Comment on "Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates" [JCP 123, 194107 (2005)]

D. T. Gillespie[a)]

Dan T Gillespie Consulting, 30504 Cordoba Place, Castaic, California 91384

L. R. Petzold

Dept. of Computer Science, University of California, Santa Barbara, California 93106

Y. Cao

Dept. of Computer Science, Virginia Tech, Blacksburg, Virginia 24061

Abstract: The slow-scale stochastic simulation algorithm (ssSSA) proposed in [*J. Chem. Phys.* **122**, 014116 (2005)] and the nested stochastic simulation algorithm (nSSA) proposed in [*J. Chem. Phys.* **123**,194107 (2005)] are closely related approximate simulation procedures aimed at speeding up the stochastic simulation of stiff chemical systems, i.e., systems that evolve through fast and slow dynamical modes with the fast mode being stable. This Comment aims to clarify some misconceptions that have arisen over the relationship between the ssSSA and the nSSA as regards both their theoretical foundations and their practical implementations.

Submitted as **T6.06.150** to *The Journal of Chemical Physics*

(Version of 15 Jan 2007)

a) Author to whom correspondence should be addressed.

Nearly a dozen papers have been published over the past five years aimed at approximately speeding up the exact stochastic simulation algorithm (SSA) for chemical systems that evolve on widely different time scales. Such systems are termed *stiff*, a term whose meaning in a stochastic context was recently elucidated in Rathinam, et al.[1] Here we focus on just two of these papers, namely Ref. 2 by the present authors, which introduced the *slow-scale stochastic simulation algorithm* (ssSSA), and Ref. 3 by E, Liu and Vanden-Eijnden, which was published almost a year later and introduced the *nested stochastic simulation algorithm* (nSSA). We first correct an incorrect claim that was made in Ref. 3 concerning the ssSSA. We then show that the misunderstanding that led the authors of Ref. 3 to make that incorrect claim also prevented them from seeing that, although there are indeed differences in how the ssSSA and the nSSA are implemented, those two algorithms share a common theoretical foundation – one that was fully established in Ref. 2.

The incorrect claim in Ref. 3 occurred in its discussion of how to speed up the simulation of the model reaction set

$$S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2 \underset{c_4}{\overset{c_3}{\rightleftharpoons}} S_3 \underset{c_6}{\overset{c_5}{\rightleftharpoons}} S_4 \tag{1}$$

when the outer rate constants (c_1, c_2, c_5 and c_6) are very much larger than the inner rate constants (c_3 and c_4). The authors of Ref. 3 noted that since all four species get changed by at least one of the fast reactions (R_1, R_2, R_5 and R_6), then this system has no slow species. From that they concluded, wrongly, that the ssSSA "would not result in any changes over the straightforward SSA." In fact, the ssSSA does not require that there be any slow species in order to be applicable or efficacious. And for this particular problem, the speedup over the SSA provided by the ssSSA is substantially greater than the speedup provided by the nSSA, as we shall now demonstrate.

Figure 1a shows the results of an exact SSA run of reactions (1) for the rate constants

$$c_1 = c_6 = 3\times10^4,\ c_2 = c_5 = 6\times10^4,\ c_3 = c_4 = 1\,. \tag{2}$$

The species populations here were plotted out after the occurrence of each *slow* reaction (R_3 or R_4). A total of 2.16×10^8 reaction events were simulated in the time span shown, but only 1,771 of those corresponded to a slow reaction.

A simulation run of these same reactions using the ssSSA[2], a procedure that explicitly simulates *only* the slow reactions, produced trajectories that are statistically indistinguishable from those shown in Fig. 1a (to save space we do not show them). By "statistically indistinguishable," we mean that like-plotted results of repeated independent runs of the SSA and the ssSSA could not be told apart. But the ssSSA run was accomplished $\sim 2.7\times10^4$ times faster (comparing actual run times) than the SSA run.

Simulating these reactions using the nSSA[3] requires one to choose a value for a certain control parameter, T_f. Larger values for T_f will make the simulation more accurate, but also cause it to run longer. Reference 3 gives little practical guidance for choosing T_f. We found that taking $T_f = 1.1\times10^{-5}$, a value which for reasons explained

below should give acceptable results, indeed caused the nSSA to produce trajectories that were likewise statistically indistinguishable from the SSA trajectories in Fig. 1a. That nSSA run was accomplished ~ 150 times faster than the SSA run, but ~ 180 times slower than the ssSSA run. A second nSSA run, this one made with T_{f} smaller by a factor of 0.1, sped things up: it was $\sim 1.5\times 10^3$ times faster than the SSA run, and ~ 18 times slower than the ssSSA run, and it produced the trajectories shown in Fig. 1b. The difference in the statistical textures of these trajectories compared to the exact SSA trajectories in Fig. 1a shows that this value for T_{f} is too small.

The fact that the authors of Ref. 3 believed the ssSSA was inapplicable to this problem suggests that our exposition in Ref. 2 was less than clear. We shall therefore give a brief recapitulation of the ssSSA and its underlying logic, and then describe how we think the nSSA fits into that picture.

The ssSSA proceeds in a series of steps.[2] The first step is to make a provisional partitioning of the *reactions* into fast and slow subsets. The criteria for making this partitioning is that the propensity functions (or stochastic rates) $a_j^{\mathrm{f}}(\mathbf{x})$ of each fast reaction R_j^{f} should usually be very much larger than the propensity function $a_j^{\mathrm{s}}(\mathbf{x})$ of any slow reaction R_j^{s}. The qualifier "usually" is needed because the value of a propensity function generally depends on the species population vector $\mathbf{x}$, and a propensity function might be very large in some regions of state space and very small in other regions. It is therefore impossible to know whether a propensity function will be large or small most of the time unless we already have a good idea about how the system behaves. But this partitioning is only *provisional*; it will later be subjected to a test that will determine whether or not it is acceptable.

Like the ssSSA, the nSSA of Ref. 3 also aims to evolve the system by simulating only the slow reactions, so it too begins by partitioning the reactions into fast and slow subsets. But the discussion of this task given in Ref. 3 glosses over the practical difficulties just described, and leaves the impression that deciding whether any reaction is fast or slow can always be done easily and confidently. This will not always be so, especially in the commonly occurring situation of a system that is only borderline-stiff. As presented in Ref. 3, the nSSA does not provide an acceptance criterion to assure the legitimacy of a fast/slow partitioning. This is problematic, because the strategy of skipping over the fast reactions and simulating only the slow ones will be sound only if the fast reactions are in some sense "less important" than the slow ones. There are chemical systems (detonation reactions for instance) in which the fast reactions are *not* less important than the slow ones; their simulation cannot be accelerated by either the ssSSA or the nSSA. The partitioning acceptance test of the ssSSA, to be described shortly, provides a way of determining whether or not it should be okay to skip over the fast reactions.

With the reactions provisionally partitioned, the second step in the ssSSA is to partition the *species* into fast and slow subsets. The criterion for doing this is simple[2]: Any species whose population *gets changed* by at least one fast reaction is classified as fast, and all other species are classified as slow. This leads to a partitioning of the state

vector into fast and slow components, $\mathbf{X}(t) = \left(\mathbf{X}^{\mathrm{f}}(t), \mathbf{X}^{\mathrm{s}}(t)\right)$. Points to note: a fast species population can get changed by a slow reaction but a slow species population cannot get changed by a fast reaction; a fast species population need not be large; and the set of slow species might be empty.

The third step in the ssSSA is to define the *virtual fast process* $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ as the fast species populations evolving under *only* the fast reactions.[2] Thus, $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ is $\mathbf{X}^{\mathrm{f}}(t)$ with all the slow reactions switched off. $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ is a fictitious process, in contrast to the "real" fast process $\mathbf{X}^{\mathrm{f}}(t)$ which consists of the same fast variables evolving under *all* the reactions. $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ will always be more tractable than $\mathbf{X}^{\mathrm{f}}(t)$.

The initial partitioning of the reactions is now deemed acceptable if and only if two conditions are satisfied:[2] First, $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ must be *stable* in the technical sense of stochastic process theory; i.e., the limit $\hat{\mathbf{X}}^{\mathrm{f}}(\infty)$ of $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ as $t \to \infty$ must exist as a well-behaved random variable, or equivalently, the initially-conditioned probability that $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ has a given value $\mathbf{y}$ must become independent of t as $t \to \infty$. Second, $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ must effectively reach this limit $\hat{\mathbf{X}}^{\mathrm{f}}(\infty)$ in a time that is *small* compared to the expected time to the *next slow* reaction. In effect, these two conditions require that $\mathbf{X}(t)$ be *stiff*; i.e., it should have fast and slow dynamical modes, and the fast mode should be stable. Systems that do not satisfy these two conditions will typically be those for which the fast reactions are no less important than the slow ones, and hence should not be skipped over. Verifying the satisfaction of these two stochastic stiffness conditions can be a challenging task, but it has been successfully done for several simple systems.[2,4]

The logical basis for the ssSSA, and we contend the nSSA too, is the *slow-scale approximation lemma*[2]: With the system in state $(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})$ at time t, let Δ_{s} be a time that is very large compared to the time it takes $\hat{\mathbf{X}}^{\mathrm{f}}(t)$ to relax to $\hat{\mathbf{X}}^{\mathrm{f}}(\infty)$, yet very small compared to the expected time to the next slow reaction. (The existence of such a Δ_{s} is guaranteed by the two stiffness conditions.) Then the probability that the slow reaction R_j^{s} will fire in the time interval $[t, t+\Delta_{\mathrm{s}})$ can be well approximated by $\overline{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})\Delta_{\mathrm{s}}$, where

$$\overline{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}}) \triangleq \sum_{\mathbf{y}^{\mathrm{f}}} \hat{P}(\mathbf{y}^{\mathrm{f}}, \infty | \mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})\, a_j^{\mathrm{s}}(\mathbf{y}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}}), \tag{3}$$

with $\hat{P}(\mathbf{y}^{\mathrm{f}}, \infty | \mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})$ being the probability that $\hat{\mathbf{X}}^{\mathrm{f}}(\infty) = \mathbf{y}^{\mathrm{f}}$ given the initial state $(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})$.

The import of this lemma can be appreciated by recalling that the definition of R_j^{s}'s real propensity function $a_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})$ is that, with the system is in state $(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})$ at time t, $a_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}})\,dt$ gives the probability that R_j^{s} will fire in the next *infinitesimal* interval $[t, t+dt)$. Since Δ_{s} is an "approximate infinitesimal" on the time scale of the slow

reactions, then the lemma implies that $\bar{a}_j^{\rm s}(\mathbf{x}^{\rm f},\mathbf{x}^{\rm s})$ defined in (3) can *approximately* be regarded as $R_j^{\rm s}$'s propensity function *on the time scale of the slow reactions*.

The slow-scale approximation lemma can be proved as follows:[2] Let $[t',t'+dt')$ be an infinitesimal subinterval of $[t,t+\Delta_{\rm s})$. The probability that $R_j^{\rm s}$ will fire in $[t',t'+dt')$ is

$$a_j^{\rm s}\left(\mathbf{X}^{\rm f}(t'),\mathbf{X}^{\rm s}(t')\right)dt' \approx a_j^{\rm s}\left(\hat{\mathbf{X}}^{\rm f}(t'),\mathbf{x}^{\rm s}\right)dt'. \tag{4}$$

The last step follows because it is very unlikely that any slow reaction will fire anywhere in the interval $[t,t+\Delta_{\rm s})$, so $\mathbf{X}^{\rm f}(t')$ can be well approximated there by $\hat{\mathbf{X}}^{\rm f}(t')$, and $\mathbf{X}^{\rm s}(t')$ can be well approximated by $\mathbf{x}_{\rm s}$. Since there is a negligibly small probability of more than one $R_j^{\rm s}$ reaction firing in $[t,t+\Delta_{\rm s})$, we can invoke the addition law of probability for mutually exclusive events to compute

$$\text{Prob}\left\{R_j^{\rm s} \text{ in } [t,t+\Delta_{\rm s})\right\} \approx \int_t^{t+\Delta_{\rm s}} a_j^{\rm s}\left(\hat{\mathbf{X}}^{\rm f}(t'),\mathbf{x}^{\rm s}\right)dt', \tag{5a}$$

$$\approx \left\{\frac{1}{\Delta_{\rm s}}\int_t^{t+\Delta_{\rm s}} a_j^{\rm s}\left(\hat{\mathbf{X}}^{\rm f}(t'),\mathbf{x}^{\rm s}\right)dt'\right\}\Delta_{\rm s}, \tag{5b}$$

$$\approx \left\{\sum_{\mathbf{y}^{\rm f}} \hat{P}(\mathbf{y}^{\rm f},\infty|\mathbf{x}^{\rm f},\mathbf{x}^{\rm s})\, a_j^{\rm s}(\mathbf{y}^{\rm f},\mathbf{x}^{\rm s})\right\}\Delta_{\rm s}. \tag{5c}$$

The quantity in braces in (5b) is, since $\Delta_{\rm s}$ is "very large" on the time scale of the *fast* reactions, the *temporal average* of $a_j^{\rm s}\left(\hat{\mathbf{X}}^{\rm f}(t'),\mathbf{x}^{\rm s}\right)$. In proceeding to (5c), we are choosing to evaluate that temporal average as the *ensemble average* $\left\langle a_j^{\rm s}\left(\hat{\mathbf{X}}^{\rm f}(\infty),\mathbf{x}^{\rm s}\right)\right\rangle$, a tactic that is ubiquitously employed in ergodic systems.[5] This proves the lemma. (The reader is invited to compare this proof to the one outlined in the Reply to this Comment.)

The strategy of the ssSSA is to simulate only the slow reactions, using the SSA with the slow-scale propensity functions (3), and generating the populations of the fast species by sampling the probability function $\hat{P}(\mathbf{y}^{\rm f},\infty|\mathbf{x}^{\rm f},\mathbf{x}^{\rm s})$. In the case of reactions (1), the virtual fast process consists of the two uncoupled reaction pairs $S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2$ and $S_3 \underset{c_6}{\overset{c_5}{\rightleftharpoons}} S_4$. The asymptotic properties of the reversible isomerization reaction are exactly calculable.[2] For the initial state $(x_1,\ldots,x_4)$, $\hat{X}_1(\infty)$ and $\hat{X}_3(\infty)$ turn out to be the independent binomial random variables,

$$\hat{X}_1(\infty) = \text{Bin}\left(\frac{c_2}{c_1+c_2}, x_1+x_2\right),\ \hat{X}_3(\infty) = \text{Bin}\left(\frac{c_6}{c_5+c_6}, x_3+x_4\right). \tag{6a}$$

The other two state variables can be computed from

$$\hat{X}_2(\infty) = x_1 + x_2 - \hat{X}_1(\infty)\,, \quad \hat{X}_4(\infty) = x_3 + x_4 - \hat{X}_3(\infty)\,. \tag{6b}$$

With $\hat{\mathbf{X}}^{\mathrm{f}}(\infty)$ thus specified, the slow-scale propensity functions for the two slow reactions R_3 and R_4 are easily computed:

$$\overline{a}_3(\mathbf{x}) = c_3 \left\langle \hat{X}_2(\infty) \right\rangle = \frac{c_3 c_1 (x_1 + x_2)}{c_1 + c_2}, \quad \overline{a}_4(\mathbf{x}) = c_4 \left\langle \hat{X}_3(\infty) \right\rangle = \frac{c_4 c_6 (x_3 + x_4)}{c_5 + c_6}\,. \tag{7}$$

As was noted in the Appendix of Ref. 2, the relaxation times of the two reversible isomerizations are $(c_1 + c_2)^{-1}$ and $(c_5 + c_6)^{-1}$, so the relaxation time of the virtual fast process here will be the larger of those two times. The average time to the next slow reaction can be estimated as the reciprocal of $\overline{a}_3(\mathbf{x}) + \overline{a}_4(\mathbf{x})$. It can be verified that, for the values in (2), the relaxation time of the virtual fast process is orders of magnitude smaller than the average time to the next slow reaction, so the stiffness conditions are satisfied. The ssSSA then proceeds by using the SSA to simulate the two slow reactions R_3 and R_4 using the propensity functions (7). Whenever a slow reaction fires, the species populations $(x_1, \ldots, x_4)$ are changed appropriately, and then the *fast* species populations (which in this case are all the populations) are "relaxed" by replacing them with random samples from Eqs. (6). The resulting trajectories are statistically indistinguishable from the SSA trajectories in Fig. 1a.

In the context of the slow-scale approximation lemma, the nSSA[3] can be understood simply as proceeding from the *temporal* average in Eq. (5b) instead of the *ensemble* average in Eq. (5c). That is, in the nSSA, the slow-scale propensity functions are estimated by simulating the virtual fast process over a sufficiently long time T_{f} and then computing from that simulation data the temporal averages

$$\overline{a}_j^{\mathrm{s}}(\mathbf{x}^{\mathrm{f}}, \mathbf{x}^{\mathrm{s}}) \approx \frac{1}{T_{\mathrm{f}}} \int_t^{t+T_{\mathrm{f}}} a_j^{\mathrm{s}}\left(\hat{\mathbf{X}}^{\mathrm{f}}(t'), \mathbf{x}^{\mathrm{s}}\right) dt'\,, \tag{8}$$

as prescribed by Eq. (5b). The required random samples of $\hat{\mathbf{X}}^{\mathrm{f}}(\infty)$ are taken to be the values found at the end of that simulation:

$$\hat{\mathbf{X}}^{\mathrm{f}}(\infty) \approx \hat{\mathbf{X}}^{\mathrm{f}}(t + T_{\mathrm{f}})\,. \tag{9}$$

But the approximations (8) and (9) will be good only if T_{f} is "sufficiently large". As was noted earlier, the relaxation time for the virtual fast process for reactions (1) is $\max\left\{(c_1 + c_2)^{-1}, (c_5 + c_6)^{-1}\right\} = 1.1 \times 10^5$. This is the value for T_{f} that we used in the nSSA simulation run that replicated the exact trajectories in Fig. 1a. But taking T_{f} to be one-tenth of the relaxation time produced the overly correlated trajectories shown in Fig. 1b.

We do not dispute the correctness or the usefulness of the nSSA; indeed, we commend its use when the ensemble averages required by the ssSSA cannot be conveniently and accurately estimated analytically. But clearly, the nSSA and the ssSSA share the same theoretical foundation: Both require partitioning the reactions and the species into fast and slow subsets, since only by doing that can we identify the virtual fast process. And both require the two stiffness conditions to be satisfied, so that the slow-

scale approximation lemma will apply in either its temporal average form (5a) or its ensemble average form (5b).

We regret that several issues continue to separate us from the authors of Ref. 3. We disagree with their claim that the slow-scale approximation lemma was proved prior to Ref. 2 in Refs. 6 and 7; because, neither of those earlier papers identified the virtual fast process $\hat{\mathbf{X}}^{\mathrm{f}}(t)$, or required it to reach a well defined form $\hat{\mathbf{X}}^{\mathrm{f}}(\infty)$ in a time that is small compared to the expected time to the next slow reaction, as hypothesized by the lemma. We contend that the same is true of the four papers newly cited in the Reply which go back three decades before Ref. 7, and which are all being claimed to have (apparently redundantly) proved the lemma. Second, it seems to us that the authors of Ref. 3 do not appreciate the crucial fact that, even if one can determine the slow-scale propensity functions exactly, as we did for reactions (1) in Eqs. (7), inaccuracies will still arise if the relaxation time of the virtual fast process is not sufficiently small compared to the expected time to the next slow reaction. Lastly, we do not agree that their generalized "slow variables" are the key "to understand[ing] how and why the [nSSA] works."[3] Those variables are simply quantities that are conserved by the virtual fast process – $x_1 + x_2$ and $x_3 + x_4$ in the case of reactions (1) – and the exploitation of such conserved quantities in the asymptotic analysis of Markov processes is commonplace, e.g., in deriving our Eqs. (6) and (7). We suggest that an understanding of how and why the nSSA works is provided not by the conservation relations obeyed by the virtual fast process, but rather by the slow-scale approximation lemma of Ref. 2 that provides a context and rationale for that process.

REFERENCES

[1] M. Rathinam, L. Petzold, Y. Cao, D. Gillespie, *J. Chem. Phys.* **119**, 12784-12794 (2003).

[2] Y. Cao, D. T. Gillespie and L. R. Petzold, *J. Chem. Phys.* **122**, 014116 (2005).

[3] W. E, D. Liu and E. Vanden-Eijnden, *J. Chem. Phys.* **123**, 194107 (2005).

[4] Y. Cao, D. T. Gillespie and L. R. Petzold, *J. Chem. Phys.* **123**, 144917 (2005).

[5] The equivalence of the temporal and ensemble averages of a stable, jump stochastic process $Y(t)$ can be seen from the chain of reasoning

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T Y(t)dt = \lim_{T,N\to\infty}\frac{1}{T}\sum_{k=1}^{N} Y(t_k)\left(\frac{T}{N}\right) = \lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} Y(\infty)_k \ ,$$

where $t_k = (T/N)k$, and $Y(\infty)_k$ is a sample of $Y(\infty)$.

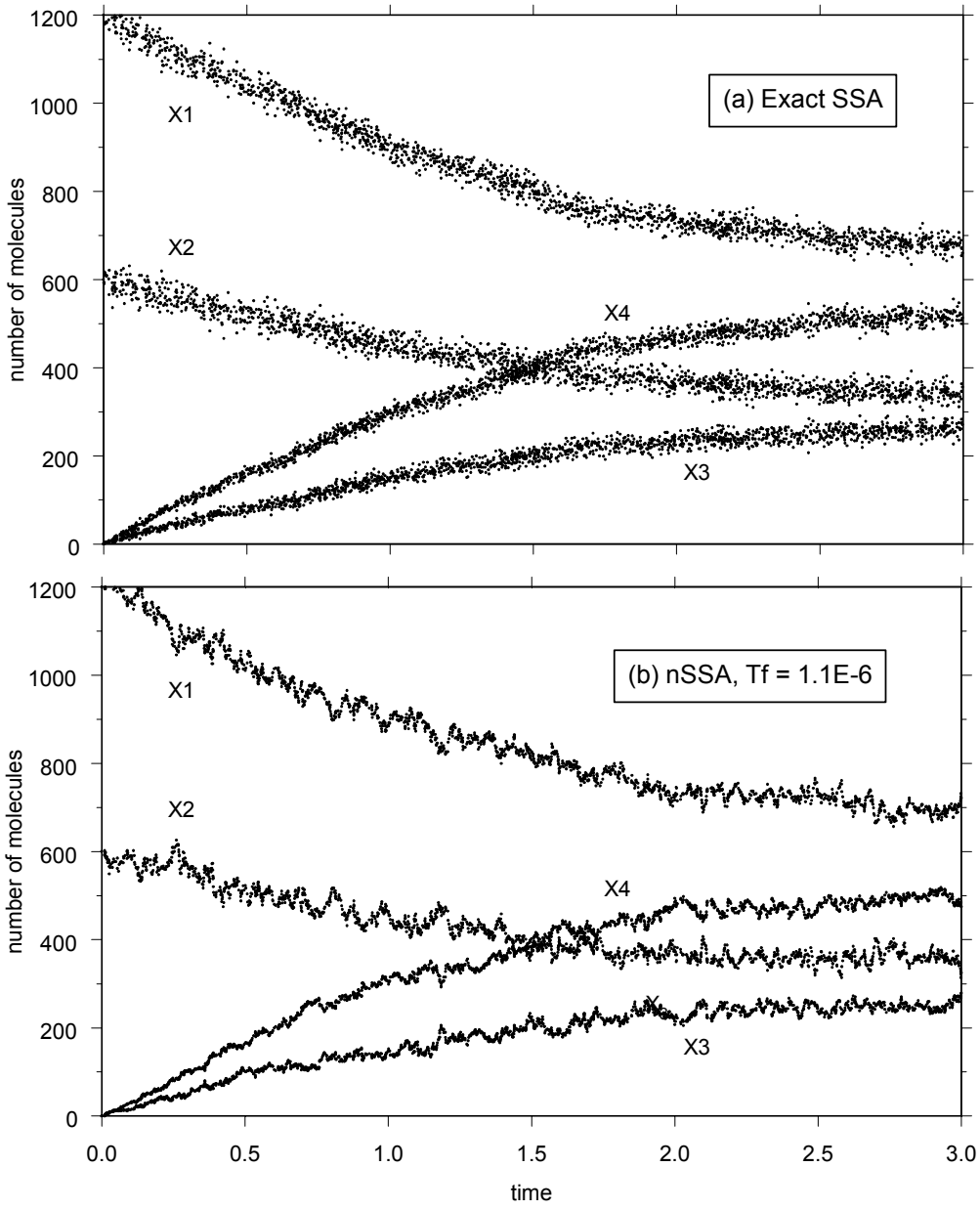
[6] C. Rao and A. P. Arkin, *J. Chem. Phys.* **118**, 4999-5010 (2003).

[7] R. Khasminskii, G. Yin, and Q. Zhang, *Q. Appl. Math.* **55**, 177 (1997).

FIGURE CAPTION

Fig. 1. **(a)** An exact SSA simulation of reactions (1) for the parameter values (2). X1 is the molecular population of species S_1, etc. The populations were plotted after the occurrence of each *slow* reaction, R_3 or R_4; these slow reactions comprised one out of every 1.2×10^5 reactions that were simulated to get this plot. Simulations using the ssSSA[2] and the nSSA[3], the latter with parameter $T_f = 1.1\times10^{-5}$, produced plotted trajectories that were statistically indistinguishable from these. Compared to the SSA run, the ssSSA run was $\sim 2.7\times10^4$ times faster, and the nSSA run was $\sim 1.5\times10^2$ times faster. **(b)** A simulation run of the nSSA with $T_f = 1.1\times10^{-6}$. Differences in the textures of the trajectories from the exact run in (a) are apparent with this smaller T_f; yet this nSSA run was ~ 18 times slower than the ssSSA run.

Fig 1


(a) Exact SSA
X1
X2
X4
X3
(b) nSSA, Tf = 1.1E-6
X1
X2
X4
X3
number of molecules
number of molecules
time
0
200
400
600
800
1000
1200
0.0
0.5
1.0
1.5
2.0
2.5
3.0