The Slow-Scale Stochastic Simulation Algorithm

Yang Cao
Department of Computer Science, University of California Santa Barbara, Santa Barbara, California (ycao@engineering.ucsb.edu)

Daniel T. Gillespie
Dan T Gillespie Consulting, Castaic, California (gillespiedt@mailaps.org)

Linda R. Petzold
Department of Computer Science, University of California Santa Barbara, Santa Barbara, California (petzold@engineering.ucsb.edu)

Abstract: Reactions in real chemical systems often take place on vastly different time scales, with "fast" reaction channels firing very much more frequently than "slow" ones. These firings will be interdependent if, as is usually the case, the fast and slow reactions involve some of the same species. An exact stochastic simulation of such a system will necessarily spend most of its time simulating the more numerous fast reaction events. This is a frustratingly inefficient allocation of computational effort when dynamical stiffness is present, since in that case a fast reaction event will be of much less importance to the system's evolution than will a slow reaction event. For such situations, this paper develops a systematic approximate theory that allows one to stochastically advance the system in time by simulating the firings of only the slow reaction events. Developing an effective strategy to implement this theory poses some challenges, but as is illustrated here for two simple systems, when those challenges can be overcome, very substantial increases in simulation speed can be realized.

JCP MS# A4.07.041

Second Revised Version (September 26, 2004)

I. INTRODUCTION

In cellular systems, where the small number of molecules of a few reactant species sometimes necessitates a stochastic description of the system's temporal behavior, chemical reactions often take place on vastly different time scales. When that happens, a chronological log of successive reaction events, such as would be obtained from an application of the stochastic simulation algorithm (SSA),[1] would reveal that the overwhelming majority of the reaction events are firings of just a few reaction channels, the so-called "fast" reactions. But it is usually the case that the less frequent "slow" reactions will have a greater impact on the behavior of the system. Since the SSA treats all reaction events alike, it will spend the great majority of its time simulating the many relatively uninteresting fast reaction events. The question arises, is there legitimate a way to skip over the fast reactions, and explicitly simulate only the slow reactions?

If the fast and slow reactions do not involve the same species this is of course very easy to do. In that case we essentially have two dynamically independent systems, and even though they evolve in the same physical space they can be numerically simulated independently of each other. Much more common though are situations in which the fast and slow reactions share some species. And when, for instance, the population of a species that is a reactant in some slow reaction gets changed by firings of one or more fast reactions, then the slow reaction will be dependent on the fast reactions. In such a situation, it is not obvious how, or even if, we can legitimately simulate the slow reactions without also simulating the fast ones.

Attempts to solve this problem have been made by other investigators.[2,3] These attempts basically approximate the fast reactions using something akin to the deterministic reaction rate equation, and then try to treat the slow reactions stochastically. In the main, we agree that this is essentially what should be done. But subtle differences in the procedures advocated thus far underscore the fact that there are unresolved fundamental questions about how we should go about untangling the fast and slow parts of a system for separate treatment: Are the descriptors "fast" and "slow" more aptly applied to reactions or to species? Can one actually identify a fast sub-system that is physically meaningful, yet also mathematically more tractable than the full system? These are just two of the questions with which we shall try to come to terms in this paper.

Our aim here will not be to propose an *ad hoc* simulation recipe whose correctness can be assessed only by comparing its predictions with those of the exact SSA for a variety of test systems. Rather, we shall try to logically deduce what an *a priori* correct multi-scale stochastic simulation algorithm *ought* to do. We shall then demonstrate for two simple systems how such an algorithm might be implemented, although implementation strategies for more complicated systems will be left as an open issue.

II. FAST AND SLOW REACTION CHANNELS

We consider a well-stirred chemical system in which N molecular species $\{S_1,\ldots,S_N\}$ interact through M elementary reaction channels $\{R_1,\ldots,R_M\}$. The *state* of the system is $X(t)\equiv\big(X_1(t),\ldots,X_N(t)\big)$, where $X_i(t)$ is the number of S_i molecules in the system at time t. Reaction channel R_j is characterized by a *propensity function*

$a_j(x)$, where $a_j(x)\,dt$ is the probability given $X(t)=x$ that one R_j reaction will occur in the infinitesimal time interval $[t,t+dt)$, and a *state-change vector* $\nu_j \equiv \left(\nu_{1j},\ldots,\nu_{Nj}\right)$, where ν_{ij} is the change in the S_i population induced by one R_j event.

Our interest here will be exclusively with systems in which some "relatively unimportant" fast reaction channels are firing much more frequently than all the other slow reaction channels. But the procedure we shall use to identify which reaction channels are "fast" and which are "slow" is subtle, and has *two stages*. In the first stage, we make a *provisional* partitioning of the reactions based the usual values of their propensity functions: Reactions whose propensity functions are usually much larger than the propensity functions of all the other reactions are called "fast", and all the other reactions are called "slow". To distinguish between these two classes of reactions, we re-label them thusly:

$$R^{\mathrm{f}} \equiv \left\{R_1^{\mathrm{f}},\ldots,R_{M_{\mathrm{f}}}^{\mathrm{f}}\right\}, \text{ the set of } \textit{fast reactions}, \tag{1a}$$

$$R^{\mathrm{s}} \equiv \left\{R_1^{\mathrm{s}},\ldots,R_{M_{\mathrm{s}}}^{\mathrm{s}}\right\}, \text{ the set of } \textit{slow reactions}, \tag{1b}$$

with $M_{\mathrm{f}}+M_{\mathrm{s}}=M$. This naturally induces a similar re-labeling of the corresponding propensity functions, the fast ones being $a_1^{\mathrm{f}},\ldots,a_{M_{\mathrm{f}}}^{\mathrm{f}}$ and the slow ones being $a_1^{\mathrm{s}},\ldots,a_{M_{\mathrm{s}}}^{\mathrm{s}}$.

The overall result of this partitioning of the reactions will be that the expected time to the occurrence of the next fast reaction will usually be very much smaller than the expected time to the occurrence of the next slow reaction. *But this partitioning of the reactions on the basis of their propensity function values is tentative and provisional.* As will be explained in Sec. V, we may later find it necessary to change some reaction assignments in order to satisfy some other more critical conditions. Securing the satisfaction of these latter conditions (stated in Sec. V) will constitute the second and final stage of our procedure for partitioning the reactions into fast and slow subsets.

III. FAST AND SLOW SPECIES

Having partitioned and re-labeled the M *reactions*, we now make a similar partitioning and re-labeling of the N *species*: $S=\left(S^{\mathrm{f}},S^{\mathrm{s}}\right)$, where

$$S^{\mathrm{f}} \equiv \left\{S_1^{\mathrm{f}},\ldots,S_{N_{\mathrm{f}}}^{\mathrm{f}}\right\}, \text{ the set of } \textit{fast species}, \tag{2a}$$

$$S^{\mathrm{s}} \equiv \left\{S_1^{\mathrm{s}},\ldots,S_{N_{\mathrm{s}}}^{\mathrm{s}}\right\}, \text{ the set of } \textit{slow species}, \tag{2b}$$

with $N_{\mathrm{f}}+N_{\mathrm{s}}=N$. This will of course give rise to a like partitioning of the state vector $X(t)=\left(X^{\mathrm{f}}(t),X^{\mathrm{s}}(t)\right)$, and also the generic state space variable $x=\left(x^{\mathrm{f}},x^{\mathrm{s}}\right)$, into fast and slow parts, and their components will be similarly subscripted. Our criterion for making this species partitioning is simple and sharp: We define a "fast species" to be any species whose population gets changed by some fast reaction, and a "slow species" to be any species whose population does not get changed by any fast reaction. Note the asymmetry in this definition: a slow species cannot get changed by a fast reaction , but a fast species can get changed by a slow reaction.

The fast and slow reaction propensity functions will depend in general on *both* fast and slow species:

$$a_j^{\mathrm{f}}(x) = a_j^{\mathrm{f}}(x^{\mathrm{f}}, x^{\mathrm{s}}), \quad j = 1, \ldots, M_{\mathrm{f}}, \tag{3a}$$

$$a_j^{\mathrm{s}}(x) = a_j^{\mathrm{s}}(x^{\mathrm{f}}, x^{\mathrm{s}}), \quad j = 1, \ldots, M_{\mathrm{s}}. \tag{3b}$$

The corresponding fast and slow reaction *state-change vectors* now appear as

$$\nu_j^{\mathrm{f}} \triangleq \left(\nu_{1j}^{\mathrm{ff}}, \ldots, \nu_{N_{\mathrm{f}} j}^{\mathrm{ff}}\right), \quad j = 1, \ldots, M_{\mathrm{f}}, \tag{4a}$$

$$\nu_j^{\mathrm{s}} \triangleq \left(\nu_{1j}^{\mathrm{fs}}, \ldots, \nu_{N_{\mathrm{f}} j}^{\mathrm{fs}}, \nu_{1j}^{\mathrm{ss}}, \ldots, \nu_{N_{\mathrm{s}} j}^{\mathrm{ss}}\right), \quad j = 1, \ldots, M_{\mathrm{s}}, \tag{4b}$$

where $\nu_{ij}^{\sigma\rho}$ is by definition the change in the number of molecules of species S_i^{σ} ($\sigma = \mathrm{f}, \mathrm{s}$) induced by one reaction R_j^{ρ} ($\rho = \mathrm{f}, \mathrm{s}$). Since by definition slow species do not get changed by fast reactions, $\nu_{ij}^{\mathrm{sf}} \equiv 0$; accordingly, we have dropped those zero components from the R_j^{f} state-change vector ν_j^{f} in Eq. (4a), and we henceforth regard ν_j^{f} to be a vector with the same dimensionality (N_{f}) as the fast species state vector X^{f}. This will be important for our later analysis.

IV. THE VIRTUAL FAST PROCESS

The full system state vector $X(t) = \left(X^{\mathrm{f}}(t), X^{\mathrm{s}}(t)\right)$ evolves as a self-contained, past-forgetting, and hence *Markovian* process; accordingly, it obeys the Markovian chemical master equation (CME), and it can be simulated by the Markovian stochastic simulation algorithm (SSA). But this will usually not be true for the individual component processes $X^{\mathrm{f}}(t)$ and $X^{\mathrm{s}}(t)$, because they are coupled; e.g., $X^{\mathrm{f}}(t)$ will be Markovian only if it evolves completely independently of $X^{\mathrm{s}}(t)$, and that will never be the case in situations of interest to us. Since non-Markovian processes are notoriously difficult to work with, we now introduce a new *virtual* fast process $\hat{X}^{\mathrm{f}}(t)$, which *is* Markovian.

By definition, $\hat{X}^{\mathrm{f}}(t)$ is composed of the same fast species state variables as $X^{\mathrm{f}}(t)$, but it evolves *only* through the fast reactions R^{f}. In other words, $\hat{X}^{\mathrm{f}}(t)$ is $X^{\mathrm{f}}(t)$ with all the slow reactions turned off. Switching off the slow reactions gives us a Markov process $\hat{X}^{\mathrm{f}}(t)$ for two reasons: First, the fast state variables now get changed *only* by the fast reactions. And second, any x^{s} that appears explicitly as an argument (for a catalyst species) in a fast reaction propensity function $a_j^{\mathrm{f}}(x^{\mathrm{f}}, x^{\mathrm{s}})$ will now be a *constant parameter* instead of a dynamical variable. $\hat{X}^{\mathrm{f}}(t)$ thus obeys the *virtual fast CME*,

$$\frac{\partial \hat{P}(x^{\mathrm{f}}, t \,|\, x_0, t_0)}{\partial t} = \sum_{j=1}^{M_{\mathrm{f}}} \Big\{ a_j^{\mathrm{f}}(x^{\mathrm{f}} - \nu_j^{\mathrm{f}}, x_0^{\mathrm{s}}) \hat{P}(x^{\mathrm{f}} - \nu_j^{\mathrm{f}}, t \,|\, x_0, t_0) - a_j^{\mathrm{f}}(x^{\mathrm{f}}, x_0^{\mathrm{s}}) \hat{P}(x^{\mathrm{f}}, t \,|\, x_0, t_0) \Big\}, \tag{5}$$

where

$$\hat{P}(x^{\mathrm{f}},t\,|\,x_0,t_0) \triangleq \Pr\left\{\hat{X}^{\mathrm{f}}(t)=x^{\mathrm{f}}\,|\,X(t_0)=x_0\right\}. \tag{6}$$

Equation (5) is an "ordinary" CME, since ν_j^{f} lies in the same space as x^{f}, and all sources of change in $\hat{X}^{\mathrm{f}}(t)$ are accounted for on the right hand side.

$\hat{X}^{\mathrm{f}}(t)$ will obviously be a more tractable process than $X(t)$ since it has fewer species and fewer reaction channels. But $\hat{X}^{\mathrm{f}}(t)$ will also be more tractable than $X^{\mathrm{f}}(t)$; that's because in practice there is usually no simpler way of solving for $X^{\mathrm{f}}(t)$ than solving for the full process $X(t)$. But $\hat{X}^{\mathrm{f}}(t)$, although Markovian, is not a physically real process, since in reality we can't "turn off" the slow reactions. We shall see shortly though that, under certain restrictive conditions which will be approximately realized for our purposes here, $\hat{X}^{\mathrm{f}}(t)$ can provide an acceptable *approximation* to $X^{\mathrm{f}}(t)$.

V. STOCHASTIC STIFFNESS

The existence of fast and slow reactions and species is quite common in real chemical systems, and in the context of a traditional deterministic ODE analysis it often gives rise to the problem of *stiffness*. Stiffness is technically defined as the presence in the system of dynamical modes that evolve on widely different time scales, with the *fastest* mode being *stable*. We can translate the latter stability requirement into the stochastic context by requiring $\hat{X}^{\mathrm{f}}(t)$ to be a "stable process"; technically this means that the limit

$$\lim_{t\to\infty}\hat{P}(x^{\mathrm{f}},t\,|\,x_0,t_0) \equiv \hat{P}(x^{\mathrm{f}},\infty\,|\,x_0) \tag{7}$$

exists independently of t and t_0. Equation (7) does not mean that the state of the system eventually stops changing with time, but only that our best estimate of the system's state eventually stops changing with time. The function $\hat{P}(x^{\mathrm{f}},\infty\,|\,x_0)$ gives the probability that $\hat{X}^{\mathrm{f}}(\infty)$ will equal x^{f}, given the initial value x_0. In principle, this function can be calculated by solving the time-stationary form of the virtual fast CME (5), namely

$$0=\sum_{j=1}^{M_{\mathrm{f}}}\left\{a_j^{\mathrm{f}}(x^{\mathrm{f}}-\nu_j^{\mathrm{f}},x_0^{\mathrm{s}})\hat{P}(x^{\mathrm{f}}-\nu_j^{\mathrm{f}},\infty\,|\,x_0)-a_j^{\mathrm{f}}(x^{\mathrm{f}},x_0^{\mathrm{s}})\hat{P}(x^{\mathrm{f}},\infty\,|\,x_0)\right\}. \tag{8}$$

As a set of purely *algebraic* equations, (8) will be much easier to work with than the set *differential* equations (5); thus, it should always be easier to compute the properties of $\hat{X}^{\mathrm{f}}(\infty)$ than the properties of $\hat{X}^{\mathrm{f}}(t)$ for some finite t. Of course, this is not to say that solving (8) exactly for $\hat{P}(x^{\mathrm{f}},\infty\,|\,x_0)$ will always be practicable.

In the preceding sections, we have described a fairly straightforward procedure for partitioning the system into fast and slow reactions and fast and slow species, and then extracting from those partitionings a virtual fast process $\hat{X}^{\mathrm{f}}(t)$. But in practice, we will be motivated to do all this *only* in circumstances where *dynamical stiffness* is causing a problem, namely, the burying of "important" slow reaction events in a vast multitude of

"unimportant" fast reaction events. In order to focus our analysis on that special circumstance – and to exclude circumstances in which the fast reaction events are no less important than the slow ones – we now impose *two key requirements* on our system. *These requirements may force us to change our initial tentative designations of fast and slow reactions based on propensity function values*.

Our first requirement is that the virtual fast process $\hat{X}^{\mathrm{f}}(t)$ must be *stable*; i.e. $\hat{P}(x^{\mathrm{f}},\infty|x_0)$ in Eq. (7) must exist as a well behaved, time-independent probability function. Our second requirement is that the relaxation of $\hat{X}^{\mathrm{f}}(t)$ to its stationary asymptotic form, $\hat{X}^{\mathrm{f}}(t)\to\hat{X}^{\mathrm{f}}(\infty)$, must happen *very quickly* on the time scale of the *slow* reactions. A more precise way of stating this second requirement is to say that "the relaxation time of the virtual fast process" must be very much less than "the expected time to the next slow reaction"; later we shall see how these two times can be quantitatively estimated for two specific systems.

The two requirements we have just imposed mean the system has to be "stiff" in a *stochastic* context; i.e., it must evolve on widely different time scales, with the fastest evolving mode being stable. We shall often refer to these two conditions as the "stiffness conditions". If satisfying these conditions can be accomplished only by making some changes in the way we originally partitioned the reactions into fast and slow subsets, then we make those changes, regardless of the values of propensity functions. On the other hand, if these conditions *cannot* be satisfied by making such changes, we shall take that as a sign that the fast reactions are no less important than the slow ones, so it is not a good idea to try skipping over the fast reactions.

VI. THE SLOW-SCALE PROPENSITY FUNCTIONS

Thus far we have simply made some definitions – of fast and slow reactions, fast and slow species, and a virtual fast process – all framed in a setting of "dynamical stiffness" as described by the two conditions set forth in Sec. V. We are now ready to establish the key result of this paper.

Recall that the R_j^{s} propensity function $a_j^{\mathrm{s}}(x^{\mathrm{f}},x^{\mathrm{s}})$ is defined so that, if the system is in the state $(x^{\mathrm{f}},x^{\mathrm{s}})$ at time t, then $a_j^{\mathrm{s}}(x^{\mathrm{f}},x^{\mathrm{s}})\,dt$ gives the probability that one R_j^{s} reaction will occur in the next *infinitesimal* time interval $[t,t+dt)$. We would like to find an "effective" R_j^{s} propensity function, one whose product with a *finite* time Δ_{s}, which is stipulated to be *very small* on the time scale of the *slow* reactions but *very large* on the time scale of the *fast* reactions, provides an acceptable approximation to the probability that one R_j^{s} reaction will occur in the next finite time interval $[t,t+\Delta_{\mathrm{s}})$.

The simplest choice for this effective propensity function would of course be the propensity function itself; i.e., we could just estimate the probability in question as $a_j^{\mathrm{s}}(x^{\mathrm{f}},x^{\mathrm{s}})\Delta_{\mathrm{s}}$. The problem with that doing that is that it ignores the evolution of the fast state variable during $[t,t+\Delta_{\mathrm{s}})$. A possible refinement would be to replace x^{f} in $a_j^{\mathrm{s}}(x^{\mathrm{f}},x^{\mathrm{s}})\Delta_{\mathrm{s}}$ with a *sample* of the random variable $\hat{X}^{\mathrm{f}}(\infty)$.

Although there will be circumstances in which either of these approximations will be satisfactory, there will also be circumstances in which neither will suffice. We shall show that a more carefully reasoned analysis leads to the following result.

The Slow-Scale Approximation: Let the system be in state $(x^{\mathrm{f}}, x^{\mathrm{s}})$ at time t. And let the fast and slow time scales of the system be well separated, in the sense that the relaxation time of the (stable) virtual fast process $\hat{X}^{\mathrm{f}}(t)$ is very small compared to the expected time to the next slow reaction. Then if Δ_{s} is a time increment that is large compared to the former time but small compared to the latter, the probability that one R_j^{s} reaction will occur in the time interval $[t, t+\Delta_{\mathrm{s}})$ can be well approximated by $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}})\,\Delta_{\mathrm{s}}$, where

$$\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}}) \triangleq \sum_{x^{\mathrm{f}\prime}} \hat{P}(x^{\mathrm{f}\prime}, \infty \,|\, x^{\mathrm{f}}, x^{\mathrm{s}})\, a_j^{\mathrm{s}}(x^{\mathrm{f}\prime}, x^{\mathrm{s}}), \tag{9}$$

$\hat{P}$ being the probability density function of $\hat{X}^{\mathrm{f}}(\infty)$.

We shall call the function $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}})$ defined in (9) the *slow-scale propensity function* for reaction channel R_j^{s}. It is evidently the average of the regular R_j^{s} propensity function over the fast variables, treated as though they were distributed according to the asymptotic virtual fast process $\hat{X}^{\mathrm{f}}(\infty)$. Note that $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}})$ depends in general on the values of *both* the fast and slow state variables at the *beginning* of the interval $[t, t+\Delta_{\mathrm{s}})$. We shall now give an analytical justification of the fundamental property that the Slow-Scale Approximation attributes to the function defined in (9). In the next section, we shall see how that property allows us to approximately simulate the evolution of the system *one slow reaction at a time*.

Justification*:* With the system in state $(x^{\mathrm{f}}, x^{\mathrm{s}})$ at time t, divide the time interval $[t, t+\Delta_{\mathrm{s}})$ into infinitesimally small subintervals, and consider a typical such subinterval, $[t', t'+dt')$. In the interval $[t, t')$ just preceding this infinitesimal subinterval, fast reactions may have been firing but slow reactions, to a good approximation, have *not*; because, by hypothesis it is very unlikely for any slow reaction to occur in the entire interval $[t, t+\Delta_{\mathrm{s}})$. Since fast reactions do not alter the populations of the slow species, we therefore have $X(t') \approx (X^{\mathrm{f}}(t'), x^{\mathrm{s}})$; hence, the probability that one R_j^{s} reaction will occur in the infinitesimal subinterval $[t', t'+dt')$ is approximately $a_j^{\mathrm{s}}\left(X^{\mathrm{f}}(t'), x^{\mathrm{s}}\right) dt'$.

Since there is a nil probability of *more* than one slow reaction occurring in the interval $[t, t+\Delta_{\mathrm{s}})$, we can regard occurrences of an R_j^{s} reaction in *all* of the infinitesimal subintervals of $[t, t+\Delta_{\mathrm{s}})$ as mutually exclusive events. We can then invoke the addition law of probability, and compute the probability that an R_j^{s} reaction will occur in *any* of those infinitesimal subintervals as the *sum* of the individual probabilities. Therefore, the

probability that one R_j^{s} reaction will occur in the entire interval $[t, t+\Delta_{\mathrm{s}})$ is (approximately)

$$\int_t^{t+\Delta_{\mathrm{s}}} a_j^{\mathrm{s}}\left(X^{\mathrm{f}}(t'), x^{\mathrm{s}}\right) dt' \approx \left(\frac{1}{\Delta_{\mathrm{s}}}\int_t^{t+\Delta_{\mathrm{s}}} a_j^{\mathrm{s}}\left(\hat{X}^{\mathrm{f}}(t'), x^{\mathrm{s}}\right) dt'\right)\Delta_{\mathrm{s}}\,.$$

The replacement of $X^{\mathrm{f}}(t')$ with $\hat{X}^{\mathrm{f}}(t')$ in the last step here is justified because those two processes will be the same if the slow reactions are not firing, and to a good approximation they are *not* over the interval $[t, t+\Delta_{\mathrm{s}})$.

Finally, we invoke the hypothesized fact that Δ_{s}, although very small on the time scale of the slow process $X^{\mathrm{s}}(t')$, is *very large* compared to the time it takes the virtual fast process $\hat{X}^{\mathrm{f}}(t')$ to relax to its asymptotic form $\hat{X}^{\mathrm{f}}(\infty)$. In that case, the quantity in parentheses on the right approximates the $\Delta_{\mathrm{s}} \to \infty$ *temporal average* of the function $a_j^{\mathrm{s}}(\hat{X}^{\mathrm{f}}(t'), x^{\mathrm{s}})$. And following a practice that is very common in statistical physics, we can estimate this *temporal* average by the *ensemble* average with respect to the *asymptotic* virtual fast process, $\hat{X}^{\mathrm{f}}(\infty)$: [4]

$$\frac{1}{\Delta_{\mathrm{s}}}\int_t^{t+\Delta_{\mathrm{s}}} a_j^{\mathrm{s}}\left(\hat{X}^{\mathrm{f}}(t'), x^{\mathrm{s}}\right) dt' \approx \sum_{x^{\mathrm{f}\prime}} \Pr\left\{\hat{X}^{\mathrm{f}}(\infty) = x^{\mathrm{f}\prime}\right\} a_j^{\mathrm{s}}(x^{\mathrm{f}\prime}, x^{\mathrm{s}})\,.$$

Since $\Pr\left\{\hat{X}^{\mathrm{f}}(\infty) = x^{\mathrm{f}\prime}\right\} = \hat{P}(x^{\mathrm{f}\prime}, \infty | x^{\mathrm{f}}, x^{\mathrm{s}})$, the quantity on the right is the function defined in (9); therefore, by virtue of the previous equation, we conclude that the probability that one R_j^{s} reaction will occur in the time interval $[t, t+\Delta_{\mathrm{s}})$ is indeed as asserted by the Slow-Scale Approximation.

VII. THE SLOW-SCALE SSA

We are assuming now that our system is such that there exists a "quasi-infinitesimal" time interval $d_{\mathrm{s}}t$, which is essentially an *infinitesimal* on the time scale of the slow reactions but *very large* compared to the relaxation time of the virtual fast process. The Slow-Scale Approximation tells us that in this circumstance, given $X(t) = (x^{\mathrm{f}}, x^{\mathrm{s}})$, the probability that one R_j^{s} reaction will occur in $[t, t+d_{\mathrm{s}}t)$ is approximately given by $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}})\, d_{\mathrm{s}}t$, where $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}})$ is defined in (9). With this result, we can now proceed using arguments that parallel those used in deriving the standard SSA: [1]

First defining

$$\overline{a}_0^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}}) \triangleq \sum_{j=1}^{M_{\mathrm{s}}} \overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}}; x^{\mathrm{f}})\,, \tag{10}$$

we can prove that the probability that the next slow reaction will occur in the quasi-infinitesimal time interval $[t+\tau, t+\tau+d_s\tau)$ *and* will be an R_j^s reaction is (approximately)

$$\overline{p}_j^s(\tau, j|x^f, x^s, t)\, d_s\tau = \exp\left(-\overline{a}_0^s(x^s;x^f)\,\tau\right)\times \overline{a}_j^s(x^s;x^f)\, d_s\tau, \quad j=1,\ldots,M_s. \tag{11}$$

From this we can go on to show that, given $X(t)=(x^f,x^s)$, the time τ to the *next slow* reaction and the index j of that reaction can be (approximately) generated by the following two formulas, wherein r_1 and r_2 are unit-interval uniform random numbers:

$$\tau = \frac{1}{\overline{a}_0^s(x^s;x^f)}\ln\left(\frac{1}{r_1}\right), \tag{12a}$$

$$j = \text{ the smallest integer satisfying } \sum_{j'=1}^{j}\overline{a}_{j'}^s(x^s;x^f) \ge r_2\,\overline{a}_0^s(x^s;x^f). \tag{12b}$$

With this ability to estimate the time to and the index of the next slow reaction, we can now construct an approximate general procedure for stochastically simulating the evolution of the system *one slow reaction at a time* (explanatory details will follow):

The Slow-Scale SSA:

Preparation: Set all parameter values. Partition the system into fast and slow reactions and species. Identify the virtual fast process, and compute on the basis of Eq. (8) its stationary probability function $\hat{P}(x^{f\prime},\infty|x^f,x^s)$.

Initialization: Given the initial state $X(t_0)=(x_0^f,x_0^s)$, initialize the time and state variables by setting $t=t_0$, $x^f=x_0^f$, and $x^s=x_0^s$.

1. With the system in state (x^f,x^s) at time t, compute $\overline{a}_j^s(x^s;x^f)$ for $j=1,\ldots,M_s$ according to Eq. (9).

2. Compute $\overline{a}_0^s(x^s;x^f)$ in Eq. (10), and then generate values for τ and j according to Eqs. (12).

3. Advance the system to the next slow reaction by replacing $t\leftarrow t+\tau$, and

$$x_i^s \leftarrow x_i^s + \nu_{ij}^{ss} \quad (i=1,\ldots,N_s), \tag{13a}$$

$$x_i^f \leftarrow x_i^f + \nu_{ij}^{fs} \quad (i=1,\ldots,N_f), \tag{13b}$$

$$x^f \leftarrow \text{sample of } \hat{P}(x^{f\prime},\infty|x^f,x^s). \tag{14}$$

4. Record $X(t)=(x^f,x^s)$ as desired. Then return to Step 1, or else stop.

The most difficult part of the above procedure will be computing $\hat{P}(x^{f\prime},\infty|x^f,x^s)$; indeed, this will usually have to be done *approximately*, since the stationary virtual CME (8) can be solved exactly for only a few very simple systems. There are two reasons why we want this function. First, it enables the computation in Step 1 of the slow-scale propensity functions according to Eq. (9); however, we shall see shortly that this never

requires knowing more than just the *first two moments* of $\hat{P}(x^{\mathrm{f}\prime},\infty|x^{\mathrm{f}},x^{\mathrm{s}})$. The second reason why we want $\hat{P}(x^{\mathrm{f}\prime},\infty|x^{\mathrm{f}},x^{\mathrm{s}})$ is to complete the updating of the fast state variables in (14). But as we shall see later, the operation (14) often can be carried out satisfactorily knowing only the first two moments of $\hat{P}(x^{\mathrm{f}\prime},\infty|x^{\mathrm{f}},x^{\mathrm{s}})$. And since the operation (14) has no effect on the rest of the algorithm, it could be omitted entirely if a read-out of the fast variables is not required.

The computation of the slow-scale propensity functions $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}};x^{\mathrm{f}})$ in Step 1 will depend on the forms of the (true) R_j^{s} propensity functions $a_j^{\mathrm{s}}(x)$. But in terms of the moments of the virtual fast state variables, there are actually only five possibilities:

- If $a_j^{\mathrm{s}}(x)$ is independent of x^{f}, then $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}};x^{\mathrm{f}})=a_j^{\mathrm{s}}(x^{\mathrm{s}})$. (15a)
- If $a_j^{\mathrm{s}}(x)=c_j^{\mathrm{s}}x_i^{\mathrm{f}}$, then $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}};x^{\mathrm{f}})=c_j^{\mathrm{s}}\left\langle\hat{X}_i^{\mathrm{f}}(\infty)\right\rangle$. (15b)
- If $a_j^{\mathrm{s}}(x)=c_j^{\mathrm{s}}x_i^{\mathrm{f}}x_{i'}^{\mathrm{s}}$, then $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}};x^{\mathrm{f}})=c_j^{\mathrm{s}}x_{i'}^{\mathrm{s}}\left\langle\hat{X}_i^{\mathrm{f}}(\infty)\right\rangle$. (15c)
- If $a_j^{\mathrm{s}}(x)=c_j^{\mathrm{s}}\frac{1}{2}x_i^{\mathrm{f}}(x_i^{\mathrm{f}}-1)$, then $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}};x^{\mathrm{f}})=c_j^{\mathrm{s}}\frac{1}{2}\left\langle\hat{X}_i^{\mathrm{f}}(\infty)\left(\hat{X}_i^{\mathrm{f}}(\infty)-1\right)\right\rangle$. (15d)
- If $a_j^{\mathrm{s}}(x)=c_j^{\mathrm{s}}x_i^{\mathrm{f}}x_{i'}^{\mathrm{f}}$ for $i\neq i'$, then $\overline{a}_j^{\mathrm{s}}(x^{\mathrm{s}};x^{\mathrm{f}})=c_j^{\mathrm{s}}\left\langle\hat{X}_i^{\mathrm{f}}(\infty)\hat{X}_{i'}^{\mathrm{f}}(\infty)\right\rangle$. (15e)

Here we have used the averaging notation

$$\left\langle f\left(\hat{X}^{\mathrm{f}}(\infty)\right)\right\rangle\triangleq\sum_{x^{\mathrm{f}\prime}}\hat{P}(x^{\mathrm{f}\prime},\infty|x^{\mathrm{f}},x^{\mathrm{s}})\,f(x^{\mathrm{f}\prime}). \tag{16}$$

The four cases (15b-e) refer, respectively, to the R_j^{s} forms $S_i^{\mathrm{f}}\to\cdots$, $S_i^{\mathrm{f}}+S_{i'}^{\mathrm{s}}\to\cdots$, $S_i^{\mathrm{f}}+S_i^{\mathrm{f}}\to\cdots$, and $S_i^{\mathrm{f}}+S_{i'}^{\mathrm{f}}\to\cdots$. An inspection of the above results shows that in order to compute *any* slow-scale propensity function, we will never need more than the first two moments of $\hat{X}^{\mathrm{f}}(\infty)$.

In Step 3, the state update is carried out separately for the slow state variables and the fast state variables. The slow state variable update formula (13a) simply increases each slow species component x_i^{s} by ν_{ij}^{ss}, reflecting the fact that one R_j^{s} reaction has occurred in $[t,t+\tau]$. Of course, many fast reactions have also occurred in that time interval, but by construction they have no effect on the slow state variables.

The updating of the fast state variables is accomplished in two stages: In (13b) the fast state variables are changed to reflect the occurrence of the one R_j^{s} reaction. Then, in (14), the fast state variables are all "relaxed" to their stationary values. Although (14) overwrites (13b), that overwrite will be influenced by the outcome of (13b). The end result of this two-step procedure is, however, *not* the fast variables *immediately* after the R_j^{s} reaction, but rather the fast variables a "short time" later than that – specifically, a time after the R_j^{s} reaction that is short on the time scale of the slow reactions but long on the time scale of the fast reactions. As will be explained more fully later, this slightly

delayed sampling of the fast variables is necessitated by the fact that times in the immediate neighborhood of a slow reaction are usually not unbiased sampling times for the fast variables.

In the following two sections, we shall illustrate the foregoing theory for two very simple virtual fast processes. For each we shall first calculate the asymptotic moments in Eqs. (15), so that we will be able to compute the slow-scale propensity function for *any* slow reaction that might accompany these fast reactions. These asymptotic moment calculations can be done exactly for our first fast process, but we shall have to resort to approximations for the second. We shall then couple each of these virtual fast processes with one or two simple slow reactions in order to illustrate how the full Slow-Scale SSA gets implemented.

VIII. EXAMPLE 1: THE FAST REVERSIBLE ISOMERIZATION

One of the simplest stable fast processes arises from the reversible isomerization,

$$S_1 \underset{c_2}{\overset{c_1}{\rightleftarrows}} S_2 \,. \tag{17}$$

Assuming these two reactions are the only "fast" reactions, the virtual fast process $\hat{X}^{\mathrm{f}}(t)$ will then be $\big(X_1(t), X_2(t)\big)$ with propensity functions and state-change vectors

$$\left.\begin{aligned} a_1(x) &= c_1 x_1, \quad & \nu_1 &= (-1,+1) \\ a_2(x) &= c_2 x_2, \quad & \nu_2 &= (+1,-1) \end{aligned}\right\}. \tag{18}$$

What distinguishes this "virtual" fast process $\hat{X}^{\mathrm{f}}(t)$ from the "real" fast process $X^{\mathrm{f}}(t)$ is that the virtual process obeys the following *conservation relation*, which expresses the constancy of the total number of isomers:

$$\hat{X}_1(t) + \hat{X}_2(t) = x_{\mathrm{T}} \quad \text{(constant)} \,. \tag{19}$$

This relation greatly simplifies the analysis of the virtual fast process, because it reduces that problem to a single independent state variable. In contrast, for $X^{\mathrm{f}}(t)$ the sum in Eq. (19) will *not* generally be constant, owing to the presence of slow channels that can change the S_1 or S_2 populations independently.

When $\hat{X}_2(t)$ is eliminated in favor of $\hat{X}_1(t)$ by means of Eq. (19), $\hat{X}_1(t)$ takes the form of a bounded "birth-death" Markov process (see Appendix A) with "stepping functions"

$$W_-(x_1') = c_1 x_1', \quad W_+(x_1') = c_2 (x_{\mathrm{T}} - x_1') \,. \tag{20}$$

By analytically iterating the recursion formula (A2) (taking $x^* = 0$), the asymptotic probability distribution of this process can be calculated exactly. In this way, $\hat{X}_1(\infty)$ is found to be the *binomial* random variable $\mathcal{B}(q, x_{\mathrm{T}})$, whose probability function is

$$\hat{P}(x_1', \infty \,|\, x_1, x_2) = \frac{x_{\mathrm{T}}!}{x_1'!\,(x_{\mathrm{T}} - x_1')!} q^{x_1'} (1-q)^{x_{\mathrm{T}} - x_1'}, \quad (x_1' = 0,1,\ldots,x_{\mathrm{T}}) \,, \tag{21}$$

where $$q \equiv \frac{c_2}{c_1 + c_2}, \tag{22}$$

and $$x_{\mathrm{T}} = x_1 + x_2 . \tag{23}$$

Note that $\hat{P}(x_1', \infty | x_1, x_2)$ does indeed depend on the fast state vector (x_1, x_2) at the "initial" time t, through the sum of its two components.

The mean and variance of the binomial random variable $\mathcal{B}(q, x_{\mathrm{T}})$ can be directly evaluated from the definition (A4), and the results are well known to be

$$\left\langle \hat{X}_1(\infty) \right\rangle = x_{\mathrm{T}} q = \frac{c_2 x_{\mathrm{T}}}{c_1 + c_2}, \tag{24a}$$

$$\mathrm{var}\left\{ \hat{X}_1(\infty) \right\} = x_{\mathrm{T}} q (1-q) = \frac{c_1 c_2 x_{\mathrm{T}}}{(c_1 + c_2)^2} . \tag{24b}$$

Using Eqs. (24), along with the relation $\hat{X}_2(\infty) = x_{\mathrm{T}} - \hat{X}_1(\infty)$, it is now a simple matter to deduce the following results for use in Eqs. (15), which allow us to compute all possible slow-scale propensity functions with respect to this virtual fast process:

$$\left\langle \hat{X}_1(\infty) \right\rangle = \frac{c_2 x_{\mathrm{T}}}{c_1 + c_2}, \quad \left\langle \hat{X}_2(\infty) \right\rangle = \frac{c_1 x_{\mathrm{T}}}{c_1 + c_2}, \tag{25a}$$

$$\left\langle \hat{X}_1(\infty) \left(\hat{X}_1(\infty) - 1 \right) \right\rangle = \frac{c_2^2}{(c_1 + c_2)^2} x_{\mathrm{T}} \left(x_{\mathrm{T}} - 1 \right), \tag{25b}$$

$$\left\langle \hat{X}_2(\infty) \left(\hat{X}_2(\infty) - 1 \right) \right\rangle = \frac{c_1^2}{(c_1 + c_2)^2} x_{\mathrm{T}} \left(x_{\mathrm{T}} - 1 \right), \tag{25c}$$

$$\left\langle \hat{X}_1(\infty) \hat{X}_2(\infty) \right\rangle = \frac{c_1 c_2}{(c_1 + c_2)^2} x_{\mathrm{T}} \left(x_{\mathrm{T}} - 1 \right). \tag{25d}$$

Notice that all of these asymptotic moments of the virtual fast process depend on the values (x_1, x_2) of the fast state variables at the initial time t through their sum x_{T}.

The stationary means in Eqs. (25a) are exactly what we would get by solving the stationary (equilibrium) deterministic *reaction rate equation* (RRE) for reactions (17), namely $c_1 X_1 = c_2 X_2$ with $X_1 + X_2 = x_{\mathrm{T}}$. This happens because reactions (17) are linear. Although the stationary RRE cannot give us the exact second-order moments in Eqs. (25b-d), it turns out that the stationary RRE solution affords excellent approximations to those in the common circumstance that $x_{\mathrm{T}} \gg 1$; because in that case, the last factor $(x_{\mathrm{T}} - 1)$ in each of Eqs. (25b-d) can be well approximated by x_{T}, and the stationary RRE solutions (25a) can then be invoked to obtain

$$x_{\mathrm{T}} \gg 1 : \begin{cases} \left\langle \hat{X}_i(\infty) \left(\hat{X}_i(\infty) - 1 \right) \right\rangle \approx \left\langle \hat{X}_i(\infty) \right\rangle^2 \qquad (i = 1, 2), \\ \left\langle \hat{X}_1(\infty) \hat{X}_2(\infty) \right\rangle \approx \left\langle \hat{X}_1(\infty) \right\rangle \left\langle \hat{X}_2(\infty) \right\rangle . \end{cases} \tag{26}$$

So we see that, in the common circumstance $x_T \gg 1$, *any* slow-scale propensity function can be well approximated in terms of the solution of the stationary RRE for reactions (17). Whether the stationary RRE will serve us as well for *nonlinear* fast processes remains to be seen.

There are several different time scales associated with this virtual fast process. Since in state (x_1, x_2) the probability that either R_1 or R_2 will fire in the next infinitesimal time dt is $(c_1 x_1 + c_2 x_2)\,dt$, then

$$\text{mean time to next fast reaction} \;=\; \frac{1}{c_1 x_1 + c_2 x_2}. \tag{27a}$$

The *asymptotic* mean time to the next fast reaction can then be obtained by replacing x_1 and x_2 here with their stationary means in (25a). That gives

$$\text{asymptotic mean time to next fast reaction} \;=\; \frac{c_1 + c_2}{2c_1 c_2 (x_1 + x_2)}. \tag{27b}$$

The other time scale of interest to us is the time scale on which the virtual fast process relaxes to its $t = \infty$ stationary form. It has been shown elsewhere[5] that the mean and variance of $\hat{X}_1(t)$ and $\hat{X}_2(t)$ for reactions (17) exponentially approach their $t = \infty$ values in a time of order

$$\text{relaxation time} \;\approx\; \frac{1}{c_1 + c_2}. \tag{28}$$

Indeed, since in this linear case the means satisfy the deterministic RRE, we have

$$\frac{d\left\langle \hat{X}_1(t) \right\rangle}{dt} = -c_1 \left\langle \hat{X}_1(t) \right\rangle + c_2 \left(x_T - \left\langle \hat{X}_1(t) \right\rangle \right) = -(c_1 + c_2) \left\langle \hat{X}_1(t) \right\rangle + c_2 x_T,$$

for which the solution is easily shown to be

$$\left\langle \hat{X}_1(t) \right\rangle = \left\langle \hat{X}_1(\infty) \right\rangle + \left(x_{01} - \left\langle \hat{X}_1(\infty) \right\rangle \right) \mathrm{e}^{-(c_1 + c_2)t},$$

in agreement with Eq. (28).

As is clear from the statement of the Slow-Scale Approximation in Sec. VI, the key requirement for applying the Slow-Scale SSA is that the relaxation time of the virtual fast process must be much smaller than the average time to the next slow reaction. Estimating the latter time as the reciprocal of the sum of the *slow-scale* propensity functions, *the condition for using the Slow-Scale SSA in this case is thus*

$$c_1 + c_2 \gg \sum_{j=1}^{M_s} \overline{a}_j^{\,s}(x). \tag{29}$$

To illustrate the application of the foregoing results, let us suppose that the fast reactions (17) are occurring in conjunction with the single slow reaction,

$$S_2 \xrightarrow{c_3} S_3, \tag{30}$$

for which

$$a_3(x) = c_3 x_2, \quad \nu_3 = (0, -1, +1). \tag{31}$$

Thus, the fast reactions are R_1 and R_2, the slow reaction is R_3, the fast species are S_1 and S_2, and the slow species is S_3. (Note that the only reactant in the slow reaction is a fast species.) By Eqs. (15b) and (25a), the slow-scale propensity function for this reaction is

$$\overline{a}_3(x_3;x_1,x_2) = c_3\left\langle \hat{X}_2(\infty)\right\rangle = \frac{c_3c_1(x_1+x_2)}{c_1+c_2}. \tag{32}$$

According to condition (29), we should be able to invoke this slow-scale propensity function whenever

$$c_1+c_2 \gg \frac{c_3c_1(x_1+x_2)}{c_1+c_2}. \tag{33}$$

Assuming this condition holds, the Slow-Scale SSA for reactions (17) and (30) then goes as follows:

Initialize: Given $X(t_0)=(x_{10},x_{20},x_{30})$, set $t\leftarrow t_0$ and $x_i\leftarrow x_{i0}$ $(i=1,2,3)$.

Step 1. In state (x_1,x_2,x_3) at time t, compute

$$\overline{a}_3(x_3;x_1,x_2) = \frac{c_3c_1(x_1+x_2)}{c_1+c_2}.$$

Step 2. Draw a unit-interval uniform random number r, and compute

$$\tau = \frac{1}{\overline{a}_3(x_3;x_1,x_2)}\ln\left(\frac{1}{r}\right).$$

Step 3. Advance to the next R_3 reaction by replacing $t\leftarrow t+\tau$ and

$$\begin{aligned}
&x_3\leftarrow x_3+1,\\
&x_2\leftarrow x_2-1,\\
&\begin{cases}
\text{With } x_{\mathrm{T}}=x_1+x_2,\\
x_1\leftarrow \text{ sample of } \mathcal{B}\left(\frac{c_2}{c_1+c_2},x_{\mathrm{T}}\right),\\
x_2\leftarrow x_{\mathrm{T}}-x_1.
\end{cases}
\end{aligned}$$

Step 4. Record (t,x_1,x_2,x_3) if desired. Then return to Step 1, or else stop.

In Step 3, the x_3 update implements (13a), the first x_2 update implements (13b), and the bracketed procedure implements (14). Note that the first x_2 update affects the subsequent x_1 and x_2 updates in brackets through the variable x_{T}. An examination of the full algorithm will reveal that the entire bracketed procedure in Step 3 can be omitted if a readout of the two fast species populations is not required. Making that omission will not affect the simulation of the slow species population, which in many cases will be the only one of practical interest.

Figure 1 shows the results of two simulation runs of reactions (17) and (30) for the following set of parameter values:

$$c_1 = 1,\ c_2 = 2,\ c_3 = 5\times10^{-5};\quad x_{10} = 1200,\ x_{20} = 600,\ x_{30} = 0\,. \tag{34}$$

Figure 1a was obtained using the exact SSA, and Fig. 1b was obtained using the approximate Slow-Scale SSA. Both figures plot the populations of all three species at the time of each R_3 event. But successive dots in Fig. 1a are separated by an average of about 76,000 simulated reactions, whereas the dots in Fig. 1b show all of the simulated reactions. On the scale of these figures, the two plots appear to be statistically indistinguishable. The accuracy of the Slow-Scale SSA run in Fig. 1b should hinge on how well condition (33) is satisfied. For the parameter values in (34), we find that the left side of (33) is initially two orders of magnitude larger than the right side, and that imbalance gradually improves as the simulation progresses (since $x_1 + x_2$ decreases by 1 with each R_3 reaction). The gain in computational efficiency of the Slow-Scale SSA over the exact SSA in this case is striking: The exact SSA run of Fig. 1a simulated over 40 million reactions, whereas the Slow-Scale SSA run in Fig. 1b simulated only 521 reactions (all R_3). The SSA simulation took over 20 minutes to execute, while the Slow-Scale SSA took only a fraction of a second.

Figures 2 and 3 show two more simulation runs of reactions (17) and (30), but now for the parameter values

$$c_1 = 10,\ c_2 = 4\times10^{4},\ c_3 = 2;\quad x_{10} = 2000,\ x_{20} = x_{30} = 0\,. \tag{35}$$

These parameter values have the interesting consequence that the average population of the fast species S_2, as computed from Eq. (25a), is initially 0.5, and it decreases as the simulation progresses (since $x_{\mathrm{T}} = x_1 + x_2$ decreases). The physical implication of this can be seen in the exact SSA run of Fig. 2a, where the populations of all three species are plotted out immediately after each R_3 reaction. Most of the time the S_2 population is 0, sometimes it is 1, occasionally it is 2, and only rarely is it anything more. An S_2 molecule here has a very short lifetime (on average $1/(c_2 + c_3) \approx 0.5$), usually turning into an S_1 molecule but occasionally (on average a fraction $c_3/(c_2 + c_3) \approx 10^{-5}$ of the time) turning into an S_3 molecule.

In Fig. 2b we show a *re-plotting* of the X_2 trajectory for this SSA run, with the S_2 population now being sampled *immediately before* each R_3 reaction; it is of course the trajectory in Fig. 2a increased by exactly 1. In Fig. 2c we show yet another re-plotting of the X_2 trajectory, with the samplings now taken at *equally spaced time intervals*. The differences in the three X_2 trajectories in Fig. 2, which again are all taken from the same SSA run, illustrate an important point: *The occurrence times of the slow reactions will not be statistically independent of the fast species populations.* In this case, an S_2 population of n will be n times more likely to experience an R_3 reaction than an S_2 population of 1. (This effect is also present in the X_1 trajectory, but it is not noticeable because that population is so large.) Although all three S_2 population plots in Fig. 2 are "correct", the equal-time plot in Fig. 2c would seem to be the most "typical."

For the parameter values (35), condition (33) is satisfied by 4 orders of magnitude initially, and even more as the simulation progresses; therefore, the Slow-Scale SSA should be applicable. Figure 3 shows the results of a Slow-Scale SSA simulation. We observe that the trajectories for the S_1 and S_3 populations match those in the exact SSA run of Fig. 2 extremely well. But whereas the SSA run had to simulate over 23 million reactions, the Slow-Scale SSA run simulated only 587, with commensurate differences in the run times. The S_2 population trajectory in Fig. 3 evidently matches the equal-time SSA trajectory in Fig. 2c very well, and much better than either of the trajectories in Figs. 2a or 2b. This agreement with the "typical" S_2 trajectory in Fig. 2c justifies our two-stage updating scheme (13b) and (14) for the fast variables in the general Slow-Scale SSA, which "relaxes" those fast variables after each slow reaction before sampling them. There does not seem to be a feasible way for the Slow-Scale SSA to accurately reproduce the fast state variables *immediately* after a slow reaction, but that should never pose a practical problem since those values are atypical anyway.

IX. EXAMPLE 2: THE FAST REVERSIBLE DIMERIZATION

Suppose the fast reactions are the reversible dimerization,

$$S_1 + S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2 \,. \tag{36}$$

Assuming these two reactions are the only fast reactions, the virtual fast process $\hat{X}^{\mathrm{f}}(t)$ will be $\left(\hat{X}_1(t), \hat{X}_2(t)\right)$, with propensity functions and state-change vectors

$$\left.\begin{aligned} a_1(x) &= c_1 \tfrac{1}{2} x_1(x_1 - 1), \quad & \nu_1 &= (-2, +1) \\ a_2(x) &= c_2 x_2, & \nu_2 &= (+2, -1) \end{aligned}\right\}. \tag{37}$$

This virtual fast process obeys the conservation relation

$$\hat{X}_1(t) + 2\hat{X}_2(t) = x_{\mathrm{T}} \quad \text{(a constant)}\,, \tag{38}$$

which simply asserts the constancy of the total number of monomeric (S_1) units.

When $\hat{X}_1(t)$ is eliminated in favor of $\hat{X}_2(t)$ by means of Eq. (38), $\hat{X}_2(t)$ takes the form of a birth-death Markov process (see Appendix A) with stepping functions

$$W_-(x_2) = c_2 x_2, \quad W_+(x_2) = c_1 \tfrac{1}{2}(x_{\mathrm{T}} - 2x_2)(x_{\mathrm{T}} - 2x_2 - 1)\,. \tag{39}$$

It follows from Eq. (38) that this birth-death process is bounded above by

$$x_{2\mathrm{Max}} = \left[x_{\mathrm{T}}/2\right], \tag{40}$$

where $[\cdots]$ denotes "the greatest integer in".

When Eqs. (39) and (40) are substituted into Eq. (A2), the resulting recursion relation for the probability density function of $\hat{X}_2(\infty)$ unfortunately does not yield a tractable analytic expression, as it did for the example in the preceding section. Nor do Eqs. (A4) yield tractable formulas for the mean and variance of $\hat{X}_2(\infty)$. But a

knowledge of $\left\langle \hat{X}_2(\infty) \right\rangle$ and $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$ is all that is needed in order to evaluate any slow-scale propensity functions that might accompany the fast reactions (36). For, since Eq. (38) implies that $\hat{X}_1(\infty) = x_\mathrm{T} - 2\hat{X}_2(\infty)$, then knowing $\left\langle \hat{X}_2(\infty) \right\rangle$ and $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$ we can compute in succession the asymptotic moments

$$\left\langle \hat{X}_1(\infty) \right\rangle = x_\mathrm{T} - 2\left\langle \hat{X}_2(\infty) \right\rangle, \tag{41a}$$

$$\mathrm{var}\left\{ \hat{X}_1(\infty) \right\} = 4\,\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}, \tag{41b}$$

$$\left\langle \hat{X}_i^2(\infty) \right\rangle = \mathrm{var}\left\{ \hat{X}_i(\infty) \right\} + \left\langle \hat{X}_i(\infty) \right\rangle^2 \quad (i = 1,2), \tag{41c}$$

$$\left\langle \hat{X}_i(\infty)\left(\hat{X}_i(\infty) - 1 \right) \right\rangle = \left\langle \hat{X}_i^2(\infty) \right\rangle - \left\langle \hat{X}_i(\infty) \right\rangle \quad (i = 1,2), \tag{41d}$$

$$\left\langle \hat{X}_1(\infty)\hat{X}_2(\infty) \right\rangle = x_\mathrm{T} \left\langle \hat{X}_2(\infty) \right\rangle - 2\left\langle \hat{X}_2^2(\infty) \right\rangle. \tag{41e}$$

On account of Eqs. (15), these are all we need to evaluate any slow-scale propensity function. But how can we obtain estimates of $\left\langle \hat{X}_2(\infty) \right\rangle$ and $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$? It turns out that there are several approximate ways of doing that, as we shall now elaborate.

The simplest way would be to assume that $\hat{X}_2(t)$ is a *deterministic* process governed by the reaction rate equation (RRE). In that case, $\left\langle \hat{X}_2(\infty) \right\rangle$ would coincide with the stationary or equilibrium solution of the RRE, and $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$ would be zero. The stationary RRE for reactions (36) reads $\frac{1}{2}c_1 X_1^2 = c_2 X_2$, where we have invoked the usual RRE assumption that the population numbers involved are $\gg 1$. With the conservation relation (38), the stationary RRE becomes

$$\tfrac{1}{2}c_1 \left(x_\mathrm{T} - 2X_2 \right)^2 = c_2 X_2. \tag{42}$$

The only root of this quadratic equation satisfying $X_2 \le x_\mathrm{T}/2$, as required by Eq. (38), is

$$X_2^{RRE} = \frac{1}{4}\left\{ \left(2x_\mathrm{T} + \frac{c_2}{c_1} \right) - \sqrt{\left(2x_\mathrm{T} + \frac{c_2}{c_1} \right)^2 - 4x_\mathrm{T}^2} \right\}. \tag{43}$$

We might therefore try approximating $\left\langle \hat{X}_2(\infty) \right\rangle \approx X_2^{\mathrm{RRE}}$. But the RRE also implies that $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\} = 0$, and that might be too inaccurate in some circumstances.

A different way of estimating $\left\langle \hat{X}_2(\infty) \right\rangle$ and $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$ would be to make use of the birth-death process formulas (A5)-(A8). As discussed in Appendix A, any *relative maximum* of the probability density function of $\hat{X}_2(\infty)$ can be computed as the greatest integer in a *down-going root* of the function

$$\begin{aligned} \alpha(x_2) &\triangleq W_+(x_2 - 1) - W_-(x_2), \\ &= 2c_1 x_2^2 - \left(c_1(2x_\mathrm{T} + 3) + c_2 \right) x_2 + \tfrac{1}{2}c_1(x_\mathrm{T} + 2)(x_\mathrm{T} + 1). \end{aligned} \tag{44}$$

Since α here is a concave-up parabola, it can have at most one down-going root, which will necessarily be the smaller one. Using the quadratic formula, we find that root to be

$$x_2^{\mathrm{dgr}} = \frac{1}{4}\left\{\left(2x_{\mathrm{T}} + 3 + \frac{c_2}{c_1}\right) - \sqrt{\left(2x_{\mathrm{T}} + 3 + \frac{c_2}{c_1}\right)^2 - 4(x_{\mathrm{T}} + 2)(x_{\mathrm{T}} + 1)}\right\}. \tag{45a}$$

Therefore, the probability density function of $\hat{X}_2(\infty)$ has a single relative maximum at

$$\hat{x}_2 = \left[x_2^{\mathrm{dgr}}\right]. \tag{45b}$$

which, by definition, is the *stable state* of the birth-death process $\hat{X}_2(t)$. A related useful result is that the *Gaussian variance* of any stable state, which is defined as the variance of the "best Gaussian fit" to the corresponding peak in the probability density function of $\hat{X}_2(\infty)$, is given by formula (A7). Using Eqs. (39) and (44), that formula is

$$\sigma_{\mathrm{G}}^2(\hat{x}_2) = \frac{c_2\hat{x}_2}{-4c_1\hat{x}_2 + c_1(2x_{\mathrm{T}} + 3) + c_2}. \tag{46}$$

Since there is *only one* stable state, we could reasonably approximate

$$\left\langle \hat{X}_2(\infty)\right\rangle \approx x_2^{\mathrm{dgr}}, \tag{47a}$$

$$\mathrm{var}\left\{X_2(\infty)\right\} \approx \sigma_{\mathrm{G}}^2(\hat{x}_2). \tag{47b}$$

The error in Eq. (47a) arises from the fact that it identifies the mean of $\hat{X}_2(\infty)$ with the most likely value of $\hat{X}_2(\infty)$; however, in many cases that error will be quite small. A comparison of Eqs. (45a) and (43) shows that those two estimates of $\left\langle \hat{X}_2(\infty)\right\rangle$ should be very close to each other in the common case that $x_{\mathrm{T}} \gg 1$. But surely, the variance estimate (47b) should be better than the estimate of zero that is predicted by the RRE.

A bonus of this "alpha-function approach" is that it gives us, through Eq. (A8), the following estimate of the time required for $\hat{X}_2(t)$ to relax to $\hat{X}_2(\infty)$:

$$\hat{t} = \frac{1}{-4c_1\hat{x}_2 + c_1(2x_{\mathrm{T}} + 3) + c_2}. \tag{48}$$

Any application of the Slow-Scale SSA will require that this time be small on the time scales of all slow reactions that accompany the fast reactions (36).

A third way to estimate $\left\langle \hat{X}_2(\infty)\right\rangle$ and $\mathrm{var}\left\{\hat{X}_2(\infty)\right\}$ is to make use of the *stationary moment equations* (B4) and (B5), which are derived in Appendix B. Equation (B4) for $i = 2$ reads

$$\begin{aligned} 0 &= \nu_{21}\left\langle a_1\right\rangle + \nu_{22}\left\langle a_2\right\rangle, \\ &= (+1)\left\langle c_1 \tfrac{1}{2}\left(\hat{X}_1(\infty)\left(\hat{X}_1(\infty) - 1\right)\right)\right\rangle + (-1)\left\langle c_2\hat{X}_2(\infty)\right\rangle. \end{aligned}$$

Substituting $\hat{X}_1(\infty) = x_{\mathrm{T}} - 2\hat{X}_2(\infty)$ from the conservation relation (38) and then simplifying, we get

$$0 = 2\left\langle \hat{X}_2^2(\infty)\right\rangle - \left(2x_\mathrm{T} - 1 + \frac{c_2}{c_1}\right)\left\langle \hat{X}_2(\infty)\right\rangle + \tfrac{1}{2} x_\mathrm{T}(x_\mathrm{T} - 1)\,. \tag{49a}$$

And Eq. (B5) for $i = i' = 2$ reads

$$\begin{aligned} 0 &= 2\left[\nu_{21}\left\langle \hat{X}_2(\infty) a_1\right\rangle + \nu_{22}\left\langle \hat{X}_2(\infty) a_2\right\rangle\right] + \nu_{21}^2\left\langle a_1\right\rangle + \nu_{22}^2\left\langle a_2\right\rangle, \\ &= 2\left[(+1)\left\langle \hat{X}_2(\infty)\, c_1 \tfrac{1}{2}\, \hat{X}_1(\infty)\left(\hat{X}_1(\infty) - 1\right)\right\rangle + (-1)\left\langle \hat{X}_2(\infty)\, c_2 \hat{X}_2(\infty)\right\rangle\right] \\ &\quad + (+1)^2\left\langle c_1 \tfrac{1}{2}\, \hat{X}_1(\infty)\left(\hat{X}_1(\infty) - 1\right)\right\rangle + (-1)^2\left\langle c_2 \hat{X}_2(\infty)\right\rangle. \end{aligned}$$

Again substituting $\hat{X}_1(\infty) = x_\mathrm{T} - 2\hat{X}_2(\infty)$ and simplifying, we get

$$\begin{aligned} 0 = 4\left\langle \hat{X}_2^3(\infty)\right\rangle - 2\left(2x_\mathrm{T} - 2 + \frac{c_2}{c_1}\right)\left\langle \hat{X}_2^2(\infty)\right\rangle \\ + \left(x_\mathrm{T}^2 - 3x_\mathrm{T} + 1 + \frac{c_2}{c_1}\right)\left\langle \hat{X}_2(\infty)\right\rangle + \frac{1}{2} x_\mathrm{T}\left(x_\mathrm{T} - 1\right). \end{aligned} \tag{49b}$$

Although Eqs. (49) are exact, they evidently constitute two equations in *three* unknowns, $\left\langle \hat{X}_2(\infty)\right\rangle$, $\left\langle \hat{X}_2^2(\infty)\right\rangle$, and $\left\langle \hat{X}_2^3(\infty)\right\rangle$. We could continue to develop higher order moment equations, but there would always be one more unknown than equation. To break this open ended chain, we will have to make some sort of *approximation*. For example, we could treat $\hat{X}_2(\infty)$ as a *normal* random variable with some mean m_2 and some variance σ_2^2: $\hat{X}_2(\infty) = \mathcal{N}(m_2, \sigma_2^2)$. Under that assumption, the first three moments of $\hat{X}_2(\infty)$ will be given in terms of the mean and variance by Eqs. (B9). When those formulas are substituted into Eqs. (49), we get the following two equations in the *two* unknowns, m_2 and σ_2^2:

$$0 = 2\left(\sigma_2^2 + m_2^2\right) - \left(2x_\mathrm{T} - 1 + \frac{c_2}{c_1}\right) m_2 + \tfrac{1}{2} x_\mathrm{T}(x_\mathrm{T} - 1)\,, \tag{50a}$$

$$\begin{aligned} 0 = 4m_2\left(3\sigma_2^2 + m_2^2\right) - 2\left(2x_\mathrm{T} - 2 + \frac{c_2}{c_1}\right)\left(\sigma_2^2 + m_2^2\right) \\ + \left(x_\mathrm{T}^2 - 3x_\mathrm{T} + 1 + \frac{c_2}{c_1}\right) m_2 + \frac{1}{2} x_\mathrm{T}\left(x_\mathrm{T} - 1\right). \end{aligned} \tag{50b}$$

Since Eqs. (50) are *nonlinear* in the unknowns, a *numerical* solution method is indicated; however, once the solutions are in hand, we simply take

$$\left\langle \hat{X}_2(\infty)\right\rangle \approx m_2\,, \tag{51a}$$

$$\mathrm{var}\left\{ \hat{X}_2(\infty)\right\} \approx \sigma_2^2\,. \tag{51b}$$

We may expect that the approximations (51) will be reasonably close to the alpha-function approximations (47), since both involve normal approximations.

Finally, we can *combine* features of the above approaches to obtain a couple of "hybrid" estimation methods: If we simply assume that the RRE estimate of the mean in Eq. (43) is sufficiently accurate, we could use it in place of $\hat{x}_2$ in Eq. (46), or in place of m_2 in Eq. (49a), to deduce the variance. Thus, combining Eqs. (43) and (46), we have the estimates

$$\left\langle \hat{X}_2(\infty) \right\rangle \approx X_2^{\mathrm{RRE}}, \tag{52a}$$

$$\mathrm{var}\left\{ \hat{X}_2(\infty) \right\} \approx \frac{c_2 X_2^{\mathrm{RRE}}}{-4c_1 X_2^{\mathrm{RRE}} + c_1\left(2x_{\mathrm{T}} + 3\right) + c_2}. \tag{52b}$$

And combining Eqs. (43) and (50a), we have the estimates

$$\left\langle \hat{X}_2(\infty) \right\rangle \approx X_2^{\mathrm{RRE}}, \tag{53a}$$

$$\mathrm{var}\left\{ \hat{X}_2(\infty) \right\} \approx -\left(X_2^{\mathrm{RRE}} \right)^2 + \frac{1}{2}\left(2x_{\mathrm{T}} - 1 + \frac{c_2}{c_1} \right) X_2^{\mathrm{RRE}} - \tfrac{1}{4} x_{\mathrm{T}}(x_{\mathrm{T}} - 1). \tag{53b}$$

We thus have *four different ways* of approximately estimating $\left\langle \hat{X}_2(\infty) \right\rangle$ and $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$. To test these, we have made some numerical calculations for the case

$$c_1 = 1, \quad c_2 = 200, \quad x_{\mathrm{T}} = 2000. \tag{54}$$

Figure 4 shows, as the curve with open circles, the *exact* probability density function $\hat{P}(x_2, \infty | x_{\mathrm{T}})$ of $\hat{X}_2(\infty)$, as computed numerically from the recursion relation (A2). This function is really defined on the entire interval $0 \le x_2 \le 1000$, but its value is exceedingly small outside the x_2 interval shown. A parallel exact evaluation of the first two moments of $\hat{X}_2(\infty)$ according to formula (A4) yielded the mean and variance values shown in the first row of Table 1. In Fig. 4, a *normal* density function with that mean and variance is shown as the solid curve, and it evidently fits the exact curve quite well; the fit is evidently much better than that provided by the *binomial* curve, shown dotted, which has the same mean and upper limit 1000. This suggests that the normal approximations that were made in the several estimation strategies discussed above should be reasonable.

The second row in Table 1 shows the estimates of the mean and variance of $\hat{X}_2(\infty)$ given by the alpha-function formulas (47), along with the estimate of the relaxation time $\hat{t}$ provided by formula (48). The third row in Table 1 shows the estimates of $\left\langle \hat{X}_2(\infty) \right\rangle$ and $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$ given by the normally approximated stationary moment equations (50)-(51). The fourth row in Table 1 shows the estimates given by the hybrid formulas (52), and the last row shows the estimates given by the hybrid formulas (53).

	$\left\langle \hat{X}_2(\infty) \right\rangle$	$\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$	$\hat{t}$
exact values	**729.811**	**113.996**	
Eqs. (47), (48)	730.477	113.796	7.8×10^{-5}
Eqs. (50)-(51)	729.811	114.055	
Eqs. (52)	729.844	113.716	
Eqs. (53)	729.844	135.078	

Table 1. For the fast reactions (36) with parameter values (54).

The results in Table 1 imply that, at least for these parameter values, the first three approximation methods, namely the alpha function method, the normally approximated stationary moment equation method, and the hybrid alpha function / RRE method, all give excellent approximations to $\left\langle \hat{X}_2(\infty) \right\rangle$ and $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$. The slight but tolerable error in the alpha function method mean (in the second row of figures in Table 1) is undoubtedly due to that estimate being the most likely value of $\hat{X}_2(\infty)$ instead of the mean. The surprisingly large error in the estimate of $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$ in the second hybrid method (in last row in Table 1) suggests that we should not use that method. An investigation of the source of this error revealed that the normally approximated second stationary moment equation (53b) is extremely sensitive to the value of X_2^{RRE}; thus, comparing the figures in rows 3 and 5 of Table 1, we see that changing the value of X_2^{RRE} in Eq. (53b) from 729.811 to 729.844, a change of only 0.005%, induces an 18% change in the value of $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$, from 114.005 to 135.078. In contrast, the alpha function formula (52b), is not nearly so sensitive to the value of X_2^{RRE}: For formula (52b), a 0.9% change in the value of X_2^{RRE} induces only a 0.07% change in the computed value of $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$.

In view of the comparable accuracy of the first three methods, computational simplicity would dictate using the third one, namely Eqs. (52), to compute $\left\langle \hat{X}_2(\infty) \right\rangle$ and $\mathrm{var}\left\{ \hat{X}_2(\infty) \right\}$ in a slow-scale simulation. Once those values are substituted into Eqs. (41), we will have all we need to evaluate *any* slow-scale propensity function associated with the fast reactions (36).

Now let us suppose that the two fast reactions (36) are occurring in conjunction with the two *slow* reactions

$$S_1 \xrightarrow{c_3} \varnothing, \quad S_2 \xrightarrow{c_4} S_3, \tag{55}$$

for which

$$\left.\begin{aligned} a_3(x) &= c_3 x_1, \quad \nu_3 = (-1,0,0) \\ a_4(x) &= c_4 x_2, \quad \nu_4 = (0,-1,1) \end{aligned}\right\}. \tag{56}$$

In words, a molecule of the monomer species S_1 can spontaneously decay, and a molecule of the unstable dimer species S_2 can spontaneously convert to a stable form S_3. The four reactions (36) and (55) have been considered in earlier works as the "dimer-decay" model,[5,6] although the reaction channels were indexed differently. In the present context, the fast reactions are R_1 and R_2, the slow reactions are R_3 and R_4, the fast species are S_1 and S_2, and the slow species is S_3. Note that all the reactants for the slow reactions happen to be fast species.

From Eqs. (15b), (52a), and (41a), we compute the slow-scale propensity functions for reactions (55) as

$$\overline{a}_3(x_3;x_1,x_2) = c_3\left\langle \hat{X}_1(\infty)\right\rangle \approx c_3\left(x_\mathrm{T} - 2X_2^{RRE}\right), \tag{57a}$$

$$\overline{a}_4(x_3;x_1,x_2) = c_4\left\langle \hat{X}_2(\infty)\right\rangle \approx c_4 X_2^{RRE}. \tag{57b}$$

Here, $x_\mathrm{T} = x_1 + 2x_2$, and X_2^{RRE} is given by Eq. (43). We are assuming here that the stationary solution of the RRE for the virtual fast process provides an acceptable approximation to the *means* of the fast variables, an assumption that the results in Table 1 support.

The key requirement for being able to apply the Slow-Scale SSA is that the relaxation time for the virtual fast process should be much smaller than the average time to the next slow reaction. We can estimate the relaxation time $\hat{t}$ of the virtual fast process by Eq. (48), replacing $\hat{x}_2$ therein with X_2^{RRE}. The mean time to the next slow-scale reaction can be estimated as the reciprocal of the sum of the propensity functions of the slow reactions, $\left(c_3x_1 + c_4x_2\right)^{-1}$, with x_1 and x_2 replaced by their respective relaxed values $x_\mathrm{T} - 2X_2^{RRE}$ and X_2^{RRE}. The condition that this latter time be much larger than $\hat{t}$ then becomes

$$-4c_1X_2^{RRE} + c_1\left(2x_\mathrm{T} + 3\right) + c_2 \gg c_3\left(x_\mathrm{T} - 2X_2^{RRE}\right) + c_4X_2^{RRE}. \tag{58}$$

Assuming condition (58) holds, the Slow-Scale SSA for reactions (36) and (55) is as follows:

Initialize: Given $X(t_0) = (x_{10}, x_{20}, x_{30})$, set $t \leftarrow t_0$ and $x_i \leftarrow x_{i0}$ $(i = 1,2,3)$.

Compute $x_\mathrm{T} = x_1 + 2x_2$, and then compute X_2^{RRE} from Eq. (43).

Step 1. In state (x_1, x_2, x_3) at time t, compute

$$\overline{a}_3(x_3;x_1,x_2) = c_3\left(x_\mathrm{T} - 2X_2^{RRE}\right), \quad \overline{a}_4(x_3;x_1,x_2) = c_4X_2^{RRE}.$$

Step 2. Compute $\overline{a}_0(x_3;x_1,x_2) = \overline{a}_3(x_3;x_1,x_2) + \overline{a}_4(x_3;x_1,x_2)$. Then, with r_1 and r_2 independent unit-interval uniform random numbers, compute

$$\tau = \frac{1}{\overline{a}_0(x_3;x_1,x_2)}\ln\left(\frac{1}{r_1}\right),$$

$$j = \begin{cases} 3, & \text{if } \overline{a}_3(x_3; x_1, x_2) \ge r_2 \overline{a}_0(x_3; x_1, x_2), \\ 4, & \text{otherwise.} \end{cases}$$

Step 3. Advance to the next slow reaction by replacing $t \leftarrow t + \tau$ and

$$x_i \leftarrow x_i + \nu_{ij} \quad (i = 1, 2, 3),$$

$$x_{\mathrm{T}} \leftarrow x_1 + 2x_2,$$

$$X_2^{\mathrm{RRE}} \leftarrow \text{ Eq. (43)},$$

$$\begin{cases} \text{compute } \operatorname{var}\left\{\hat{X}_2(\infty)\right\} \text{ from Eq. (52b)}, \\ x_2 \leftarrow \text{ round}\left\{\text{sample } \mathcal{N}\left(X_2^{\mathrm{RRE}}, \operatorname{var}\left\{\hat{X}_2(\infty)\right\}\right)\right\} \text{ in } [0, [x_{\mathrm{T}}/2]], \\ x_1 \leftarrow x_{\mathrm{T}} - 2x_2. \end{cases}$$

Step 4. Record (t, x_1, x_2, x_3) if desired. Then return to Step 1, or else stop.

Some clarifying comments are in order regarding the procedures in Step 3: The x_i update implements (13a) and (13b). As a consequence of those updates for $i = 1$ and 2 (the fast species), the recalculation of x_{T} results in x_{T} getting reduced by 1 if $j = 3$, or 2 if $j = 4$, and this necessitates the re-evaluation of X_2^{RRE}. The bracketed procedure implements (16), under the assumption that $\hat{X}_2(\infty)$ can be decently approximated as a *normal* random variable with mean (52a) and variance (52b) – an assumption that is supported by the results in Fig. 4. But to keep x_1 and x_2 nonnegative integers, we round the normal sample value for x_2 to the nearest integer and then force it to be not less than 0 and not greater than the greatest integer in $x_{\mathrm{T}}/2$.

Notice that the evolution of x_3 depends on x_1 and x_2 *only* through the quantity x_{T}, a quantity that does not get changed by the bracketed procedure. Therefore, the entire bracketed procedure can be omitted if a readout of the two fast species populations is not required. In that case, the coding can be slightly simplified by deleting all references to the variables x_1 and x_2, and treating x_{T} as a state variable that gets reduced by 1 whenever R_3 fires or 2 whenever R_4 fires.

To test the foregoing simulation algorithm, we choose the reaction constant values

$$c_1 = 1, \quad c_2 = 200, \quad c_3 = 0.02, \quad c_4 = 0.004, \tag{59a}$$

along with the initial values

$$x_{10} = 540, \quad x_{20} = 730, \quad x_{30} = 0. \tag{59b}$$

These initial values for the S_1 and S_2 population give $x_{\mathrm{T}} = 2000$, and they approximately satisfy the RRE for the virtual fast process; thus, we initially have $X_2^{\mathrm{RRE}} \approx 730$. The LHS of (58) then evaluates to ≈ 1283, and the RHS of (58) evaluates to ≈ 13. Condition (58) is therefore reasonably well satisfied, at least initially, so the Slow-Scale SSA should be applicable.

We first show, in Fig. 5a, the results of an exact SSA run for this system. The populations of the three species are plotted out immediately after each occurrence of a slow reaction (R_3 or R_4). During the time interval shown in the figure, there were 2.483×10^7 reactions in all, and 1,742 of those were slow reactions; thus, successive slow reactions are separated by, on average, 1.4×10^4 fast reactions.

Figure 5b shows the results of a simulation made using the Slow-Scale SSA as detailed above. The populations here are plotted out after *every simulated* reaction. The trajectories in Figs. 5a and 5b are, for all practical purposes, statistically indistinguishable. But whereas the exact SSA simulation in Fig. 5a took 17 minutes to simulate 2.483×10^7 reactions, the Slow-Scale SSA in Fig. 5b took a fraction of a second to simulate 1,741 reactions.

X. SUMMARY AND CONCLUSIONS

Our focus in this paper has been exclusively on chemical systems that exhibit a wide range of dynamical modes, the fastest of which is stable. The operational meanings of these terms are spelled out in Secs. II – V. We provisionally identify the fast reactions as those whose propensity functions have much larger values, at least most of the time, than the propensity functions of all the other slow reactions. We next identify the fast and slow species by declaring a fast species to be any whose population gets changed by at least one fast reaction; all the other species are called slow. Several subtleties attend these definitions of fast and slow reactions and species: A slow species cannot get changed by a fast reaction but a fast species can get changed by a slow reaction; the propensity functions of *both* fast and slow reactions can depend on *both* fast and slow species; and the population of a fast species need *not* be "large".

Our next step is to define the *virtual fast system* (or virtual fast process) to be the imaginary system composed of all the fast species and *only* the fast reactions; i.e., the *virtual* fast system is the *real* fast system with all the slow reactions switched off. Unlike the real fast process, the virtual fast process is Markovian, and hence potentially tractable. But for the systems of interest to us, this virtual fast process must satisfy two critical conditions: First, it must be *stable*; i.e., its $t \to \infty$ probability distribution must exist and be independent of t. And second, in the current state, the relaxation time of the virtual fast process must be very much less than the expected time to the next slow reaction. If satisfying these conditions requires modifying our initial provisional roster of fast and slow reactions, then we do that (regardless of propensity function values). But we expect that these two conditions can be satisfied by any system whose deterministic reaction rate equations exhibit pronounced *stiffness*, and we know from experience that many important real world systems fall into that category.

In the context of the foregoing definitions and conditions, the key result of our work here is the Slow-Scale Approximation in Sec. VI. It asserts that for each slow reaction we can define a "slow-scale propensity function" (9) – as the *average* of the regular propensity function with respect to the *asymptotic virtual fast process* – which *approximately replaces* the regular propensity function on the time scale of the slow reactions. In Sec. VII we showed how these slow-scale propensity functions can be used to simulate the evolution of the system *one slow reaction at a time*, a simulation procedure that we dubbed the *Slow-Scale SSA*. This algorithm offers the potential for

substantial gains in computational speed over the exact SSA whenever the time scales of the fast and slow reactions are widely separated.

In Secs. VIII and IX we illustrated the use of the Slow-Scale SSA on two simple systems. The virtual fast processes for these two systems were, respectively, the reversible isomerization reactions and the reversible dimerization reactions. The asymptotic properties of the reversible isomerization were calculated exactly, but the asymptotic properties of the reversible dimerization could only be calculated approximately. For the latter effort, we proposed four different approximation procedures, three of which were found to be very accurate for the numerical example being considered. Coupling these two fast reactions with some simple slow reactions, we showed that for system parameter values that satisfy the hypothesis of the Slow-Scale Approximation, simulations carried out using the Slow-Scale SSA produced results that were practically indistinguishable from the results of the exact SSA, but did so roughly a thousand times faster (see Figs. 1, 2, 3, and 5).

Our two example fast processes were of course chosen for their simplicity, so that we could show how the Slow-Scale SSA works when the asymptotic properties of the virtual fast process can be estimated analytically. But fast reversible isomerizations and dimerizations are actually rather common in real cellular systems. For example, in the lambda-phage model of Arkin, *et al.*,[7] two particular fast reversible dimerizations sometimes account for over 95% of the reaction activity. So it is worth emphasizing that the formulas we have derived here for the first two moments of fast reversible isomerization and dimerization reactions, such as Eqs. (24)-(25) for the isomerization reactions, allow evaluation of *any* slow-scale propensity function relative to those fast processes. In a separate paper now nearing completion, the present authors will show how the formulas obtained here for the fast reversible dimerization can be used to speed up the lambda-phage model simulation. That paper will also describe an alternative simulation-based procedure for generating random samples of the stationary virtual fast process for use in operation (14) of the Slow-Scale SSA.

Our work here has many parallels with the path-breaking papers of Haseltine and Rawlings,[2] and Rao and Arkin;[3] for instance, our provisional grouping of the reactions into fast and slow categories on the basis of propensity function values follows directly in the footsteps of Haseltine and Rawlings.[2] But our virtual fast system is defined differently, and we think more simply, than the fast system used by both Haseltine and Rawlings[2] and Rao and Arkin,[3] which was the *real* fast system *conditioned* on the slow system. The non-Markovian nature of that process makes a reliable analysis problematic.[8] In contrast, our virtual fast system, being Markovian, is much easier to analyze. And our Slow-Scale Approximation shows precisely how to make use of the stationary (asymptotic) properties of that virtual fast system to construct a reliable Slow-Scale SSA.

In some cases, our Slow-Scale SSA will be practically the same as the simulation strategies that emerge from References 2 and 3, after all the approximations invoked by those two approaches have been made. In these cases of commonality, the fast species populations effectively get replaced by the solution of the deterministic reaction rate equation for a system that is essentially our virtual fast system. Indeed, in the simulations reported in the preceding two sections, this is pretty much all that was done insofar as the

evolution of the slow species is concerned. But we believe that our derivation of the Slow-Scale Approximation makes the theory underlying that replacement much more transparent. It provides a rational basis for deciding beforehand whether or not such a replacement is warranted, and it tells us what we should do when it is not. For instance, it enabled us to treat with confidence the system of Figs. 2 and 3, in which the critical fast species S_2 has an average population that is less than one. By contrast, treating the fast reactions in that circumstance using the chemical Langevin equation or the reaction rate equation, along the lines suggested by Haseltine and Rawlings,[2] would have seemed questionable since those equations usually require the species populations to be large.

On the subject of population sizes, the simulation results in Figs 1 and 5 illustrate an often unappreciated point: The size of the fluctuations in a species population is not tied in any simple way to the size of the mean population of that species. Although it is true that larger populations tend to exhibit smaller relative fluctuations, the trajectories in Figs. 1 and 5 show that fluctuation ranges of individual species depend strongly on the details of the specific reactions involved, and there is no "universal critical population level" above which fluctuations can always be ignored and below which they cannot.

Another interesting feature of the simulation results in Figs. 1 and 5 is the way in which the population of the slow species S_3 remains relatively smooth in spite of the large fluctuations in the population of the fast species S_2 that gives rise to S_3. A similar phenomenon was observed some time ago by Kurata, et al.[9] in their studies of the heat shock response mechanism in *E. coli*, and it was pointed out[10] that that system seemed to be subjecting the fluctuations of one sparsely populated but critical fast species to a kind of "low pass filter". We believe the reason for this filtering effect can be discerned from our proof in Sec. VI of the Slow-Scale Approximation: It shows that the ensemble average of the slow reaction propensity function in Eq. (9) actually represents a *time integral* over that propensity function (which in turn arose by invoking the addition law of probability), and as is well known, temporal integration tends to filter out high frequency fluctuations. The integral is smoother than the integrand. So while it is true that slow-scale propensity functions depend on the fast variables, that dependence is through a time integral over the fast variables, which smoothes out their high frequency fluctuations.

The relation of the Slow-Scale SSA to leaping methods[5,6] remains to be fully explored, but one thing in that regard is already clear: For *stiff* systems, the Slow-Scale SSA is far superior to *explicit* leaping methods.[6] By way of illustration, we found that an explicit tau-leaping simulation of the reactions in Fig. 5, made with the accuracy control parameter chosen large enough to admit noticeable differences from an exact SSA simulation, gave initial leaps that spanned less than 100 reactions, as compared to leaps spanning over 10^4 reactions in the more accurate Slow-Scale SSA run of Fig. 5b. This is not surprising since it is now recognized[5] that explicit leaping methods perform poorly on stiff systems – and the parameter values (59a) make reactions (36) and (55) very stiff. Explicit leaps are limited by stability considerations to the time scale of the fastest mode in the system. *Implicit* tau-leaping, however, is another matter, since it is expressly designed to accommodate stiff systems.[5] Our future work will explore the connection between implicit tau-leaping and the Slow-Scale SSA.

Another topic for future work will be to explore how the several methods described in Sec. IX for computing the asymptotic properties of a virtual fast process can be extended to more complicated processes, such as processes with more than one independent state variable. Success in this effort will be critical to making the Slow-Scale SSA a broadly applicable methodology.

ACKNOWLEDGEMENTS: We are grateful to Muruhan Rathinam, Hana El-Samad, and John Doyle for insightful suggestions at various stages of this work. We thank Carol Gillespie for assistance in performing the numerical calculations and creating the figures. This work was supported in part by the U.S. Air Force Office of Scientific Research and the California Institute of Technology under DARPA award No. F30602-01-2-0558. DG received additional support from the Molecular Sciences Institute under contract No. 244725 with the Sandia National Laboratories and the Department of Energy's "Genomes to Life" Program. YC and LP received additional support from the Department of Energy under DOE award No. DE-FG03-00ER25430, and from the National Science Foundation under NSF award No. CTS-0205584, and from the Institute for Collaborative Biotechnologies though grant DAAD19-03-D-0004 from the U.S. Army Research Office.

APPENDIX A: STATIONARY PROPERTIES OF UNIVARIATE BIRTH-DEATH MARKOV PROCESSES

A *univariate birth-death Markov process* $X(t)$ is by definition a scalar jump Markov process that is confined to the non-negative integers and changes state only in steps of ± 1.[11] The dynamics of such a process are governed by two "stepping functions" $W_{\pm}$. They are defined so that, if $X(t) = x$, then $W_{\pm}(x)dt$ gives the probability that $X(t+dt)$ will equal $x \pm 1$ for any infinitesimal $dt > 0$. The stationary form of the master equation for such a process reads

$$0 = \left[W_{-}(x+1)P(x+1,\infty) - W_{-}(x)P(x,\infty)\right] + \left[W_{+}(x-1)P(x-1,\infty) - W_{+}(x)P(x,\infty)\right].$$

By rearranging the terms in this equation we can deduce that

$$W_{-}(x)P(x,\infty) - W_{+}(x-1)P(x-1,\infty) = \text{constant},$$

and a consideration of the case $x = 0$ shows that the constant must be zero; thus, the stationary solution of the master equation, when it exists, satisfies

$$W_{-}(x)P(x,\infty) = W_{+}(x-1)P(x-1,\infty). \tag{A1}$$

Equation (A1) is called the *detailed balance* relation. It is evidently a recursion relation, since it allows us to calculate the values of $P(x,\infty)$ for all x in terms of its value at any arbitrarily chosen $x = x^*$ by

$$P(x,\infty) = \begin{cases} \dfrac{W_{-}(x+1)}{W_{+}(x)} P(x+1,\infty), & \text{for } x = x^*-1,\ldots,0, \\ \dfrac{W_{+}(x-1)}{W_{-}(x)} P(x-1,\infty), & \text{for } x = x^*+1,\ldots,L. \end{cases} \tag{A2}$$

This expresses every $P(x,\infty)$ as some x-dependent factor times $P(x^*,\infty)$, and the value of the latter can then determined by imposing the normalization condition,

$$\sum_{x=0}^{L} P(x,\infty) = 1. \tag{A3}$$

To avoid computational underflow in numerically iterating the recursion (A2), one should choose x^* to be at or near a relative maximum of P and initially take $P(x^*,\infty) = 0.1$. The upper limit L assumed in Eqs. (A2) and (A3) could, from a strictly mathematical point of view, be ∞; however, in the practical chemical problems with which we shall be concerned, where x represents the number of molecules of some species, L will always be finite. The *moments* of $X(\infty)$ can be calculated from $P(x,\infty)$ as

$$\left\langle X^n(\infty) \right\rangle \equiv \sum_{x=0}^{L} x^n P(x,\infty) \quad (n = 1,2,\ldots). \tag{A4}$$

Sometimes Eq. (A2) can be iterated analytically, as in Example 1 of the text. Other times, the upper x-limit L may be small enough that the iteration can be accomplished numerically. More often than not, though, the iteration gives results that are too complicated to be of practical use. But it is possible to extract from Eq. (A2) some

relatively simple formulas that give the locations and widths of the relative maximums of $P(x,\infty)$, and in the case of unimodal distributions these often provide acceptable estimates of the mean and variance of $X(\infty)$.

A *relative maximum* of $P(x,\infty)$ is called a *stable state* of $X(t)$. It can be proved from Eq. (A1) that a relative maximum of $P(x,\infty)$ can always be computed as the greatest integer in a down-going root of the function[11]

$$\alpha(x) \triangleq W_+(x-1) - W_-(x); \tag{A5}$$

i.e., $\hat{x}$ (a non-negative integer) will a stable state of $X(t)$ if and only if, for some $\delta \in [0,1)$,

$$\alpha(\hat{x}+\delta) = 0 \quad \text{and} \quad \alpha'(\hat{x}+\delta) < 0. \tag{A6}$$

These defining conditions for a stable state $\hat{x}$ can usually be approximated to $\alpha(\hat{x}) = 0$ and $\alpha'(\hat{x}) < 0$. And if $X(t)$ has *only one* stable state, we can usually put $\langle X(\infty)\rangle \approx \hat{x}$.

It can also be shown from Eq. (A2) that the *Gaussian variance* of stable state $\hat{x}$, which is defined as the variance of the Gaussian function $G(x)$ that satisfies $G(\hat{x}) = P(\hat{x},\infty)$, $G'(\hat{x}) = P'(\hat{x},\infty) = 0$, and $G''(\hat{x}) = P''(\hat{x},\infty)$, is given by[11]

$$\sigma_{\mathrm{G}}^2(\hat{x}) = \frac{W_-(\hat{x})}{-\alpha'(\hat{x})}. \tag{A7}$$

In words, $\sigma_{\mathrm{G}}^2(\hat{x})$ is the variance of the "best Gaussian fit" to the peak in $P(x,\infty)$ at $x = \hat{x}$. If $X(t)$ has *only one* stable state $\hat{x}$, we can usually put $\mathrm{var}\{X(\infty)\} \approx \sigma_{\mathrm{G}}^2(\hat{x})$.

It is also useful to have some idea of how large $t - t_0$ needs to be in order for $P(x,t|x_0,t_0)$ to be well approximated by $P(x,\infty)$, or equivalently, how long it takes $X(t_0)$ to *relax* to $X(\infty)$. In cases where there is more than one stable state this relaxation time will be roughly the average time it takes the process to visit *all* of its stable states at least once, and that time (which may be quite long) will be difficult to compute in general. But if there is *only one* stable state, which is the case for the simple examples that we are considering here, the relaxation time will be of the order of the time it takes $\langle X(t)\rangle$ to relax to $\langle X(\infty)\rangle$. We can estimate that time by reasoning as follows:

The time-evolution equation for the *mean* of a birth-death Markov process $X(t)$ reads[11]

$$\frac{d\langle X(t)\rangle}{dt} = \langle W_+(X(t)) - W_-(X(t))\rangle \approx \langle \alpha(X(t))\rangle,$$

where the last step has invoked the definition (A4) together with the *assumption* that the values of $X(t)$ are typically large compared to 1 (which is usually the case). Expanding $\alpha(x)$ in a Taylor series about the stable state value $\hat{x}$, and assuming that we can confine our attention to a region around $\hat{x}$ that is small enough that α can be linearly approximated there, we get

$$\frac{d\langle X(t)\rangle}{dt} \approx \langle \alpha(\hat{x}) + \alpha'(\hat{x})\big(X(t)-\hat{x}\big)\rangle = \alpha'(\hat{x})\big(\langle X(t)\rangle - \hat{x}\big).$$

Putting $u(t) \equiv \langle X(t)\rangle - \hat{x}$, we thus see that $du(t)/dt \approx \alpha'(\hat{x})u(t)$. The solution of this differential equation is $u(t) \approx u(0)\exp\big(\alpha'(\hat{x})t\big)$; therefore,

$$\langle X(t)\rangle \approx \hat{x} + \big(X(0)-\hat{x}\big)\exp\big(\alpha'(\hat{x})t\big).$$

Recalling that $\alpha'(\hat{x}) < 0$, we thus conclude that $\langle X(t)\rangle$ relaxes to $\langle X(\infty)\rangle \approx \hat{x}$ in a time of order

$$\hat{t} \approx \frac{1}{-\alpha'(\hat{x})}. \tag{A8}$$

If there is *only one* stable state $\hat{x}$, this can usually be taken as a reasonable estimate of the time it takes for a birth-death process $X(t)$ to relax to its stationary form $X(\infty)$.

To test the usefulness of formulas (A6)-(A8), let's see how closely they reproduce the exact results found in Sec. VIII.A for the reversible isomerization (17). Substituting Eqs. (20) into the definition (A4), we find that the birth-death process $\hat{X}_1(t)$ has

$$\alpha(x_1) = c_2\big(x_T - (x_1 - 1)\big) - c_1 x_1 = c_2(x_T + 1) - x_1(c_1 + c_2). \tag{A9}$$

This linear function evidently has a single down-going root at $x_1 = c_2(x_T+1)/(c_1+c_2)$, so by Eq. (A6) the stable state of $X_1(t)$ is

$$\hat{x}_1 = \left[\frac{c_2(x_T+1)}{(c_1+c_2)}\right], \tag{A10}$$

where the brackets signify "greatest integer in". In the usual case that $x_T \gg 1$, this value for $\hat{x}_1$ evidently provides an excellent approximation to $\langle \hat{X}_1(\infty)\rangle$ in Eq. (24a). From Eq. (A7) we compute

$$\sigma_G^2(\hat{x}) = \frac{c_1\hat{x}_1}{c_1+c_2} = \frac{c_1}{c_1+c_2}\left[\frac{c_2(x_T+1)}{(c_1+c_2)}\right], \tag{A11}$$

where the last step invokes Eq. (A10). In the case that $x_T \gg 1$, this value for $\sigma_G^2(\hat{x})$ evidently provides an excellent approximation to $\mathrm{var}\{\hat{X}_1(\infty)\}$ in Eq. (24b). And finally, substituting Eq. (A9) into Eq. (A8) gives

$$\hat{t} \approx \frac{1}{c_1+c_2}, \tag{A12}$$

which agrees exactly with the relaxation time estimate (28).

So if we had not been able to analytically solve the recursion relation (A2) to get the results (21)-(24), and the dynamical equations for $\langle \hat{X}_1(t)\rangle$ and $\mathrm{var}\{\hat{X}_1(t)\}$ to get the result (28), we could have obtained very good approximations to all those results by using the much simpler formulas (A6)-(A8).

APPENDIX B: THE STATIONARY MOMENT EQUATIONS

The generic N-species, M-reaction stationary chemical master equation reads

$$0=\sum_{j=1}^{M}\left\{a_j(x-\nu_j)P(x-\nu_j,\infty)-a_j(x)P(x,\infty)\right\}, \tag{B1}$$

where $x\equiv(x_1,\ldots,x_N)$ and $\nu_j\equiv(\nu_{1j},\ldots,\nu_{Nj})$. The stationary average of any function of state f is

$$\left\langle f\left(X(\infty)\right)\right\rangle\equiv\sum_x f(x)\,P(x,\infty), \tag{B2}$$

where the summation extends over all values of all components of x. If we multiply Eq. (B1) by $f(x)$ and then sum over x we get

$$0=\sum_{j=1}^{M}\sum_x f(x)\,a_j(x-\nu_j)P(x-\nu_j,\infty)-\sum_{j=1}^{M}\sum_x f(x)\,a_j(x)P(x,t).$$

But since

$$\sum_x f(x)\,a_j(x-\nu_j)P(x-\nu_j,\infty)=\sum_x f(x+\nu_j)\,a_j(x)P(x,\infty),$$

this is

$$0=\sum_{j=1}^{M}\sum_x\left[f(x+\nu_j)-f(x)\right]a_j(x)P(x,\infty),$$

or, using Eq. (B2),

$$0=\sum_{j=1}^{M}\left\langle\left[f\left(X(\infty)+\nu_j\right)-f\left(X(\infty)\right)\right]a_j\left(X(\infty)\right)\right\rangle. \tag{B3}$$

Putting $f(x)=x_i$ in Eq. (B3) gives

$$0=\sum_{j=1}^{M}\left\langle\left[\left(X_i(\infty)+\nu_{ij}\right)-X_i(\infty)\right]a_j\left(X(\infty)\right)\right\rangle,$$

whence

$$0=\sum_{j=1}^{M}\nu_{ij}\left\langle a_j\left(X(\infty)\right)\right\rangle\quad(i=1,\ldots,N). \tag{B4}$$

And putting $f(x)=x_ix_{i'}$ in Eq. (B3) gives

$$0=\sum_{j=1}^{M}\left\langle\left[\left(X_i(\infty)+\nu_{ij}\right)\left(X_{i'}(\infty)+\nu_{i'j}\right)-X_i(\infty)X_{i'}(\infty)\right]a_j\left(X(\infty)\right)\right\rangle,$$

whence

$$0=\sum_{j=1}^{M}\nu_{ij}\left\langle X_{i'}(\infty)\,a_j\left(X(\infty)\right)\right\rangle+\sum_{j=1}^{M}\nu_{i'j}\left\langle X_i(\infty)\,a_j\left(X(\infty)\right)\right\rangle$$
$$+\sum_{j=1}^{M}\nu_{ij}\nu_{i'j}\left\langle a_j\left(X(\infty)\right)\right\rangle\quad(i=1,\ldots,N;\,i'=i,\ldots,N). \tag{B5}$$

If all the propensity functions are no more than linear in the state variables, meaning that none of the R_j reactions involves more than one reactant molecule, then the N equations (B4) can be solved for the N stationary first moments $\langle X_i(\infty)\rangle$, and the $\frac{1}{2}N(N+1)$ equations (B5) can be solved for the $\frac{1}{2}N(N+1)$ stationary second moments $\langle X_i(\infty)X_{i'}(\infty)\rangle$. An example is provided by the reversible isomerization reaction defined in Eqs. (17) and (18): Equation (B4) gives, using the conservation relation (19),

$$\begin{aligned} 0 &= \nu_{11}\langle a_1(X(\infty))\rangle + \nu_{21}\langle a_2(X(\infty))\rangle, \\ &= (-1)\langle c_1 X_1(\infty)\rangle + (+1)\langle c_2(x_\mathrm{T} - X_1(\infty))\rangle, \\ 0 &= c_2 x_\mathrm{T} - (c_1 + c_2)\langle X_1(\infty)\rangle. \end{aligned}$$

This gives the exact result (24a) for $\langle X_1(\infty)\rangle$. And Eq. (B5) gives for $i = i' = 1$,

$$\begin{aligned} 0 &= 2\big(\nu_{11}\langle X_1(\infty)a_1(X(\infty))\rangle + \nu_{12}\langle X_1(\infty)a_2(X(\infty))\rangle\big) \\ &\qquad + \nu_{11}\nu_{11}\langle a_1(X(\infty))\rangle + \nu_{12}\nu_{12}\langle a_2(X(\infty))\rangle, \\ &= 2\Big((-1)\langle X_1(\infty)c_1X_1(\infty)\rangle + (+1)\langle X_1(\infty)c_2(x_\mathrm{T} - X_1(\infty))\rangle\Big) \\ &\qquad + (-1)^2\langle c_1X_1(\infty)\rangle + (+1)^2\langle c_2(x_\mathrm{T} - X_1(\infty))\rangle, \\ 0 &= -2(c_1 + c_2)\langle X_1^2(\infty)\rangle + (2c_2x_\mathrm{T} + c_1 - c_2)\langle X_1(\infty)\rangle + c_2x_\mathrm{T}. \end{aligned}$$

Using the previously obtained result for $\langle X_1(\infty)\rangle$ this last equation can be reduced to

$$\langle X_1^2(\infty)\rangle = \langle X_1(\infty)\rangle^2 + \frac{c_1c_2x_\mathrm{T}}{(c_1 + c_2)^2},$$

which gives the exact result (24b) for $\mathrm{var}\{X_1(\infty)\}$.

More often, though, at least one reaction will be bimolecular, so its propensity function will be quadratic in the components of X. In that case, Eqs. (B4) will contain at least one second-order moment $\langle X_i(\infty)X_{i'}(\infty)\rangle$, and Eqs. (B5) will contain at least one third-order moment $\langle X_i(\infty)X_{i'}(\infty)X_{i''}(\infty)\rangle$, and so on. The set of stationary moment equations will then be infinitely open-ended, and hence not solvable. Usually, the best we can do in such a situation is to make some sort of approximating assumption that closes the moment equations (B4) and (B5).

The crudest closure approximation is the *deterministic approximation*:

$$\langle X_i^2(\infty)\rangle = \langle X_i(\infty)\rangle^2, \quad \forall i. \tag{B6}$$

This implies that $\mathrm{var}\{X_i(\infty)\} = 0$ for all i, and hence that $X_i(\infty)$ is a sure variable, with no fluctuations. The fact that $\mathrm{cov}\{X_i(\infty), X_{i'}(\infty)\}$ is bounded in absolute value by the product of $\sqrt{\mathrm{var}\{X_i(\infty)\}}$ and $\sqrt{\mathrm{var}\{X_{i'}(\infty)\}}$ means that, under the deterministic

approximation (B6), all the covariances vanish as well; thus, $\left\langle X_i(\infty)X_{i'}(\infty)\right\rangle = \left\langle X_i(\infty)\right\rangle\left\langle X_{i'}(\infty)\right\rangle$. Equation (B4) thus simplifies to

$$0 = \sum_{j=1}^{M} \nu_{ij}\, a_j\left(\left\langle X(\infty)\right\rangle\right) \quad (i = 1,\ldots,N)\,. \tag{B7}$$

This says that the $\left\langle X_i(\infty)\right\rangle$ are just the solutions of the standard *stationary reaction rate equation*. There is no need for Eqs. (B5) in this approximation, since (B6) has the effect of approximating all higher order moments as simple products of the first-order moments.

As crude as the deterministic approximation is, there are many circumstances in which it will be adequate. Usually this will happen if all the species populations are large compared to 1, which is not an uncommon situation in chemical kinetics. But circumstances can also arise in which the deterministic approximation will not be adequate, and in those cases a more sophisticated approximation strategy for closing the stationary moment equations must be used.

In the special case that $X(\infty)$ is a scalar random variable whose probability density function has a *single peak*, which is to say a single stable stationary state, it is sometimes possible to approximate $X(\infty)$ as a *normal* random variable,

$$X(\infty) = \mathcal{N}(m,\sigma^2)\,. \tag{B8}$$

The mean m and variance σ^2 are to be chosen to satisfy the first two moment equations (B4) and (B5). This is possible since, even though those two equations typically involve moments of $X(\infty)$ higher than the second, for the *normal* random variable all moments are given explicitly in terms of the two parameters m and σ^2. In particular,

$$\left\langle X(\infty)\right\rangle = m\,, \tag{B9a}$$

$$\left\langle X^2(\infty)\right\rangle = m^2 + \sigma^2\,, \tag{B9b}$$

$$\left\langle X^3(\infty)\right\rangle = m\left(m^2 + 3\sigma^2\right). \tag{B9c}$$

The strategy, then, is to substitute Eqs. (B9) into Eqs. (B4) and (B5), solve those two equations simultaneously for the two parameters m and σ^2, and finally evaluate the first two moments from Eqs. (B9a) and (B9b). Since the resulting two equations are nonlinear in the unknowns, a numerical solution procedure will usually be required. For obvious physical reasons we can accept only solutions with both m and σ^2 non-negative; indeed, the entire approximation will be suspect if it is not also true that $m - \sigma > 0$.

REFERENCES

1 D. T. Gillespie, *J. Comput. Phys.* **22**, 403 (1976); *J. Phys. Chem.* **81**, 2340 (1977); *Physica A* **188**, 404 (1992).

2 E. L. Haseltine and J. B. Rawlings, *J. Chem. Phys.* **117**, 6959 (2002).

3 C. V. Rao and A. P. Arkin, *J. Chem. Phys.* **118**, 4999 (2003).

4 The concepts of temporal and ensemble averages in are discussed in some detail in F. Reif, *Fundamentals of Statistical and Thermal Physics* (McGraw-Hill, New York, 1965, see Sec. 15.14). Briefly, the argument is that for any *stable* process $Y(t)$ and *sufficiently large* T and N, we can write the "temporal average" of $Y(t)$ as

$$\frac{1}{T}\int_0^T Y(t)dt \approx \frac{1}{T}\sum_{i=1}^{N} Y(t_{i-1})\left(\frac{T}{N}\right) \approx \frac{1}{N}\sum_{i=1}^{N} Y(\infty)_i \ .$$

The first equality invokes a partitioning of the interval $[0,T]$ into N subintervals of size T/N according to $t_i = (T/N)i \ \ (i = 0,\ldots,N)$. The second equality follows from the fact that, when T is sufficiently large, the random variable $Y(t)$ can be replaced almost everywhere by its asymptotic form $Y(\infty)$, so the values $Y(t_{i-1})$ become, in aggregate, just N random samples of $Y(\infty)$. The final expression approximates, for sufficiently large N, the "ensemble average" of the random variable $Y(\infty)$; that average can also be computed as $\sum_{\text{all } y} \Pr\{Y(\infty)=y\}\, y$.

5 M. Rathinam, L. R. Petzold, Y. Cao, D. T. Gillespie, *J. Chem. Phys.* **119**, 12784 (2003).

6 D. T. Gillespie, *J. Chem. Phys.* **115**, 1716 (2001); D. T. Gillespie and L. R. Petzold, *J. Chem. Phys.* **119**, 8229 (2003).

7 A. Arkin, J. Ross, and H. H. McAdams, *Genetics* **149**, 1633 (1998).

8 That the components of a multivariate Markov process usually are not individually Markovian is widely recognized. For example, in Brownian motion the Brownian particle's velocity $V(t)$ and position $X(t)$ together form a bivariate Markov process $\big(V(t),X(t)\big)$; however, $X(t)$ is not by itself a Markov process, and $V(t)$ will be Markovian only if there is no position-dependent force field. It should therefore come as no surprise that our fast and slow processes $X^{\mathrm{f}}(t)$ and $X^{\mathrm{s}}(t)$, which together form the Markov process $X(t) = \big(X^{\mathrm{f}}(t), X^{\mathrm{s}}(t)\big)$, are not individually Markovian. The reason can be understood heuristically as follows: Each of these two processes supplies to the other, through the connections indicated by Eqs. (3) and (4b), information about the other's past values that goes *beyond* what is conveyed by the other's present value. Since that information gets used in advancing the two processes in time, the processes by themselves are not "past-forgetting" in the Markov sense. More formally, a process will be Markovian if and only if it satisfies the Chapman-Kolmogorov condition – an integral

relation of which the familiar master equation is a special differential form for processes of the jump type. It follows that a jump process cannot be Markovian unless it satisfies a canonical form master equation, like our Eq. (5). $X^{\mathrm{f}}(t)$ and $X^{\mathrm{s}}(t)$ separately do not satisfy such an equation, so they are not Markovian. References 2 and 3 take for their fast process the (real) fast process *conditioned* on the slow process. That process is shown in Ref. 3 to satisfy *its* Eq. (5) – an equation that *almost* has the canonical form of *our* Eq. (5), but not quite; it falls short in that the slow variables in the two propensity functions on the right hand side have different values. This implies that the fast process conditioned on the slow process is not Markovian. To carry out a proper analysis of a non-Markovian process on its own terms – i.e., without first "embedding" it in a larger Markov process and then analyzing that larger process – is horrendously difficult, largely because one must ensure the simultaneous satisfaction of infinitely many coupled integral equations; see D. Gillespie, W. Alltop, and J. Martin, *Chaos* **11**, 548 (2001).

9 H. Kurata, H. El-Samad, T.-M. Yi, J. Khammash, and J. Doyle, "Feedback regulation of the heat shock response in E. coli", *Proc. 40th IEEE Conf. on Decision and Control*, 2001.

10 John Doyle, private communication.

11 D. T. Gillespie, *Markov Processes: An Introduction for Physical Scientists* (Academic Press, San Diego, 1992). See especially Sec. 6.4 and Appendix D; in the latter it is shown that the "effective width" of a Gaussian peak is $\sqrt{2\pi}$ times the standard deviation.

FIGURE CAPTIONS

Fig. 1. Results of simulating reactions (17) and (30) for the parameter values (34), made using (a) the exact SSA, and (b) the approximate Slow-Scale SSA. In both runs, points were plotted immediately after each firing of the slow reaction (30). The exact run in (a) had to simulate over 40 million reaction events, whereas the approximate run in (b) simulated only 521 reaction events.

Fig. 2. An exact SSA simulation of reactions (17) and (30) for parameter values (35), which values give an initial average S_2 population of only 0.5. In (a), samplings of all three populations are plotted *immediately after* each firing of the slow reaction (30). More than 23 million reactions in all were simulated here. In (b) the S_2 population for the same run is sampled *immediately before* each slow reaction, and in (c) the S_2 population for the same run is sampled *at equal time intervals* of $\Delta t = 1.167$. The differences in the three X_2 trajectories highlight the fact that R_3 firing times are highly dependent on the S_2 population. The equal-time-sampling trajectory (c) should be the most "typical."

Fig. 3. A simulation of reactions (17) and (30) for parameter values (35) made using the approximate Slow-Scale SSA. The points are plotted out just after each firing of the slow reaction (30), except that the X_1 and X_2 trajectories have been "relaxed" to a very short time after that. For reasons explained in the text, the X_2 trajectory here matches the "typical" SSA X_2 trajectory in Fig. 2c better than the either of the X_2 trajectories in Figs. 2a or 2b.

Fig. 4. Plots of $P(x_2, \infty | x_T)$-versus-x_2 for the reversible dimerization reactions (36), using the parameter values (54). The *open circles* show the *exact* function, as computed numerically from the recursion relation (A2). The *solid* curve is the *normal* distribution with the same mean and variance as the exact curve. The *dashed* curve is the *binomial* distribution with the same mean and upper limit $[x_T/2]$ as the exact curve. The distribution is actually defined on the interval $[0,1000]$, but all three curves are *very* close to zero outside the peak area; however, a semi-log plot would reveal substantial differences among the three curves outside the peak in just *how* close to zero they are.

Fig. 5. Results of simulating reactions (36) and (55) for the parameter values (59), made using (a) the exact SSA, and (b) the approximate Slow-Scale SSA. In both runs, points were plotted immediately after the firing of *either* of the slow reactions (55). The exact run in (a) simulated 2.483×10^7 reaction events in about 17 minutes, whereas the approximate run in (b) simulated 1,741 reaction events in a fraction of a second.

Fig 1

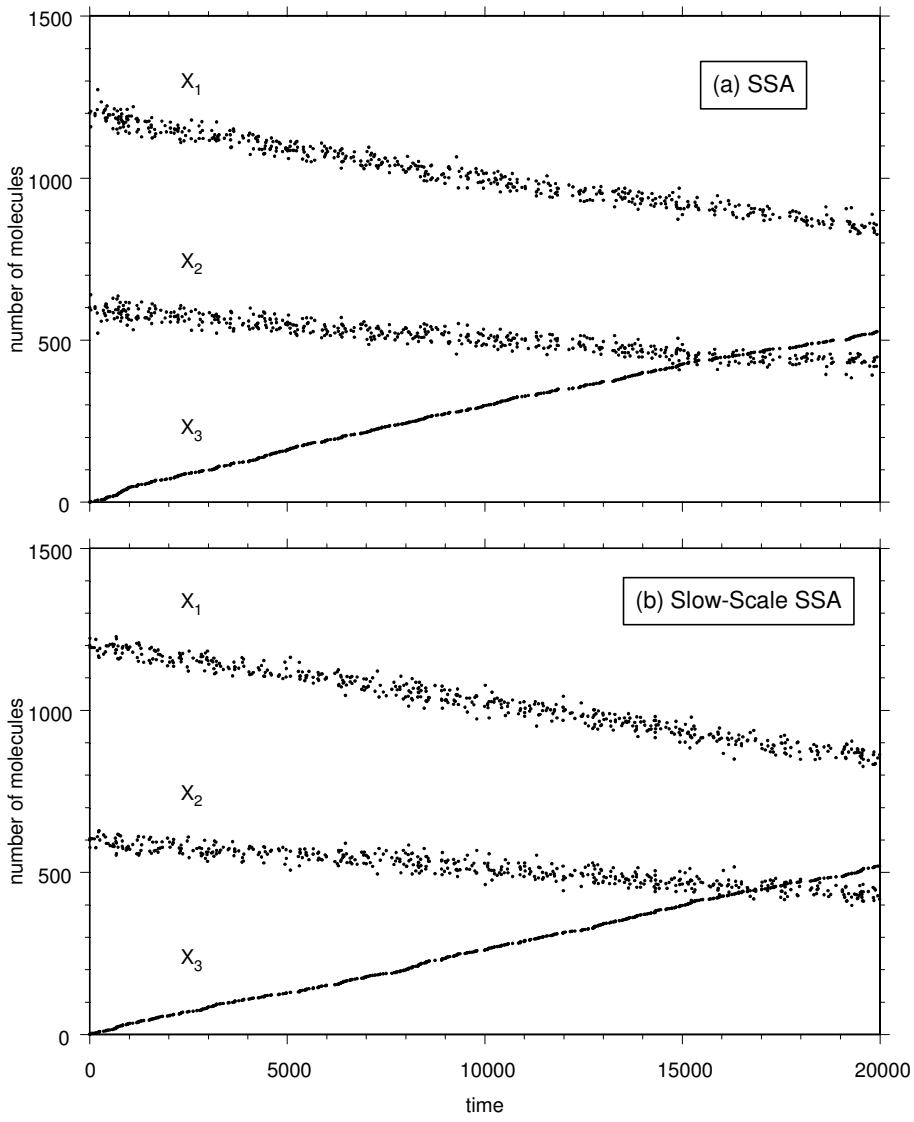

1500
1000
500
0
number of molecules
X1
X2
X3
(a) SSA
(b) Slow-Scale SSA
0
5000
10000
15000
20000
time

Fig 2

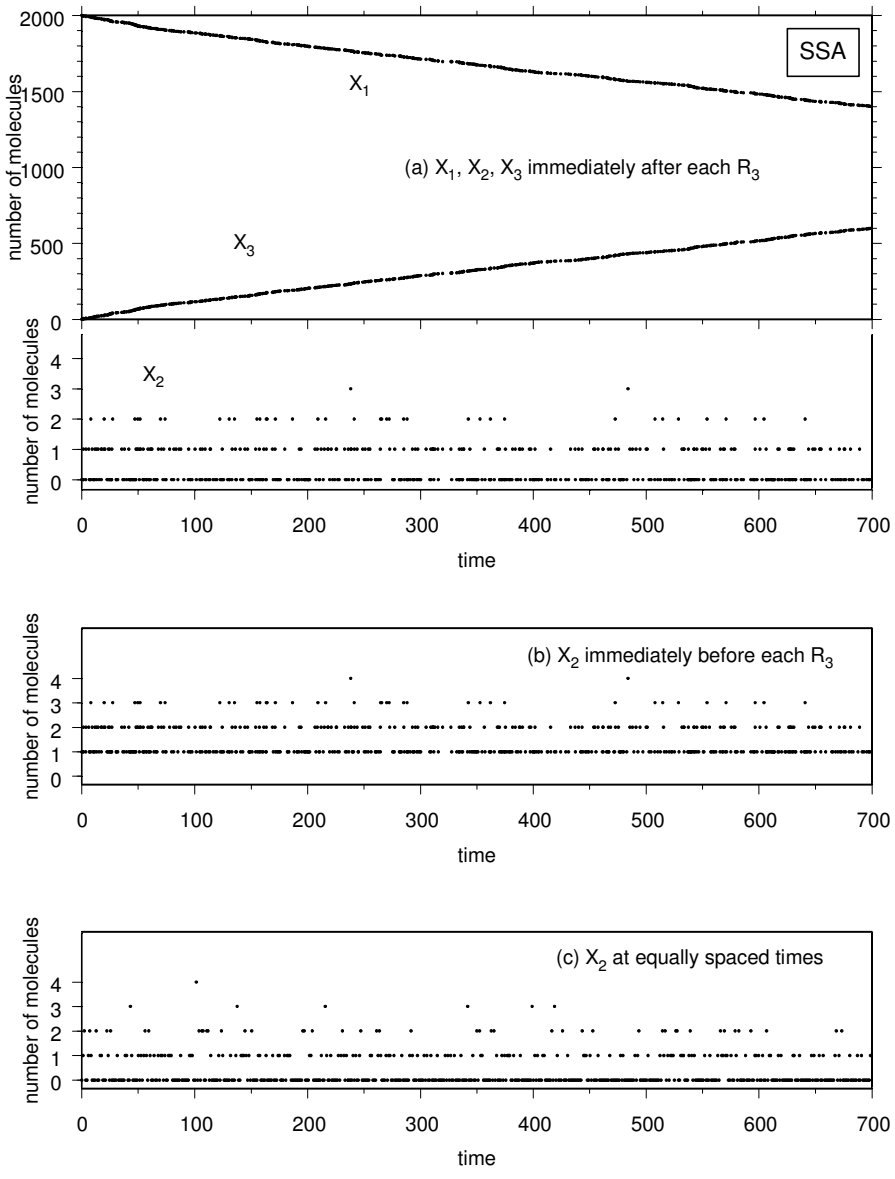

SSA
X1
X3
X2
(a) X1, X2, X3 immediately after each R3
(b) X2 immediately before each R3
(c) X2 at equally spaced times
number of molecules
time
0
100
200
300
400
500
600
700
500
1000
1500
2000
1
2
3
4

Fig 3

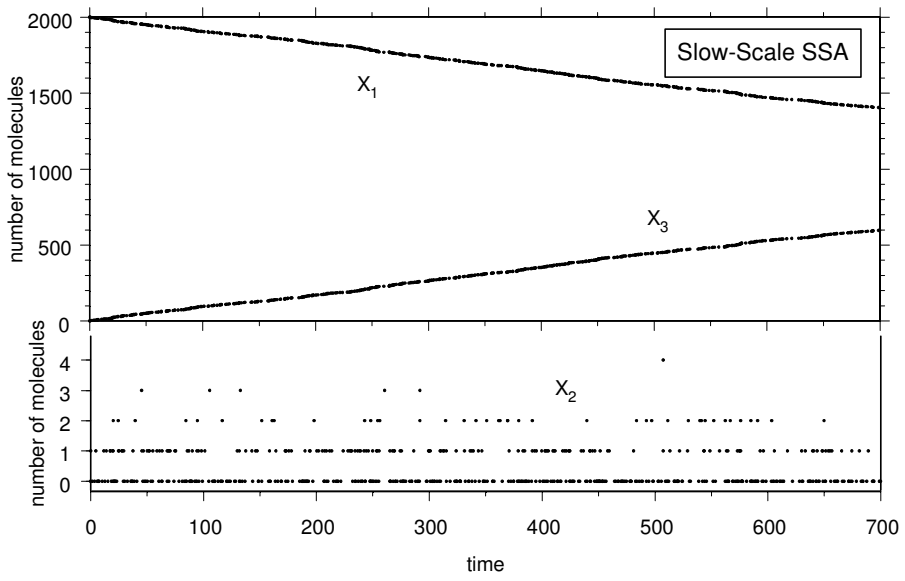

Slow-Scale SSA
X1
X3
X2
number of molecules
number of molecules
time
2000
1500
1000
500
0
4
3
2
1
0
0
100
200
300
400
500
600
700

Fig 4

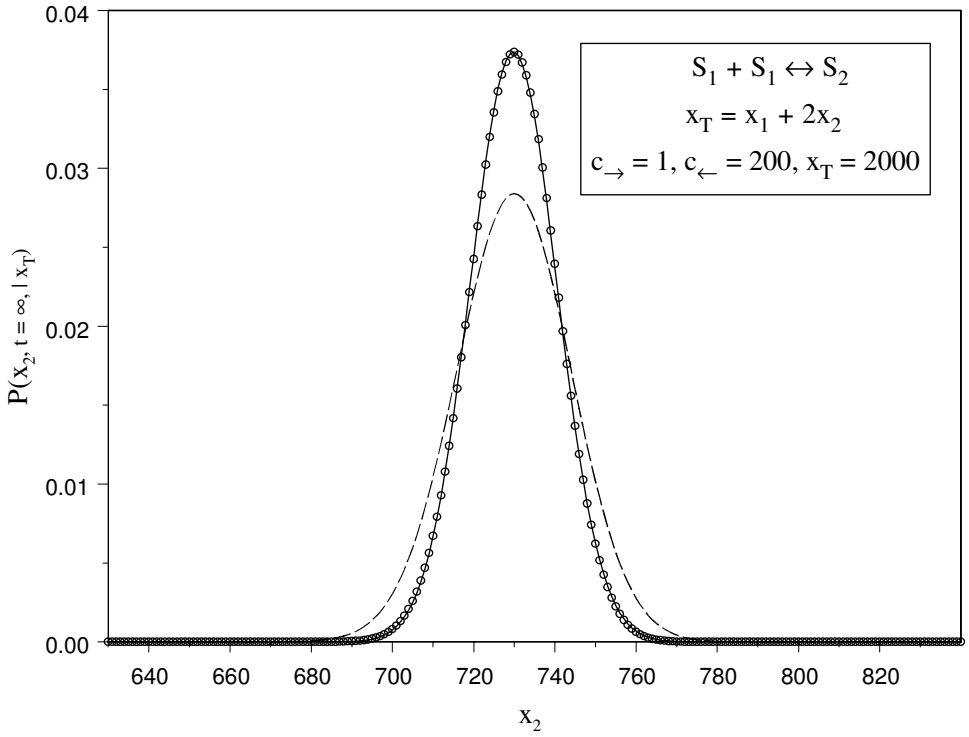
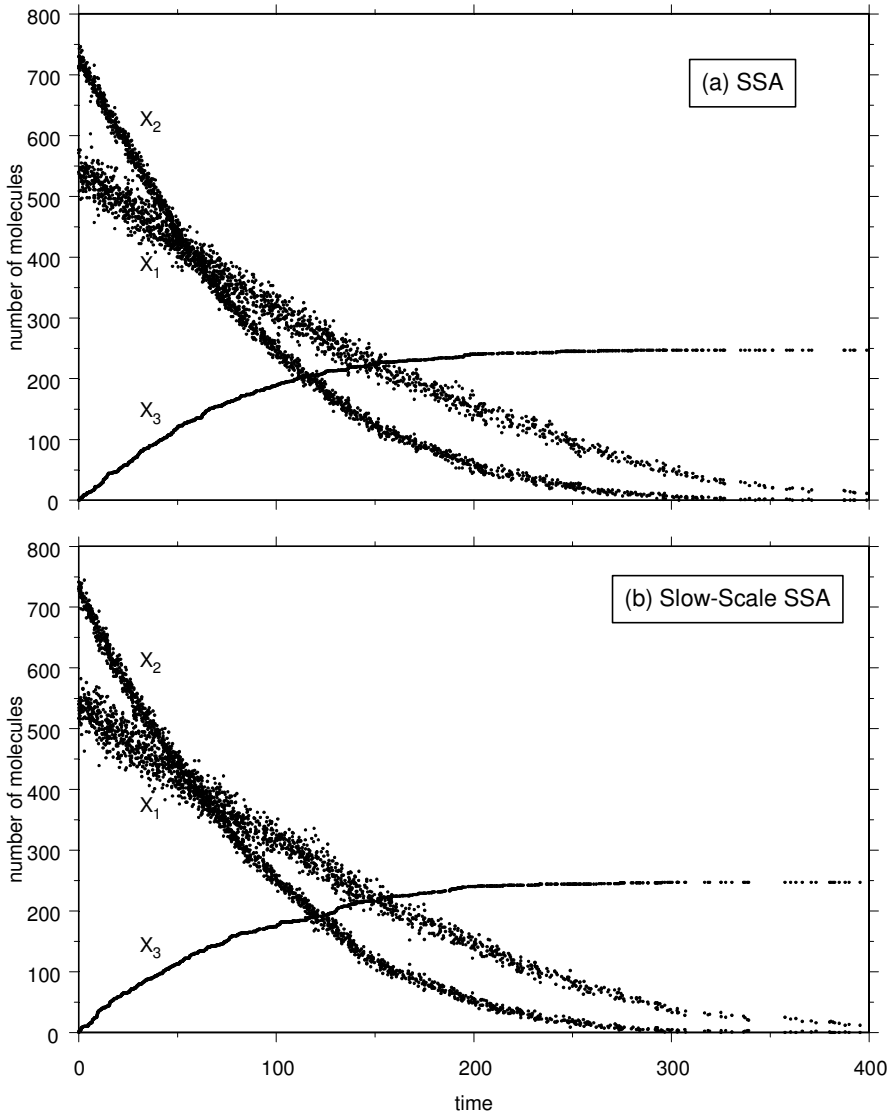

0.04
0.03
0.02
0.01
0.00
640
660
680
700
720
740
760
780
800
820
$P(x_2, t = \infty, | x_T)$
x_2
$S_1 + S_1 \leftrightarrow S_2$
$x_T = x_1 + 2x_2$
$c_{\rightarrow} = 1, c_{\leftarrow} = 200, x_T = 2000$

Fig 5


(a) SSA
(b) Slow-Scale SSA
number of molecules
time
X1
X2
X3
0
100
200
300
400
500
600
700
800