq}$.

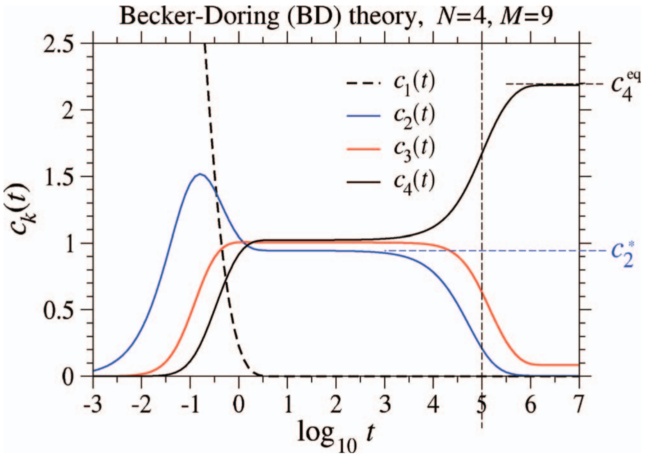

FIG. 2. Numerical solution of Eqs. (1) for the cluster concentrations $c_k(t)$, plotted as a function of $\log_{10}t$. In this example, $N = 4$, $M = 9$, and $\varepsilon = 10^{-5}$. Analytic expressions for both the intermediate-time metastable concentrations c_k^*, and the long-time equilibrium concentrations $c_k^{\rm eq}$ (indicated at the right for $k = 2$ and $k = 4$, respectively) can be found in Sec. 1 of Appendix. The typical coarsening time $t_c \sim \varepsilon^{-1}$ is indicated in this plot by the vertical dashed line at $\log_{10}(\varepsilon^{-1}) = 5$. In the mass-action BD theory, $c_k(t)/M$ is nearly independent of M (see Fig. 3(c)).

The metastable concentrations c_k^* can be computed exactly in the $\varepsilon \to 0^+$ asymptotic limit by directly setting $\varepsilon = 0$ in Eqs. (1). In this case, Eqs. (1) can be linearized via a proper time rescaling, effectively eliminating the multiplicative term c_1 from the RHS. The mathematical details are outlined in Sec. 1 of Appendix, which also shows that $c_k^* \propto M$.

The equilibrium concentrations $c_k^{\rm eq}$ for small ε can be found by setting $\dot c_k = 0$ in Eqs. (1) and perturbatively computing the resulting algebraic equations. As described in Sec. 1 of Appendix this leads to

$$c_k^{\rm eq} \approx \frac{\varepsilon}{2}\left(\frac{2M}{\varepsilon N}\right)^{k/N}\left[1 - \frac{k(N-1)}{N^2}\left(\frac{\varepsilon N}{2M}\right)^{1/N} + \ldots\right]. \tag{2}$$

At equilibrium, we find $c_k^{\rm eq} \gg c_{k-1}^{\rm eq}$, and for $\varepsilon \ll 1$, the maximal cluster of size N dominates with concentration $c_N^{\rm eq} \approx M/N$, while $c_{k<N}^{\rm eq} \sim \varepsilon^{1-k/N} \approx 0$ as $\varepsilon \to 0^+$. Since clusters are arrested from further growth at size N and detachment is slow, the system is driven towards having almost all of its mass in the largest clusters.

Upon closer inspection, a fundamental inconsistency emerges. The solution $c_k^{\rm eq} \approx (M/N)\delta_{k,N}$ cannot be appropriate if $M < N$, when there is insufficient mass to ever form a single maximal cluster. Since the classic BD model does not restrict cluster sizes ($N \to \infty$), the emergence of asymptotically large clusters requires either the system volume or the mass M to diverge. If volume diverges, spatial homogeneity and the well-mixed assumption can easily break down. If on the other hand, one sets the mass $M > N$, there is a question as to how to take the limits $M \to \infty$ and $N \to \infty$. Recognizing these obvious deficiencies in BD-type mass-action equations, we now perform a more careful analysis of the fully discrete stochastic master equation to reveal dramatic differences between mass-action and stochastic results, even in the large system limit.

III. STOCHASTIC ANALYSIS

Consider the probability density $P(\{n\}; t) \equiv P(n_1, n_2, \ldots, n_N; t)$ of the system being in a state with n_1 monomers, n_2 dimers, n_3 trimers, …, n_N N-mers. Using the above definition $\varepsilon \equiv q/p$, the full stochastic master equation describing the time evolution of $P(\{n\}; t)$ is[31,32]

$$\begin{aligned}\dot P(\{n\};t) = &-\Lambda(\{n\})P(\{n\};t)\\ &+\frac{1}{2}(n_1+2)(n_1+1)W_1^+W_1^+W_2^-P(\{n\};t)\\ &+\varepsilon(n_2+1)W_2^+W_1^-W_1^-P(\{n\};t)\\ &+\sum_{i=2}^{N-1}(n_1+1)(n_i+1)W_1^+W_i^+W_{i+1}^-P(\{n\};t)\\ &+\varepsilon\sum_{i=3}^{N}(n_i+1)W_1^-W_{i-1}^-W_i^+P(\{n\};t),\end{aligned} \tag{3}$$

where $P(\{n\}, t) = 0$ if any $n_i < 0$, $\Lambda(\{n\}) = \frac{1}{2}n_1(n_1-1) + \sum_{i=2}^{N-1} n_1 n_i + \varepsilon\sum_{i=2}^{N} n_i$ is total rate out of configuration $\{n\}$, and $W_j^\pm$ is the unit raising/lowering operator on the number of clusters of size j. For example,

$$\begin{aligned}&W_1^+W_i^+W_{i+1}^-P(\{n\};t)\\ &\quad\equiv P(n_1+1,\ldots,n_i+1,n_{i+1}-1,\ldots;t).\end{aligned} \tag{4}$$

To be consistent with the analysis of the BD equations of Sec. II, we will assume that all the mass is initially in the form of monomers: $P(\{n\}; t = 0) = \delta_{n_1,M}\delta_{n_2,0}\cdots\delta_{n_N,0}$. By construction, the stochastic dynamics described by Eq. (3) obey the total mass conservation constraint

$$M = \sum_{k=1}^{N} k n_k. \tag{5}$$

The mean numbers of clusters of size k, defined by $\langle n_k(t)\rangle \equiv \sum_{\{n\}} n_k P(\{n\}; t)$ are the direct counterparts to the mean cluster "concentrations" $c_k(t)$ obtained from numerical solutions of the BD equations in Sec. II.

We first simulate the stochastic master equation using a kinetic Monte-Carlo (KMC) or residence time algorithm as described by Bortz *et al.*[34] and detailed in Sec. 2 of Appendix. Figure 3 plots mean cluster numbers $\langle n_k(t)\rangle$ derived from simulations of Eq. (3) with $N = 8$, $M = 16$, 17, and $\varepsilon = 10^{-6}$. Also shown for comparison are the mean-field results $c_k(t)$ (thin, dashed curves) for expected cluster concentrations $c_k(t)$ derived from numerical solutions of the BD equations. Our results show that during short and intermediate times $t \lesssim \varepsilon^{-1}$, there is little difference between the predictions for $M = 16$ and $M = 17$. Moreover, the mass-action concentrations $c_k(t) \approx \langle n_k(t)\rangle$.

The most striking differences between the $M = 16$ and $M = 17$ cases occur at long times $t \gg \varepsilon^{-1}$. For $M = 2N = 16$ (Fig. 3(a)) the mass-action solution $c_k^{\rm eq}$ roughly approximates $\langle n_k(t)\rangle$, while for $M = 17$, there is a dramatic difference between $c_k^{\rm eq}$ and the asymptotically exact mean numbers $\langle n_k^{\rm eq}\rangle$. Figure 3(c) highlights the differences between $c_k^{\rm eq}$ and $\langle n_k^{\rm eq}\rangle$, particularly for $k = N = 8$ (red curves). The approximation

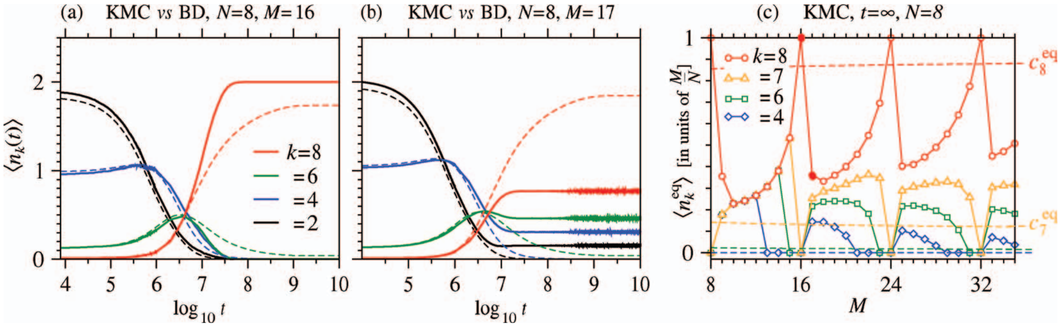

FIG. 3. Mean cluster sizes $\langle n_k(t)\rangle$ obtained from averaging 10^6 KMC simulations of a stochastic self-assembly process with $\varepsilon = 10^{-6}$. Only $k = 2, 4, 6, 8$ are displayed. (a) For $N = 8$ and $M = 16$, nearly all the mass is concentrated in $\langle n_8^{eq}\rangle \approx 2$ at equilibrium. (b) For $N = 8$, $M = 17$, a much broader equilibrium mean cluster distribution arises. For comparison, the numerical solution for $c_k(t)$ from the BD equations is displayed by the dashed curves. The simulation and mean-field results agree well with each other, but diverge at long times as equilibrium is approached, particularly when the total mass M is indivisible by N. Interestingly, even though the BD equations are most accurate for short and intermediate times, it is during the intermediate metastable regime that the variations in the cluster concentrations are largest across simulation trajectories (see Sec. 2 of Appendix). (c) The difference between c_k^{eq} and $\langle n_k^{eq}\rangle$ (plotted in units of M/N here) is highlighted as a function of M. The red dashed line corresponds to c_8^{eq} (which is nearly independent of M), while the open circles correspond to $\langle n_8^{eq}\rangle$ found from Monte-Carlo simulation. Note that $c_8^{eq} \sim \langle n_8^{eq}\rangle$ only when $\varepsilon \to 0^+$ and M is divisible by $N = 8$, or when $M \to \infty$. The filled red circles correspond to $M = 16$ and $M = 17$ as detailed in (a) and (b), respectively. A few other mean concentrations, $\langle n_{4,6,7}^{eq}\rangle$, along with the corresponding $c_{4,6,7}^{eq}$ (dashed lines) are also plotted for reference.

$c_k^{eq} \sim \langle n_k^{eq}\rangle$ is qualitatively reasonable only when M is divisible by N, or when M is very large. Even in this case, c_k^{eq} converges slowly to $\langle n_k^{eq}\rangle$ as $\varepsilon \to 0^+$, especially for large k. For example, c_N^{eq} converges to $\langle n_N^{eq}\rangle$ but with a remaining error term of order $\varepsilon^{1/N}$.

To quantitatively understand these differences, and how stochastic and finite-size effects influence the self-assembly process, we must find analytic ways of computing the relevant configurational probabilities $P(\{n\}; t \to \infty)$ obeying the master equation. Since we assume that detachment is slow, the most highly weighted equilibrium configurations are those with the fewest total number of clusters. For each set $\{M, N\}$, we can thus enumerate the states with the lowest number of clusters and use detailed balance to compute their relative weights.

As an explicit example, consider the possible states for the simple case $N = 4$, $M = 9$ shown in Fig. 4. When $0 < \varepsilon \ll 1$, nearly all the weight settles into states with the lowest number of clusters ($\mathcal{N}_{\min} = 3$ here). As detailed in Sec. 3 of Appendix, applying detailed balance between the $\mathcal{N}_{\min} = 3$ and $\mathcal{N}_{\min} + 1 = 4$ states, neglecting corrections of $O(\varepsilon)$, we find $\langle n_1\rangle \approx \langle n_2\rangle \approx 6/13$, $\langle n_3\rangle \approx 9/13$, and $\langle n_4\rangle \approx 18/13$.

To extend our results to general M and N, we start from the state with the highest possible number of maximum-sized clusters, given by the integer part of $[M/N]$, and distribute the remaining particles among the smaller ones. The number of largest clusters is then successively reduced until all mass is exhausted. In this way, we inductively enumerate all states with near minimal total numbers of clusters. We then use detailed balance to compute the relative equilibrium weights of these few-cluster states and find closed-form solutions for the mean equilibrium cluster numbers $\langle n_k^{eq}\rangle$. In order to write our analytic solutions we consider the following representation of mass $M = \sigma N - j$ where σ denotes the maximum possible number of largest clusters, and $0 \leq j \leq N - 1$ represents the remainder of M/N. We thus arrive at one of the main results of this paper: exact solutions to the expected equilibrium cluster numbers in the $\varepsilon \to 0^+$ limit:

$$\langle n_N^{eq}\rangle = \frac{\sigma(\sigma-1)}{(\sigma+j-1)}, \tag{6}$$

$$\langle n_{N-k}^{eq}\rangle = \frac{\sigma(\sigma-1)j(j-1)\ldots(j-k+1)}{(\sigma+j-1)(\sigma+j-2)\ldots(\sigma+j-k-1)}.$$

These expressions are valid for $0 \leq j < N - 1$ and all k. Note that within all minimum cluster number states, the smallest cluster that can be formed is of size $N - j$. Therefore, $\langle n_{N-k}^{eq}\rangle = 0$ for $k > j$, which is automatically satisfied by Eq. (6). In the special case $j = N - 1$, the total mass can also be expressed as $M = \sigma N - (N - 1) = (\sigma - 1)N + 1$. Therefore, $j = N - 1$ corresponds to *adding* a single monomer to a

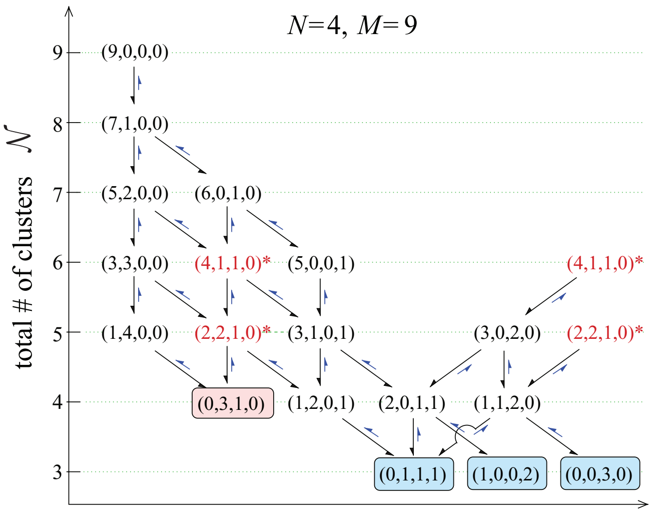

FIG. 4. Enumeration of the configurations (n_1, n_2, n_3, n_4) for the $N = 4$, $M = 9$ case. Only three distinct states with a minimum number of clusters $\mathcal{N}_{\min} = 3$ arise. These states are all connected by monomer attachment/detachment steps with states having $\mathcal{N}_{\min} + 1 = 4$ clusters. By identifying the states that connect the three minimum cluster states, we can compute the weights of each of these states in the $\varepsilon \to 0^+$ limit. If $\varepsilon = 0$, the system loses ergodicity and appreciable probability will be forever trapped in state (0, 3, 1, 0), leading to a very different metastable, pre-coarsening distribution $\langle n_k^*\rangle$.

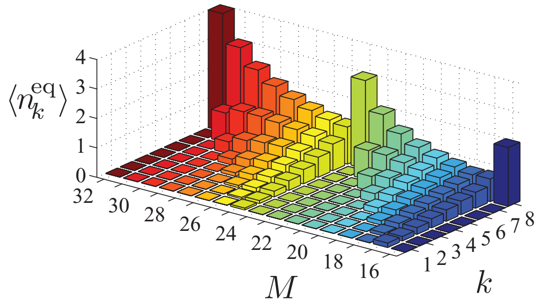

FIG. 5. The expected equilibrium cluster numbers $\langle n_k^{\mathrm{eq}} \rangle$ in the $\varepsilon \to 0^+$ limit, plotted as functions of $1 \le k \le N = 8$ and total mass M.

system with initially $M = (\sigma - 1)N$ monomers. In this case, the formulae for the equilibrium concentrations must take into account combinatoric factors of 2 that arise when monomers appear in the populated configurations. A careful enumeration for $j = N - 1$ yields

$$\begin{aligned} \langle n_1^{\mathrm{eq}} \rangle &= \frac{2(N-1)!}{D(\sigma, N-1)}, \\ \langle n_{N-k}^{\mathrm{eq}} \rangle &= \frac{\prod_{\ell=1}^{k}(N-\ell)\prod_{i=1}^{N-k-1}(\sigma - 2 + i)}{D(\sigma, N-1)}, \\ \langle n_N^{\mathrm{eq}} \rangle &= (\sigma - 1)\frac{D(\sigma - 1, N-1)}{D(\sigma, N-1)}, \end{aligned} \tag{7}$$

where $D(\sigma, j) \equiv j! + \prod_{\ell=1}^{j-1}(\sigma + \ell)$. Equations (6) and (7) are asymptotically exact in the $\varepsilon \to 0^+$ limit. We verified that results derived from our Monte-Carlo simulations match closely to these formulae for small ε.

Figure 5 plots our exact result $\langle n_k^{\mathrm{eq}} \rangle$ for $N = 8$ and varying total mass M. The previously analyzed cases $M = 16$, 17 are represented by the last two rows to the right. When $j = 0$, such as in the $M = 16$ case, the maximum cluster size N is commensurate with the total mass M, so all the mass is deposited into the largest clusters. The inaccuracy of the BD mass-action equations can be readily seen by considering the incommensurate case where $j > 0$ and the total monomer number M is not an integer multiple of the maximum cluster size N. Recall that from Eq. (2) the $\varepsilon \to 0^+$ limit of the mass-action equations always give $c_{k<N}^{\mathrm{eq}} \sim \varepsilon^{1-k/N} \ll 1$ and $c_N^{\mathrm{eq}} \approx M/N$. However, If M is not divisible by N, there will necessarily be remaining monomers that form smaller clusters. The number of ways these remaining monomers can be combined into smaller clusters can become large, thereby generating a broad distribution of cluster sizes. Thus, whereas the BD results imply the predominance of the largest cluster size, our stochastic analysis shows the emergence of a much broader distribution.

For example, let us add a single monomer to the previously analyzed (Fig. 3(a)) state $N = 8$, $M = 16$ ($\sigma = 2, j = 0$). When $M = 17$, our formulae in Eqs. (7) can be used by setting $\sigma = 3$ and $j = N - 1 = 7$. Note that by adding just a *single* monomer, the mean cluster size distribution, which for $M = 16$ was concentrated into the largest cluster, disperses and nearly uniformly populates all cluster sizes. This dispersal effect arises from the incommensurability of the ratio of total mass to maximum cluster size. In our $N = 8$, $M = 17$ example, (1, 0, 0, 0, 0, 0, 0, 2) is clearly one possible state with the lowest number of clusters $\mathcal{N}_{\min} = 3$. However, as long as dissociation is allowed ($\varepsilon > 0$) a large number (in this case 7) of additional nontrivial 3-cluster states are possible:

$$\begin{array}{lcl} (0, 1, 0, 0, 0, 0, 1, 1) & & (0, 0, 0, 1, 0, 1, 1, 0) \\ (0, 0, 1, 0, 0, 1, 0, 1) & \text{and} & (0, 0, 0, 0, 2, 0, 1, 0) \\ (0, 0, 0, 1, 1, 0, 0, 1) & & (0, 0, 1, 0, 0, 0, 2, 0) \\ & & (0, 0, 0, 0, 1, 2, 0, 0). \end{array} \tag{8}$$

The equilibrium weights of these 8 new states are comparable, resulting in a very flat mean cluster size distribution as shown in Fig. 5. When $j < N - 1$, or when σ is large, this dispersal effect diminishes. From Eqs. (7), $\langle n_{N-1}^{\mathrm{eq}} \rangle / \langle n_N^{\mathrm{eq}} \rangle \sim N^2/M$, showing that the BD result $c_k^{\mathrm{eq}} \sim (M/N)\delta_{k,N}$ is asymptotically accurate when $M \gg N^2$, or equivalently, when $\sigma \gg N$. Thus, the periodically varying curve $(N/M)\langle n_k^{\mathrm{eq}} \rangle$ in Fig. 3(c) asymptotes to the mass-action result ($\varepsilon \to 0^+$) as $M/N^2 \to \infty$.

IV. SUMMARY AND CONCLUSIONS

We have formulated the fully stochastic model of homogeneous nucleation and self-assembly. From the associated stochastic master equation, analytic results for the expected equilibrium and cluster numbers were computed, verified with KMC simulations, and compared with those derived from the mass-action Becker-Döring equations. At intermediate times, we find that the mean cluster concentrations derived from mean-field (mass-action BD equations) and stochastic treatments are qualitatively similar. Ironically, in this regime, stochastic fluctuations are strongest (Sec. 2 of Appendix). Recursion relations for computing the expected cluster numbers in the intermediate-time, metastable regime are also presented in Sec. 4 of Appendix.

Surprisingly, stochastic and mass-action results differ dramatically in the coarsened, equilibrium regime where stochastic fluctuations are small, but where mass incommensurability arises. Our results indicate a dramatic broadening of the mean cluster size distribution when the total mass M is indivisible by N and is just above a multiple of N. This "dispersal" effect is especially prevalent for small σ, even if the total mass M or system size is large. In general, the mean cluster numbers increase with cluster size, but in the special case $M = N + 1$, the cluster size distribution can even be "inverted" ($\langle n_k^{\mathrm{eq}} \rangle > \langle n_\ell^{\mathrm{eq}} \rangle$, when $k < \ell$). Our findings originate from the analysis of the discrete stochastic master equation and are completely neglected by mean field theories.

Our discrete stochastic model also allows us to consider self-assembly in systems of differing total mass and maximum cluster size. Table I lists regimes of validity and results for three different models: mass-action Becker-Döring equations without an imposed maximum cluster size, Becker-Döring equations with a fixed finite maximum cluster size N, and the fully stochastic master equation. Three different ways of taking the large system limits $M, N \to \infty$ are considered. The first case of $N = \infty$ with M finite corresponds to nucleation with unbounded cluster sizes. All models yield a single cluster of size M, but give different rates of coarsening. If $M \gg N^2$, the finite$-N$ BD model matches the asymptotically exact stochastic model where all the mass is concentrated into the largest allowed cluster. In this case,

TABLE I. Accuracy and validity regimes for equilibrium cluster numbers of different nucleation models in the $\varepsilon \ll 1$. The results indicated by * or † match in the $\varepsilon \to 0^+$ limit, but they approach their common result very differently as $\varepsilon \to 0$.

Equilibrium cluster numbers ($\varepsilon \ll 1$)	$\frac{M}{N} \to 0$	$\frac{M}{N}$ finite	$\frac{M}{N} \gg N$
BD ($N = \infty$)	Eq. (2)*	...	...
BD (finite N)	Eq. (2)*	Eq. (2)	Eq. (2)†
Stochastic model	Eq. (6)*	Eqs. (6) and (7)	Eqs. (6) and (7)†

concentrations $c^{\rm eq}_{k<N}$ from mass-action models will vanish as $\sim\varepsilon^{1\,-\,k/N}$, which will slowly converge in ε, especially for large N and $k \lesssim N$. However, the stochastic counterparts $\langle n^{\rm eq}_{k<N}\rangle \sim \varepsilon$ converge much faster to the common result $(M/N)\delta_{k,\,N}$. Therefore, although results from mass-action models (where applicable) and the discrete stochastic approach match in the $\varepsilon \to 0^+$ limit, they approach their common result very differently. Finally, in the intermediate scaling regime where $M \sim N$, we find the novel incommensurability effect discussed in this paper, and highlighted in Figs. 3(c) and 5. Our findings suggest that for many applications, where the effective M/N is finite, mean-field models of self-assembly fail and a discrete stochastic treatment is required.

Experimentally, it may be possible to design a small, closed system in which molecular or mesoscopic-particle self-assembly occurs.[35] By adjusting the total particle numbers M and the maximum aggregate size N so that $M/N = O(1)$, the intermediate scaling regime where commensurability plays a role may be accessible. Another experimental system in which our kinetic model might be applicable is protein aggregates which are comprised of smaller clusters of molecules.[36] These smaller clusters within the larger aggregates have been shown to be out-of-equilibrium due to fluctuations in the sizes of the large aggregates. It would be interesting to adapt our analysis to this problem by treating each large aggregate as system with fluctuating total mass M, and finding the distribution of the smaller clusters within a single fluctuating aggregate.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (NSF) through Grant Nos. DMS-1032131 (T.C.), DMS-1021818 (T.C.), DMS-0719462 (M.D.), and DMS-1021850 (M.D.). T.C. is also supported by the Army Research Office through Grant No. 58386MA. M.D. is also supported by an ARO MURI grant, No. W911NF-11-1-0332.

APPENDIX: TECHNICAL DETAILS

In this section, we provide a few technical details of our analysis.

1. Mean-field analysis

The mass-action Becker-Döring equation associated with the stochastic master equation (Eq. (3)) can be derived by first ensemble averaging the cluster numbers n_k over the distribution $P(\{n\};\,t)$. Upon multiplying Eq. (3) by n_k and summing over all possible states that obey the mass conservation constraint (Eq. (5)), one finds after some algebra

$$
\begin{aligned}
\langle \dot{n}_1(t)\rangle &= -2\left\langle \frac{n_1(n_1-1)}{2}\right\rangle - \sum_{j=2}^{N-1}\langle n_1 n_j\rangle + 2\varepsilon\langle n_2\rangle + \varepsilon\sum_{j=3}^{N}\langle n_j\rangle,\\
\langle \dot{n}_2(t)\rangle &= -\langle n_1 n_2\rangle + \left\langle \frac{n_1(n_1-1)}{2}\right\rangle + \varepsilon\langle n_3\rangle - \varepsilon\langle n_2\rangle,\\
\langle \dot{n}_k(t)\rangle &= -\langle n_1 n_k\rangle + \langle n_1 n_{k-1}\rangle - \varepsilon\langle n_k\rangle + \varepsilon\langle n_{k+1}\rangle,\\
\langle \dot{n}_N(t)\rangle &= \langle n_1 n_{N-1}\rangle - \varepsilon\langle n_N\rangle. \qquad \text{(A1)}
\end{aligned}
$$

Although dimerization destroys two monomers, there are $n_1(n_1-1)/2$ ways of choosing two monomers. Equations (A1) are the first in a hierarchy of equations describing the moments and correlations of the cluster numbers. Their lowest order mean-field level closure is implemented by assuming that the cluster numbers are uncorrelated: $\langle n_i n_j\rangle = \langle n_i\rangle\langle n_j\rangle$, and that the number of monomers remains large $\langle n_1 - 1\rangle \approx \langle n_1\rangle$. Under these approximations, and identifying $\langle n_k(t)\rangle$ with $c_k(t)$, Eqs. (A1) reduce to the well-studied Becker-Döring Eqs. (1).

In steady state, analytic progress can be made on the Becker-Döring equations. In the limit $\varepsilon \to 0^+$, the equilibrium concentrations $c_k^{\rm eq}$ can be found by first defining the effective fugacity $z \equiv c_1^{\rm eq}/\varepsilon$. Substitution of $c_1^{\rm eq}$ into the equilibrium limit of Eqs. (1) gives $c^{\rm eq}_{k>1} = \varepsilon z^k/2$. Further using the constraint $M = \sum_{k=1}^{N} k c_k^{\rm eq}$ yields an algebraic equation whose real root determines z:

$$
\varepsilon\frac{Nz^{N+2} - (N+1)z^{N+1} + z(z^2 - 2z + 2)}{2(z-1)^2} = M. \qquad \text{(A2)}
$$

The root z can be expressed as a power series in εN:

$$
z \approx \left(\frac{2M}{\varepsilon N}\right)^{1/N} - \frac{N-1}{N^2} + O\left((\varepsilon N)^{1/N}\right), \qquad \text{(A3)}
$$

giving the asymptotic approximation to $c_k^{\rm eq}$ shown in Eq. (2).

The values c_k^* in the "quenched" metastable regime can also be accurately approximated by the steady-state limit of Eqs. (1) with the detachment ε set to zero. In this case, the model forbids monomers to detach once they have attached to a cluster. Since all aggregation terms are now proportional to c_1, we define a new "time" variable

$$
\tau(t) \equiv \int_0^t c_1(t')\mathrm{d}t' \qquad \text{(A4)}
$$

which transforms our original Eqs. (1) (with $\varepsilon = 0$) into

$$
\begin{aligned}
\frac{\mathrm{d}c_1}{\mathrm{d}\tau} &= -c_1 - \sum_{j=2}^{N-1} c_j,\\
\frac{\mathrm{d}c_2}{\mathrm{d}\tau} &= -c_2 + \frac{1}{2}c_1,\\
\frac{\mathrm{d}c_k}{\mathrm{d}\tau} &= -c_k + c_{k-1},\\
\frac{\mathrm{d}c_N}{\mathrm{d}\tau} &= c_{N-1}.
\end{aligned} \qquad \text{(A5)}
$$

Upon using the Laplace transform $\tilde{c}_k(s) = \int_0^\infty c_k(\tau)e^{-s\tau}\mathrm{d}\tau$, and the initial condition $c_k(t=0) = M\delta_{k,1}$, Eqs. (A5) can be solved to find

$$\tilde{c}_1(s) = \frac{2Ms(s+1)^{N-2}}{(2s^2+2s+1)(s+1)^{N-2}-1}. \tag{A6}$$

This expression can be inverse Laplace transformed to find $c_1(\tau)$. By defining τ_* as the rescaled time at which monomers are depleted, $c_1(\tau_*) = 0$, we can find the quenched concentrations of all the other larger clusters $c_k(\tau_*)$ by Laplace inverting the expressions

$$\tilde{c}_k(s) = \frac{Ms(s+1)^{N-k-1}}{(2s^2+2s+1)(s+1)^{N-2}-1},$$

$$\tilde{c}_N(s) = \frac{M}{(2s^2+2s+1)(s+1)^{N-2}-1}, \tag{A7}$$

and evaluating the results at τ_*. Clearly, the solutions to the metastable concentrations c_k^* are proportional to the total mass M. In addition, if $N \leq 5$, the poles of $\tilde{c}_1(s)$ can be found analytically, and τ_* and $c_k(\tau_*)$ can be found as solutions to simple transcendental equations. For $N = 4$, we find $\tau_* = 1.7124$, and

$$c_1^* \equiv 0,\ \ c_2^* = 0.9445,\ \ c_3^* = 1.0058,\ \ c_4^* = 1.0234. \tag{A8}$$

The mean-field concentrations stay close to these values until a time of order ε^{-1} when detachment and mass redistribution allows the clusters to coarsen towards their final equilibrium values c_k^{eq}.

2. Kinetic Monte-Carlo simulations

To check our analytic results and explore parameter regimes, we performed extensive KMC simulations of the self-assembly process (Eq. (3)) using the Bortz-Kalos-Lebowitz continuous-time algorithm. For each combination of M and N, $N_{\mathrm{runs}} = 10^6$ separate simulations were performed with data for the time evolution of the cluster populations aggregated across the simulations. Specifically, simulation data was recorded every 0.01 time units up to a simulation time of 100, and then recorded every 100 time units thereafter. A simulation time of 100 was generally sufficient for the process to reach the long-lived metastable state. Time series for the average cluster populations $\langle n_k(t)\rangle$ were computed by averaging the independent simulations according to

$$\langle n_k(i\ \Delta t)\rangle = N_{\mathrm{runs}}^{-1}\sum_{r=1}^{N_{\mathrm{runs}}} n_k^{(r)}(i\Delta t), \tag{A9}$$

where r indexes the simulation run, i specifies the time step, and Δt is the magnitude of each time step. Estimates for equilibrium populations were provided by $\langle n_k(t)\rangle$ with $t \approx 10^{10}$.

All simulations used to generate our results were started from $n_1 = M$, and thorough insensitivity of the equilibrium cluster populations to initial conditions was verified for a number of random cases. Our simulations also provide the entire distribution of cluster numbers, which can be used to compute higher moments of cluster concentrations n_k for each size k. For example, we define the variance of the concentrations of each k-cluster by

$$\sigma_k(i\Delta t) \equiv N_{\mathrm{runs}}^{-1}\sum_{r=1}^{N_{\mathrm{runs}}}\left(n_k^{(r)}(i\Delta t) - \langle n_k(i\Delta t)\rangle\right)^2, \tag{A10}$$

and plot them in Figs. 6. Note that the variance is largest in the metastable regime and diminishes in the equilibrium regime if M is divisible by N. For incommensurate masses, significant variance remains. Provided sufficient trajectories are sampled, our simulations can also be used to accurately construct the full distribution of each concentration n_k. In Fig. 7, we plot the probability Prob($n_5 = m$) of observing m pentamers for eight time slices. Figure 8 shows the distributions of each cluster size at equilibrium. Note the dispersal, or "emulsification" of the system when $M = 33 = 4 \times 8 + 1$.

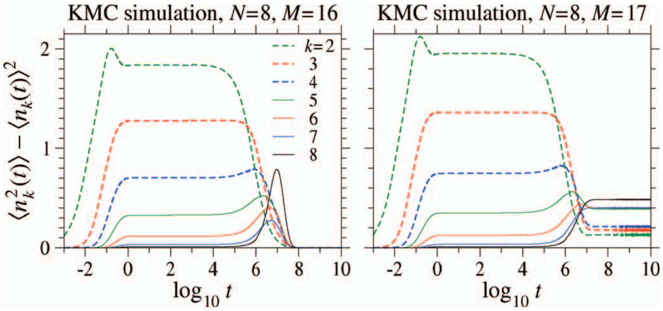

FIG. 6. (a) The variance (Eq. (A10)) for $N = 8$, $M = 16$, and $\varepsilon = 10^{-6}$. (b) The variance when $M = 17$. The variance is large during the metastable regime and is insensitive to $\varepsilon \to 0^+$. At equilibrium, the variance is extremely small when M is divisible by N, but is measurable when there is an extra monomer.

3. Equilibrium solution $\langle n_k^{\mathrm{eq}}\rangle$

Figure 4 displays all accessible states for an $N = 4$, $M = 9$ self-assembly process started with all monomers ($n_1(t = 0) = M = 9$). For the slow detachment (small ε) process nearly all the weight at equilibrium is distributed among states with the lowest number of clusters, three in this case. To

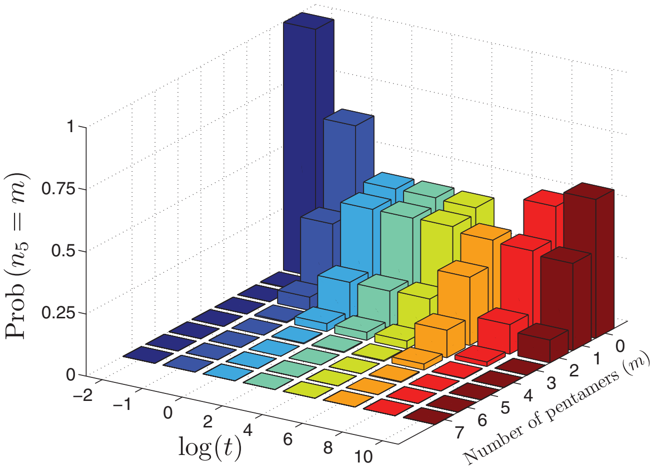

FIG. 7. The fraction of simulation trajectories with $n_5 = m$ at various times. The simulation parameters are $N = 8$, $M = 33$, and $\varepsilon = 10^{-6}$.

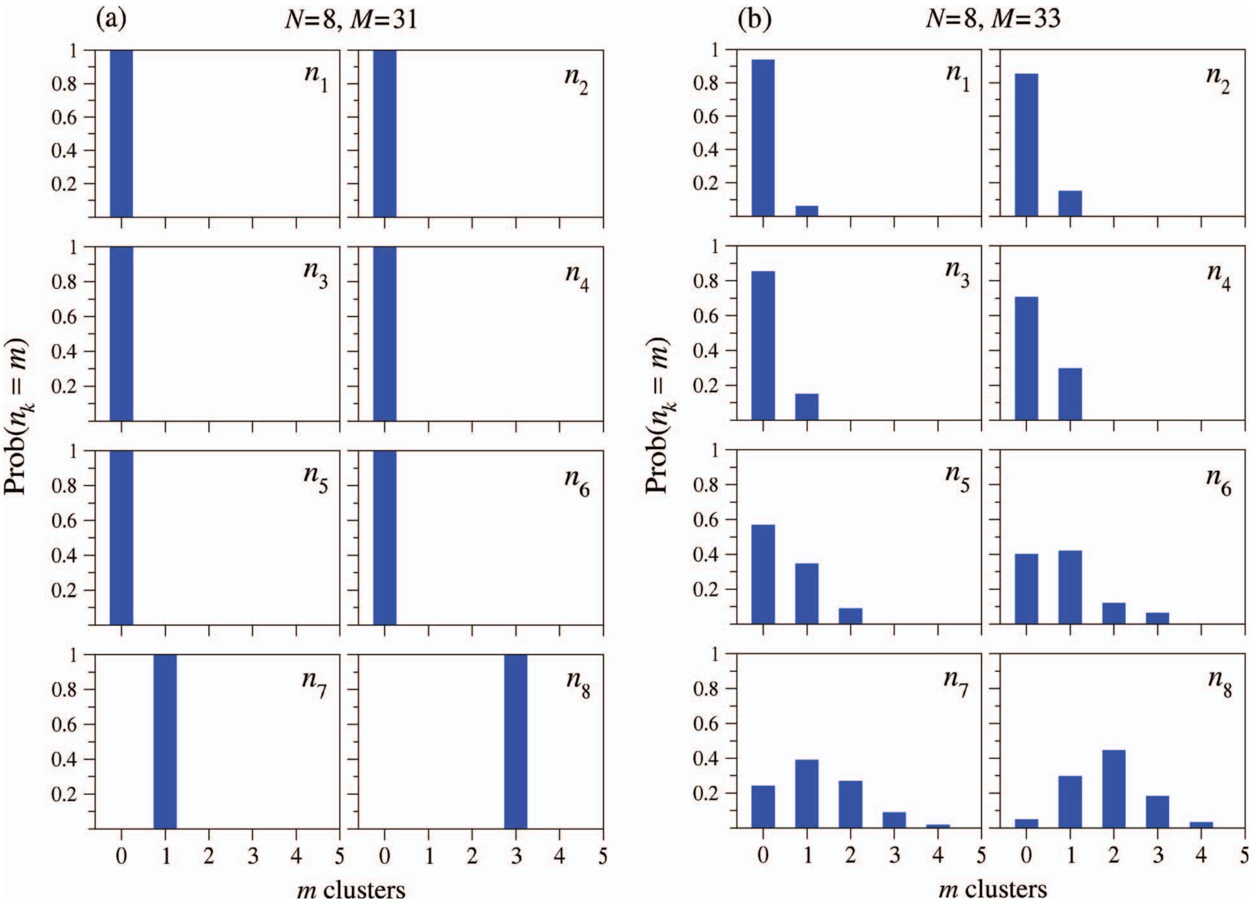

FIG. 8. The distributions of each cluster size at equilibrium determined from KMC simulations with $N = 8$ and $\varepsilon = 10^{-6}$. (a) For $M = 31$ nearly all the mass resides in three $k = 8$ clusters, with one remaining $k = 7$ cluster. For $M = 32$, all the distributions of n_k are peaked at zero, except for Prob($n_8 = 4$) = 1 (not shown). (b) Upon adding yet one more monomer ($M = 33$), the mean cluster concentration distribution disperses significantly and smaller clusters arise. The broader distribution of cluster concentrations arises from indivisibility of the total mass by the maximum cluster size, resulting in a proliferation of allowable configurations as illustrated by the states listed in Eq. (8).

compute the partitioning, we consider transitions only among these low cluster states, as illustrated in Fig. 9.

The transition rates are determined by the number of ways each transition can occur, and are labeled in Fig. 9. Therefore, detailed balance requires $P_{\rm eq}(2, 0, 1, 1) = \varepsilon P_{\rm eq}(0, 1, 1, 1)$, $2P_{\rm eq}(2, 0, 1, 1) = 2\varepsilon P_{\rm eq}(1, 0, 0, 2)$, $2P_{\rm eq}(1, 1, 2, 0) = \varepsilon P_{\rm eq}(0, 1, 1, 1)$, and $P_{\rm eq}(1, 1, 2, 0) = 3\varepsilon P_{\rm eq}(0, 0, 3, 0)$. Imposing normalization of the three-cluster states, neglecting corrections of $O(\varepsilon)$, we find $P_{\rm eq}(0, 1, 1, 1) \approx P_{\rm eq}(1, 0, 0, 2) \approx 6/13$ and $P_{\rm eq}(0, 0, 3, 0) \approx 1/13$. These probabilities yield $\langle n_1^{\rm eq}\rangle \approx 6/13$, $\langle n_2^{\rm eq}\rangle \approx 6/13$, $\langle n_3^{\rm eq}\rangle \approx 9/13$, and $\langle n_4^{\rm eq}\rangle \approx 18/13$. In general, we extend this analysis to general M and N by enumerating all low cluster number states. This is done by identifying the state containing the highest possible number of largest clusters, and distributing the remaining mass into smaller clusters. The number of largest clusters is then successively decreased and only states with the lowest number of total clusters are counted. States generated this way are connected through detailed balance to find the relative probabilities of the dominant states. After normalization we find $P_{\rm eq}(\{n\})$ and $\langle n_k^{\rm eq}\rangle$ as presented in Eqs. (6) and (7).

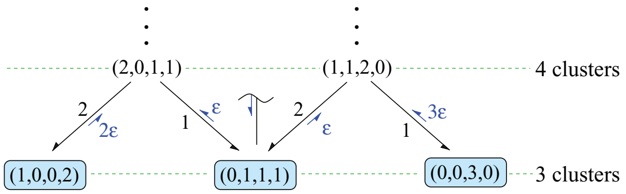

FIG. 9. Enumeration of the transitions between the two lowest cluster number states (3 and 4 clusters) for the $N = 4$, $M = 9$ case. Only three distinct states with a minimum number of clusters $\mathcal{N}_{\rm min} = 3$ arise. These states are all connected by monomer attachment/detachment steps with states having $\mathcal{N}_{\rm min} + 1 = 4$ clusters. By identifying the states that connect the three minimum cluster states, we can compute the weights of each of these states in the $\varepsilon \to 0^+$ limit.

4. Recursion for metastable concentrations $\langle n_k^*\rangle$

Besides Monte-Carlo simulations and our main analytic results for the mean equilibrium cluster concentrations $\langle n_k^{\rm eq}\rangle$, we can also derive a recursion relation that can be used to compute the expected cluster numbers for the irreversible assembly process where $\varepsilon = 0$. In this case, some probability is trapped in configurations that cannot further evolve (e.g., the state (0, 3, 1, 0) in Fig. 4). These "quenched" probabilities $P^*(\{n\})$ lead to a good approximation to the expected, metastable cluster densities $\langle n_k^*\rangle$ when $\varepsilon > 0$ is small. Further coarsening of these metastable cluster size distributions arise only after $t \gg \varepsilon^{-1}$ when detachment and mass redistribution has had time to occur.

We consider irreversible dynamics ($\varepsilon = 0$) and derive a recursion relation for the frozen steady-state probability density $P^*(n_1, \ldots, n_N)$ which emerges either when $n_1 = 0$ or when $n_1 = 1$ and $n_{k \neq 1, N} = 0$. In the first case, all monomers have been depleted, while in the second one, there are no incomplete clusters available to accept the single remaining monomer. Note that the latter case arises only when $(M - 1)$ is a multiple of N.

In order to construct the quenched steady state we start from the initial configuration when all mass is in the form of monomers $n_1 = M$ and $P^*(M, 0, \ldots, 0) = 1$. From this initial state, the only other state that can be constructed is that containing one dimer, which leads also to $P^*(M - 2, 1, 0, \ldots, 0) = 1$. This means that our initial free monomer probability will "migrate" in its entirety to a new configuration with one dimer and $M - 2$ monomers. There are no other choices for this first step. We now find the relative weights of the second generation states ($(M - 4, 2, 0, \ldots, 0)$ and $(M - 3, 0, 1, \ldots, 0)$) formed from $(M - 2, 1, 0, \ldots, 0)$:

$$\begin{tikzcd} & (M-4,2,0,\ldots,0) \\ (M-2,1,0,\ldots,0) \arrow[ur] \arrow[dr] & \\ & (M-3,0,1,\ldots,0) \end{tikzcd} . \qquad \text{(A11)}$$

This process is continued until the quenched steady state is reached where $n_1 = 0$, or $n_1 = 1$ and $n_{k \neq 1, N} = 0$. For the simple case of $N = 3$, only two possibilities exist: monomer aggregation leads to the formation of a dimer, or a monomer and a dimer can give rise to a trimer. The recursion relation that determines the quenched probabilities is

$$P^*(n_1, n_2, n_3) = \frac{M-2n_2-3n_3+1}{M-3n_3-1} P^*(n_1+2, n_2-1, n_3)\big|_{n_2 \geq 1} + \frac{2(n_2+1)}{M-3n_3+2} P^*(n_1+1, n_2+1, n_3-1)\big|_{n_3 \geq 1}, \qquad \text{(A12)}$$

where we have explicitly taken into account the fact that $M = n_1 + 2n_2 + 3n_3$. The prefactors can be calculated by balancing the flux of material into each state. For example, the first term in Eq. (A12) arises from the following consideration. State (n_1, n_2, n_3) arises only via states $(n_1 + 2, n_2 - 1, n_3)$ (dimer formation) and $(n_1 + 1, n_2 + 1, n_3 - 1)$ (trimer formation). To calculate the contribution from state $(n_1 + 2, n_2 - 1, n_3)$, we note that the latter equilibrates only with states (n_1, n_2, n_3) and $(n_1 + 1, n_2 - 2, n_3 + 1)$. There are $(n_1 + 2)(n_1 + 1)/2$ ways of creating a new dimer from $(n_1 + 2)$ monomers. Similarly, there are $(n_1 + 2)(n_2 - 1)$ ways to create a trimer from $(n_1 + 2)$ monomers and $(n_2 - 1)$ dimers. The relative weight w_1 for material transferring from state $(n_1 + 2, n_1 - 1, n_3)$ to (n_1, n_2, n_3) is thus given by

$$w_1 = \frac{(n_1+2)(n_1+1)}{(n_1+2)(n_1+1) + 2(n_1+2)(n_2-1)}. \qquad \text{(A13)}$$

Upon simplifying and using the mass conservation constraint we find that w_1 corresponds to the first prefactor in Eq. (A12). Similarly, to calculate the contribution from state $(n_1 + 1, n_2 + 1, n_3 - 1)$, we note that it equilibrates with states (n_1, n_2, n_3) and $(n_1 - 1, n_2 + 2, n_1 - 1)$. There are $(n_1 + 1)n_1/2$ ways of creating a new dimer from $n_1 + 1$ monomers. Similarly, there are $(n_1 + 1)(n_2 + 1)$ ways to create a trimer from $(n_1 + 1)$ monomers and $(n_2 + 1)$ dimers. The relative weight w_2 for material transferring from state $(n_1 + 1, n_2 + 1, n_3 - 1) \to (n_1, n_2, n_3)$ is given by

$$w_2 = \frac{2(n_1+1)(n_2+1)}{2(n_1+1)(n_2+1) + (n_1+1)n_1}, \qquad \text{(A14)}$$

which is exactly the second prefactor in Eq. (A12). We can apply the same reasoning for general N to find relationships among the steady-state probabilities in the irreversible case $\varepsilon = 0$:

$$\begin{aligned} P^*(n_1, n_2, \ldots, n_N) = {} & \frac{(n_1+1)}{n_1 - 1 + 2\sum_{k=2}^{N-1} n_k} P^*(n_1+2, n_2-1, n_3, \ldots, n_N)|_{n_2 \geq 1} \\ & + \sum_{j=2}^{N-2} \frac{2(n_j+1)}{n_1 + 2\sum_{k=2}^{N-1} n_k} P^*(n_1+1, \ldots, n_j+1, n_{j+1} - 1, \ldots, n_N)\big|_{n_{j+1} \geq 1} \\ & + \frac{2(n_{N-1}+1)}{n_1 + 2 + 2\sum_{k=2}^{N-1} n_k} P^*(n_1+1, n_2, \ldots, n_{N-1}+1, n_N - 1)\big|_{n_N \geq 1}. \end{aligned} \qquad \text{(A15)}$$

It is understood that all occupation numbers must be nonnegative and the process stops whenever $n_1 = 0$ or $n_1 = 1$ and $n_{k \neq 1, N} = 0$.

This algorithm is difficult to implement for large M because the number of computations grows exponentially with M. However, we can straightforwardly apply it to our small-system examples. For example, for the $N = 4$, $M = 9$ case illustrated in Fig. 4, we find that for $\varepsilon = 0$, the system settles to three configurations with the following weights:

$$P^*(0, 3, 1, 0) = \frac{921}{5488}, \quad P^*(0, 0, 3, 0) = \frac{2873}{24696},$$

$$P^*(0, 1, 1, 1) = \frac{4015}{7056}, \quad P^*(1, 0, 0, 2) = \frac{259}{1764}.$$

These probabilities yield the mean cluster numbers

$$\langle n_1^* \rangle = 0.14683, \ \langle n_2^* \rangle = 1.07248, \ \langle n_3^* \rangle = 1.08584,$$
$$\text{and } \langle n_4^* \rangle = 0.86267.$$

From the recursion relations, it is clear that the exact results $\langle n_k^{\mathrm{eq}} \rangle$ depend on M in a nontrivial manner. Although the results for c_k^* derived from BD equations are simply proportional to M (see Eqs. (A6) and (A7)), they cannot be exact. However, comparing the values of c_k^{eq} in Eqs. (A8) with the values for $\langle n_k^* \rangle$ above, they form a reasonable approximation to the exact results. We have also shown that the c_k^* give reasonable approximations to $\langle n_k^* \rangle$ for all values N and M that we have investigated. This good approximation holds despite the large variances in the mean cluster numbers in metastable regime as shown in Fig. 6.

[1] D. Kashchiev, *Nucleation: Basic Theory With Applications* (Butterworth-Heinemann, Oxford, 2000).

[2] C. J. Hernandez and T. G. Mason, J. Phys. Chem. C **111**, 4477 (2007).

[3] R. Groß and M. Dorigo, Proc. IEEE **96**, 1490 (2008).

[4] J. G. Amar, M. N. Popescu, and F. Family, Mater. Res. Soc. Symp. Proc. **570**, 3 (1999).

[5] M. N. Popescu, J. G. Amar, and F. Family, Phys. Rev. B **64**, 205404 (2001).

[6] P. A. Mulheran and M. Basham, Phys. Rev. B **77**, 075427 (2001).

[7] T. P. J. Knowles, C. A. Waudby, G. L. Devlin, S. I. A. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, Science **326**, 1533 (2009).

[8] D. Sept and A. J. McCammon, Biophys. J. **81**, 667 (2001).

[9] J. Miné, L. Disseau, M. Takahashi, G. Cappello, M. Dutreix, and J.-L. Viovy, Nucleic Acids Res. **35**, 7171 (2007).

[10] L. Edelstein-Keshet and G. Ermentrout, Bull. Math. Biol. **60**, 449 (1998).

[11] M. F. Bishop and F. A. Ferrone, Biophys. J. **46**, 631 (1984).

[12] E. T. Powers and D. L. Powers, Biophys. J. **91**, 122 (2006).

[13] D. Endres and A. Zlotnick, Biophys. J. **83**, 1217 (2002).

[14] A. Zlotnick, J. Mol. Biol. **366**, 14 (2007).

[15] B. Sweeney, T. Zhang, and R. Schwartz, Biophys. J. **94**, 772 (2008).

[16] J. Z. Porterfield and A. Zlotnick, "An overview of capsid assembly kinetics," in *Emerging Topics in Physical Virology*, edited by P. G. Stockley and R. Twarock (Imperial College, London, 2010), pp. 131–158.

[17] A. Yu. Morozov, R. F. Bruinsma, and J. Rudnick, J. Chem. Phys. **131**, 155101 (2009).

[18] K. A. Brogden, Nat. Rev. Microbiol. **3**, 238 (2005).

[19] G. L. Ryan and A. D. Rutenberg, J. Bacteriol. **189**, 4749 (2007).

[20] M. Ehrlich, W. Boll, A. van Oijen, R. Hariharan, K. Chandran, M. L. Nibert, and T. Kirchhausen, Cell **118**, 591 (2004).

[21] B. I. Shraiman, Biophys. J. **72**, 953 (1997).

[22] L. Foret and P. Sens, Proc. Natl. Acad. Sci. U.S.A. **105**, 14763 (2008).

[23] M. Torkkeli, R. Serimaa, O. Ikkala, and M. Linder, Biophys. J. **83**, 2240 (2002).

[24] P. L. Krapivsky, E. Ben-Naim, and S. Redner, *Statistical Physics of Irreversible Processes* (Cambridge University Press, Cambridge, England, 2010)

[25] O. Penrose, J. Stat. Phys. **89**, 305 (1997).

[26] J. A. D. Wattis and J. R. King, J. Phys. A **31**, 7169 (1998).

[27] P.-E. Jabin and B. Niethammer, J. Differ. Equations **191**, 518 (2003).

[28] B. Niethammer, J. Nonlinear Sci. **13**, 115 (2008).

[29] P. Smereka, J. Stat. Phys. **132**, 519 (2008).

[30] S. N. Majumdar, S. Krishnamurthy, and M. Barma, Phys. Rev. Lett. **81**, 3691 (1998).

[31] J. S. Bhatt and I. J. Ford, J. Chem. Phys. **118**, 3166 (2003).

[32] F. Schweitzer, L. Schimansky-Geier, W. Ebeling, and H. Ulbricht, Physica A **150**, 261 (1988).

[33] T. Chou and M. R. D'Orsogna, Phys. Rev. E **84**, 011608 (2011).

[34] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, J. Comput. Phys. **17**, 10 (1975).

[35] G. M. Whitesides and M. Boncheva, Proc. Natl. Acad. Sci. U.S.A. **99**, 4769 (2002).

[36] W. Pan, P. G. Vekilov, and V. Lubchenko, J. Phys. Chem. B **114**, 7620 (2010).