

## The Subtle Business of Model Reduction for Stochastic Chemical Kinetics

**Dan T. Gillespie**<sup>a)</sup>

*Dan T Gillespie Consulting, 30504 Cordoba Pl., Castaic, California 91384*

**Yang Cao**

*Department of Computer Science, Virginia Tech, Blacksburg, Virginia 24061*

**Kevin R. Sanft and Linda R. Petzold**

*Department of Computer Science, University of California Santa Barbara, Santa Barbara California 93106*

**Abstract:** This paper addresses the problem of simplifying chemical reaction networks by adroitly reducing the number of reaction channels and chemical species. The analysis adopts a discrete-stochastic point of view, and focuses on the model reaction set  $S_1 \rightleftharpoons S_2 \rightarrow S_3$ , whose simplicity allows all the mathematics to be done exactly. The advantages and disadvantages of replacing this reaction set with a single  $S_3$ -producing reaction are analyzed quantitatively using novel criteria for measuring simulation accuracy and simulation efficiency. It is shown that in all cases in which such a model reduction can be accomplished accurately and with a significant gain in simulation efficiency, a procedure called the slow-scale stochastic simulation algorithm provides a robust and theoretically transparent way of implementing the reduction.

To appear in *Journal of Chemical Physics*

Final version (22 Dec 2008)

<sup>a)</sup> Author to whom correspondence should be addressed. [GillespieDT@mailaps.org](mailto:GillespieDT@mailaps.org)

## I. INTRODUCTION

Biochemical systems typically contain networks of many chemical reaction channels involving many molecular species. This circumstance encourages attempts to construct simpler but equivalent “reduced” reaction networks. A well known example of such a reduction is the Michaelis-Menten abridgment of the enzyme-substrate reactions<sup>1,2</sup>, which has been the subject of many refinements over the years<sup>3,4</sup> and which continues to play an important role in biochemistry today<sup>5</sup>.

Typically, an abridgment replaces the given reaction network with a network that involves fewer reaction channels and fewer chemical species. Perhaps the simplest reaction set that presents the opportunity for doing that, one that has several features in common with the enzyme-substrate reactions but is mathematically more tractable, is



where we assume that  $c_1$  and  $c_3$  are both non-zero. It is tempting to cut to the chase and replace this set of three three-species reactions with one two-species reaction, such as



where the reaction constant  $c$  is given some “suitable” value. Our focus in this paper will be to determine the conditions under which it is advisable to make such a replacement, and to show how the replacement should be implemented. Of course, if a modeler deliberately chooses to model the production of  $S_3$  molecules from  $S_1$  molecules by reaction (2) instead of by reactions (1), then this issue is moot. But we are assuming here that the modeler believes that reactions (1) really describe what is going on physically, and therefore wants any abridgement of (1), such as reaction (2), to mimic the salient effects of reactions (1) with reasonable accuracy. A modeler might choose to use reaction (2) instead of reactions (1) because the values of the rate constants  $c_1$ ,  $c_2$  and  $c_3$  in (1) are not all known. But choosing an appropriate value for  $c$  in (2) inevitably makes assumptions about those three rate constants; thus, it might be better to use (1) with those assumptions made explicitly and openly, since that would not only preserve the topology of reactions (1) but also make it easy to incorporate later new information about the unknown rate constants.

The most obvious advantage in replacing reactions (1) with a single  $S_3$ -producing reaction like (2) is the reduction in the numbers of reactions and species that we have to contend with. Another advantage might be speeding up the numerical simulation of reactions (1). By simulation we mean here *stochastic* simulation, since stochasticity often plays a role in cellular systems. But there are two potential drawbacks to such a reduction: First, as will be elaborated on below, this is always an approximation, since it is simply not possible for any single reaction to exactly mimic reactions (1) in all respects. Second, if we want to have the option of embedding reactions (1) in a larger network of reactions, some of which may involve species that get removed in the model

reduction, as  $S_2$  has in (2), then it may be impossible to simulate those other reactions when using the reduced model.

In this paper we will address these matters in detail for reaction set (1). We will begin by presenting some novel perspectives on simulation *efficiency* and simulation *accuracy*. We will show that these new perspectives imply that a one-reaction abridgment of (1) will be advisable in some circumstances, but not in others. We will then show that, in all cases where a model reduction can be done accurately and with a significant gain in stochastic simulation efficiency, implementing the reduction will be more involved than just swapping reactions (1) for reaction (2). Finally, we will establish a new perspective on the results of two recent papers, namely, the slow-scale stochastic simulation algorithm (ssSSA) of Cao et al.<sup>6</sup>, and the stochastic quasi-steady-state approximation singular perturbation analysis (sQSPA) of Mastny et al.<sup>7</sup>.

The reaction network (1) we are focusing on here is obviously very simple. But that simplicity allows all the mathematics, which even in this case is non-trivial, to be done exactly, and thus all issues to be explored thoroughly. We believe that a clarification of these issues in the context of reactions (1) can lead to a better understanding of how these issues play out in more complicated reaction networks.

## II. QUANTIFYING THE GAIN IN SIMULATION EFFICIENCY

Well-stirred chemical systems with discrete molecular populations and stochastic reaction dynamics can be exactly simulated by the well-known stochastic simulation algorithm (SSA).<sup>8</sup> The only downside is that the SSA is usually quite slow: The SSA simulates every reaction event, so the time required to make an SSA simulation run is proportional to the number of reaction events that occur.

Replacing reactions (1) with a single  $S_3$ -producing reaction, such as reaction (2), would evidently have the consequence that a new  $S_3$  molecule would be produced by each reaction event. In contrast, the creation of a new  $S_3$  molecule via reactions (1) usually requires more than one reaction event. Therefore, a fair measure of the gain in simulation efficiency realized by such an abridgment would be *the average number of reaction events that are needed by reactions (1) to produce one  $S_3$  molecule*. That number turns out to be surprisingly easy to compute.

Suppose a molecule starts out as an  $S_1$  molecule, or as we shall say, starts “in state  $S_1$ ”. On each visit to state  $S_2$ , the molecule has probability  $c_3/(c_2 + c_3)$  of going on to state  $S_3$ , and probability  $c_2/(c_2 + c_3)$  of going back to  $S_1$ . So in  $n$  visits to state  $S_2$ , the molecule would go on to state  $S_3$  an average of  $nc_3/(c_2 + c_3)$  times; thus, in order to get an average of one visit to state  $S_3$ , the molecule needs to visit state  $S_2$  a total of  $n_1$  times, where  $n_1 c_3/(c_2 + c_3) = 1$ . It follows that  $n_1 = (c_2 + c_3)/c_3$ . But each of these  $n_1$  visits to state  $S_2$  requires exactly two reaction events, namely the reaction  $R_1$  that brings the molecule to state  $S_2$ , and either reaction  $R_2$  or  $R_3$  that takes the molecule away.

Therefore, the average number of reaction events required for the molecule to get from state  $S_1$  to state  $S_3$  via the reaction set (1) is  $2n_1 = 2(c_2 + c_3)/c_3$ . If the molecule had started in state  $S_2$  instead of state  $S_1$ , it would be exactly one reaction event closer to state  $S_3$ , so the average number of reaction events for a molecule in state  $S_2$  to reach state  $S_3$  would be  $2n_1 - 1$ , a difference that is not significant for our purposes here. We thus conclude that the *gain in simulation efficiency* achieved by replacing reactions (1) with a single  $S_3$ -producing reaction is approximately

$$G = 2 \left( \frac{c_2 + c_3}{c_3} \right). \quad (3)$$

That is to say, simulating the production of  $S_3$  molecules via some single reaction like (2) will be approximately  $G$  times faster than simulating via reactions (1). The overall gain in simulation efficiency would actually be less than this if reactions (1) were embedded in a larger reaction network that is also simulated.

Two things are noteworthy about the result (3). First, the gain depends on  $c_2$  and  $c_3$ , but not on  $c_1$ . And second, we will have

$$G \gg 1 \text{ if and only if } c_2 \gg c_3. \quad (4)$$

The assertion of (3) that  $G = 2$  when  $c_2 = 0$  is an obvious result; the assertion that  $G = 4$  when  $c_2 = c_3$  is perhaps less obvious. But both of those efficiency gains are modest compared to the gains achievable when  $c_2$  is one or more orders of magnitude larger than  $c_3$ . It follows that one should carefully examine one's goals in reducing reactions (1) when the strong inequality  $c_2 \gg c_3$  is not satisfied.

### III. ACCURACY: THE IMPORTANCE OF BEING EXPONENTIAL

Let  $T(x_1, x_2)$  be the time to the next firing of reaction  $R_3$  in (1) when there are  $x_1$   $S_1$  molecules and  $x_2$   $S_2$  molecules. Obviously,  $T(x_1, x_2)$  will be some kind of random variable. In this section, we will show that a necessary condition for reactions (1) to be *accurately* replaceable by a single  $S_3$ -producing reaction, such as for example reaction (2), is that  $T(x_1, x_2)$  be approximately an *exponential* random variable. In the next section we will calculate the exact probability density function of  $T(x_1, x_2)$ , with the aim of finding out under what conditions this exponentiality requirement is satisfied.

Stochastic chemical kinetics, which encompasses both the SSA and the chemical master equation, assumes that the dynamics of reaction (2) are described by<sup>8</sup>

$$cdt = \text{the probability that a given } S_1 \text{ molecule in reaction (2) will} \\ \text{become an } S_3 \text{ molecule in the next infinitesimal time interval } dt. \quad (5)$$

Verification of this critical condition is awkward to accomplish directly. A more convenient but completely equivalent condition is afforded by the following theorem, which is proved in Appendix A.

**Theorem 1:** Condition (5) is equivalent to saying that the time required for a given  $S_1$  molecule to become an  $S_3$  molecule via reaction (2) is an *exponential* random variable with mean  $1/c$ .

We recall that the exponential random variable with mean  $c^{-1}$  is defined to be the random variable on  $0 \leq t < \infty$  which has probability density function (pdf)  $c \exp(-ct)$ , and cumulative distribution function (cdf)  $1 - \exp(-ct)$ . It follows from Theorem 1 that reaction (2) cannot be simulated using the SSA, nor analyzed via the chemical master equation, unless the time required for a given  $S_1$  molecule to become an  $S_3$  molecule via reaction (2) is, at least approximately, an exponential random variable. This result has motivated some recent molecular dynamics studies of excluded volume effects in simple well-stirred one- and two-dimensional chemical systems.<sup>9</sup>

It might be thought that this exponential requirement, being stochastic, would not apply if we were content to describe reaction (2) in terms of traditional deterministic chemical kinetics. However, that is not true. To see why, recall that the traditional reaction rate equation for (2), written in terms of the number  $X_i(t)$  of  $S_i$  molecules at time  $t$ , and assuming  $X_3(0) = 0$ , is

$$\frac{dX_3(t)}{dt} = cX_1(t) = c(X_1(0) - X_3(t)). \quad (6a)$$

The solution to this equation is

$$X_3(t) = X_1(0)(1 - e^{-ct}). \quad (6b)$$

Consistency requires that  $X_3(t)$  in Eq. (6b) should accurately describe the behavior of the *average* number of  $S_3$  molecules in the *stochastic* formulation (note that we are dealing here with a linear, first-order reaction). Let  $f(t)$  be the cdf for the time-to-reaction  $\tau$  of any particular  $S_1$  molecule; i.e.,  $f(t)$  is the probability that  $\tau \leq t$ , and hence the probability that an  $S_1$  molecule will have become an  $S_3$  molecule by time  $t$ . Then since the  $S_1$  molecules react independently of each other, the probability that exactly  $n$  of them will have become an  $S_3$  molecule by time  $t$  is

$$\frac{X_1(0)!}{n![X_1(0) - n]!} [f(t)]^n [1 - f(t)]^{X_1(0) - n}.$$

This implies that the number of  $S_3$  molecules created in time  $t$  is the binomial random variable  $\mathcal{B}(f(t), X_1(0))$ . Since the mean of that random variable is  $f(t)X_1(0)$ , then agreement with Eq. (6b) requires that

$$X_1(0)(1 - e^{-ct}) = f(t)X_1(0),$$

or

$$f(t) = 1 - e^{-ct}. \quad (7)$$

But this is the cdf of the exponential random variable with mean  $1/c$ . Thus we conclude that the time  $\tau$  to reaction (2) for each individual  $S_1$  molecule must be exponentially distributed in order for the deterministic rate equations (6) to be valid.

For an example of a non-exponential  $\tau$ -distribution that is clearly inconsistent with Eq. (6b), suppose that  $\tau$  were uniformly distributed in the interval  $[c^{-1} - \varepsilon, c^{-1} + \varepsilon)$  for some  $\varepsilon < c^{-1}$ . Then the mean time-to-reaction for each  $S_1$  molecule would indeed be  $c^{-1}$ . But with this lifetime distribution, the number of  $S_3$  molecules would obviously stay at zero until time  $c^{-1} - \varepsilon$ , and then rise roughly linearly to  $X_1(0)$  in a time  $2\varepsilon$ . This is clearly not the behavior predicted by formula (6b).

The relevance of the foregoing result to the problem of replacing reactions (1) with some single  $S_3$ -producing reaction such as (2) can be understood as follows. If there are  $x_1$   $S_1$  molecules in the system, then (5) and the addition law of probability imply that the probability that reaction (2) will fire in the next  $dt$  is  $x_1 \times c dt = cx_1 dt$ . More generally, *any* single reaction that produces one  $S_3$  molecule will have the property that, for some state-dependent function  $a$  which is called the reaction's *propensity* function,  $adt$  gives the probability that the reaction will fire in the next  $dt$ . This implies, by the same reasoning that led to Theorem 1, that the time to the next firing of that reaction will be exponentially distributed with mean  $a^{-1}$ . Therefore, if this reaction is to be a surrogate for reactions (1) – a replacement that approximately replicates the way in which reactions (1) produce  $S_3$  molecules – then the time  $T(x_1, x_2)$  to the next firing of reaction  $R_3$  in (1) must be, at least approximately, exponentially distributed. If that turns out to be so, then an approximating surrogate reaction for (1) should exist. But if  $T(x_1, x_2)$  is found to be clearly *non-exponential*, then we must conclude that reactions (1) *cannot* be accurately replaced by a single  $S_3$ -producing reaction.

#### IV. DISTRIBUTION OF THE TIME TO THE NEXT $R_3$ REACTION

In Appendix B, we prove that the probability density function  $P(t; x_1, x_2)$  of  $T(x_1, x_2)$ , the time to the next  $R_3$  reaction in (1) when there are  $x_1$   $S_1$  molecules and  $x_2$   $S_2$  molecules, is given by Eq. (B18). In that formula, the four functions  $p(\beta, t | \alpha, 0)$  for  $\alpha, \beta = 1, 2$  are given explicitly by Eqs. (B13) and (B14), and

$$\lambda_{\pm} \equiv \frac{1}{2} \left[ (c_1 + c_2 + c_3) \pm \sqrt{(c_1 + c_2 + c_3)^2 - 4c_1c_3} \right]. \quad (8)$$

That  $P(t; x_1, x_2)$  in Eq. (B18) is not generally exponential can be seen by noting that its form for  $T(1, 0)$ , the time for a single  $S_1$  molecule to become an  $S_3$  molecule via reactions (1), turns out to be<sup>10</sup>

$$P(t; 1, 0) = \frac{c_1 c_3}{(\lambda_+ - \lambda_-)} \left[ e^{-\lambda_- t} - e^{-\lambda_+ t} \right]. \quad (9)$$

This pdf is obviously not exponential; indeed, it vanishes at  $t = 0$ , whereas the pdf of any exponential random variable has its maximum at  $t = 0$ . It also follows from (B18) that the pdf of the time for a single  $S_2$  molecule to become an  $S_3$  molecule via reactions (1) is

$$P(t; 0, 1) = \frac{c_3}{(\lambda_+ - \lambda_-)} \left[ (c_1 - \lambda_-) e^{-\lambda_- t} + (\lambda_+ - c_1) e^{-\lambda_+ t} \right]. \quad (10)$$

Although this pdf achieves its maximum at  $t = 0$ , it still does not generally have a simple exponential form. Plots of the two pdfs (9) and (10) for  $c_1 = c_3 = 1$  and  $c_2 = 0.1$  are shown in Fig. 1 on a semi-log scale, where a truly exponential pdf would appear as a down-sloping straight line. The non-exponential character of  $P(t; 1, 0)$  is obvious; that of  $P(t; 0, 1)$  is evinced by a gradual change of slope around  $t = 2$ .

The consequences of the non-exponential form of the pdf  $P(t; 1, 0)$  in Eq. (9) are illustrated in Fig. 2. The jagged solid curve shows a single  $X_3(t)$  trajectory obtained in an exact SSA run of reactions (1), using the parameter values  $c_1 = c_3 = 1$ ,  $c_2 = 0.1$  and the initial condition  $X_1(0) = 300$ ,  $X_2(0) = X_3(0) = 0$ . The dashed curve shows the *average* of 10,000 such trajectories. It can be shown that the mean of the pdf in Eq. (9) is  $(c_1 + c_2 + c_3)/c_1 c_3$ , which in this case equals 2.1; i.e., the average time for an  $S_1$  molecule to become an  $S_3$  molecule via reactions (1) in this case is 2.1. If we made the usual deterministic assumption that formula (6a) applies with  $c = (2.1)^{-1}$ , then Eq. (6b) would give the trajectory shown as the *dotted* curve in Fig. 2. The mismatch between that curve and the dashed curve illustrates the inappropriateness of replacing reactions (1) with reaction (2) when the time between  $R_3$  reactions is not exponentially distributed.

The additional revelation in Fig. 1 that  $P(t; 1, 0)$  is not the same curve as  $P(t; 0, 1)$  illustrates another potential problem for model reduction: While any acceptable single-reaction abridgment of reactions (1) will accurately replicate the time-evolution of the  $S_3$  population, and hence also the time-evolution of the *total*  $S_1$  and  $S_2$  population, the abridgment might not accurately replicate the time-evolutions of the  $S_1$  and  $S_2$  populations separately; e.g., reaction (2) gives us no indication of the  $S_2$  population. Therefore, if  $P(t; x_1, x_2)$  depends on  $x_1$  and  $x_2$  *individually*, and not just on their sum

$$x_1 + x_2 \equiv x_{12}, \quad (11)$$

as Fig. 1 shows happens when  $c_1 = c_3 = 1$ ,  $c_2 = 0.1$  and  $x_{12} = 1$ , then the lack of information about the individual values of  $x_1$  and  $x_2$  could make using the abridged reaction in a real simulation impossible, even if  $P(t; x_1, x_2)$  were exponential.

A close inspection of Eqs. (B18), (B13) and (B14) reveals that  $P(t; x_1, x_2)$  is in general a polynomial in  $e^{-\lambda_- t}$  and  $e^{-\lambda_+ t}$ . From Eq. (8), it can be shown that when both  $c_1$  and  $c_3$  are positive, as we are assuming here, then  $0 < \lambda_- \leq \lambda_+$ . Therefore, a necessary condition for  $P(t; x_1, x_2)$  to be approximately exponential is for the rate constants to be such that

$$\lambda_- \ll \lambda_+; \quad (12)$$

because then, all terms involving  $e^{-\lambda_+ t}$  will be negligibly small for  $t \gg 1/\lambda_+$ , and we can hope that the  $t$ -dependence for  $t \gg 1/\lambda_+$  will be given by some power of  $e^{-\lambda_- t}$ .

When  $0 < \lambda_- \leq \lambda_+$ , the extreme inequality (12) will be satisfied if and only if

$$\frac{\lambda_+ \lambda_-}{(\lambda_+ + \lambda_-)^2} \ll 1. \quad (13)$$

Since Eq. (8) implies that  $\lambda_+ \lambda_- = c_1 c_3$  and  $\lambda_+ + \lambda_- = c_1 + c_2 + c_3$ , then condition (13) is the same as

$$\frac{c_1 c_3}{(c_1 + c_2 + c_3)^2} \equiv \left( \frac{c_1}{c_1 + c_2 + c_3} \right) \left( \frac{c_3}{c_1 + c_2 + c_3} \right) \ll 1. \quad (14)$$

Since each factor in the middle of (14) is less than 1, then the right inequality in (14) can be satisfied if and only if at least one of those two factors is  $\ll 1$ . The first factor will be  $\ll 1$  if and only if  $c_2 + c_3 \gg c_1$ , which is the same as either  $c_2 \gg c_1$  or  $c_3 \gg c_1$ . And the second factor will be  $\ll 1$  if and only if  $c_1 + c_2 \gg c_3$ , which is the same as either  $c_1 \gg c_3$  or  $c_2 \gg c_3$ . Thus we conclude that *condition (12) will be satisfied if and only if at least one of the following four conditions holds:*

$$(a) c_2 \gg c_1; \quad (b) c_3 \gg c_1; \quad (c) c_1 \gg c_3; \quad (d) c_2 \gg c_3. \quad (15)$$

Note that these four conditions are not mutually exclusive; e.g., the condition  $c_2 \gg c_3 \gg c_1$  satisfies both (d) and (b). Nor are these conditions collectively exhaustive; e.g., the condition  $c_1 = c_2 = c_3$  satisfies none of conditions (15). But satisfaction of at least one of conditions (15) is necessary, and as we shall see shortly sufficient, for  $P(t; x_1, x_2)$  to be exponential.

Assume now that *at least one* of conditions (15) is satisfied. Then the strong inequality (14) will also be satisfied, so we will have from (8) that

$$\lambda_{\pm} = \frac{(c_1 + c_2 + c_3)}{2} \left[ 1 \pm \sqrt{1 - \frac{4c_1c_3}{(c_1 + c_2 + c_3)^2}} \right] \approx \frac{(c_1 + c_2 + c_3)}{2} \left[ 1 \pm \left( 1 - \frac{1}{2} \frac{4c_1c_3}{(c_1 + c_2 + c_3)^2} \right) \right],$$

whence

$$\lambda_+ \approx c_1 + c_2 + c_3, \quad \lambda_- \approx \frac{c_1c_3}{c_1 + c_2 + c_3}. \quad (16)$$

When (16) is substituted into Eqs. (B13) and (B14), and the results are substituted into Eq. (B18), we obtain since  $\lambda_- \ll \lambda_+$ ,

$$P(t; x_1, x_2) \approx \frac{c_1c_3x_1}{c_1 + c_2 + c_3} e^{-\lambda_-(x_1+x_2)t} \left( \frac{c_1 + c_2}{c_1 + c_2 + c_3} \right)^{x_2} + \frac{c_1c_3x_2}{c_1 + c_2 + c_3} e^{-\lambda_-(x_1+x_2)t} \left( \frac{c_1 + c_2}{c_1 + c_2 + c_3} \right)^{x_2-1} \quad (t \gg 1/\lambda_+). \quad (17)$$

The restriction on  $t$  here ensures that all terms involving  $e^{-\lambda_+t}$  have become negligibly small. Again, this approximation assumes that at least one of conditions (15) is satisfied.

Now let us examine (17) for the individual conditions (15). First, if either condition (a) or (b) holds, so that  $c_2 + c_3 \gg c_1$ , then Eq. (16) gives  $\lambda_+ \approx c_2 + c_3$  and  $\lambda_- \approx c_1c_3/(c_2 + c_3)$ . Equation (17) simplifies slightly, in that  $c_1$  gets dropped from all denominators. Further simplification of (17) follows from the observation that the condition  $c_2 + c_3 \gg c_1$  implies that reaction  $R_1$ , which creates  $S_2$  molecules, will occur much less frequently than reactions  $R_2$  and  $R_3$ , which destroy  $S_2$  molecules. The  $S_2$  population will thus usually be very small, and a reasonable approximation would be to set  $x_2 \approx 0$ , and hence  $x_1 \approx x_{12}$ . With those approximations, the second term in Eq. (17) is effectively removed and the equation finally reduces to

$$P(t; x_1, x_2) \approx \left( \frac{c_1c_3x_{12}}{c_2 + c_3} \right) e^{-\left( \frac{c_1c_3x_{12}}{c_2 + c_3} \right)t} \quad (t \gg (c_2 + c_3)^{-1}, \text{ (a) or (b)}). \quad (18)$$

Since this pdf has the exponential form, an accurate single-reaction abridgment should be possible. And the decay constant in (18) will be the propensity function of the surrogate reaction. The fact that this decay constant depends on  $x_1$  and  $x_2$  only through their sum  $x_{12}$  suggests that the reduced model should be amenable to simulation.

Now suppose that either condition (c) or condition (d) holds. Then  $c_1 + c_2 \gg c_3$ , and Eqs. (16) give  $\lambda_+ \approx c_1 + c_2$  and  $\lambda_- \approx c_1c_3/(c_1 + c_2)$ . The relation  $c_1 + c_2 \gg c_3$  implies that  $c_3$  can be dropped from all denominators in Eq. (17). That equation then reduces to

$$P(t; x_1, x_2) \approx \left( \frac{c_1c_3x_{12}}{c_1 + c_2} \right) e^{-\left( \frac{c_1c_3x_{12}}{c_1 + c_2} \right)t} \quad (t \gg (c_1 + c_2)^{-1}, \text{ (c) or (d)}). \quad (19)$$

Again this pdf has the exponential form, with the decay constant depending on  $x_1$  and  $x_2$  only through their sum  $x_{12}$ . Therefore, replacing reactions (1) with a single  $S_3$ -producing reaction, whose propensity function is the decay constant in (19), should be feasible. Note that (19) does not assume, as (18) does, that  $x_2 \approx 0$ .

## V. IMPLEMENTING THE REDUCED MODEL

We showed in the preceding section that an accurate replacement of reactions (1) by a single  $S_3$ -producing reaction should be possible in the four cases (a)-(d) in (15). More specifically, the result in Eq. (18) shows that in cases (a) and (b) the  $S_3$ -producing reaction should have the propensity function

$$a = \frac{c_1 c_3 x_{12}}{c_2 + c_3} \quad (\text{for (a) } c_2 \gg c_1 \text{ or (b) } c_3 \gg c_1), \quad (20)$$

with the understanding that  $x_2 \approx 0$ , and that we are not interested in phenomena occurring on timescales of order  $(c_2 + c_3)^{-1}$  or smaller. And the result in Eq. (19) shows that in cases (c) and (d) the  $S_3$ -producing reaction should have propensity function

$$a = \frac{c_1 c_3 x_{12}}{c_1 + c_2} \quad (\text{for (c) } c_1 \gg c_3 \text{ or (d) } c_2 \gg c_3), \quad (21)$$

with no restrictions on  $x_2$ , but with the understanding that we are not interested in phenomena occurring on timescales of order  $(c_1 + c_2)^{-1}$  or smaller. But exactly how should the replacement reaction be framed in these cases?

First let us dispose of two “obvious” cases in which  $c_2 \approx 0$ , and reaction  $R_2$  practically never fires. The first case couples that condition with condition (b),  $c_3 \gg c_1$ : In a short time of order  $c_3^{-1}$ , all  $S_2$  molecules become  $S_3$  molecules via reaction  $R_3$ ; thereafter, the  $S_1$  molecules convert to  $S_3$  molecules essentially via reaction (2) with the approximate propensity function  $a = c_1 x_1$ , since each  $R_1$  firing will practically always be followed immediately by an  $R_3$  firing. This result also follows by putting  $c_2 \approx 0$  in Eq. (20), and remembering that  $x_{12} \approx x_1$  since (20) assumes that  $x_2 \approx 0$ . The other “obvious” case couples condition  $c_2 \approx 0$  with condition (c),  $c_1 \gg c_3$ : In a short time of order  $c_1^{-1}$ , all  $S_1$  molecules become  $S_2$  molecules via reaction  $R_1$ ; thereafter, the  $S_2$  molecules convert to  $S_3$  molecules via reaction  $R_3$  with the approximate propensity function  $a = c_3 x_2 = c_3 x_{12}$ . This result also follows by putting  $c_2 \approx 0$  in Eq. (21). In both of these “obvious” cases, the simulation speedup factor realized by the abridgment is about 2, which is rather modest.

A more interesting case arises by conjoining conditions (a) and (b), and requiring that *both*  $c_2$  and  $c_3$  be  $\gg c_1$ , a condition that we will write  $c_2, c_3 \gg c_1$ . This condition has been analyzed in detail by Mastny et al.<sup>7</sup> using a reduction method which they call the stochastic quasi-steady-state approximation, singular perturbation analysis (sQSPA). The conclusion of their analysis (see Ref. 7, Table II) expressed in our notation here is that reactions (1) can be replaced by reaction (2) with propensity function  $a = c_1 c_2 x_1 / (c_2 + c_3)$ . This is the same as our result in Eq. (20), since the assumption  $c_2, c_3 \gg c_1$  in Mastny et al.<sup>7</sup> implies  $x_2 \approx 0$ , and hence  $x_1 \approx x_{12}$ . Our first-passage-time analysis thus confirms the Mastny et al.<sup>7</sup> result, including the proviso which is implicit in their derivation that this approximation is valid only on timescales larger than  $(c_2 + c_3)^{-1}$ . The resulting gain in simulation efficiency (3) will be large or small according to whether  $c_2/c_3$  is large or small. But we note that the condition  $x_2 \approx 0$  would appear to pose a problem if we wanted to embed the abridged reaction in a network of other reactions, some of which create or consume  $S_2$  molecules.

Another interesting case in (15) is (d),  $c_2 \gg c_3$ . We showed in Sec. II this is the condition for a truly substantial speedup in stochastic simulation. But it turns out that simply replacing reactions (1) with reaction (2) using the propensity function (21) has some limitations. To illustrate, we have used the exact SSA to simulate each of reactions (1) and (2) for parameter values

$$c_1 = 3, \quad c_2 = 2, \quad c_3 = 10^{-4} \quad (22a)$$

and the initial condition

$$X_1(0) = 300, \quad X_2(0) = X_3(0) = 0. \quad (22b)$$

Figure 3 shows the results of the SSA simulation of reactions (1). In this figure, the species populations have been plotted out immediately after the occurrence of each  $R_3$  reaction, so only 300 points get plotted in the conversion of the 300  $S_1$  molecules. But approximately  $1.2 \times 10^7$  reaction events had to be simulated in order to get those 300  $R_3$  reactions, so there were on average  $4 \times 10^4$   $R_1$  and  $R_2$  reaction events between successive  $R_3$  reaction events, a figure that agrees with formula (3).

Figure 4 shows the results of the SSA simulation of the surrogate reaction (2) using the propensity function in Eq. (21) and the same parameter values (22) as used in Fig. 3. Here the populations have been plotted out after each simulated reaction. Since only 300 reaction events were simulated in this run, compared to the  $1.2 \times 10^7$  events that were simulated to produce Fig. 3, the gain in simulation efficiency achieved by using the surrogate reaction (2) is truly large. Comparing Fig. 4 with Fig. 3 shows that the surrogate reaction (2) does give a satisfactory representation of the  $X_3(t)$  trajectory, just as we expect on the basis of our analysis. But reaction (2) evidently does not provide a satisfactory representation of the  $X_1(t)$  trajectory; furthermore, it gives us no information

at all about the  $X_2(t)$  trajectory. The explanation for these shortcomings is not hard to fathom: When we stop simulating reactions  $R_1$  and  $R_2$ , as we do when we substitute reaction (2) for reactions (1), we lose the ability to accurately track the populations of species  $S_1$  and  $S_2$ .

If we were interested in only  $X_3(t)$ , and if reactions (1) were the only reactions in the system that involve species  $S_1$  and  $S_2$ , then we might be satisfied with this state of affairs. But we are often concerned with situations in which reactions (1) take place concurrently with other reactions, some of which have one or both of species  $S_1$  and  $S_2$  as reactants. With no reliable information about the instantaneous populations of  $S_1$  and  $S_2$  when using reaction (2), how are we to evaluate the propensity functions of those other reactions in order to simulate their firings along with the firings of reaction (2)? Evidently, simply replacing reactions (1) with reaction (2) when  $c_2 \gg c_3$  will not be satisfactory if there are other reactions in the system that have  $S_1$  and  $S_2$  as reactants, or if we want to see how species  $S_1$  and  $S_2$  behave on the timescale of reaction  $R_3$ .

## VI. THE SLOW-SCALE SSA: A ROBUST RECIPE FOR CONDITION (d)

We will show in this section that, under condition (d),

$$c_2 \gg c_3, \quad (23)$$

replacing reactions (1) with a single  $S_3$ -producing reaction can be accurately and robustly accomplished using a procedure called the *slow-scale stochastic simulation algorithm* (ssSSA). Designed more generally for “stiff” stochastic systems (systems with a wide separation of timescales with the fastest mode being stable), the ssSSA was introduced in Ref. 6 by some of the present authors, and is basically a refinement of ideas introduced earlier by Haseltine and Rawlings<sup>11</sup> and Rao and Arkin<sup>12</sup>. Instead of replacing reactions (1) with a single new reaction like (2), the ssSSA eliminates reactions  $R_1$  and  $R_2$ , and then uses a modified propensity function for reaction  $R_3$ .

We should note that condition (23) differs from, and thus *corrects*, the condition advertised in Ref. 6 for applying the ssSSA to reactions (1).<sup>13</sup> Also, as will be explained in the next paragraph, condition (23) does *not* need to be supplemented by  $c_1 \gg c_3$ .

When condition (23) is satisfied, an  $S_2$  molecule is much more likely to become an  $S_1$  molecule than an  $S_3$  molecule. Thus, successive occurrences of reaction  $R_3$  will usually be separated by very many occurrences of reactions  $R_1$  and  $R_2$ ; indeed, as we showed in Sec. II, there will be on average  $2(c_2 + c_3)/c_3$   $R_1$  and  $R_2$  reactions occurring between successive  $R_3$  reactions, and that number will be  $\gg 1$  when condition (23) holds. Since  $R_1$  and  $R_2$  will be firing much more frequently than  $R_3$ , we call  $R_1$  and  $R_2$  “fast reactions,” and  $R_3$  a “slow reaction.” Notice that the designation of  $R_1$  as a fast

reaction under condition (23) is justified regardless of the size of  $c_1 > 0$ ; because, between two successive  $R_3$  reactions, there will inevitably be as many  $R_1$  firings as  $R_2$  firings. And it is the *average frequency of firing*, not the size of the reaction rate constant, that determines whether a reaction is “fast” or “slow” for the ssSSA. Species  $S_1$  and  $S_2$  are then designated as “fast species” because their populations *get changed* by a fast reaction, and  $S_3$  is called a “slow species” because its population does not.

The fast species populations evolving under only the fast reactions, i.e.,  $S_1 \xrightleftharpoons[c_2]{c_1} S_2$ , constitutes what is called the *virtual fast process*. We denote it by  $(\hat{X}_1(t), \hat{X}_2(t))$ , using the hat to distinguish it from the real fast process  $(X_1(t), X_2(t))$  which evolves under all three reactions in (1). For the virtual fast process (but not for the real fast process), we have the conservation relation

$$\hat{X}_1(t) + \hat{X}_2(t) = x_{12} \quad (\text{a constant}); \quad (24)$$

therefore, the virtual fast process has only one independent variable. We choose it to be  $\hat{X}_2(t)$ , and then take  $\hat{X}_1(t) = x_{12} - \hat{X}_2(t)$ . The process  $\hat{X}_2(t)$  thus evolves according to the propensity functions

$$\hat{a}_1(x_2) = c_1(x_{12} - x_2), \quad \hat{a}_2(x_2) = c_2 x_2,$$

with  $\hat{X}_2(t)$  increasing by 1 each time  $R_1$  fires, and decreasing by 1 each time  $R_2$  fires. This simple stochastic process has been well studied.<sup>14</sup> It can be shown that its  $t \rightarrow \infty$  limit  $\hat{X}_2(\infty)$  is the binomial random variable with parameters  $c_1/(c_1 + c_2)$  and  $x_{12}$ :

$$\hat{X}_2(\infty) = \mathcal{B}(c_1/(c_1 + c_2), x_{12}). \quad (25)$$

Since  $\mathcal{B}(p, N)$  has mean  $Np$  and variance  $Np(1-p)$ , then

$$\langle \hat{X}_2(\infty) \rangle = \frac{c_1 x_{12}}{c_1 + c_2} \quad \text{and} \quad \text{var} \{ \hat{X}_2(\infty) \} = \frac{c_1 c_2 x_{12}}{(c_1 + c_2)^2}. \quad (26a)$$

It then follows from Eq. (24) (or by symmetry) that

$$\langle \hat{X}_1(\infty) \rangle = \frac{c_2 x_{12}}{c_1 + c_2} \quad \text{and} \quad \text{var} \{ \hat{X}_1(\infty) \} = \frac{c_1 c_2 x_{12}}{(c_1 + c_2)^2}. \quad (26b)$$

Notice that the asymptotic distribution of the virtual fast process depends on the current state  $(x_1, x_2, x_3) \equiv \mathbf{x}$  only through the quantity  $x_1 + x_2 = x_{12}$ . That these few facts about  $\hat{X}_2(t)$  are all that is needed to construct a computationally viable abridgment of reactions (1) when  $c_2 \gg c_3$  is a consequence of the following theorem.

**Theorem 2.** Given condition (23), let the system be in state  $(x_1, x_2, x_3) \equiv \mathbf{x}$  at time  $t$ . Then for any  $\delta t$  that is *large* compared to the expected time to the next  $R_1$  or  $R_2$

reaction, but *small* compared to the expected time to the next  $R_3$  reaction, the probability that reaction  $R_3$  will fire in  $[t, t + \delta t)$  is approximately  $\bar{a}_3(\mathbf{x}) \delta t$ , where

$$\bar{a}_3(\mathbf{x}) \equiv c_3 \langle \hat{X}_2(\infty) \rangle. \quad (27)$$

Furthermore,  $\hat{X}_2(\infty)$  and  $x_{12} - \hat{X}_2(\infty)$  provide good estimates of the populations of species  $S_2$  and  $S_1$  at any time after  $t + \delta t$  but before the next  $R_3$  reaction occurs.

This theorem is proved in Appendix C. It says, first of all, that  $\bar{a}_3(\mathbf{x})$  as defined in (27) is the “effective propensity function” of reaction  $R_3$  on the timescale of that (slow) reaction. This is so because the defining attribute of a propensity function is that its product with an “effectively infinitesimal” time span gives the probability that the reaction will occur in that time span. With Eq. (26a), Eq. (27) takes the explicit form

$$\bar{a}_3(\mathbf{x}) = \frac{c_3 c_1 x_{12}}{c_1 + c_2}. \quad (28)$$

Note that this is the same as the propensity function (21) that our first-passage-time analysis gave for condition (23). Theorem 2 also tells us that the  $S_2$  and  $S_1$  populations at any time greater than  $\delta t$  after the last  $R_3$  reaction can be estimated by drawing a sample  $x_2$  of the random variable  $\hat{X}_2(\infty)$  in (25), and then taking  $X_2 = x_2$  and  $X_1 = x_{12} - x_2$ .

The critical assumption used in proving Theorem 2 (see Appendix C) is that between successive firings of reaction  $R_3$  there will typically be many firings of reactions  $R_1$  and  $R_2$ . We showed in Sec. II that this will always be so if condition (23) holds. To see that the result (28) is consistent with this fact, we reason as follows: Since  $\bar{a}_3(\mathbf{x}) \delta t$  is (approximately) the probability that  $R_3$  will fire in the next  $\delta t$ , then the mean time to the next firing of  $R_3$  will be (approximately)

$$\frac{1}{\bar{a}_3(\mathbf{x})} = \frac{c_1 + c_2}{c_3 c_1 x_{12}}. \quad (29a)$$

And since the average probability that either  $R_1$  or  $R_2$  will fire in the next  $dt$  is  $c_1 \langle \hat{X}_1(\infty) \rangle dt + c_2 \langle \hat{X}_2(\infty) \rangle dt$ , then the average mean time to the next firing of either  $R_1$  or  $R_2$  will be (approximately)

$$\frac{1}{c_1 \langle \hat{X}_1(\infty) \rangle + c_2 \langle \hat{X}_2(\infty) \rangle} = \frac{c_1 + c_2}{2c_2 c_1 x_{12}}, \quad (29b)$$

where the last step follows upon substituting from Eqs. (26). Now observe that, under condition  $c_2 \gg c_3$ , the time (29b) will indeed be very much smaller than the time (29a); moreover, no other condition is required to ensure this.

The strategy of the ssSSA is to use the standard SSA to simulate *only* reaction  $R_3$ , but taking the propensity function for that reaction to be the function (28) instead of  $c_3x_2$ . At each firing of  $R_3$ , the ssSSA increases the  $S_3$  population by 1 and decreases  $x_{12}$  by 1. The ssSSA then “waits” for a time of order  $\delta t$ , which is very small on the timescale of reaction  $R_3$  but nevertheless large enough for the fast species populations to “relax” to their  $t = \infty$  values, and it then estimates the populations of the fast species by sampling the binomial random variable (25). The full ssSSA simulation procedure for reactions (1) thus proceeds as follows:

1. In state  $(x_1, x_2, x_3)$  at time  $t$ , and with  $x_{12} = x_1 + x_2$ , evaluate  $\bar{a}_3$  in (28).
2. Draw a unit-interval uniform random number  $r$  and compute the time to the next  $R_3$  reaction,  $\tau = (1/\bar{a}_3) \ln(1/r)$ .
3. Advance time to the next  $R_3$  reaction by replacing  $t \leftarrow t + \tau$ . Then actualize that reaction by replacing  $x_3 \leftarrow x_3 + 1$  and  $x_{12} \leftarrow x_{12} - 1$ .
4. Generate the “relaxed” populations of the fast species by taking  $x_2$  to be a sample of the binomial random variable (25), and  $x_1 = x_{12} - x_2$ .
5. Record  $(t, x_1, x_2, x_3)$  if desired. Then return to step 1, or else stop.

Figure 5 shows the results of a ssSSA run made in this way for the parameter values (22). The results are seen to be practically indistinguishable (in a statistical sense) from the exact SSA results in Fig. 3. But whereas the SSA run took about 6 minutes to execute, the ssSSA run took only a fraction of a second. Notice that the ssSSA remedies the deficiencies of the reaction (2) simulation in Fig. 4 as regards species  $S_1$  and  $S_2$ .

What happens if reactions (1) are embedded in a network of other reactions, some of which involve the fast species  $S_1$  and  $S_2$  as reactants? The answer to this question depends on whether the other reactions are “fast” or “slow”. If any of the other reactions are as fast or faster than reactions  $R_1$  and  $R_2$ , then we must start the analysis all over by finding, if possible, a new virtual fast process that is asymptotically stable. But if all of the new reactions are *slow* – i.e., they occur infrequently relative to reactions  $R_1$  and  $R_2$  – then they can easily be accommodated in the above simulation procedure. For example, the additional slow reaction  $R_4$ ,  $S_1 + S_4 \xrightarrow{c_4} S_5$ , which has propensity function  $a_4(\mathbf{x}) = c_4x_1x_4$ , would be assigned the effective propensity function

$$\bar{a}_4(\mathbf{x}) = c_4 \left\langle \hat{X}_1(\infty) \right\rangle x_4 = \frac{c_4 c_2 x_1 x_4}{c_1 + c_2},$$

where the last equality follows from (26b). And the additional slow reaction  $S_1 + S_2 \xrightarrow{c_5} S_6$  with propensity function  $a_5(\mathbf{x}) = c_5x_1x_2$  would be assigned the effective propensity function

$$\bar{a}_5(\mathbf{x}) = c_5 \langle \hat{X}_1(\infty) \hat{X}_2(\infty) \rangle = \frac{c_5 c_1 c_2}{(c_1 + c_2)^2} x_{12} (x_{12} - 1).$$

The last step here follows by first writing

$$\langle \hat{X}_1(\infty) \hat{X}_2(\infty) \rangle = \langle (x_{12} - \hat{X}_2(\infty)) \hat{X}_2(\infty) \rangle = x_{12} \langle \hat{X}_2(\infty) \rangle - \langle \hat{X}_2^2(\infty) \rangle,$$

then using  $\langle \hat{X}_2^2(\infty) \rangle = \langle \hat{X}_2(\infty) \rangle^2 + \text{var} \langle \hat{X}_2(\infty) \rangle$ , and finally invoking Eqs. (26). Thus, any new *slow* reactions involving the fast species  $S_1$  and  $S_2$  can be accommodated by the ssSSA, despite the fact that we have no sure knowledge of the instantaneous populations of those fast species.

The status of the fast species populations in the ssSSA merits further comment: Although the values for  $x_1$  and  $x_2$  generated in Step 4 get plotted in Step 5, those values are not used in the computations that drive the simulation; therefore, if plots of the fast species populations are not needed, Step 4 can be omitted without any impairment to simulation accuracy. The fact is that  $x_1$  and  $x_2$  are not individually “tracked” by the ssSSA, because the ssSSA does not simulate reactions  $R_1$  and  $R_2$ . Step 4 merely estimates how the values of  $x_1$  and  $x_2$  might appear on the slow timescale. But the sum  $x_1 + x_2 = x_{12}$  is accurately tracked, and that sum is all that we need to implement reaction  $R_3$ , or any other *slow* reaction that involves one or both of  $S_1$  and  $S_2$  as reactants.

## VII. SUMMARY AND CONCLUSIONS

In this paper we have shown that replacing reactions (1) with a single reaction that produces  $S_3$  cannot be done accurately unless the time to the next creation of an  $S_3$  molecule via reactions (1) can be well approximated by an *exponential* random variable. We showed that this applies even to the associated deterministic reaction rate equations. The specific requirement for accuracy is that  $P(t; x_1, x_2)$ , the probability density function of the time to the next  $R_3$  event in (1) given  $x_1$   $S_1$  molecules and  $x_2$   $S_2$  molecules, should be well approximated by the canonical exponential form  $a e^{-at}$ . If that is so, then a surrogate reaction that accurately mimics the production of  $S_3$  molecules should exist, and  $a$  will be its propensity function. If, however, the surrogate reaction is unable to accurately track the  $S_1$  and  $S_2$  populations individually, then even if the exponential approximation obtains, a model reduction will be feasible only if  $a$  depends on  $x_1$  and  $x_2$  only through their sum  $x_{12}$ .

Against this background, we derived using first-passage-time theory an exact formula for  $P(t; x_1, x_2)$ . We then showed that there are only four situations in which that function satisfies the foregoing conditions: (a)  $c_2 \gg c_1$ , (b)  $c_3 \gg c_1$ , (c)  $c_1 \gg c_3$ , and (d)  $c_2 \gg c_3$ . We found that if either of conditions (a) or (b) holds, then, under the reasonable

assumption that the  $S_2$  population is practically always zero, the propensity function of the surrogate reaction will have the form  $a = c_1 c_3 x_{12} / (c_2 + c_3)$ . And we found that if either of conditions (c) or (d) holds, then the propensity function of the surrogate reaction will have the form  $a = c_1 c_3 x_{12} / (c_1 + c_2)$ , with no assumptions being made regarding the  $S_2$  population. Note that conditions (a) - (d) are not mutually exclusive; e.g., if  $c_2 \gg c_3 \gg c_1$ , then conditions (d) and (b) are both satisfied, and each of the two different formulas for  $a$  in those two cases reduces to the same result,  $c_1 c_3 x_{12} / c_2$ .

We pointed out that abridgment solely for the sake of reducing the size of the model is not always prudent. Abridging a set of reactions is always an approximation, so there is always some loss of accuracy. In particular, although we can be confident that in the scenarios (a) - (d) the true behavior of the  $S_3$  population in reactions (1) will be accurately replicated by the surrogate reaction, that might not be so for the  $S_1$  and  $S_2$  populations, since most model reductions will eliminate or severely constrain those two species. That might not matter if reactions (1) occur in isolation, in which case it would be a clear benefit of the abridgment. But it could give rise to a serious problem if reactions (1) are embedded in a larger network of reactions, some of which have  $S_1$  and  $S_2$  as reactants or products.

Since stochastic simulation is usually the tool of choice for analyzing complex cellular reaction networks, one reasonable goal of model reduction is to make stochastic simulation run faster. We showed that the maximum speedup factor in any single-reaction abridgment of reactions (1) is  $2(c_2 + c_3)/c_3$ . This implies that, of the four cases (a) - (d), the only one for which abridgment has a chance of producing a significant gain in simulation speed is case (d),  $c_2 \gg c_3$ . If that condition is satisfied, the speedup factor will be  $\gg 1$ . If it is not satisfied, the speedup factor will typically be rather small, and possibly not large enough to compensate for the loss of accuracy and robustness that invariably attends model reduction.

We showed that condition  $c_2 \gg c_3$  is the sole requirement for accurately applying the ssSSA procedure of Cao et al.<sup>6</sup> to reactions (1), contrary to earlier assertions.<sup>13</sup> We emphasized that the ssSSA implements a single-reaction abridgment of reactions (1) in a way that overcomes several shortcomings that arise if reactions (1) are simply replaced by reaction (2): In the ssSSA, the  $S_1$  and  $S_2$  populations are accurately represented on the timescale of reaction  $R_3$ , and additional slow reactions that involve  $S_1$  and  $S_2$  as reactants can easily be accommodated.

Finally, we showed that our first-passage-time analysis provides a framework which unites the abridgment under condition  $c_2 \gg c_3$  given by the ssSSA of Cao et al.<sup>6</sup>, and the abridgment under condition  $c_2, c_3 \gg c_1$  given by the sQSPA procedure of Mastny et al.<sup>7</sup> Furthermore, our first passage-time analysis enables us to identify *all* of the conditions under which a single-reaction abridgment of reactions (1) is possible.

## ACKNOWLEDGEMENTS

The authors thank Sotiria Lampoudi for some helpful discussions, and also the Journal's anonymous reviewer for some pertinent observations. The authors gratefully acknowledge financial support as follows: DG was supported by the California Institute of Technology through Consulting Agreement 102-1080890 pursuant to Grant R01GM078992 from the National Institute of General Medical Sciences, and through Contract 82-1083250 pursuant to Grant R01EB007511 from the National Institute of Biomedical Imaging and Bioengineering, and also from the University of California at Santa Barbara under Consulting Agreement 054281A20 pursuant to funding from the National Institutes of Health. YC was supported by the National Science Foundation under award CCF-0726763, and also the National Institutes of Health under awards GM073744 and GM078989. KS and LP were supported by Grant R01EB007511 from the National Institute of Biomedical Imaging and Bioengineering, Pfizer Inc., DOE DE-FG02-04ER25621, NSF IGERT DG02-21715, and the Institute for Collaborative Biotechnologies through Grant DFR3A-8-447850-23002 from the U.S. Army Research Office. KS was also supported by a National Science Foundation Graduate Research Fellowship. The content of this work is solely the responsibility of the authors, and does not necessarily reflect the official views of any of the aforementioned institutions.

## APPENDIX A: Proof of Theorem 1

(Necessity) Given (5), let  $P_0(\tau)$  denote the probability that the  $S_1$  molecule will *not* react during the next time span  $\tau$ . By the laws of probability, this function must satisfy  $P_0(\tau+d\tau) = P_0(\tau) \times [1 - cd\tau]$ , where the last factor is the probability that the  $S_1$  molecule, having not reacted in time  $\tau$ , will not react in  $[\tau, \tau + d\tau]$ . This implies the differential equation  $dP_0(\tau)/d\tau = -cP_0(\tau)$ . The solution of that equation for the initial condition  $P_0(0) = 1$  is  $P_0(\tau) = \exp(-c\tau)$ . Therefore, the probability that a given  $S_1$  molecule will react in the infinitesimal time interval  $[\tau, \tau + d\tau]$  is {the probability that it will *not* react in  $[0, \tau]$ } times {the probability that it *will* react in the next  $d\tau$ }:  $P_0(\tau) \times cd\tau = c \exp(-c\tau) d\tau$ . This implies that the probability density function (pdf) of the time for the  $S_1$  molecule to react is  $c \exp(-c\tau)$ , which is precisely the pdf of the exponential random variable with mean  $c^{-1}$ .

(Sufficiency) Given that the pdf of the time until the  $S_1$  molecule reacts is  $c \exp(-c\tau)$ , it follows that probability that the molecule will react in the time interval  $[\tau, \tau + d\tau]$  is  $c \exp(-c\tau) d\tau$ . Therefore, the probability that the molecule will react in the *next*  $d\tau$ , i.e., in the time interval  $[0, d\tau]$ , is  $c \exp(0) d\tau = cd\tau$ , as asserted in (5).

### APPENDIX B: PDF of the Time to the Next $R_3$ Event

Regard (1) as depicting the transitions of a single “random walker” among three “states”  $S_1$ ,  $S_2$  and  $S_3$ . We will first derive formulas for the pdfs  $P_{\alpha \rightarrow 3}(t)$  of the times  $T_{\alpha \rightarrow 3}$  ( $\alpha = 1$  or  $2$ ) required for the walker, starting in state  $S_\alpha$ , to *first* reach state  $S_3$ . To that end, let  $p(n, t | \alpha, 0)$  be the probability that the walker, having started at time 0 in state  $S_\alpha$  ( $\alpha = 1$  or  $2$ ), will be found at time  $t \geq 0$  to be in state  $S_n$  ( $n = 1, 2, 3$ ). Since according to (1) the walker will remain in state  $S_3$  after it arrives there, then

$$p(3, t | \alpha, 0) = \text{Prob}\{T_{\alpha \rightarrow 3} \leq t\}. \quad (\text{B1})$$

The probability on the right is, by definition, the cdf of the random variable  $T_{\alpha \rightarrow 3}$ . Since the derivative of the cdf with respect to  $\tau$  gives the corresponding pdf, then

$$P_{\alpha \rightarrow 3}(t) = \frac{dp(3, t | \alpha, 0)}{dt}. \quad (\text{B2})$$

The laws of probability give us the following relations among the functions  $p(n, t | \alpha, 0)$  at any time  $t$  and any infinitesimally later time  $t + dt$ :

$$p(1, t + dt | \alpha, 0) = p(1, t | \alpha, 0) \times [1 - c_1 dt] + p(2, t | \alpha, 0) \times c_2 dt, \quad (\text{B3a})$$

$$p(2, t + dt | \alpha, 0) = p(1, t | \alpha, 0) \times c_1 dt + p(2, t | \alpha, 0) \times [1 - c_2 dt - c_3 dt], \quad (\text{B3b})$$

$$p(3, t + dt | \alpha, 0) = p(3, t | \alpha, 0) + p(2, t | \alpha, 0) \times c_3 dt. \quad (\text{B3c})$$

For example, Eq. (B3a) is the statement that {the probability of being in state  $S_1$  at time  $t + dt$ } is equal to the *sum* of {the probability of being in state  $S_1$  at time  $t$  and then *not* jumping away in the next  $dt$ } plus {the probability of being in state  $S_2$  at time  $t$  and then jumping to state  $S_1$  in the next  $dt$ }. This logic ignores routes to state  $S_1$  at time  $t + dt$  that involve more than one jump in time  $[t, t + dt)$ , but that is permissible here since the probabilities for those paths will be of higher order than 1 in  $dt$ . Analogous reasoning gives Eqs. (B3b) and (B3c). By algebraically rearranging each of these equations, dividing through by  $dt$ , and then taking the limit  $dt \rightarrow 0$ , we obtain the following set of coupled ODEs, which constitute the “master equation” for this stochastic process:

$$\frac{dp(1, t | \alpha, 0)}{dt} = -c_1 p(1, t | \alpha, 0) + c_2 p(2, t | \alpha, 0), \quad (\text{B4a})$$

$$\frac{dp(2, t | \alpha, 0)}{dt} = c_1 p(1, t | \alpha, 0) - (c_2 + c_3) p(2, t | \alpha, 0), \quad (\text{B4b})$$

$$\frac{dp(3, t | \alpha, 0)}{dt} = c_3 p(2, t | \alpha, 0). \quad (\text{B5})$$

Note that Eqs. (B4a) and (B4b) constitute a closed pair of ODEs for  $p(1,t|\alpha,0)$  and  $p(2,t|\alpha,0)$ . Once that pair of equations has been solved,  $p(3,t|\alpha,0)$  can be obtained either by solving Eq. (B5), or more simply from the fact that

$$p(3,t|\alpha,0) = 1 - p(1,t|\alpha,0) - p(2,t|\alpha,0). \quad (\text{B6})$$

Combining Eqs. (B2) and (B5), we see that the function  $P_{\alpha \rightarrow 3}(t)$  can be computed as

$$P_{\alpha \rightarrow 3}(t) = c_3 p(2,t|\alpha,0). \quad (\text{B7})$$

Equations (B4) can be solved in a standard way that begins by writing them as

$$\frac{d\mathbf{p}_\alpha(t)}{dt} = -\mathbf{A} \cdot \mathbf{p}_\alpha(t), \quad (\text{B8})$$

where

$$\mathbf{p}_\alpha(t) \equiv \begin{pmatrix} p(1,t|\alpha,0) \\ p(2,t|\alpha,0) \end{pmatrix} \quad \text{and} \quad \mathbf{A} \equiv \begin{pmatrix} c_1 & -c_2 \\ -c_1 & c_2 + c_3 \end{pmatrix}. \quad (\text{B9})$$

The solution to (B8) turns out to involve the eigenvalues  $\lambda_+$  and  $\lambda_-$  of  $\mathbf{A}$ . These are evidently the solutions of the quadratic equation

$$(c_1 - \lambda_\pm)(c_2 + c_3 - \lambda_\pm) - c_1 c_2 = 0, \quad (\text{B10})$$

and are easily found to be

$$\lambda_\pm \equiv \frac{1}{2} \left[ (c_1 + c_2 + c_3) \pm \sqrt{(c_1 + c_2 + c_3)^2 - 4c_1 c_3} \right]. \quad (\text{B11})$$

A little algebra shows that the quantity under the radical here is never negative, so

$$0 \leq \lambda_- \leq \lambda_+. \quad (\text{B12})$$

We shall not belabor the process by which the solutions to Eqs. (B4) for  $\alpha = 1$  and 2 are obtained, because it can be verified by simple differentiation that the functions below satisfy Eqs. (B4). And it is also easy to verify that they satisfy the required initial conditions.

$$p(1,t|1,0) = \frac{1}{(\lambda_+ - \lambda_-)} \left[ (\lambda_+ - c_1) e^{-\lambda_- t} + (c_1 - \lambda_-) e^{-\lambda_+ t} \right], \quad (\text{B13a})$$

$$p(2,t|1,0) = \frac{c_1}{(\lambda_+ - \lambda_-)} \left[ e^{-\lambda_- t} - e^{-\lambda_+ t} \right]; \quad (\text{B13b})$$

$$p(1,t|2,0) = \frac{c_2}{(\lambda_+ - \lambda_-)} \left[ e^{-\lambda_- t} - e^{-\lambda_+ t} \right], \quad (\text{B14a})$$

$$p(2,t|2,0) = \frac{1}{(\lambda_+ - \lambda_-)} \left[ (c_1 - \lambda_-) e^{-\lambda_- t} + (\lambda_+ - c_1) e^{-\lambda_+ t} \right]. \quad (\text{B14b})$$

The pdf's of the first passage times  $T_{1 \rightarrow 3}$  and  $T_{2 \rightarrow 3}$  can now be obtained simply by substituting Eqs. (B13b) and (B14b) into formula (B7). However, our main concern here is with the more general case in which  $x_1$  random walkers are initially in state  $S_1$  and  $x_2$  are initially in state  $S_2$ . The pdf  $P(t; x_1, x_2)$  of the time  $T(x_1, x_2)$  required for the *first* of those walkers to reach state  $S_3$  can be computed by reasoning as follows: Since the walkers evolve independently of each other, the probability that *none* of them will reach state  $S_3$  *earlier* than time  $t$  is

$$\text{Prob}\{T(x_1, x_2) > t\} = \left(\text{Prob}\{T_{1 \rightarrow 3} > t\}\right)^{x_1} \left(\text{Prob}\{T_{2 \rightarrow 3} > t\}\right)^{x_2}. \quad (\text{B15})$$

This is equivalent to

$$\text{Prob}\{T(x_1, x_2) \leq t\} = 1 - \left(1 - \text{Prob}\{T_{1 \rightarrow 3} \leq t\}\right)^{x_1} \left(1 - \text{Prob}\{T_{2 \rightarrow 3} \leq t\}\right)^{x_2}.$$

Using (B1), this last equation can be written

$$\text{Prob}\{T(x_1, x_2) \leq t\} = 1 - \left(1 - p(3, t|1, 0)\right)^{x_1} \left(1 - p(3, t|2, 0)\right)^{x_2}. \quad (\text{B16})$$

But the left side of (B16) is, by definition, the cdf of the random variable  $T(x_1, x_2)$ . Therefore, the derivative of (B16) with respect to  $t$  gives the pdf of  $T(x_1, x_2)$ :

$$P(t; x_1, x_2) = \frac{d}{dt} \left[ 1 - \left(1 - p(3, t|1, 0)\right)^{x_1} \left(1 - p(3, t|2, 0)\right)^{x_2} \right]. \quad (\text{B17})$$

Upon evaluating this derivative with the help of Eqs. (B5) and (B6), we get

$$\begin{aligned} P(t; x_1, x_2) &= x_1 c_3 p(2, t|1, 0) \left(p(1, t|1, 0) + p(2, t|1, 0)\right)^{x_1 - 1} \left(p(1, t|2, 0) + p(2, t|2, 0)\right)^{x_2} \\ &\quad + x_2 c_3 p(2, t|2, 0) \left(p(1, t|1, 0) + p(2, t|1, 0)\right)^{x_1} \left(p(1, t|2, 0) + p(2, t|2, 0)\right)^{x_2 - 1}. \end{aligned} \quad (\text{B18})$$

Since all the  $p$ -functions on the right side of (B18) are given explicitly by Eqs. (B13) and (B14), we have in (B18) an exact, explicit formula for the pdf of the first-passage time  $T(x_1, x_2)$ .

### APPENDIX C: Proof of Theorem 2.

That it is possible, when  $c_2/c_3 \gg 1$ , to choose a time span  $\delta t$  that contains very many  $R_1$  and  $R_2$  events but practically no  $R_3$  events, follows from the fact established in Sec. II that successive  $R_3$  reactions will, on average, be separated by  $(c_2 + c_3)/c_3$  pairs of  $R_1$  and  $R_2$  reactions. Let  $[t', t' + dt')$  be an infinitesimal subinterval of the interval  $[t, t + \delta t)$ . The probability that  $R_3$  will fire in  $[t', t' + dt')$  is  $c_3 X_2(t') dt'$ . But

$$c_3 X_2(t') dt' \approx c_3 \hat{X}_2(t') dt', \quad (\text{C1})$$

because the dearth of  $R_3$  events in  $[t, t + \delta t)$  implies that the real fast process  $X_2(t')$  can be well approximated there by the virtual fast process  $\hat{X}_2(t')$ . The probability that  $R_3$  will fire *anywhere* in the interval  $[t, t + \delta t)$  can now be computed by summing the probabilities (C1) over all the  $dt'$  subintervals of  $[t, t + \delta t)$ :

$$\int_t^{t+\delta t} c_3 \hat{X}_2(t') dt' \equiv c_3 \left\{ \frac{1}{\delta t} \int_t^{t+\delta t} \hat{X}_2(t') dt' \right\} \delta t. \quad (\text{C2})$$

This invocation of the addition law of probability for mutually exclusive events is justified since the probability for more than one  $R_3$  firing in  $[t, t + \delta t)$  is practically zero. Now let  $K$  be an integer that is roughly equal to the expected number of firings of  $R_1$  and  $R_2$  in  $[t, t + \delta t)$ , a number that will be  $\gg 1$ . Subdividing  $[t, t + \delta t)$  into  $K$  subintervals of equal length  $\delta t/K$ , we can approximate the integral in braces in (C2) as

$$\frac{1}{\delta t} \int_t^{t+\delta t} \hat{X}_2(t') dt' \approx \frac{1}{\delta t} \sum_{k=1}^K \hat{X}_2(t_k) \left( \frac{\delta t}{K} \right) = \frac{1}{K} \sum_{k=1}^K \hat{X}_2(t_k), \quad (\text{C3})$$

where  $t_k$  ( $k=1, \dots, K$ ) locates the center of the  $k^{\text{th}}$  subinterval. After the first few  $R_1$  and  $R_2$  firings, the process  $\hat{X}_2(t)$  will effectively “de-correlate” and “relax” to its time-independent form  $\hat{X}_2(\infty)$ ; thus, the  $K$  values  $\hat{X}_2(t_1), \dots, \hat{X}_2(t_K)$  in (C3) can collectively be approximated by  $K$  *sample* values  $\hat{X}_2(\infty)^{(1)}, \dots, \hat{X}_2(\infty)^{(K)}$  of the random variable  $\hat{X}_2(\infty)$ . Equation (C3) then becomes

$$\frac{1}{\delta t} \int_t^{t+\delta t} \hat{X}_2(t') dt' \approx \frac{1}{K} \sum_{k=1}^K \hat{X}_2(\infty)^{(k)} \approx \langle \hat{X}_2(\infty) \rangle. \quad (\text{C4})$$

Substituting (C4) into (C2), we conclude that the probability that reaction  $R_3$  will fire in  $[t, t + \delta t)$  is approximately equal to  $c_3 \langle \hat{X}_2(\infty) \rangle \delta t$ . That is the first assertion of Theorem 2. The second assertion follows from the fact that, for any  $t' > t + \delta t$  prior to the next  $R_3$  event,  $\hat{X}_2(t')$  can be approximated by  $\hat{X}_2(\infty)$ .

## REFERENCES

- <sup>1</sup> L. Michaelis and M. L. Menten, *Biochem. Z.* **49**, 333 (1913).
- <sup>2</sup> G. E. Briggs and J. B. S. Haldane, *Biochem. J.* **19**, 338 (1925); D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry* (Freeman, 2005).
- <sup>3</sup> J. A. Borghans, R. J. deBoer, and L. A. Segel, *Bull. Math. Biol.* **58**, 43 (1996).
- <sup>4</sup> A. R. Tzafiriri and E. R. Edelman, *J. Theor. Biol.* **226**, 303 (2004).
- <sup>5</sup> D. Barik, M. Paul, W. Baumann, Y. Cao, and J. Tyson, *Biophysical Journal* (in press).
- <sup>6</sup> Y. Cao, D. T. Gillespie, and L. R. Petzold, *J. Chem. Phys.* **122**, 014116 (2005).
- <sup>7</sup> E. A. Mastny, E. L. Haseltine, and J. B. Rawlings, *J. Chem. Phys.* **127**, 094106 (2007).
- <sup>8</sup> For a review of the stochastic simulation algorithm, the chemical master equation, and related topics, see D. T. Gillespie, *Annu. Rev. Phys. Chem.* **58**, 35 (2007).
- <sup>9</sup> D. T. Gillespie, S. Lampoudi, and L. R. Petzold, *J. Chem. Phys.* **126**, 034302 (2007); S. Lampoudi, D. T. Gillespie, and L. R. Petzold, submitted.
- <sup>10</sup> If  $c_2 = 0$  and  $c_3 \geq c_1$ , Eqs. (8) give  $\lambda_+ = c_3$  and  $\lambda_- = c_1$ . Putting those results into Eq. (9) gives a formula that is indeterminate when  $c_3 = c_1$ . But applying L'Hospital's rule to that indeterminate form, taking derivatives with respect to  $c_3$ , yields the pdf  $c_1^2 t e^{-c_1 t}$ . This non-exponential form, which goes to zero as  $t \rightarrow 0$ , is the pdf of the gamma random variable  $\Gamma(c_1, 2)$ , which is defined as the sum of two statistically independent exponentials with the same mean  $c_1^{-1}$ . And this is exactly what we should expect for the time for an  $S_1$  to  $S_3$  conversion via reactions (1) when  $c_2 = 0$  and  $c_3 = c_1$ .
- <sup>11</sup> E. L. Haseltine and J. B. Rawlings, *J. Chem. Phys.* **117**, 6959 (2002).
- <sup>12</sup> C. Rao and A. P. Arkin, *J. Chem. Phys.* **118**, 4999 (2003).
- <sup>13</sup> In Ref. 6, it was stated that the condition for applying the ssSSA to reactions (1) is  $(c_1 + c_2)^2 \gg c_1 c_3 x_{12}$ . That is *incorrect*, as it arises from comparing a single-walker timescale with a many-walker timescale. The correct condition is simply  $c_2 \gg c_3$ , as can be seen not only from the result (3), but also from the argument at Eqs. (29). The reason why it is not necessary to supplement the condition  $c_2 \gg c_3$  with the condition  $c_1 \gg c_3$  is explained in the second paragraph of Sec. VI.
- <sup>14</sup> D. A McQuarrie, *J. Chem. Phys.* **38**, 433 (1963); I. G. Darvey, B. W. Ninham, and P. J. Staff, *J. Chem. Phys.* **45**, 2145 (1966).

## FIGURE CAPTIONS

**Fig. 1.** Semi-log plots of  $P(t; x_1, x_2)$  for  $c_1 = c_3 = 1$  and  $c_2 = 0.1$  for two cases: The solid curve is for  $x_1 = 1$  and  $x_2 = 0$ , from Eq. (9). The dashed curve is for  $x_1 = 0$  and  $x_2 = 1$ , from Eq. (10). Neither pdf has the straight-line form of an exponential pdf (there is a gradual change of slope in the dashed curve around  $t = 2$ ). The figure also shows that  $P(t; x_1, x_2)$  in this case depends on  $x_1$  and  $x_2$  individually, and not just on their sum.

**Fig. 2.** The solid curve shows a single  $X_3(t)$  trajectory obtained in an SSA run of reactions (1) with  $c_1 = c_3 = 1$ ,  $c_2 = 0.1$ ,  $X_1(0) = 300$ ,  $X_2(0) = X_3(0) = 0$ . The dashed curve shows the average of 10,000 such trajectories. The dotted curve plots the function (6b) with  $c = (2.1)^{-1}$ , which corresponds to the same mean  $S_1 \rightarrow S_3$  conversion time. The mismatch between the dashed curve and the dotted curve indicates the error that results from replacing reactions (1) with reaction (2) in this non-exponential case.

**Fig. 3.** A “true” picture of reactions (1) for the parameter settings (22) is provided by this SSA run of those reactions. Here the species populations have been plotted out only after each  $R_3$  event. Since the  $S_3$  population remains constant between successive  $R_3$  reactions, this plotting strategy reveals the full trajectory of  $X_3(t)$ . But of course, the populations of species  $S_1$  and  $S_2$  are not constant between successive  $R_3$  reactions.

**Fig. 4.** An SSA simulation of the surrogate reaction (2) with propensity function (21), using the same settings (22) as in Fig. 3. Only 300 reaction events were simulated here, as compared to  $1.2 \times 10^7$  reaction events in Fig. 3, so the gain in computational speed over reactions (1) is truly enormous. The  $X_3(t)$  trajectory has been accurately rendered. But the  $X_1(t)$  trajectory has not, and the  $X_2(t)$  trajectory has been completely lost.

**Fig. 5.** An ssSSA simulation of reactions (1) using the same settings (22) as used in Figs. 3 and 4. Here the “fast reactions”  $R_1$  and  $R_2$  have been skipped over, and only firings of the “slow reaction”  $R_3$  have been simulated, using however the modified propensity function (28) or (21). As in Fig. 4, only 300 reaction events were simulated in this run (but here those were “modified  $R_3$ ” reactions), and the population of the “slow species”  $S_3$  has been accurately rendered. But this ssSSA run evidently gives a much more accurate picture of the behavior of the “fast species”  $S_1$  and  $S_2$  than does the run in Fig. 4. Notice also that the initial rapid relaxation in Fig. 3 of  $X_1$  (from 300) and  $X_2$  (from 0) is accurately replicated in this ssSSA run.

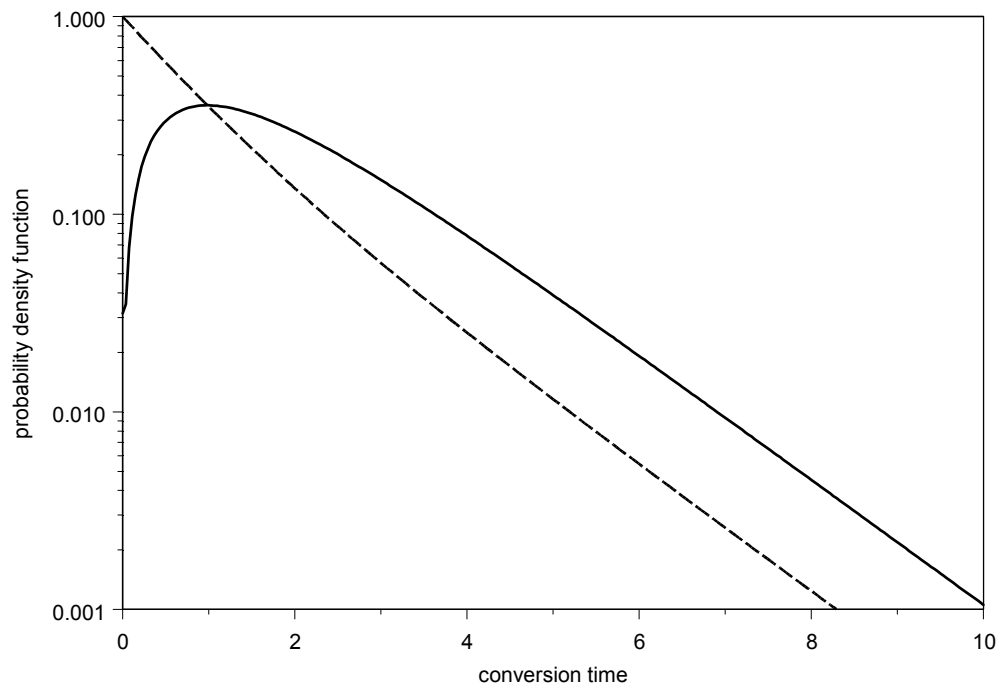


Figure 1

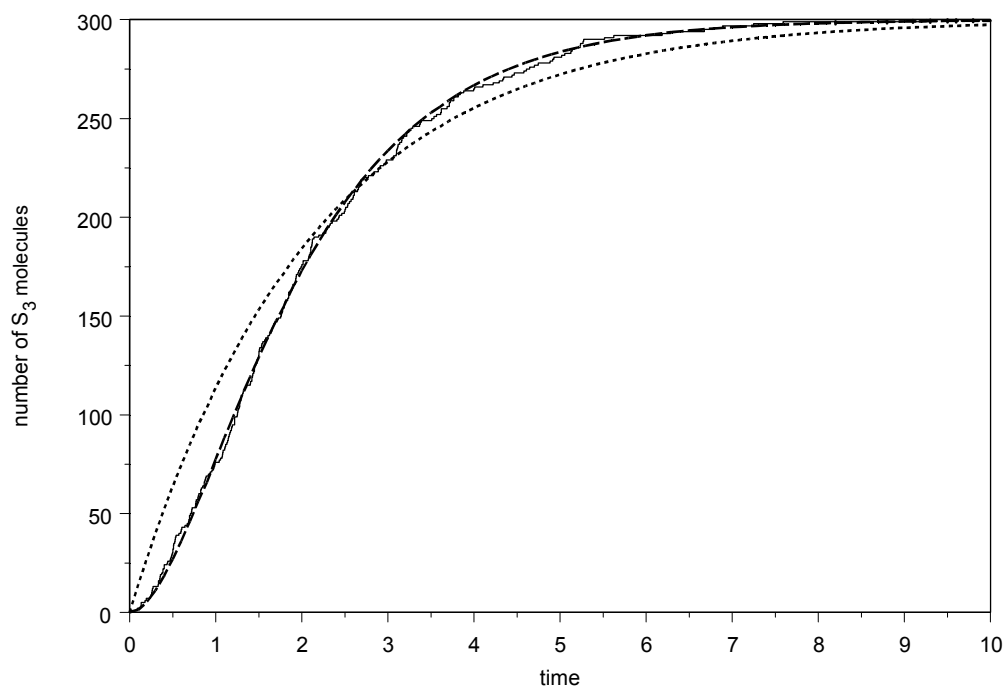


Figure 2

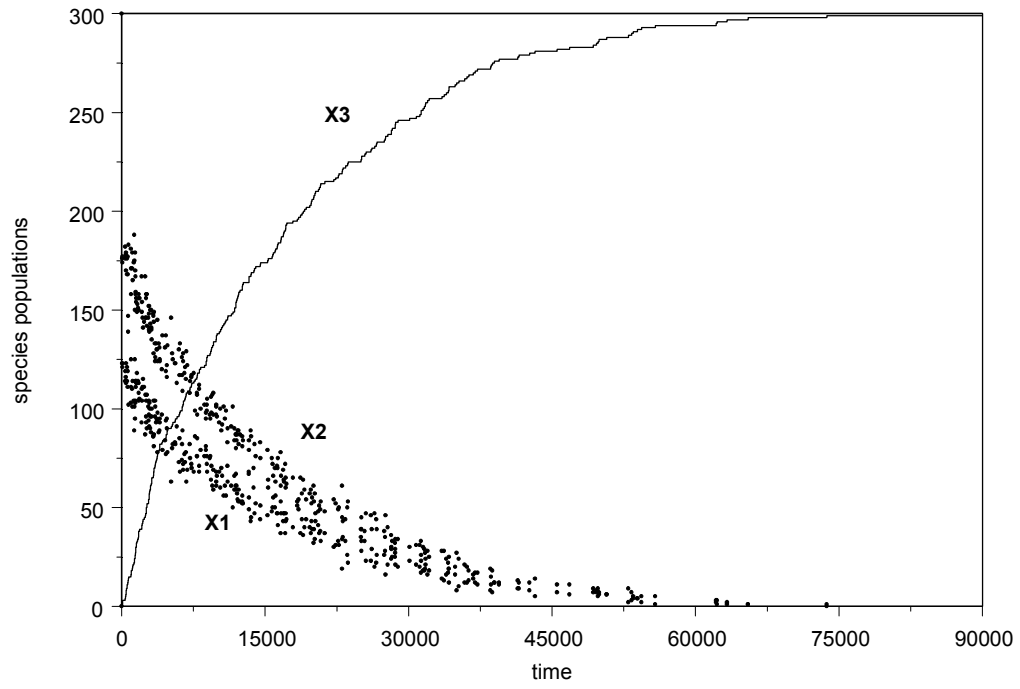


Figure 3

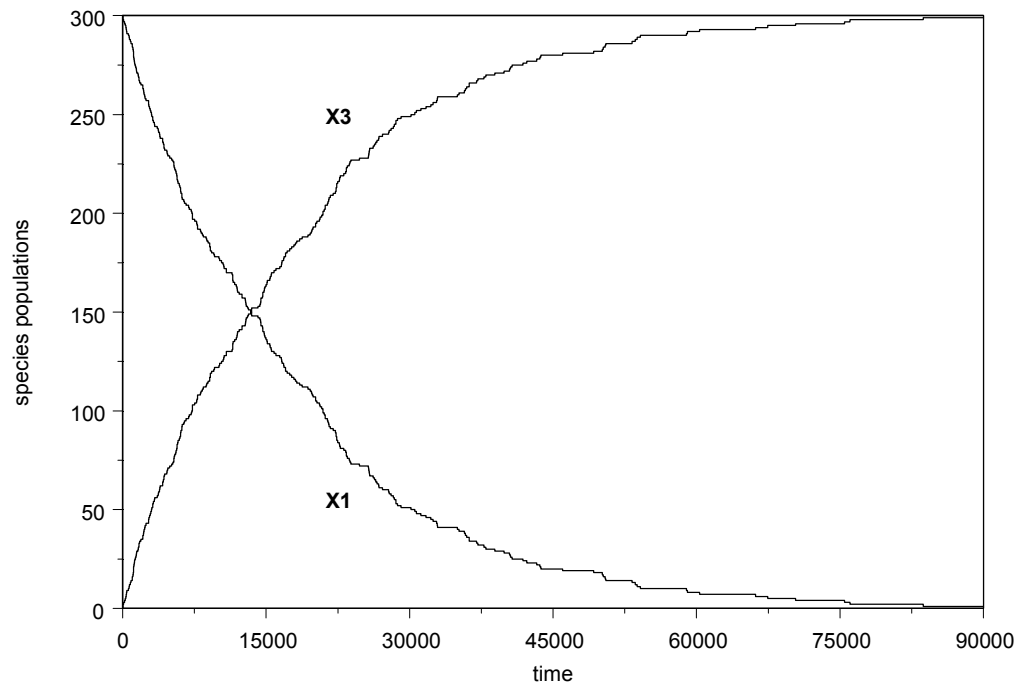


Figure 4

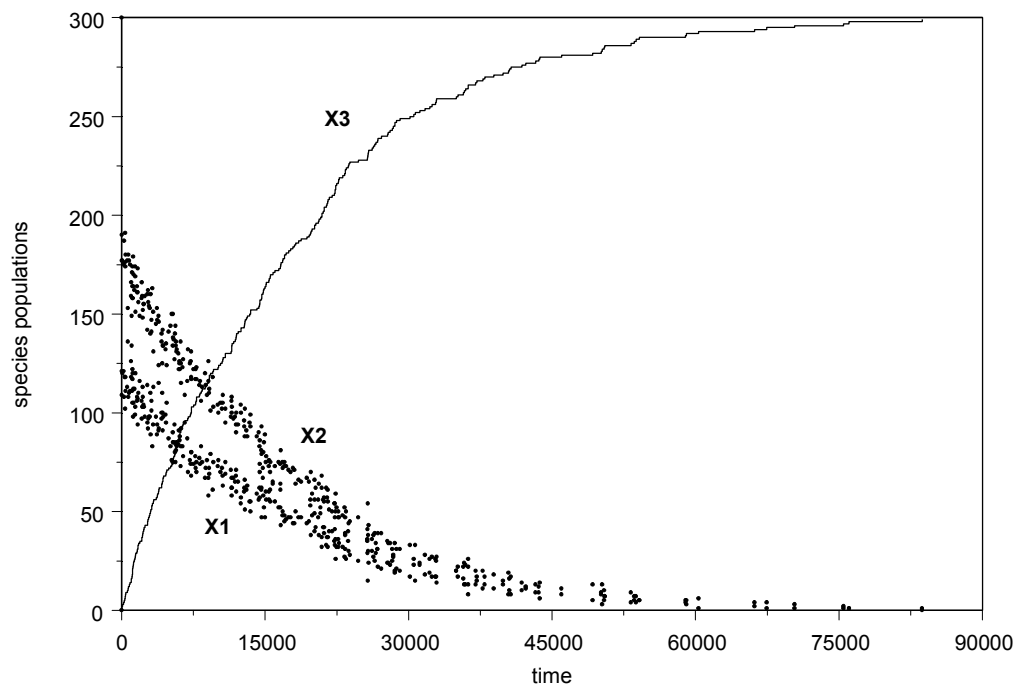


Figure 5