Experiments in Stochastic Self Assembly

Yuliy Baryshnikov[1] Ed Coffman[2] Boonsit Yimwadsana[2]

[1] Bell Labs, Lucent Technologies, Murray Hill, NJ 07974
[2] Electrical Engineering Dept., Columbia University, New York, NY 10027

Abstract

The reaction rate equations of chemical kinetics provide a useful model for molecular self assembly, which can be exhibited as a continuous-time, continuous-space limit under the LLN (law of large numbers) scaling. The model takes the form of a set of first order, usually nonlinear ODEs describing the concentrations of the various reactants. With explicit solutions in mind, these systems have rarely been tractable; even numerical approaches have been limited within the parameter spaces of practical systems. Thus, studies of algorithmic self-assembly have turned to experimental tools based on discrete-event simulation. This paper presents such a model argued from first principles and elementary collision theory. We prove first that in the hydrodynamic limit the equations underlying the discrete model become reaction rate equations; and second, we verify by a number of experiments that the accuracy of our experimental results is strikingly good, even for systems that are very small, i.e., with a relatively small population of molecules.

1 Introduction

We adopt the reference model of self assembly (SA) that we presented in [1], which is cast in the framework of polymerization: the SA process starts with a homogeneous set consisting only of monomers. As the process evolves, monomers combine with each other and with other self-assembled polymers to become larger polymers. The process stops when there are no two polymers in the system which can combine, i.e., bond. The reaction rate equations (RREs) of chemical kinetics (see [3] for many references) provide a useful hydrodynamic (fluid) model of molecular SA. In this essentially deterministic model, the system is described by a system of first-order, usually nonlinear, ODEs giving the rates of change of concentrations of the various polymers in terms of the concentrations themselves and the reaction rates. The reaction rates are determined by such factors as the temperature of the system and the bonding energy between colliding monomers. In our application, it is convenient to regard a reaction rate as a bonding probability, i.e., the probability that a pair of colliding polymers bond together to form a new polymer.

The general SA model is called *reversible* SA and includes events where polymers decompose into smaller polymers. Decomposition rates are defined in analogy with reaction rates. Although our approach applies more generally, we take a specific class of SA processes to illustrate the method: we study *incremental* SA where polymers grow and decompose one monomer at a time. The RREs describing the system are

$$\begin{aligned}\frac{dx_i(t)}{dt} &= 2x_1(t)x_{i-1}(t)p_{1,i-1} - 2x_1(t)x_i(t)p_i + x_{i+1}(t)q_i - x_i(t)q_{i-1},\ i > 2\\ &= x_1^2(t)p_1 - 2x_1(t)x_i(t)p_i + x_{i+1}(t)q_i - x_i(t)q_{i-1},\quad i = 2\\ &= -2\sum_{l\geq 1} p_1x_1(t)x_l(t) + 2x_2(t)q_1 - \sum_{j\geq 2} x_{j+1}(t)q_j,\quad i = 1\end{aligned} \tag{1}$$

where $x_i(t),\ i \geq 1$, denotes the concentration of i-polymers, i.e., those made up of i monomers; p_j is the bonding probability of a monomer colliding with a j-polymer; and where q_j is the probability that a $(j+1)$-polymer decomposes into a monomer and a j- polymer.

In systems where the maximum allowable size of polymers is moderately small, the RRE system above can be solved numerically using standard mathematical software. However, these tools become inadequate when the size limit of polymers becomes large, so one typically comes to rely on simulation experiments. A discrete-event simulation model of SA is introduced in the next section. To illustrate the fact that the underlying equations for these models, which we call *polymerization equations*, pass to the RRE limit under the LLN scaling, we focus on the irreversible case of incremental SA, where all decomposition rates are 0. Then to test the accuracy of the simulations, in Section 3 we compare the simulation results with the results for RRE systems solvable numerically.

The approach of Markovian birth-and-death equations to chemical kinetics is well studied, although it is often justified by its more realistic modeling of random molecular collision/bonding phenomena. Here, our interest in simulation is confined to its use as a means of studying systems too large for other techniques, e.g., the numerical solution of systems of ODEs; in particular, we are interested in scalings where the RREs are very accurate models, but the number of equations is so large that exact results are computationally infeasible. Our derivation of the polymerization equations is notable for its simplicity. And our results include useful estimates of convergence to the RREs, which show that the simulation of a relatively, very small number of molecules (many orders of magnitude smaller than the number in the systems being modeled) is surprisingly close to the RRE limit.

2 The discrete-event SA model

In our discrete Markov collision-event model of SA, times between successive pair-wise collisions of polymers are independent with a time-dependent rate parameter $R(t)$ which reflects reactant concentrations at time t. Consistent with elementary collision theory [2], we take $R(t)$ as proportional to the concentrations of reactant polymers in a host fluid, i.e., $R(t) = \lambda\eta^2(t)$, where the constant, λ, aggregates various underlying physical parameters affecting the collision rate between polymer pairs, and where $\eta(t)$ is the total number (the unnormalized concentration) of polymers in the system at time t. The parameter p_j is the probability that a given collision between a monomer and a j polymer is *effective*, i.e., results in a reaction that produces a new polymer.

Thus, a trajectory of irreversible incremental SA is created by a sequence of independent samples from the exponential distribution with parameter $R(t_k)$, where $t_k,\ k = 1, 2, \ldots$ is the time of the k-th two-polymer collision. At each such collision, the colliding polymers are chosen independently from the distribution $\{\eta_1(t)/\eta(t),\ \ldots,\ \eta_N(t)/\eta(t)\}$. The two polymers, say i- and j-polymers bond only if at least one is a monomer, and then only with probability p_j or p_i, depending on which is a monomer. Note that a maximum polymer size, say N, is implicit in the model by stipulating that $p_j = 0$ for all $j \geq N$.

To develop an ODE for $\eta_i(t)$ in irreversible incremental SA, consider the evolution of the number of i-polymers in the interval $[t, t + \Delta t]$ given that a single collision occurs. At time $t + \Delta t$ in incremental SA, the number of i-polymers will increase by 1 if the collision is between a monomer and an $(i - 1)$-polymer, and they bond to produce an i-polymer; the number decreases by 1 if the collision is between a monomer and an i-polymer, and they bond to produce an $(i + 1)$-polymer; and the number stays the same otherwise.

Draw two samples from the polymer population to determine a *collision sample*; if, say, the first sample is an i-polymer and the second a j-polymer, then the sample is denoted by (i, j). For $i \neq j$, the samples (i, j) and (j, i) are identical in effect, i.e., they constitute the same collision, so the probability of such a collision is simply twice the probability of either ordered sample (i, j) or (j, i) alone. For simplicity, we denote collisions by $\{i, j\}$, with $i \leq j$, and their probabilities by

$$\Pr\{i, j\} = \begin{cases} 2\frac{\eta_i(t)}{\eta(t)}\frac{\eta_j(t)}{\eta(t)} & i \neq j, \\ \left(\frac{\eta_i(t)}{\eta(t)}\right)^2 & i = j \end{cases}$$

Now let $\Psi_i,\ i > 1$, denote the set of all polymer collisions except those involving a monomer and either an i-polymer or an $(i - 1)$-polymer. Then for $i > 1$ with $\eta_i(t) > 0$, we have in the usual way, by the properties of the collision process,

$$\begin{aligned} \eta_i(t + \Delta t) = {} & \eta_i(t)(1 - R(t)\Delta t) + (\eta_i(t) + 1)R(t)\Delta t \Pr\{1, i - 1\}p_{1,i-1} \\ & + \eta_i(t)R(t)\Delta t[\sum_{\{k,l\}\in\Psi_i} \Pr\{k, l\} + \Pr\{1, i - 1\}(1 - p_{1,i-1}) \\ & + \Pr\{1, i\}(1 - p_{1,i})] + (\eta_i(t) - 1)R(t)\Delta t \Pr\{1, i\}p_{1,i} \end{aligned}$$

Applying $\sum_{\{k,l\}} \Pr\{k,l\} = 1$, ignoring terms on the order of $o(\Delta t)$, and substituting for $R(t) = \lambda\eta(t)^2$ and the conditional collision probabilities, we rearrange to obtain, in the limit $\Delta t \to 0$ the desired *incremental* polymerization equations for $\eta_i(t)$. (This is the special case of (2) below, with all decomposition rates set to 0.) After introducing the LLN scaling $x_i^n(t) := \frac{\eta_i(nt)}{n}$, where n is the initial number of monomers in the system, and then taking the limit $x_i(t) := \lim_{n\to\infty} x_i^n(t)$ we obtain, as claimed, the RREs given earlier in (1) with decomposition rates equal to 0.

For the reversible case, we introduce in addition to $R(t)$ a total rate $D(t)$ of polymer decompositions which is proportional to the concentration of reactant polymers. In particular, in analogy with reactions, an elementary model stipulates that $D(t) = \mu\eta(t)$, where, as before, the constant μ accounts for physical parameters affecting decompositions, and where the intervals between consecutive decompositions are exponentially distributed with the rate parameter $D(t)$. Hence, events occur at a total rate of $R(t)+D(t)$, and the times between successive events are modeled as exponentials with mean $1/(R(t)+D(t))$. The earlier simulation algorithm extends in the obvious way.

Keeping with the normalization $\lambda = 1$, a routine extension of the earlier analysis then yields the general incremental polymerization equations

$$\frac{d\eta_i(t)}{dt} = \begin{cases} 2\eta_1(t)(\eta_{i-1}(t)p_{i-1} - \eta_i(t)p_i) + \mu(\eta_{i+1}(t)q_i - \eta_i(t)q_{i-1}) & i > 2, \\ \eta_1^2(t)p_1 - 2\eta_1(t)\eta_2(t)p_2 + \mu(\eta_{i+1}(t)q_i - \eta_i(t)q_{i-1}) & i = 2 \end{cases} \tag{2}$$

The limit under the LLN scaling again gives the appropriate RREs, as claimed.

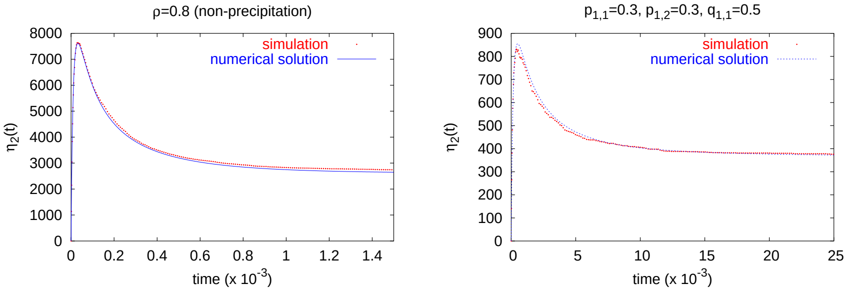
3 Experimental Results

To get an idea of the accuracy of the experimental approach,[1] consider the case of $\eta(0) = 50,000$ monomers, a number many orders of magnitude smaller than the number of elementary particles in the laboratory experiments of molecular SA being modeled. Figure 1 compares the experimental results estimating the function $\eta_2(t),\ t \geq 0$, for both irreversible and reversible self-assembly for the time intervals shown, with those obtained numerically from solutions to the corresponding systems of ODEs. As can be seen, the scaling of the figures shown here makes it visually, nearly impossible to distinguish between the exact and experimental results.

These results beg the question of just how small $\eta(0)$ can be and still have highly accurate results. A partial answer to this question was obtained by investigating the difference (given as a percentage of $\eta(0)$) in yields obtained from experimental and numerical results for the parameters shown in the tables below, where

---

[1] These results are but a small sample of those generated, and available.



(a) $\eta_2(t)$ (b) $\eta_2(t)$

Figure 1: Plots for the simulation and numerical results for $\eta_2(t)$ with $N = 3$ for irreversible and reversible self-assembly.

$\eta(0)$ ranges from 500 to 30,000. These data suggest the remarkable fact that $\eta(0)$ values as low as 10,000 yield excellent accuracy throughout.

	$N = 3, p_2 = 0.5$			$N = 4, p_3 = p_2 = 0.5$		
$p_1/\eta(0)$	0.25	0.5	0.875	0.25	0.5	0.875
500	0.922%	0.702%	0.754%	0.72%	0.831%	0.866%
1000	0.842%	2.182%	1.284%	0.09%	0.211%	0.834%
8000	0.267%	0.272%	0.475%	0.2%	0.076%	0.162%
30000	0.024%	0.041%	0.227%	0.09%	0.062%	0.044%

References

[1] Y. Baryshnikov, E. Coffman, and P. Momčilović. Phase transitions and control in self assembly. In *Proc. Foundations of Nanoscience: Self-Assembled Architectures and Devices*, Snowbird, UT, April 2004.

[2] Kenneth A. Connors. *Chemical Kinetics: The Study of Reaction Rates in Solution*. Wiley-VCH, 1990.

[3] D. Gillespie and L. Petzold. *System Modelling in Cellular Biology*. MIT Press, 2006. Z. Szallasi and J. Stelling and V. Perival, editors, Chapter title: Numerical Simulation for Biochemical Kinetics.