Automatic identification of model reductions for discrete stochastic simulation

Sheng Wu,[1,a)] Jin Fu,[1,b)] Hong Li,[2,c)] and Linda Petzold[1,d)]
[1]*Department of Computer Science, University of California Santa Barbara, Santa Barbara, California 93106, USA*
[2]*Teradata Inc., El Segundo, California 90245, USA*

(Received 27 February 2012; accepted 21 June 2012; published online 17 July 2012)

Multiple time scales in cellular chemical reaction systems present a challenge for the efficiency of stochastic simulation. Numerous model reductions have been proposed to accelerate the simulation of chemically reacting systems by exploiting time scale separation. However, these are often identified and deployed manually, requiring expert knowledge. This is time-consuming, prone to error, and opportunities for model reduction may be missed, particularly for large models. We propose an automatic model analysis algorithm using an adaptively weighted Petri net to dynamically identify opportunities for model reductions for both the stochastic simulation algorithm and tau-leaping simulation, with no requirement of expert knowledge input. Results are presented to demonstrate the utility and effectiveness of this approach. © *2012 American Institute of Physics*. [http://dx.doi.org/10.1063/1.4733563]

I. INTRODUCTION

Cellular chemical systems often exhibit dynamic behaviors that are both stochastic and discrete, due to low populations of key reactant species.[1–3] The stochastic simulation algorithm (SSA) (Ref. 4) is a useful numerical simulation tool that can capture these characteristics. However, because it must simulate every reaction event, the SSA, as well as many accelerated variants, is inefficient for most realistic models. One of the most important reasons is the presence of vastly different time scales and/or species population scales.

The explicit tau-leaping method[5,6] is a general approximate method to address the efficiency challenge due to different species population scales. It uses a Poisson approximation to leap over many reactions in one step, given that the reaction propensities remain relatively constant. However, as an explicit method, it is also inefficient for stiff systems, where vastly different time scales are involved.[7] Numerous model reductions which exploit time scale separation have been proposed for either the SSA or the explicit tau-leaping method, including the slow-scale SSA (ssSSA),[8] the stochastic quasi-steady state approximation (sQSSA),[9–11] the stochastic Michaelis-Menten model reduction (M-M),[9,10,12,13] and the time-dependent solution method.[14]

Among these methods, the ssSSA (Ref. 8) exploits the stochastic partial equilibrium of so-called virtual fast processes to treat multiple scales of the frequencies of reactions. A virtual fast process is a subsystem consisting of only fast reactions, which reaches a stochastic partial equilibrium very quickly between consecutive slow reactions. The propensities of the slow reactions can then be approximated based on the partial equilibrium distribution of the virtual fast process. The slow-scale tau-leaping method[15] is a natural combination of the ssSSA and the tau-leaping method to further accelerate simulation of such systems. Alternatively, the sQSSA (Refs. 9–11) aims at the multiple time scales of the changes of species' populations. It approximates the fast-changing species with a near stationary distribution over a relatively long time period by their quasi-steady state. The stochastic M-M approximation is designed for enzyme-substrate systems and can be derived from the sQSSA approach[9,10] or the ssSSA approach[12,13] under different conditions. Additionally, the time-dependent solution method[14] is an extension of the sQSSA. It gives more accurate results for fast changing species, regardless of whether they are in quasi-steady state or not.

With all of these different model reduction techniques, one of the major unsolved problems in practice is to efficiently identify situations where a model reduction can be safely applied and lead to a gain in performance. Usually, model reduction opportunities are identified manually and globally. By globally we mean that there is a global threshold of fast scales for the entire model and that all of the reactions or species of fast scales are grouped together throughout the simulation time span. There are several shortcomings to this approach. First, it is a manual process and requires expert knowledge of the dynamic behavior of the model, or trial-and-error which is expensive. It can be prone to error if the model does not behave as expected.[13,16] Second, it is difficult to determine a global threshold, as there may be different levels of fast scales. Third, there can be multiple disconnected fast subsystems (we will define connectivity more precisely in Sec. III). To group these disconnected subsystems together is not efficient. Finally, as the system evolves dynamically the fast subsystem may change over time. It is not easy to

a)Electronic mail: sheng@cs.ucsb.edu.
b)Electronic mail: fujin@cs.ucsb.edu.
c)Electronic mail: hong.li@teradata.com.
d)Electronic mail: petzold@cs.ucsb.edu.

apply these model reductions, particularly to large models. Note that automatic model reductions have been proposed for deterministic models yet for different purposes, namely, skeletal model reduction which permanently removes unimportant species or reactions whose contribution to species of interest is negligible.[17–22] Also note that Ref. 11 suggests an algorithm to automatically identify sQSSA opportunities for tau-leaping. But the algorithm cannot be generalized to detect model reduction opportunities other than the application of sQSSA to single fast-changing species in tau-leaping (for example, stochastic M-M). Moreover, Kuwahara *et al.* have proposed and implemented automated stochastic model reduction software based on several network motifs.[23–25] However, the method and the software are limited to the proposed static network motifs.

In this paper, we propose an automatic algorithm for the analysis of stochastic models, to identify situations where specific model reductions may be deployed safely, efficiently, and dynamically. Specifically, the algorithm uses adaptively weighted Petri nets at different time points as "snapshots" of the dynamic structure of a chemically reacting system. It analyzes these snapshots using graph theory, and decomposes the system at the "weak" nodes (slow reactions in the SSA simulation or slowly changing species in the tau-leaping simulation). The algorithm defines fast and slow as relative local properties. It locates and separates disconnected fast subsystems dynamically. More importantly, the algorithm does not need any input other than the model itself.

The outline of the paper is as follows. In Secs. II A and II B, we briefly review the SSA, the tau-leaping method, and the model reduction methods under consideration here. In Sec. II C, we introduce the Petri net,[26,27] a graphical tool to describe and analyze concurrent chemical processes. In Sec. III, we describe our model reduction detection technique in detail, for both the SSA and tau-leaping versions. In Sec. IV, we apply the algorithm to two realistic models: a heat shock response (HSR) model for *E. Coli* (Ref. 28 and 29) and a blood coagulation model.[30] We also demonstrate how fast subsystems can dynamically change, by analyzing the blood coagulation model. In Sec. V, we give some thoughts on the slow-scale tau-leaping method and propose an alternative method to better accelerate tau-leaping simulation.

II. BACKGROUND

A. SSA and tau-leaping

We begin with a well-stirred chemical reaction system with n molecular species $S_1, \ldots, S_n$ and m reaction channels $R_1, \ldots, R_m$. We assume the system is confined to a constant volume Ω at a constant temperature. Let $x_i(t)$ denote the population of species S_i at time t. Then the state of the system at time t is given by the state vector $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))^T$. Each reaction R_j is assumed to be characterized by two quantities: the probability $a_j(\mathbf{x})\,dt$ that one R_j reaction will occur in the next infinitesimal time interval $[t, t + dt)$, given $\mathbf{x}(t) = \mathbf{x}$, where $a_j(\mathbf{x})$ is termed the reaction's propensity function; and ν_j, the change to the system's state vector induced by one R_j reaction. ν_j is called the stoichiometry vector of R_j. For a unimolecular mass action reaction $S_1 \to P$, $a_j(\mathbf{x})$ has the form $c_j x_1$, where c_j is a constant. For a bimolecular mass action reaction $S_1 + S_2 \to P$, $a_j(\mathbf{x})$ has the form $c_j x_1 x_2$ (or $c_j \frac{1}{2} x_1(x_1 - 1)$ if $S_1 = S_2$), where c_j is a constant.

The dynamics of the system are given by the chemical master equation (CME) (Ref. 31)

$$\frac{\partial}{\partial t} P(\mathbf{x}, t|\mathbf{x}_0, t_0) = \sum_{j=1}^{m} [a_j(\mathbf{x} - \nu_j) P(\mathbf{x} - \nu_j, t \mid \mathbf{x}_0, t_0) - a_j(\mathbf{x}) P(\mathbf{x}, t \mid \mathbf{x}_0, t_0)], \quad (1)$$

where $P(\mathbf{x}, t|\mathbf{x}_0, t_0)$ is the probability that $\mathbf{x}(t) = \mathbf{x}$ given $\mathbf{x}(t_0) = \mathbf{x}_0$. Direct solution of the CME is not practical for all but the simplest systems, although approximate methods have been proposed.[32–36] The SSA (Ref. 4) is an exact method to numerically solve the chemical master equation by simulating a large number of possible trajectories of the system. The SSA is a kinetic Monte Carlo method based on the distribution of the time τ to the next reaction and the distribution of the index r of the next reaction. The former is an exponential distribution with mean $1/a_0(\mathbf{x})$ where $a_0(\mathbf{x}) = \sum_{j=1}^{m} a_j(\mathbf{x})$, while the latter is an integer random variable with probability mass function $P(r = j|\tau) = a_j(\mathbf{x})/a_0(\mathbf{x})$. At each step, the SSA generates two random numbers from these two distributions and advances the system by firing the chosen reaction at the chosen time. By simulating a large number of trajectories, the SSA can be used to approximate the distribution of the system vector at any given time t.

In principle, the SSA is able to simulate all well-stirred chemically reacting systems. However in practice, inefficiency of SSA simulation is a problem for most realistic models. Numerous exact approaches have been proposed to speed up the SSA, including the optimized direct method,[37] the next reaction method,[38] and the composition rejection algorithm.[39] However, these are exact methods and must simulate every reaction event in the system, so their efficiencies are limited by the number of reaction events. This can be a serious restriction for many systems. Thus, approximate methods have been proposed, among which the explicit tau-leaping method is widely used.

Instead of simulating one reaction at each step, tau-leaping[5] selects a time interval τ, fires multiple reactions during this interval, and advances the system by all the reactions that have fired during this interval. The basic idea is that if the propensity functions of all the reactions are nearly constant during the time interval $[t, t + \tau)$, the number of firings of each reaction can be approximated by a Poisson random number $\mathcal{P}(a_j(\mathbf{x})\tau)$. Then the state of the system can be advanced by the formula

$$\mathbf{x}(t + \tau) \approx \mathbf{x}(t) + \sum_{j=1}^{m} \mathcal{P}(a_j(\mathbf{x})\tau)\, \nu_j. \quad (2)$$

The requirement for the propensities to be nearly constant during the time interval is called the leap condition: for some $\varepsilon \ll 1$,

$$|\Delta_\tau a_j(\mathbf{x})/a_j(\mathbf{x})| \le \varepsilon, \quad \text{for all } j = 1, \ldots, m, \quad (3)$$

where $\Delta_\tau a_j$ is the change of a_j during the time interval $[t, t+\tau)$. The stepsize τ must be small enough to satisfy the leap condition. The consistency, stability, and accuracy of tau-leaping simulation have been studied.[40–42] It is easy to see that the strategy to select τ is critical to the accuracy and speed of tau-leaping simulation. The most widely used strategy for mass action reactions is due to Cao *et al.*[6] In that strategy, the leap condition is written in terms of the changes in species' populations rather than the changes in propensities, and the expression for τ is given by

$$\tau = \min_i \left\{ \frac{\max\{\varepsilon x_i/g_i, 1\}}{\left|\sum_j \nu_{ij} a_j(x)\right|}, \frac{\max\{\varepsilon x_i/g_i, 1\}^2}{\left|\sum_j \nu_{ij}^2 a_j(x)\right|} \right\}, \quad (4)$$

where $\varepsilon \ll 1$ is the preset accuracy control parameter, ν_{ij} are the stoichiometric coefficients, and g_i is the highest order of reaction in which species S_i appears as a reactant. The objective is to limit both the means and the variances of the changes in species' populations.

B. Model reductions

Neither the SSA or explicit tau-leaping are efficient for models with vastly different time scales. Numerous model reductions have been proposed to further accelerate simulation of systems with specific dynamic features. For our purposes, we will review the ssSSA,[8] the sQSSA,[9–11] the stochastic M-M approximation,[9,10,12,13] and the time-dependent solution method.[14]

The ssSSA aims at systems with fast subsystems that go to stochastic partial equilibrium.[8] The method first categorizes all reactions into fast reactions and slow reactions, assuming some knowledge of the system. Then it defines the virtual fast process to be the subsystem consisting of fast reactions, with all the slow reactions turned off. Defined in this way, the virtual fast process is a Markov process whose evolution is given by a CME. If there is a non-trivial limit (partial equilibrium distribution) of the CME of the virtual fast process as $t \to \infty$, and the time for the fast subsystem to reach the partial equilibrium distribution is much faster than the time to the firing of the next slow reaction, ssSSA uses the partial equilibrium distribution of the virtual fast process to approximately calculate the dynamics of the slow reactions. To calculate the effective propensities of the slow reactions, the populations of the fast species appearing in the propensity functions can often be well approximated by the expectations of their partial equilibrium distributions in the virtual fast process.[8]

The sQSSA, on the other hand, reduces the model by taking some species to their stochastic quasi-steady state.[9–11] A species S_i is said to be in stochastic quasi-steady state if

$$\frac{dP(x_i|\mathbf{x}_s)}{dt} \approx 0, \quad (5)$$

where $\mathbf{x}_s$ denotes the species that are not in the quasi-steady state. Assuming that $x_i|\mathbf{x}_s$ with fixed $\mathbf{x}_s$ is Markovian, then by applying the quasi-steady state approximation (5), the distribution of $P(x_i|\mathbf{x}_s)$ with fixed $\mathbf{x}_s$ can be obtained via the steady state master equation of $P(x_i|\mathbf{x}_s)$. Consequently, the quasi-steady state species x_i can be eliminated from the CME by summing over its conditional probability distribution over fixed $\mathbf{x}_s$. The reduced CME can be simulated with either SSA (Ref. 9) or tau-leaping,[11] where x_i is approximated by its quasi-steady state distribution. The sQSSA does not require a subsystem to reach stochastic partial equilibrium and is an alternative to the ssSSA for systems with different dynamic features.[10]

The stochastic M-M approximation[9,10,12,13] replaces the set of three reactions

$$E + S \underset{c_2}{\overset{c_1}{\rightleftharpoons}} C \xrightarrow{c_3} E + P, \quad (6)$$

with the single M-M reaction

$$S \xrightarrow{c} P. \quad (7)$$

It has been shown that the stochastic M-M approximation has the same form and validity conditions as the deterministic M-M approximation, and can be derived from either the ssSSA approach or the sQSSA approach under different conditions.[13] It is also interesting to note that the M-M reduction speeds up SSA and tau-leaping simulation differently under different conditions.[43]

The time-dependent solution method[14] is an extension to the sQSSA. It is inspired by the application of the sQSSA to tau-leaping simulation. Instead of approximating the conditional probability distribution of $P(x_i|\mathbf{x}_s)$ with fixed $\mathbf{x}_s$ by their quasi-steady state limit (5), the method uses the analytical time-dependent solution of the conditional probability distribution of those species. The method does not require the conditional probability distribution of $P(x_i|\mathbf{x}_s)$ with fixed $\mathbf{x}_s$ to be in quasi-steady state. It will provide more accurate results than the sQSSA when both methods apply. In general, the time-dependent solution is hard to derive, but for some common motifs simple analytical solutions do exist. Reference 14 gives the time-dependent solution of fast-changing species for several common motifs, automates the algorithm, and demonstrates the power of this method on a complex real-world model.

C. Petri nets

A Petri net is a graphical tool to model and study systems with concurrent processes.[26] It is named after Carl Adam Petri who invented the Petri net for the purpose of describing discrete processes.[27] Much research has been done on both theoretical developments and applications of the Petri net, including applying the Petri net to the study of biochemical reaction systems.[44,45] It is an alternative representation of stochastic biochemical reaction models and is equivalent to the CME. The Petri net describes the biochemical system in a graphical way that can be helpful in understanding the structure of the system, making use of knowledge from graph theory.

A Petri net consists of states(species), transitions (reactions), and directed edges. Species and reactions are interconnected by directed edges which show the directions of reaction flows. For example, consider a simple system consisting of 3 species and 3 reactions

$$S_1 \underset{c_2}{\overset{c_1}{\rightleftharpoons}} S_2 \xrightarrow{c_3} S_3. \quad (8)$$

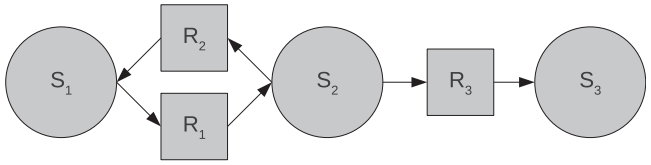


FIG. 1. The Petri net of the simple chemical system (8).

The Petri net of the system is shown in Fig. 1. Species S_1, S_2, and S_3 are denoted by round symbols and reactions R_1, R_2, and R_3 are denoted by rectangular symbols. For reaction R_1 converting S_1 to S_2, there is a directed edge coming from its reactant S_1, as well as a directed edge going to its product S_2. Directed edges are defined for all the reactions. Properly weighted, the Petri net can be a powerful tool to describe and to visualize the dynamics of chemical systems.

It is important to note that the Petri net, as well as its close variants, has been used to analyze the dynamics of chemical systems in the existing literature, yet for purposes which are different from our objectives.[46–51] In this paper, we propose a new use of Petri nets: to identify situations where specific model reductions can be safely and efficiently deployed.

III. METHODS

In this section, we explain how to use Petri nets to automate the identification of opportunities for model reduction, in the context of SSA or tau-leaping simulation. First, we review the general methodology of model reduction, in the context of both SSA and tau-leaping simulation. Next, we define relationships among reactions and species which can help us to understand the dynamic network structure of a chemically reacting system. Then we describe the algorithms (one for SSA and one for tau-leaping) and explain why they are able to capture the important dynamic features to determine when approximations are both justified and beneficial. At the end of this section, we present details of the algorithms.

A. Model reduction methodology

Many of the model reduction methods for discrete stochastic simulation exploit the presence of different time scales. In other words, there are "slow" properties embedded in "fast" elements in various chemically reacting systems. Often it is the slow properties that determine the dynamics of the system. By evaluating the slow properties properly instead of simulating the fast elements directly, model reductions can substantially increase the efficiency of stochastic simulation, without losing much accuracy.

For the ssSSA,[8] the fast element is the virtual fast process, while the slow property is the partial equilibrium distribution of the virtual fast process. The partial equilibrium distribution of the virtual fast process can be reached shortly after one slow reaction fires, and will not change until the next slow reaction fires. The ssSSA approximates the dynamics of the fast reactions by calculating the partial equilibrium distribution of the virtual fast process after each slow reaction. In this way, the fast time scale is removed from the simulation.

In the sQSSA,[9–11] the fast elements are the fast-changing species that are in stochastic quasi-steady state. The slow properties are the quasi-steady states of these fast changing species. The quasi-steady states of the fast changing species are determined by the slowly changing species involved in the production or consumption reactions.

The stochastic M-M approximation can be derived from either the sQSSA approach[9,10] or the ssSSA approach[12,13] under different conditions. For the enzyme-substrate system (6), if $E_T \ll S_0 + K_m$, where $E_T = E(t) + C(t)$, $S_0 = S(0)$, and $K_m = (c_2 + c_3)/c_1$, the enzyme-substrate complex C is changing fast but after a short time it is in stochastic quasi-steady state. Thus, the sQSSA can be applied and the fast changing species C is removed from the simulation. On the other hand, if $c_2 \gg c_3$, then reactions R_1 and R_2 form the virtual fast process, so the system can be reduced by the ssSSA. The results of the two methods agree when both of the conditions are satisfied.[13]

To exploit multiple time scales, one has to quantify "slow" and "fast" for specific chemically reacting systems. Generally, this has been pre-defined manually and globally in both deterministic and stochastic model reductions.[8–10,12,13,52–54] However, slow and fast are actually relative properties and should be defined locally and dynamically. We should also note that for the purpose of accelerating simulation, the concepts of fast and slow differ for different simulation methods. For example, as mentioned above, fast and slow usually refer to frequencies of reactions in SSA, while in tau-leaping it is the relative rates of change of species' populations (which restricts the stepsize) that matters. A Petri net describes the network structure of chemically reacting systems. Thus, it is very well suited for performing local property analysis. By assigning weights to reaction or species nodes of the Petri net, we are able to define fast and slow locally and dynamically for a given chemically reacting system, for SSA or tau-leaping.

B. Relationships among reactions and species

To make use of the Petri net, we must first define relationships between reactions and species. Recall that both SSA and tau-leaping are based on the assumption that the dynamics of a chemically reacting system is governed by propensity functions, and that the propensity function of a reaction is solely determined by the populations of its reactants and kinetic constants. Hence, reactions are linked via changes of species' populations, and vice versa. Accordingly, we categorize the relationships between reactions and species into unlinked, weakly linked, and strongly linked.

Definition 1: Reaction R and species S are ***linked*** *if there is at least one edge connecting them in the Petri net.*

In other words, reaction R and species S are linked if species S serves as a reactant or product or enzyme in

reaction R. There are two different types of links between reactions and species.

*Definition 2: A link between reaction R and species S is a **strong link** if each time a reaction R fires, the population of species S is changed.*

*Definition 3: A link between reaction R and species S is a **weak link** if each time a reaction R fires, the population of species S is unchanged.*

We can easily see that a link between reaction R and species S is weak if species S serves as an enzyme of the reaction (its stoichiometric coefficient is 0). Otherwise the link is strong.

Based on the links between reactions and species, we can define connections among reactions and connections among species, as well as the reaction connection graph and the species connection graph as follows.

*Definition 4: Reaction R_1 and reaction R_2 are **connected** if all of the following conditions are met:*

1. *There is a species S that they are both linked to;*
2. *At least one of the links is a strong link;*
3. *Species S serves as a reactant or enzyme in at least one of the reactions.*

*The **reaction connection graph** G_R of a chemically reacting system is an undirected graph whose nodes correspond to all the reactions in the system, and whose edges correspond to all the connections among reactions.*

Note that the reaction connection graph is very similar to the dependency graph commonly used to optimize the implementations of the SSA and its variants.[37–39]

*Definition 5: Species S_1 and species S_2 are **connected** if there is a reaction R that they are both linked to, and at least one of the species serves as a reactant or enzyme in reaction R.*

*The **species connection graph** G_S of a chemically reacting system is an undirected graph whose nodes correspond to all the species in the system, and whose edges correspond to all the connections among species.*

We also need to define the induced reaction connection subgraph, the induced species connection subgraph, and the induced sub-Petri net:

*Definition 6: A **reaction connection subgraph** H_R is a subgraph of the reaction connection graph G_R of a system. The subgraph H_R is **induced by reactions** $\{R_1, \ldots, R_k\}$ if the nodes of H_R correspond to reactions $\{R_1, \ldots, R_k\}$ and H_R has exactly the same edges that appear in G_R over $\{R_1, \ldots, R_k\}$.*

*Definition 7: A **species connection subgraph** H_S is a subgraph of the species connection graph G_S of a system. The subgraph H_S is **induced by species** $\{S_1, \ldots, S_l\}$ if the nodes of H_S correspond to species $\{S_1, \ldots, S_l\}$ and H_S has exactly the same edges that appear in G_S over $\{S_1, \ldots, S_l\}$.*

*Definition 8: Let G_P be the Petri net of a system. Then a Petri net H_P is called a **sub-Petri net** of the system if H_P is a subgraph of G_P.*

- *The sub-Petri net H_P is **induced by reactions** $\{R_1, \ldots, R_k\}$ if it is the Petri net corresponding to the subsystem consisting of reactions $\{R_1, \ldots, R_k\}$ and their linked species in G_P.*
- *The sub-Petri net H_P is **induced by species** $\{S_1, \ldots, S_l\}$ if it is the Petri net corresponding to the subsystem consisting of species $\{S_1, \ldots, S_l\}$ and their linked reactions in G_P.*

Note that the propensity of a reaction changes only if one of its linked reactions fires. Thus, the reaction connection graph can be very helpful while analyzing fast and slow reactions for SSA: if two fast reactions are not connected in the reaction connection subgraph induced by the fast reactions, they will belong to two different virtual fast processes. We will elaborate on this in Sec. III C.

Also note that the propensity of a reaction will change relatively rapidly if and only if the populations of some of its reactants or enzymes change relatively fast. Recall that tau-leaping uses the relative rates of change of species' populations to determine the stepsize. To accelerate tau-leaping simulation, we analyze the motifs of the species connection subgraphs induced by the fast-changing species, where by motifs we mean the graph structure of the connected components of the species connection subgraph. If two fast-changing species are not connected in the species connection subgraph induced by the fast-changing species, they will belong to two different fast-changing species groups and can be analyzed separately. We will refine the analysis process in Sec. III D.

For the enzyme-substrate system (6), we show the Petri net, the reaction connection graph, and the species connection graph of the system in Fig. 2. Model analysis using the graphs is illustrated in Secs. III C–III E.

C. Model analyzer for SSA

The SSA simulates every reaction event. It is the fast reactions that restrict the stepsize and simulation efficiency

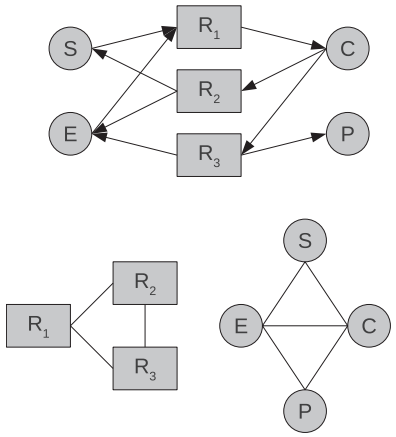


FIG. 2. The Petri net and connection graphs of the enzyme-substrate system: (upper) the Petri net, (lower left) the reaction connection graph, and (lower right) the species connection graph.

of SSA. We note that, for any reaction R, until R or one of its connected reactions fires, the propensity of R will remain unchanged. This feature was exploited in the next reaction method.[38] Consequently, if several fast reactions are connected together and all the other reactions connected to them are slow reactions, these fast reactions will form a virtual fast process (recall that a virtual fast process is defined to be the fast subsystem with slow reactions turned off). If this process can reach stochastic partial equilibrium quickly, then the ssSSA can be applied. Note that in this way, there can exist multiple disconnected virtual fast processes simultaneously.

To analyze models to accelerate SSA simulation, one has to locate all of the virtual fast processes. We already have the stationary network structure (reaction connection graph) derived from the Petri net. Now we need to quantify the dynamic "fastness" across reactions:

Definition 9: Given a chemically reacting system consisting of reactions $\{R_1, \ldots, R_m\}$, *the functions* $\{f_j(t), j = 1, \ldots, m\}$ *are called the* ***fastness functions of reactions*** $\{R_1, \ldots, R_m\}$ *if*

$$f_j(t)\,dt = C\,E[\#\,of\ \ firings\ of\ R_j\ in\ [t, t+dt)],\ \ j = 1, \ldots, m,$$

where C *is an arbitrary constant, and* E *denotes the expectation.*

With the fastness functions, we can compare the fastness of connected reactions:

Definition 10: Let $\{f_j(t), j = 1, \ldots, m\}$ *be the fastness functions of reactions* $\{R_1, \ldots, R_m\}$ *in a chemically reacting system. Then for any connected reaction pair* R_l *and* R_k, *assuming* $f_k(t') \neq 0$, *we say that at time* $t = t'$,

1. R_l is ***faster than*** R_k if $f_l(t')/f_k(t') \geq \lambda$;
2. R_l is ***of similar speed (similarly fast) as*** R_k if $\lambda^{-1} < f_l(t')/f_k(t') < \lambda$;
3. R_l is ***slower than*** R_k if $f_l(t')/f_k(t') \leq \lambda^{-1}$,

where λ *is the* ***distinguishing factor*** *greater than* 1, *and varies depending on the accuracy requirements and the speed gain that one hopes to achieve by the model reduction(s).*

It is natural to choose the propensity functions $\{a_j(t), j = 1, \ldots, m\}$ as the fastness functions of reactions. However, this can be a bad choice in practice if the propensity functions are very "noisy" due to fluctuations of species' populations, in which case the fast and slow comparison among connected reactions may change much more frequently than we want. Thus, it is more practical to use smoother functions as fastness functions, such as the running average of propensities or reaction counts in the time period of interest.

We can clearly see from Definition 10 that fast and slow are local properties in the reaction connection graph. Reaction R_1 can be faster compared to reaction R_2 but slower compared to reaction R_3.

Based on the fastness comparison between connected reaction pairs, the model analyzer for SSA partitions all the reactions into similarly fast reaction groups (SFRG) and identifies all of the virtual fast processes. The details of the partitioning algorithm will be given in Sec. III E. Once we have identified all of the virtual fast processes, we can determine whether model reductions will be applicable to them. For example, for ssSSA, for a specific virtual fast process we need to determine whether it will reach a stochastic partial equilibrium. Necessary and sufficient conditions for this are given in Ref. 50 and Ref. 46 and 55, respectively:

Necessary condition. For a virtual fast process to be able to reach stochastic partial equilibrium, the sub-Petri net induced by the virtual fast process must be consistent. A sub-Petri net is said to be consistent if there exists a column vector $\mathbf{v} = [v_1, \ldots, v_{m'}]^T$ *such that* $\Gamma\mathbf{v} = 0$ *and* $v_i > 0$ *for all* i, *where* $\Gamma = [\nu_1, \ldots, \nu_{m'}]$ *is the stoichiometric matrix of the sub-Petri net.*[50]

Sufficient condition. A virtual fast process is able to reach stochastic partial equilibrium if the chemical reaction network (CRN) of the virtual fast process is weakly reversible and of deficiency zero.[46,55]

For the detailed definition of CRN, as well as weak reversibility and deficiency of a CRN, please refer to Ref. 46 or Ref. 55. In this paper, we use only a weaker sufficient condition: a virtual fast process consisting of only reversible reaction pairs is able to reach stochastic partial equilibrium. With these two conditions, we can categorize virtual fast processes into three classes: ssSSA reducible, ssSSA irreducible, and possibly ssSSA reducible.

To illustrate these ideas, consider again the enzyme-substrate system (6). Suppose $c_1 = 1$, $c_2 = 100$, $c_3 = 1$, and the initial conditions are $S(0) = 1 \times 10^5$, $E(0) = 1 \times 10^3$, $C(0) = P(0) = 0$. After a transient period, the populations of E and C will be ~ 1 and $\sim 1 \times 10^3$, respectively. The propensity for each reaction will be $a_1 \sim 1 \times 10^5$, $a_2 \sim 1 \times 10^5$, and $a_3 \sim 1 \times 10^3$.

The reaction connection graph of the system is shown in Fig. 2. By setting the fastness functions of the reactions to be the expected propensity of each reaction, we can easily see that R_1 and R_2 will be partitioned into one SFRG Y_1, while R_3 will be partitioned into another SFRG Y_2 (details of the partitioning algorithm will be given in Sec. III E). Since Y_1 is the fastest SFRG, it will be identified as a virtual fast process. The subsystem induced by R_1 and R_2 consists of reactions R_1, R_2 and species S, E, and C. The subsystem induced by R_1 and R_2 consists of only reversible reaction pairs, and thus satisfies the sufficient condition to apply ssSSA. The ssSSA can be applied to R_1 and R_2, and speeds up the SSA simulation significantly. This result agrees with the partitioning which was used in Ref. 13 but has been determined here in an automated manner.

D. Model analyzer for tau-leaping

Model analysis for tau-leaping is similar to that for SSA, except that the analysis is performed on the species connection graph. The reason is that, as described in Sec. II A, tau-leaping leaps through multiple reactions during one step, and the stepsize is controlled by the relative

change of each species. Thus, for tau-leaping, the efficiency of the simulation is restricted by species that change relatively fast. Most model reductions for tau-leaping work by removing the species that change the fastest from the stepsize calculation: the sQSSA (Refs. 9, 10, and 11) removes species in stochastic quasi-steady state (the species' population changes fast but the distribution of the species' population remains relatively constant) from the stepsize calculation; and the time-dependent solution method[14] gives explicitly the time-dependent solution of the distributions of the populations of the fast-changing species, for common network subgraph motifs. Due to the nature of stepsize selection in tau-leaping, when treating fast-changing species it is safe to approximate the connected slowly changing species as constant during one "slow" step, where a "slow" stepsize is the stepsize dictated by the rates of changes of only the slowly changing species. Thus, we need to treat fast-changing species as a group only when they are connected. To approximate these fast-changing species groups, we use a species connection graph weighted by the fastness of species to analyze the model, locate all such fast-changing species groups, and compare them with known motifs to identify model reductions that can be applied. To do so, we first define the fastness function of species:

Definition 11: Given a chemically reacting system consisting of species $\{S_1, \ldots, S_n\}$, *the functions* $f_i(t), i = 1, \ldots, n$ *are called the* ***fastness functions of species*** $\{S_1, \ldots, S_n\}$ *if*

$$f_i(t)\,dt = C\ E\left[\tau_i^{-1} \text{ in } [t, t+dt)\right],\ i = 1, \ldots, n,$$

where C is an arbitrary constant, E denotes the expectation, and

$$\tau_i = \min\left\{\frac{\max\{\varepsilon x_i/g_i, 1\}}{\left|\sum_j \nu_{ij} a_j(x)\right|}, \frac{\max\{\varepsilon x_i/g_i, 1\}^2}{\left|\sum_j \nu_{ij}^2 a_j(x)\right|}\right\}$$

is the stepsize restriction due to species S_i *in the leap condition* (4).

We can easily see from the definition that the fastness function $f_i(t)$ of species S_i is proportional to the average number of steps that tau-leaping needs to take in a given time period due to the restriction of species S_i in the leap condition (4). The larger $f_i(t)$ is, the faster the population of S_i changes.

With the fastness function of species, we can compare fastness of connected species in the same way as we did for reactions via Definition 10, on the species connection graph of the system. Then the model analyzer partitions all the species into similarly fast species groups (SFSG) and identifies all the fast-changing species groups, as will be described in Sec. III E. For each fast-changing species group, we do not try to determine if it can reach stochastic partial equilibrium. Rather, we compare the induced sub-Petri net with known motifs to identify model reductions that can be applied. For example, to determine whether the sQSSA is applicable to a single fast-changing species S_i, we calculate the ratio of τ_{mean} (the first term in the leap condition (4)) to τ_{variance} (the second term in the leap condition (4)) of S_i, in other words, the ratio of the stepsize restriction due to the expected mean of the change of S_i to the stepsize restriction due to the expected variance of the change of S_i

$$\rho_i = \frac{\tau_{\text{mean}}}{\tau_{\text{variance}}} = \frac{\max\{\varepsilon x_i/g_i, 1\}/\left|\sum_j \nu_{ij} a_j(x)\right|}{\max\{\varepsilon x_i/g_i, 1\}^2/\left|\sum_j \nu_{ij}^2 a_j(x)\right|}. \quad (9)$$

If $\rho_i \gg \max_j\{|\nu_{ij}|\}$ (note that $\max_j\{|\nu_{ij}|\}$ is usually 1 or 2), then the population of S_i is fluctuating a lot but the mean of the population of S_i is not changing much, and the sQSSA will be applicable to S_i. There are three reasons to define such ρ's as the indicators for sQSSA opportunities. First, in Appendix A we show that $\rho_i \gg \max_j\{|\nu_{ij}|\}$ indeed indicates that S_i is in quasi-steady state. Second, τ_{mean} and τ_{variance} are already calculated in each step in tau-leaping simulation. There is essentially no overhead to calculate the ρ's. Third, as sQSSA removes the stepsize restriction due to τ_{variance} but not the stepsize restriction due to τ_{mean}, if $\rho_i \not\gg 1$, the tau-leaping simulation will not gain much speedup by applying sQSSA even if S_i is in quasi-steady state. (One such example would be the case where the population of S_i is much larger than the "fluctuation" of the population of S_i, so that the τ_{variance} of S_i is not really restricting the tau-leaping stepsize.) With the same argument, for multiple fast-changing species in a subnetwork, we can test their ρ's to determine whether the sQSSA can be applied to them simultaneously.

The most well-known motif of two fast-changing species connected together is a M-M type enzyme-substrate subsystem. In Ref. 13, the authors showed that for the stochastic M-M model reduction to be valid in a larger network, in addition to satisfying the validity condition $E_T \ll S_0 + K_m$, there must be sufficient time scale separation between the (faster) enzyme-substrate reaction and the other (slower) reactions in which the substrate S is involved. In tau-leaping simulation, this means that the substrate S must be changing much more slowly than the enzyme-substrate complex C. Thus, the stepsize restriction due to S must be much more relaxed than the stepsize restriction due to C: $\tau_S \gg \tau_C$. In Appendix B, we show that when $\min\{\rho_E, \rho_C\} \gg 1$, E and C are in quasi-steady state. Note that in this system, the stochastic M-M is valid if C is involved only in these three reactions and there is sufficient time scale separation between the stepsize restriction due to S and that due to C: $\tau_S \gg \tau_C$. We use these conditions as the criteria to identify opportunities for the stochastic M-M model reduction: $\min\{\rho_E, \rho_C\} \gg 1$, $\tau_S \gg \tau_C$, and C is involved only in these three reactions. The comparison between τ_S and τ_C is already done in the partitioning process, thus it adds no overhead to the identification of sQSSA opportunities.

An example analysis can be performed on enzyme-substrate system (6) with the same parameters as in Sec. III C. For tau-leaping, E will be the species associated with the greatest stepsize restriction, before most of the substrate S is consumed. The reactions linked with E are R_1, R_2, and R_3. R_1 consumes E, while R_2 and R_3 produce E. By approximating the population of all other species as constant during a "slow" step, R_1, R_2, and R_3 become either zeroth or first order. The ρ of E (as defined in (9)) will be very large since $a_1 \sim a_2 + a_3$. We can easily see that the sQSSA or the time-dependent solution method can be applied to E. On the

other hand, since $\min\{\rho_E, \rho_C\} \gg 1$ and $\tau_S \gg \tau_C$ (S participates in fewer reactions than does C yet the population of S is much larger than the population of C, thus $\tau_S \gg \tau_C$), the stochastic M-M can be applied to remove both E and C from the tau-leaping stepsize restriction.

E. Partitioning algorithm

Now that we have both the stationary network structure and the dynamic fastness comparison, we can construct the model analyzer for SSA (tau-leaping). The objective of the algorithm is to identify all of the SFRGs (SFSGs), from the fastest to the slowest, while we perform a breadth-first search of similarly fast reactions (species) to find each SFRG (SFSG) and to connect slower SFRGs (SFSGs) to faster SFRGs (SFSGs). This is all done on the reaction connection graph (species connection graph) of the system, whose nodes are weighted by a set of fastness functions of reactions (species). The analysis requires one realization of SSA (tau-leaping) simulation of the system. The analysis process can be repeated each time that the fastness comparison between any connected reaction (species) pair is changed, to capture all the dynamical structure change of the system. Alternatively, it can be repeated at different sampling time points for faster but coarser analysis. The analysis algorithm at time t is outlined as below:

1. Sort all of the reactions (species) based on the fastness functions of reactions (species) $f_i(t)$, so that $f_1(t) \geq f_2(t) \geq \cdots \geq f_m(t)$.
2. Let $A = \phi$ denote the reactions (species) already analyzed, $B = \phi$ denote the reactions (species) to be analyzed, and $C = \{1, 2, \ldots, m\}$ denote the reactions (species) not yet analyzed.
3. Choose the reaction R_i(species S_i) in C with the largest fastness function. Let $B = B \cup \{i\}$, $C = C \backslash \{i\}$. Form a new SFRG (SFSG) $Y_{l+1} = \{i\}$, if there are already l SFRGs (SFSGs). Let $Z_{l+1} = \phi$ denote the faster SFRGs (SFSGs) that are connected to this SFRG (SFSG).
4. Choose the reaction R_j(species S_j) in B with the largest fastness function. Let $A = A \cup \{j\}$, $B = B \backslash \{j\}$. Then for each reaction R_k(species S_k) connected to $R_j(S_j)$,
   (a) add connected similarly fast reactions (species) to SFRG (SFSG) Y_{l+1}: if $R_k(S_k)$ is similarly fast as $R_j(S_j)$ and not yet analyzed, let $B = B \cup \{k\}$, $C = C \backslash \{k\}$, and $Y_{l+1} = Y_{l+1} \cup \{k\}$.
   (b) record another SFRG (SFSG) Y_p that is connected to SFRG (SFSG) Y_{l+1}: if $R_k(S_k)$ is faster than $R_j(S_j)$ and already analyzed, assuming that $k \in Y_p$, then let $Z_{l+1} = Z_{l+1} \cup \{Y_p\}$.
   (c) record SFRG (SFSG) Y_{l+1} as a faster SFRG (SFSG) connected to a slower SFRG (SFSG) Y_q: if $R_k(S_k)$ is slower than $R_j(S_j)$ and already analyzed, assuming that $k \in Y_q$, then let $Z_q = Z_q \cup \{Y_{l+1}\}$.
   (d) otherwise do nothing.
5. Repeat 4 while $B \neq \phi$.
6. Repeat 3 while $C \neq \phi$.

The algorithm categorizes all the reactions (species) into different SFRGs (SFSGs) $\{Y_l\}$. It also records Z_l for each Y_l denoting faster SFRGs (SFSGs) connected to Y_l. Then we process the SFRGs (SFSGs) from the fastest to the slowest, to identify virtual fast processes (fast-changing species groups) to which the ssSSA (sQSSA or stochastic M-M) is applicable and beneficial: for each SFRG (SFSG) Y_l, we recursively include connected faster SFRGs (SFSGs) into the virtual fast process (fast-changing species groups) induced by Y_l. The validity conditions for these model reductions are given in Subsections III C and III D, while the beneficial conditions examine whether the model reductions speed up the simulation. For SSA, it is beneficial to apply the ssSSA to a virtual fast process Y_l only if Y_l is responsible for more reaction events than all the slow reactions combined. For tau-leaping, it is beneficial to apply the sQSSA or stochastic M-M to species in a fast-changing species group Y_l only if the stepsize τ increases significantly (for example, by 10 times) after removing the species in Y_l from the stepsize calculation.

For example, for the enzyme-substrate system (6) with the same parameters as in Sec. III C, the model analyzer for SSA partitions the reactions into $Y_1 = \{R_1, R_2\}$ and $Y_2 = \{R_3\}$, while the connected faster SFRGs are $Z_1 = \phi$ and $Z_2 = \{Y_1\}$, respectively. We have shown in Sec. III C that the ssSSA can be applied to Y_1. On the other hand, the model analyzer for tau-leaping partitions the species into $Y_1 = \{E\}$, $Y_2 = \{C\}$, $Y_3 = \{S\}$, and $Y_4 = \{P\}$ (since P does not participate in the stepsize calculation of tau-leaping, $\tau_P = \infty$), while the connected faster SFSGs are $Z_1 = \phi$, $Z_2 = \{Y_1\}$, $Z_3 = \{Y_1, Y_2\}$, and $Z_4 = \{Y_1, Y_2\}$, respectively. By comparing the induced sub-Petri net to known motifs, we can see that the sQSSA can remove Y_1 from the system (and hence the stepsize restriction). Alternatively, the stochastic M-M approximation can remove Y_1 and Y_2 simultaneously from the system (and hence the stepsize restriction), if $\{Y_2\} \cup Z_2$ is also identified as a fast-changing species group.

IV. NUMERICAL EXAMPLES

Here, we present several examples that demonstrate the applicability and power of our approach. The model analyzers were developed in the framework of StochKit2.0.[56] The Graphviz drawing library[57] was used for visualization of Petri nets. The simulations were performed on an Intel i7-2600 Linux workstation with 8GB RAM.

A. Analyzing heat-shock response model for SSA

We first applied the model analyzer for SSA to the HSR model of *E. Coli*.[28,29] The HSR model describes the mechanism of how *E. Coli* responds to stress. In *E. Coli*, RNA polymerase (RNAP) binds to sigma factor σ_{70} to transcribe genes necessary for growth at normal temperature. When *E. Coli* is exposed to high temperatures or other stresses that cause protein to unfold, sigma factor σ_{32} is rapidly induced and binds to RNAP. The resultant complex σ_{32}:RNAP initiates the transcription of genes that encode a variety of chaperone enzymes. These chaperones treat the unfolded proteins, either

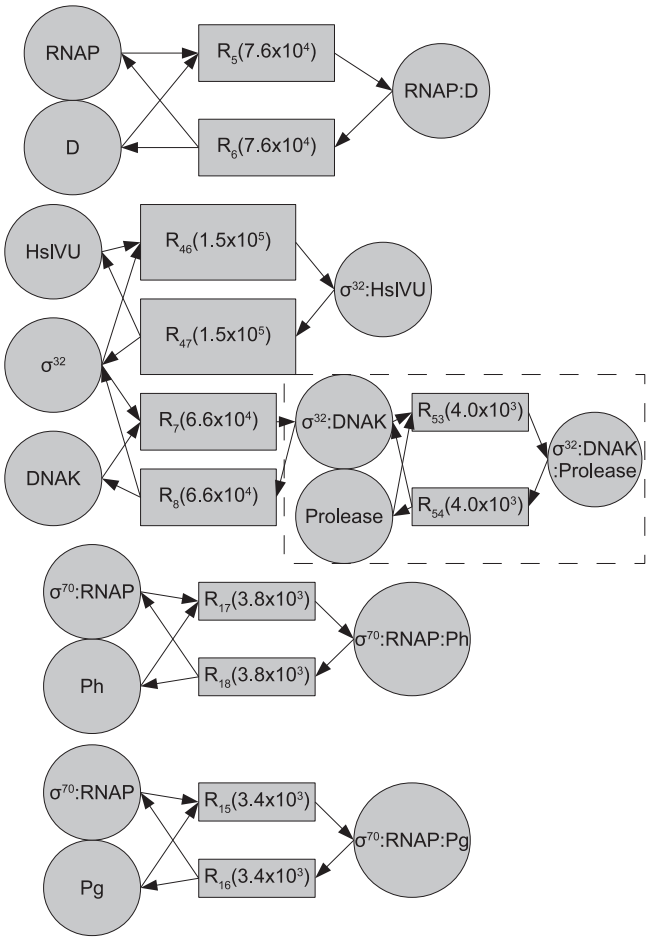


FIG. 3. The four virtual fast processes of the HSR system, with reaction nodes' weights in parentheses in the reactions boxes at $t = 100$. The reaction nodes' weights are the actual reaction counts of each reaction from $t = 98.4$ to $t = 104.2$ in one SSA simulation.

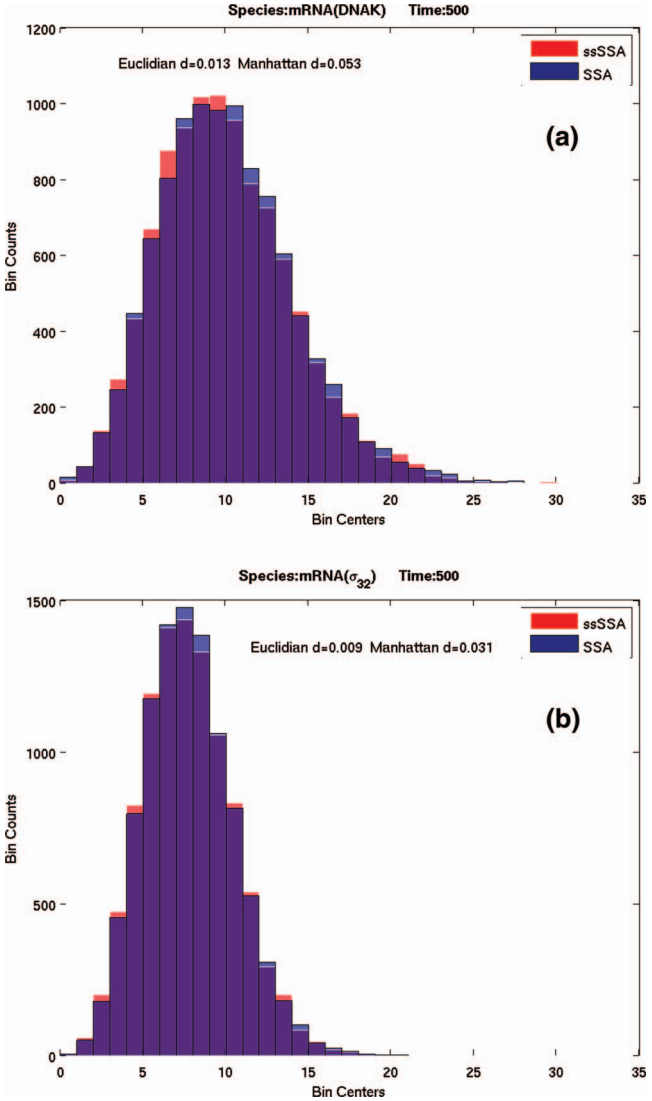


FIG. 4. Comparison of histograms (10 000 samples) of two slow species mRNA (DNAK) and mRNA (σ_{32}) at $t = 500$ solved by the original SSA and the ssSSA. The Euclidian distance and Manhattan distance in the figures are respectively L2 norm and L1 norm of the histogram distance.

by refolding them or by degrading them so that they will not aggregate. On the other hand, one of the chaperone enzymes, DNAK, is more likely to sequester a σ_{32} when there is not enough unfolded protein to bind. This precludes σ_{32}'s binding to RNAP and down-regulates the heat shock response. The details of the deterministic model for the HSR system can be found in Ref. 28, and a stochastic version was discussed in Ref. 29. We use the same stochastic model as in Ref. 37, with 28 species participating in 61 chemical reactions.

Our proof of principal implementation of the model reduction works as follows. Based on the results of one SSA simulation, with model analysis applied after every 61 000 reaction firings (so that on average each reaction fires 1000 times), we were able to identify four virtual fast processes for the ssSSA (as shown in Fig. 3), which we applied for the entire simulation time (from $t = 0$ to $t = 500$), as the model analysis results suggested. Note that these are exactly the 12 fastest reactions hand-picked in Ref. 54, except that here they have been automatically selected, verified, and divided into four virtual fast processes. The ssSSA requires means of solving for the partial equilibrium states of the virtual fast processes. We did this by solving a closed form solution for each virtual fast process. We simulated a total of 10 000 samples of the HSR model using the ssSSA. Then we simulated a total of 10 000 samples of the HSR model using the SSA. The average SSA simulation took 30 s and the average ssSSA simulation took only 0.40 s. Figure 4 shows the accuracy of the ssSSA approximation. We note that, up to this time, using the ssSSA in StochKit2.0 has required quite a bit of work in the form of providing a code to solve for the partial equilibrium. This is about to change, and will be handled automatically in a new, innovative implementation due to Sanft *et al.*[58]

B. Analyzing the coagulation model for tau-leaping

We applied the model analyzer for tau-leaping to a blood coagulation model.[30] The coagulation model describes the extrinsic blood coagulation system including pro- and anti-coagulants. In this model, one of the most important coagulation factors, thrombin (factor IIa), is formed through a cascade of reactions which are initiated when tissue factor (TF) is exposed to blood due to vessel injury: TF activates factor

VII; then the complex TF_VIIa activates factor X and factor IX; activated factor Xa in turn activates factor V, factor VIII, and factor IX; the complex factor Xa_Va, as well as activated Xa, activates prothrombin II to thrombin IIa. In the meantime, there are always anti-coagulants in the blood down-regulating activated pro-coagulants, maintaining the fluidity of the blood. Anti-coagulants antithrombin-III (ATIII) and tissue factor pathway inhibitor (TFPI) are included in the model. This model, with proper initial conditions, can be used to describe the initiation and propagation of the extrinsic blood coagulation pathway.[30] The detailed model can be found in Ref. 30, and involves 34 species and 43 reactions.

We used the same initial conditions as in Ref. 30, and set the initial concentration of TF to 2.5×10^{-11} mol/L (one of the initial concentration choices of TF in Ref. 30). We simulated the system until $t = 700$, sampled the system every 10 000 tau-leaping steps, and chose the species node weights $f_i(S)$ (fastness functions of species) to be the average of τ_i^{-1} for species S_i, during the time since the previous sampling point. We set the distinguishing factor $\lambda = 5$ to define slow and fast between connected species. Note that $\max_{i,j}\{|\nu_{ij}|\} = 1$ in this model. Thus, the ρ's as defined in Eq. (9) were compared to 1 to identify sQSSA and stochastic M-M opportunities. We were able to dynamically identify multiple motifs in the system to which model reductions can be applied (the identification of time-dependent solution approach opportunities is introduced in Ref. 14):

1. From $t = 0$ until $t \approx 30$ (the first 10 000 tau-leaping steps), species TF_VIIa is the fastest changing species and restricts the simulation efficiency. By examining the average ρ of TF_VIIa during this period, we found that $\rho \approx 67$. Thus, the sQSSA can be applied to species TF_VIIa during this time period. This is because in the very beginning the population of TF_VIIa is still small, but it participates in 12 reactions, mostly serving as the enzyme of enzyme-substrate systems. It is indeed in quasi-steady state.
2. From $t \approx 30$ to $t \approx 180$, species Xa_Va and Xa_Va_II form the fastest changing species group. The sub-Petri net induced by these two species is shown in the dashed box in Fig. 5. We can clearly see that part of the subsystem is a M-M type enzyme-substrate system. By examining the ρ's of Xa_Va and Xa_Va_II, we found that the sQSSA can be applied to them. Since R_{29}, R_{30}, and R_{31} matches the motif of a M-M type subsystem, the stochastic M-M can be applied to them first. Alternatively, the time-dependent solution approach can be applied to these two species. Since the next fastest changing species has a stepsize restriction of about 20 times greater than the stepsize restriction of Xa_Va and Xa_Va_II, either approximation can speed up the simulation by a large amount.
3. From $t \approx 180$ to $t \approx 280$, species Xa_Va, Xa_Va_II, Xa, IXa_VIIIa, and IXa_VIIIa_X form the fastest changing species group. The sub-Petri net induced by these five species is shown in Fig. 5. This is a cascade of two enzyme-substrate systems linked by a reversible reaction pair. By examining the ρ's, we found all of the 5 species to be in stochastic quasi-steady state. Also, both enzyme-substrate subsystems match the stochastic M-M motif. Thus, either the sQSSA or the stochastic M-M can be applied.
4. From $t \approx 280$ to $t \approx 430$, species IXa_VIIIa and IXa_VIIIa_X form the fastest changing species group, and species Xa_Va_II forms the second fastest changing species group. The sub-Petri net induced by these three species is shown in Fig. 5, where the enzyme-substrate system induced by IXa_VIIIa and IXa_VIIIa_X is circled by the dotted box. What happens is that factor Va and Xa are produced (activated) rapidly and so is their complex Xa_Va. Thus, the populations of Va, Xa, and Xa_Va become much larger than the populations of IXa_VIIIa, IXa_VIIIa_X, and Xa_Va_II. Although the populations of Va, Xa, and Xa_Va still change as rapidly as the populations of IXa_VIIIa, IXa_VIIIa_X, and Xa_Va_II, the changes are relatively slow compared to the large populations of Va, Xa, and Xa_Va. The stochastic M-M approximation can be applied to IXa_VIIIa and IXa_VIIIa_X, while the sQSSA can be applied to Xa_Va_II. For a more accurate result, the

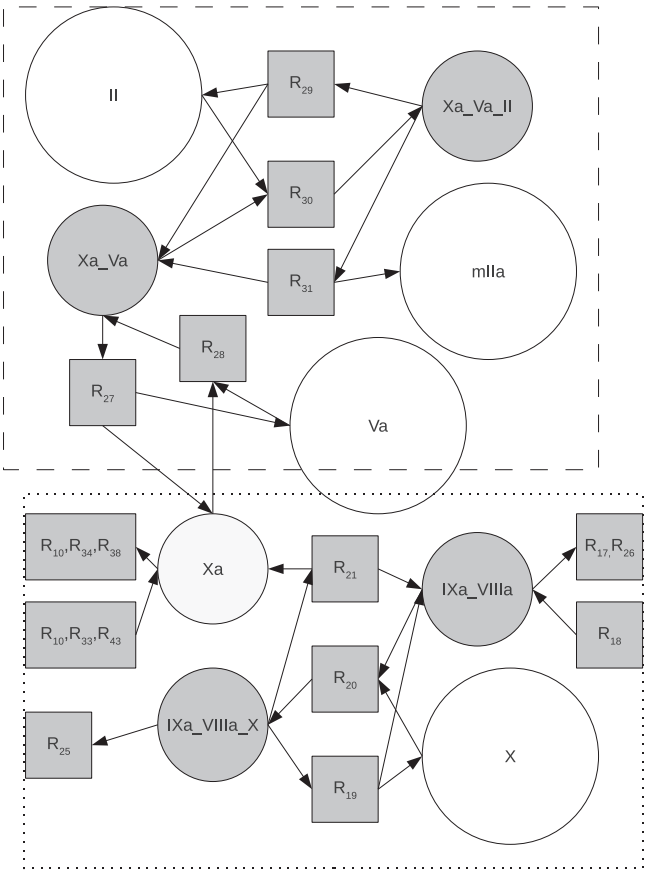


FIG. 5. The sub-Petri net induced by Xa_Va, Xa_Va_II, Xa, IXa_VIIIa, and IXa_VIIIa_X. Note that species II, mIIa, Va, and X have large populations (denoted by large circles) and are all slowly changing species. Other associated slowly changing species are not drawn because they can be safely approximated as constant during one tau-leaping step. The sub-Petri net circled by the dashed line is the enzyme-substrate system induced by Xa_Va and Xa_Va_II; the sub-Petri net circled by the dotted line is the enzyme-substrate system induced by IXa_VIIIa and IXa_VIIIa_X.

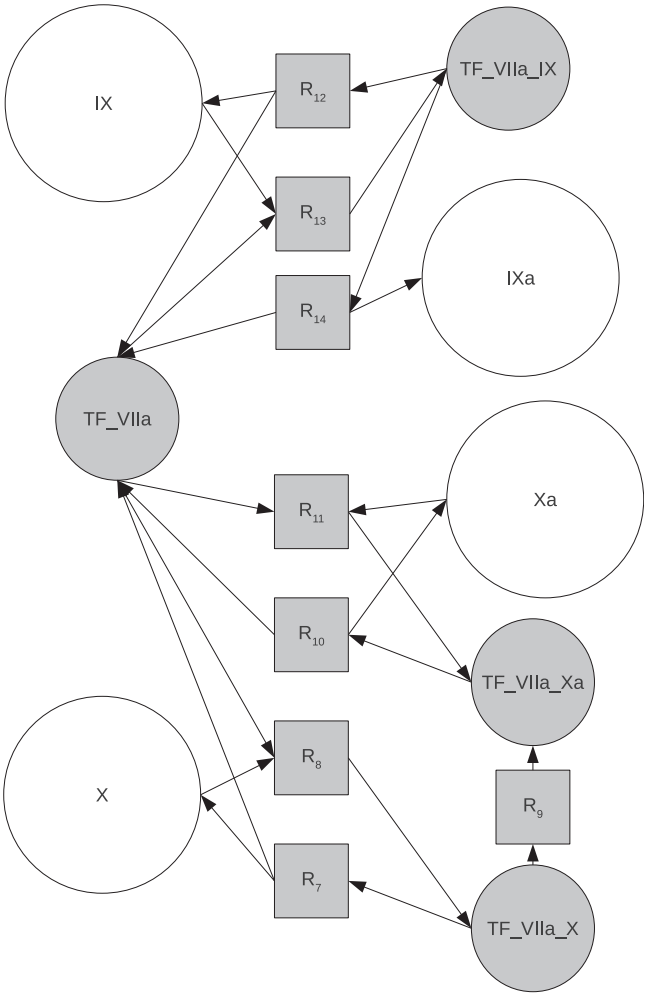


FIG. 6. The sub-Petri net induced by TF_VIIa, TF_VIIa_X, TF_VIIa_Xa, and TF_VIIa_IX. Note that X, Xa, IX, and IXa are slowly changing species.

time-dependent solution approach can be applied to both groups.

5. From $t \approx 430$ to $t \approx 700$, species IXa_VIIIa and IXa_VIIIa_X are still changing the fastest, while the second fastest changing species group becomes TF_VIIa, TF_VIIa_X, TF_VIIa_Xa, and TF_VIIa_IX. The sub-Petri net induced by TF_VIIa, TF_VIIa_X, TF_VIIa_Xa, and TF_VIIa_IX is shown in Fig. 6. We can see that TF_VIIa, TF_VIIa_X, TF_VIIa_Xa, and TF_VIIa_IX belong to two enzyme-substrate systems with the same enzyme. Note that the enzyme-substrate system activating factor X to Xa does not match the motif of a M-M type enzyme-substrate system (the reaction network structure is different). Thus, the algorithm recognizes only the enzyme-substrate system activating factor IX to IXa as stochastic M-M applicable.

After having identified model reduction opportunities automatically for a single tau-leaping simulation with model analysis applied after every 10 000 tau-leaping steps, we applied these model reductions to the ensemble simulation at the indicated time periods. For simplicity purposes, we applied only the stochastic M-M reductions and compared the results. From the model analysis results, we know there are a total of four enzyme-substrate systems for which stochastic M-M can be applied to and beneficial:

$$\mathrm{II} \xrightarrow{\mathrm{X_a\_Va}} \mathrm{mIIa}, \tag{10}$$

$$\mathrm{X} \xrightarrow{\mathrm{IX_a\_VIIIa}} \mathrm{Xa}, \tag{11}$$

$$\mathrm{IX} \xrightarrow{\mathrm{TF\_VIIa}} \mathrm{IXa}, \tag{12}$$

$$\mathrm{X} \xrightarrow{\mathrm{TF\_VIIa}} \mathrm{Xa}. \tag{13}$$

To show that it is important to apply model reductions dynamically (only when it is valid), first we applied stochastic M-M reductions at indicated time periods and called it the dynamically reduced model; that is, we applied stochastic M-M to (10) from $t = 30$ to $t = 430$, to (11) from $t = 180$ to $t = 700$, to (12) and (13) from $t = 430$ to $t = 700$. We simulated a total of 10 000 samples of the dynamically reduced model. Then we applied stochastic M-M to all four enzyme-substrate systems throughout the entire simulation time (from $t = 0$ to $t = 700$) and called it the fully reduced model. We simulated a total of 10 000 samples of the fully reduced model as well. Finally, we simulated a total of 10 000 samples of the original full model and compared simulation results. The average full model simulation took 17 s, the average fully reduced model simulation took 0.22 s, and the average dynamically reduced model simulation took 0.93 s. Figure 7 shows the accuracy of the fully reduced model and the dynamically reduced model, compared with the full model. It is easy to see that the dynamically reduced model is much more accurate than the fully reduced model.

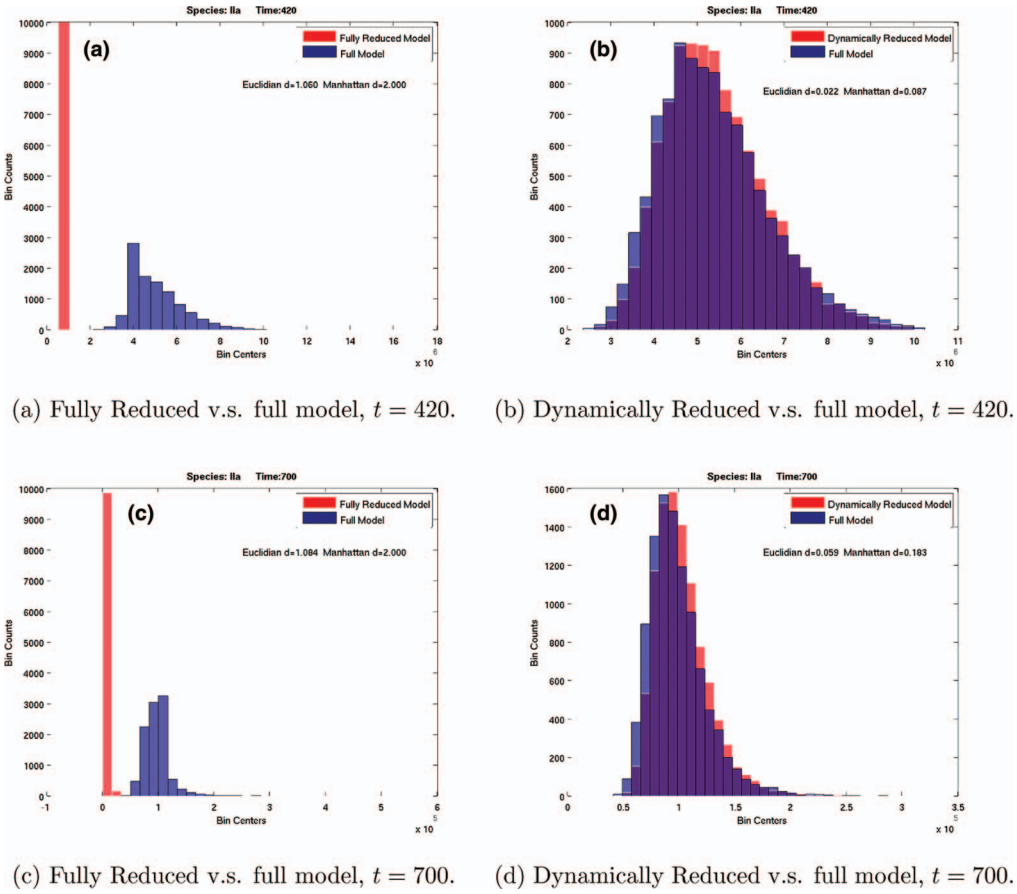
C. Model analyzer software

Our model analyzer StochMA (StochKit Model Analyzer) takes a model file in StochKit2.0 XML (Ref. 56) format (StochKit2.0 includes a converter from SBML to this format), simulates a realization of the model using SSA or tau-leaping, analyzes the model dynamically, and outputs the model reduction suggestions at different time points, in a Petri net graph format. The output examples are shown in Fig. 8–11. The results agree with our analysis in Subsections IV A and IV B.

The examples above demonstrate that the weighted Petri net based model analyzer for SSA or tau-leaping is able to dynamically identify the fastest virtual fast processes (SSA) or the fastest changing species groups (tau-leaping) to which model reductions are applicable and beneficial, with no expert knowledge input. It provides easy access to the dynamic information of an unfamiliar chemically reacting system.

V. SOME THOUGHTS ON THE SLOW-SCALE TAU-LEAPING METHOD

The slow-scale tau-leaping method[15] approximates a chemically reacting system in a manner similar to the ssSSA.[8]



(a) Fully Reduced v.s. full model, $t = 420$. (b) Dynamically Reduced v.s. full model, $t = 420$.

(c) Fully Reduced v.s. full model, $t = 700$. (d) Dynamically Reduced v.s. full model, $t = 700$.

FIG. 7. Comparison of histograms (10 000 samples) of thrombin (IIa) at different times, simulated by the fully reduced model, the dynamically reduced model, and the original full model. (a) Fully reduced vs. full model, $t = 420$. (b) Dynamically reduced vs. full model, $t = 420$. (c) Fully reduced vs. full model, $t = 700$. (d) Dynamically reduced vs. full model, $t = 700$. The Euclidian distance and Manhattan distance in the figures are respectively L2 norm and L1 norm of the histogram distance.

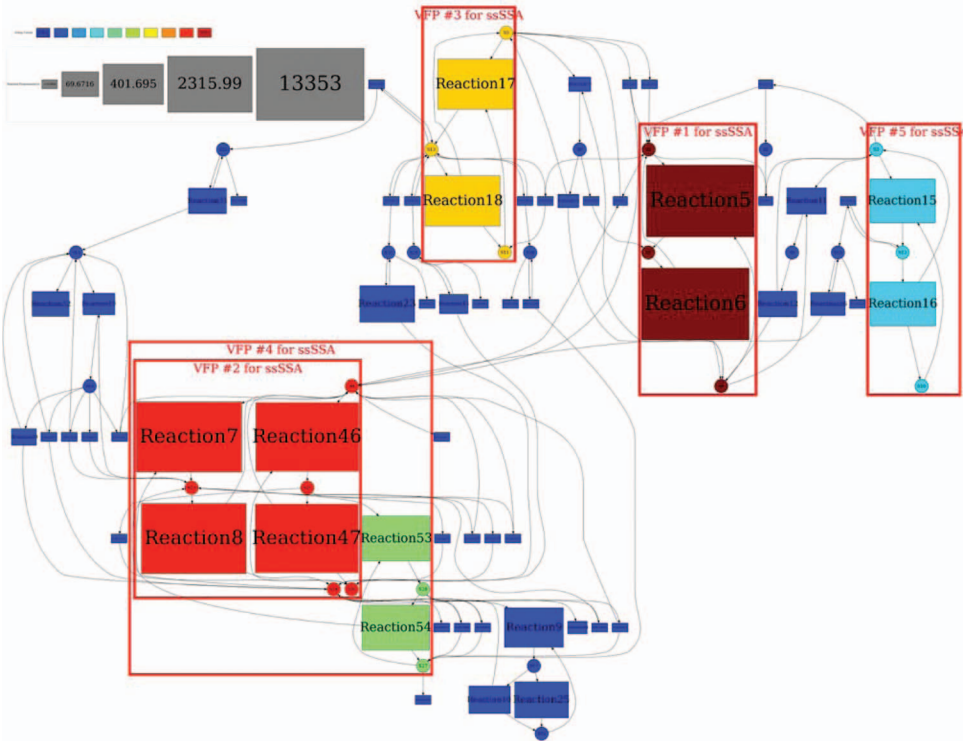


FIG. 8. The model reduction suggestions for the HSR model for SSA at t = 0. The sizes of reaction nodes are proportional to the log scale of their firing frequencies. The virtual fast processes to which ssSSA is applicable and beneficial are enclosed in red rectangles. Different virtual fast processes are depicted in different colors.

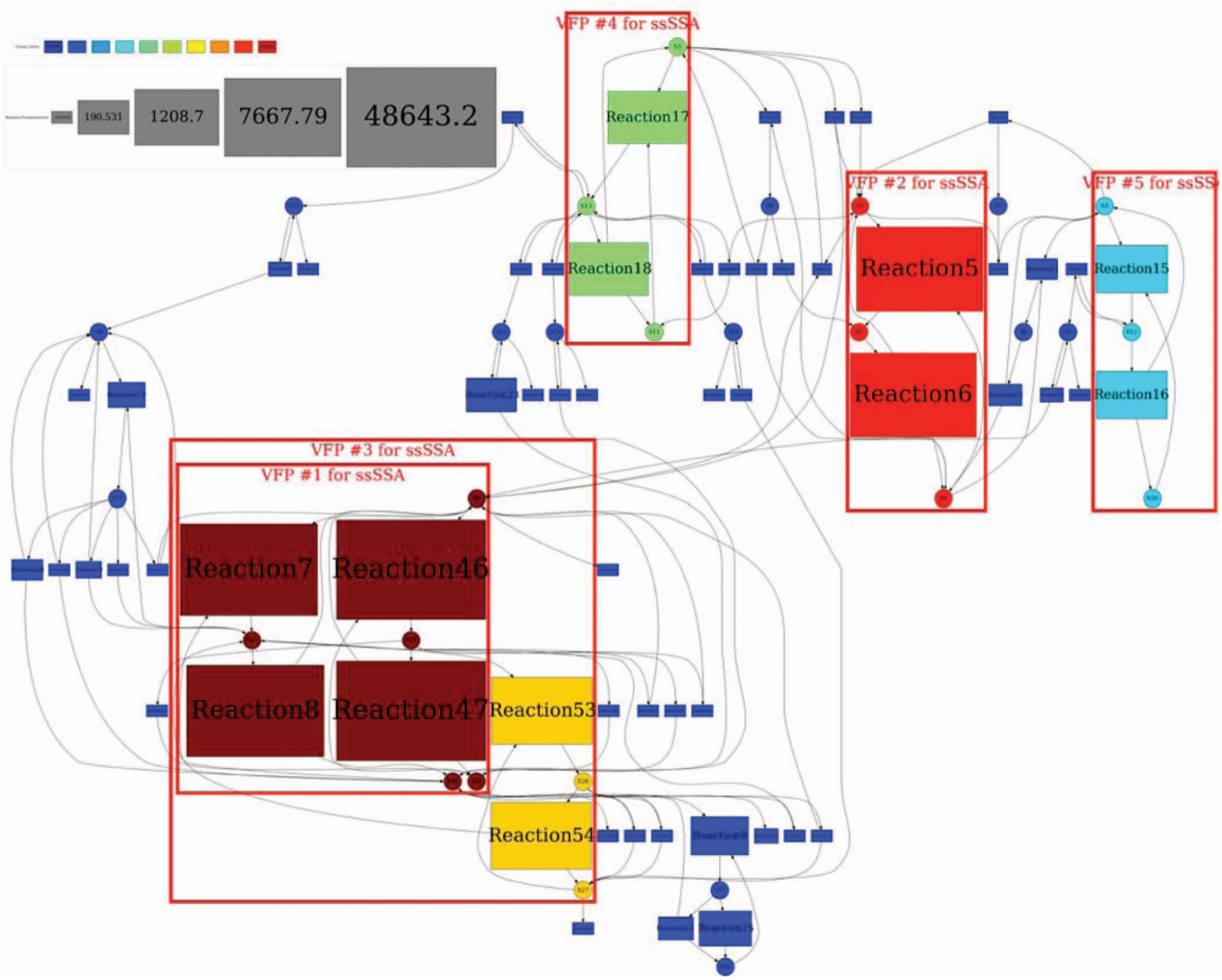


FIG. 9. The model reduction suggestions for the HSR model for SSA at t = 300. The sizes of reaction nodes are proportional to the log scale of their firing frequencies. The virtual fast processes to which ssSSA is applicable and beneficial are enclosed in red rectangles. Different virtual fast processes are depicted in different colors.

The method looks for the virtual fast process that contains only fast reactions and will reach stochastic partial equilibrium rapidly, before the next slow reaction fires. Thus, the criteria to apply both the ssSSA and the slow-scale tau-leaping are the same. However, while the ssSSA can always speed up SSA simulation by a considerable amount, in some cases the slow-scale tau-leaping does not benefit tau-leaping simulation nearly as much. Also there are cases, such as for the coagulation model, tau-leaping simulation can benefit a lot from model reductions such as stochastic M-M reductions but slow-scale tau-leaping does not apply. The discrepancy of model reduction benefits for the SSA and tau-leaping simulation has been discussed separately in Refs. 59 and 43. Here, we suggest a systematic alternative to the slow-scale tau-leaping method that better fits the dynamic features of tau-leaping simulation.

Consider an example extracted from the coagulation model. While applying the SSA model analyzer to the coagulation model, we found that from $t \approx 0$ to $t \approx 300$, the fastest reaction group contains the following 4 reactions:

$$\begin{aligned} &\mathrm{Xa\_Va} + \mathrm{II} \underset{R_{29}}{\overset{R_{30}}{\rightleftharpoons}} \mathrm{Xa\_Va\_II} \xrightarrow{R_{31}} \mathrm{Xa\_Va} + \mathrm{mIIa}, \\ &\mathrm{mIIa} + \mathrm{ATIII} \xrightarrow{R_{39}} \mathrm{mIIa\_ATIII}. \end{aligned} \tag{14}$$

The sub-Petri net induced by the 4 reactions is shown in Fig. 12. It is not hard to see that it contains the same enzyme-substrate system that also appears in the dashed box in Fig. 5. Since the sub-Petri net does not satisfy the necessary condition to apply ssSSA, this 4-reaction group will not be able to reach stochastic partial equilibrium. This is obvious because species II and ATIII are constantly consumed while species mIIa_ATIII are constantly produced. Thus, the slow-scale approximation will not be applicable to this 4-reaction group. On the other hand, the sQSSA or the stochastic M-M can be applied to the enzyme-substrate system and will substantially speed up tau-leaping simulation, as shown in Sec. IV B. The reason is that there is no time scale separation in the fastness

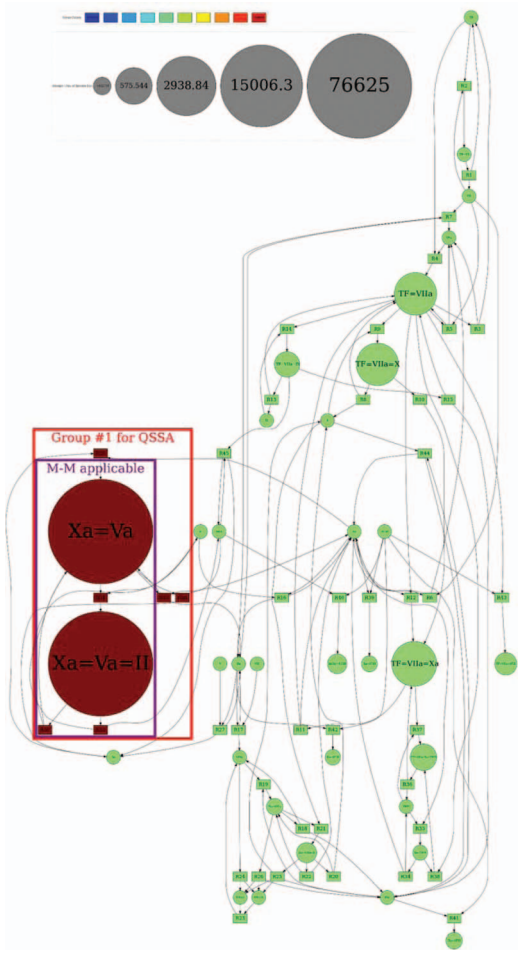


FIG. 10. The model reduction suggestions for the coagulation model for tau-leaping at t = 50. The sizes of species nodes S_i are proportional to $\log(1/\tau_i)$ (the larger the node, the smaller the stepsize). The fast-changing species groups to which sQSSA is applicable and beneficial are enclosed in red rectangles. The enzyme-substrate systems to which stochastic M-M is applicable and beneficial are enclosed in purple rectangles. Different fast-changing species groups are depicted in different colors.

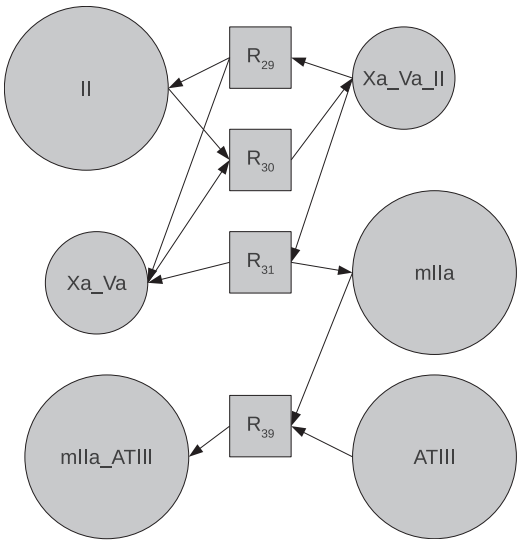


FIG. 12. The sub-Petri net induced by the fastest reaction group from $t \approx 0$ to $t \approx 300$ for the coagulation model.

of reactions, however, there is a significant time scale separation in the relative rates of change of species' populations.

Note that for the ssSSA or slow-scale tau-leaping to be valid, the fast subsystem must be able to reach stochastic partial equilibrium on the time scale of the slow part of the system, so that the propensities of the neighboring slow reactions can be accurately approximated by their expectations over the partial equilibrium distribution of the fast subsystem. Now suppose that there is a slow reaction that has mIIa_ATIII as a reactant. Then we can see how this condition will fail: mIIa_ATIII will not reach partial equilibrium because the population will always increase.

We inspected this system closely, and found that the populations of species II, mIIa, ATIII, and mIIa_ATIII were large. Thus, although all four of the reactions are fast, they do not

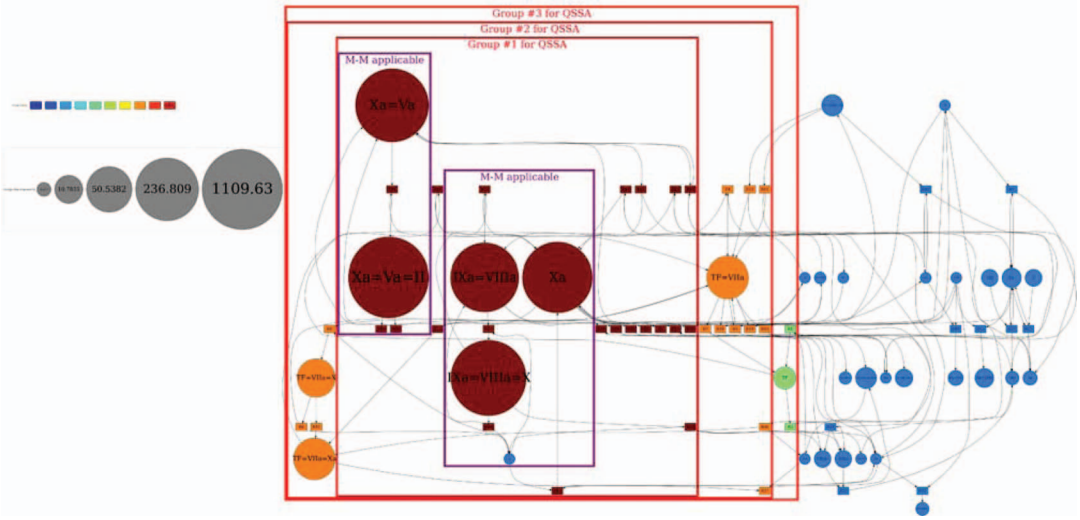


FIG. 11. The model reduction suggestions for the coagulation model for tau-leaping at t = 200. The sizes of species nodes S_i are proportional to $\log(1/\tau_i)$ (the larger the node, the smaller the stepsize). The fast-changing species groups to which sQSSA is applicable and beneficial are enclosed in red rectangles. The enzyme-substrate systems to which stochastic M-M is applicable and beneficial are enclosed in purple rectangles. Different fast-changing species groups are depicted in different colors.

result in a rapid change of the relative populations of these species. Recall that there is another piece of information that the ssSSA does not make use of: the populations of species or the values of propensities of reactions. What is relevant for the ssSSA is the absolute change of the species' populations, or the absolute change of the propensities of reactions. This fits very well with SSA since this is exactly what SSA computes. On the other hand, tau-leaping is concerned more with the relative change of both. Thus, the slow-scale approximation is sometimes not quite a good fit for tau-leaping. On the contrary, the sQSSA focuses more on the relative change of species' populations. It is applicable when the expectation of a species' population changes much more slowly than the actual change of that species' population.

In this particular example, suppose that there is a slow reaction that has mIIa_ATIII as a reactant. Because the population of mIIa_ATIII is always increasing, we cannot find a good approximation to the propensity of the slow reaction unless we know that the population of mIIa_ATIII is large and the relative change of the population of mIIa_ATIII between two slow reactions is very small. In fact, tau-leaping is already efficient in this situation, which is why mIIa_ATIII or mIIa were not identified as fast-changing species that are constraining the stepsize when we applied the model analyzer for tau-leaping to the coagulation model. What restricts the efficiency of tau-leaping simulation is species with small populations that participate in fast reactions, such as Xa_Va and Xa_Va_II. The sQSSA or the time-dependent solution approximation can yield substantial speed-ups in this situation.

As a result, we suggest an alternative to the slow-scale tau-leaping method. Instead of looking for fast reactions first and defining all the species associated with fast reactions as fast species, we look for fast-changing species first and define all the reactions associated with fast-changing species as fast-changing reactions. Instead of looking for stochastic partial equilibrium of the fast subsystem, we look for fast-changing species that are in quasi-steady state. Then we apply the sQSSA to the actual fast scale restriction in tau-leaping: fast-changing species.

VI. CONCLUSIONS

We have proposed automatic model analyzers for SSA and tau-leaping simulation. The automatic model analyzers can dynamically identify situations where model reductions can be safely and efficiently applied, with no expert knowledge input. We have demonstrated the effectiveness of the model analyzers to identify ssSSA opportunities for SSA simulation, as well as to identify sQSSA and stochastic M-M opportunities for tau-leaping simulation.

Besides these applications, the weighted reaction connection graph and species connection graph are powerful graphical tools that can be extended to identify other model reduction opportunities or interesting dynamic motifs in simulation. Automatic model reductions in simulation are also possible using this graphical tool. One shortcoming of the current model analyzer is that, since it bases its analysis on time snapshots from one SSA or tau-leaping simulation, it may not capture motifs that do not become active during that realization of the simulation. We are currently pursuing the implementation of a dynamic watcher of system dynamics which, coupled with dynamic instantiation of the model reduction, could achieve automatic model reduction for both stochastic simulation and deterministic simulation of chemically reacting systems.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the U.S. Department of Energy (DOE) through Grant No. DEFG02-04ER25621, the U.S. Army Research Office (USARO) through Grant No. W911NF, and the Institute for Collaborative Biotechnologies (ICB) through Grant No. W911NF-09-0001 from the U.S. Army Research Office.

APPENDIX A: CRITERIA TO IDENTIFY SQSSA OPPORTUNITIES IN TAU-LEAPING

Define ρ_i of species S_i as the ratio of τ_{mean} to τ_{variance} of S_i in tau-leaping. In other words, ρ_i is the ratio of the stepsize restriction due to the expected mean of the change of S_i to the stepsize restriction due to the expected variance of the change of S_i

$$\rho_i = \frac{\tau_{\text{mean}}}{\tau_{\text{variance}}} = \frac{\max\{\varepsilon x_i/g_i, 1\}/\left|\sum_j \nu_{ij} a_j(x)\right|}{\max\{\varepsilon x_i/g_i, 1\}^2/\left|\sum_j \nu_{ij}^2 a_j(x)\right|}. \tag{A1}$$

Claim: If $\rho_i \gg \max_j|\nu_{ij}|$, then species S_i is in quasi-steady state.

Proof: Let $R^{i+} = \{R_j|\nu_{ij} > 0\}$ denote all the reactions that produce S_i, and $R^{i-} = \{R_j|\nu_{ij} < 0\}$ denote all the reactions that consume S_i. Then

$$\left|\sum_j \nu_{ij} a_j(x)\right| = \left|\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) - \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)\right|, \tag{A2}$$

$$\left|\sum_j \nu_{ij}^2 a_j(x)\right| \le \max_j|\nu_{ij}|\left|\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) + \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)\right|. \tag{A3}$$

Thus,

$$\begin{aligned}\rho_i &= \frac{1}{\max\{\varepsilon x_i/g_i, 1\}}\frac{\left|\sum_j \nu_{ij}^2 a_j(x)\right|}{\left|\sum_j \nu_{ij} a_j(x)\right|} \\ &\le \frac{\max_j|\nu_{ij}|}{\max\{\varepsilon x_i/g_i, 1\}}\frac{\left|\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) + \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)\right|}{\left|\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) - \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)\right|} \\ &\le \max_j|\nu_{ij}|\frac{\left|\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) + \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)\right|}{\left|\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) - \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)\right|}.\end{aligned} \tag{A4}$$

Thus,

$$\begin{aligned}
&\rho_i \gg \max_j |\nu_{ij}| \\
&\Rightarrow \frac{\left|\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) + \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)\right|}{\left|\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) - \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)\right|} \gg 1 \\
&\Rightarrow \left|1 - \frac{2\sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)}{\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) + \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)}\right| \ll 1 \\
&\Rightarrow \frac{2\sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)}{\sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) + \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x)} \approx 1 \\
&\Rightarrow \sum_{R_j\in R^{i+}} |\nu_{ij}|a_j(x) \approx \sum_{R_j\in R^{i-}} |\nu_{ij}|a_j(x). \qquad \text{(A5)}
\end{aligned}$$

The production rate and consumption rate of S_i are almost equal. Thus, S_i is in quasi-steady state. □

APPENDIX B: CRITERIA TO IDENTIFY STOCHASTIC M-M OPPORTUNITIES IN TAU-LEAPING

Claim: For the enzyme-substrate system

$$E + S \underset{c_2}{\overset{c_1}{\rightleftharpoons}} C \xrightarrow{c_3} E + P, \qquad \text{(B1)}$$

independent of whether it is part of a larger network, the stochastic M-M can be applied if all of the following conditions are met:

1. $\min\{\rho_E, \rho_C\} \gg 1$;
2. $\tau_S \gg \tau_C$;
3. *C is involved only in these three reactions.*

Proof: As suggested in Ref. 13, the stochastic M-M can be applied to the enzyme-substrate system when $E_T \ll S_0 + K_m$. When it is part of a larger network, there is the additional condition of sufficient time scale separation between the (faster) enzyme-substrate system and the other (slower) reactions that the substrate S is involved in. We have already shown in Sec. III D that the additional condition is satisfied when $\tau_S \gg \tau_C$. Here, we need only to show that $E_T \ll S_0 + K_m$ also holds.

Let x_S, x_E, and x_C denote the populations of species S, E, and C, respectively. Then by Appendix A,

$$\begin{aligned}
&\rho_C \gg 1 \\
&\Rightarrow\ a_1(x) \approx a_2(x) + a_3(x) \\
&\Rightarrow\ c_1 x_S x_E \approx (c_2 + c_3)x_C. \qquad \text{(B2)}
\end{aligned}$$

Note that $S_0 = x_S$, $K_m = (c_2 + c_3)/c_1$, and $E_T = x_E + x_C$. Thus,

$$K_m \approx x_S \frac{E_T - x_C}{x_C}, \qquad \text{(B3)}$$

$$S_0 + K_m \approx \frac{x_S E_T}{x_C}. \qquad \text{(B4)}$$

If $x_S \gg x_C$, then $E_T \ll S_0 + K_m$. Thus, we must show $x_S \gg x_C$.

Note that C is involved in only these three reactions. As $\rho_C \gg 1$, recall that g_C is the highest order of reaction in which species C appears as a reactant. Thus, we have

$$\tau_C = \frac{\max\{\varepsilon x_C/g_C, 1\}^2}{a_1(x) + a_2(x) + a_3(x)} = \frac{\max\{\varepsilon x_C, 1\}^2}{a_1(x) + a_2(x) + a_3(x)}. \qquad \text{(B5)}$$

Let $a_{other}(x) = \sum \nu_{ij}^2 a_j(x)$, where the sum is over all the other reactions that S is involved in. Note $a_1(x) \approx a_2(x) + a_3(x)$. Then,

$$\begin{aligned}
\tau_S &\le \frac{\max\{\varepsilon x_S/g_S, 1\}^2}{a_1(x) + a_2(x) + a_{other}(x)} \\
&\le \frac{2\max\{\varepsilon x_S/g_S, 1\}^2}{a_1(x) + a_2(x) + a_3(x) + a_{other}(x)}. \qquad \text{(B6)}
\end{aligned}$$

Thus,

$$\begin{aligned}
&\tau_S \gg \tau_C \\
&\Rightarrow \frac{2\max\{\varepsilon x_S/g_S, 1\}^2}{a_1(x) + a_2(x) + a_3(x) + a_{other}(x)} \\
&\gg \frac{\max\{\varepsilon x_C, 1\}^2}{a_1(x) + a_2(x) + a_3(x)} \\
&\Rightarrow \max\{\varepsilon x_S/g_S, 1\} \gg \max\{\varepsilon x_C, 1\} \\
&\Rightarrow \begin{cases} \varepsilon x_S/g_S \gg 1 \\ \varepsilon x_S/g_S \gg \varepsilon x_C \end{cases} \\
&\Rightarrow\ x_S \gg g_S x_C \ge x_C. \qquad \text{(B7)}
\end{aligned}$$

The conclusion follows. □

[1] H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. U.S.A. **94**, 814 (1997).
[2] A. Arkin, J. Ross, and H. McAdams, Genetics **149**, 1633 (1998).
[3] H. McAdams and A. Arkin, Trends Genet. **15**, 65 (1999).
[4] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[5] D. T. Gillespie, J. Chem. Phys. **115**, 1716 (2001).
[6] Y. Cao, D. T. Gillespie, and L. R. Petzold, J. Chem. Phys. **124**, 044109 (2006).
[7] Y. Cao, D. T. Gillespie, and L. R. Petzold, J. Chem. Phys. **126**, 224101 (2007).
[8] Y. Cao, D. T. Gillespie, and L. R. Petzold, J. Chem. Phys. **122**, 014116 (2005).
[9] C. Rao and A. P. Arkin, J. Chem. Phys. **118**, 4999 (2003).
[10] E. A. Mastny, E. L. Haseltine, and J. B. Rawlings, J. Chem. Phys. **127**, 094106 (2007).
[11] Y. Pu, L. T. Watson, and Y. Cao, J. Chem. Phys. **134**, 054105 (2011).
[12] D. T. Gillespie, Y. Cao, K. R. Sanft, and L. R. Petzold, J. Chem. Phys. **130**, 064103 (2009).
[13] K. R. Sanft, D. T. Gillespie, and L. R. Petzold, IET Syst. Biol. **5**, 58 (2011).
[14] J. Fu, S. Wu, and L. R. Petzold, "Time dependent solution for acceleration of tau-leaping," J. Comp. Phys. (submitted).
[15] Y. Cao and L. R. Petzold, Comput. Methods Appl. Mech. Eng. **197**, 3472 (2008).
[16] D. Barik, M. R. Paul, W. T. Baumann, Y. Cao, and J. J. Tyson, Biophys. J. **95**, 3563 (2008).
[17] W. Zhu and L. Petzold, AIChE J. **45**, 869 (1999).
[18] B. Bhattacharjee, D. A. Schwer, P. I. Barton, and W. H. Green, Jr., Combust. Flame **135**, 191 (2003).
[19] T. Lu and C. K. Law, Proc. Combust. Inst. **30**, 1333 (2005).
[20] M. Valorani, F. Creta, D. A. Goussis, J. C. Lee, and H. N. Najm, Combust. Flame **146**, 29 (2006).
[21] P. Pepiot-Desjardins and H. Pitsch, Combust. Flame **154**, 67 (2008).
[22] T. Nagy and T. Turnyi, Combust. Flame **156**, 417 (2009).

[23]H. Kuwahara, C. Myers, M. Samoilov, N. Barker, and A. Arkin, "Automated abstraction methodology for genetic regulatory networks," in *Transactions on Computational Systems Biology VI*, edited by C. Priami, and G. Plotkin (Springer-Verlag, Berlin, 2006), pp. 150–175.
[24]H. Kuwahara and C. J. Myers, J. Comput. Biol. **15**, 779 (2008).
[25]H. Kuwahara, C. J. Myers, and M. S. Samoilov, PLOS Comput. Biol. **6**, e1000723 (2010).
[26]J. L. Peterson, *Petri Net Theory and the Modeling of Sytems* (Prentice-Hall, Englewood Cliffs, New Jersey, 1981).
[27]C. A. Petri, "Communication with automata," Technical Report No. RADC-TR-65-377 (Applied Data Research, Princeton, NJ, 1966).
[28]H. Kurata, H. El-Samad, T.-M. Yi, M. Khammash, and J. Doyle, in *Proceedings of the 40th IEEE Conference on Decision and Control* (IEEE, 2001), Vol. 1, p. 837.
[29]H. Kurata, M. Khammash, and J. Doyle, in *Proceedings of the Third International Conference on Systems Biology, 2002*.
[30]M. F. Hockin, K. C. Jones, S. J. Everse, and K. G. Mann, J. Biol. Chem. **277**, 18322 (2002).
[31]D. T. Gillespie, Annu. Rev. Phys. Chem. **58**, 35 (2007), a review of the stochastic simulation algorithm, chemical master equation, and related topics.
[32]K. Burrage, M. Hegland, S. MacNamara, and R. B. Sidje, in *150th Markov Anniversary Meeting*, edited by A. Langville, and W. Stewardt (Boson Books, Raleigh, NC, 2006), pp. 21–38.
[33]B. Munsky and M. Khammash, J. Chem. Phys. **124**, 044104 (2006).
[34]L. Ferm, P. Lötstedt, and P. Sjöberg, BIT **46**, 61 (2006).
[35]M. Mateescu, V. Wolf, F. Didier, and T. Henzinger, IET Syst. Biol. **4**, 441 (2010).
[36]T. Jahnke and T. Udrescu, J. Comput. Phys. **229**, 5724 (2010).
[37]Y. Cao, H. Li, and L. R. Petzold, J. Chem. Phys. **121**, 4059 (2004).
[38]M. A. Gibson and J. Bruck, J. Chem. Phys. **104**, 1876 (2000).
[39]A. Slepoy, A. P. Thompson, and S. J. Plimpton, J. Chem. Phys. **128**, 205101 (2008).
[40]M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, Multiscale Model. Simul. **4**, 867 (2005).
[41]T. Li, Multiscale Model. Simul. **6**, 417436 (2007).
[42]D. F. Anderson, A. Ganguly, and T. G. Kurtz, Ann. Appl. Probab. **21**, 2226–2262 (2011).
[43]S. Wu, J. Fu, Y. Cao, and L. R. Petzold, J. Chem. Phys. **134**, 134112 (2011).
[44]C. Chaouiya, Briefings Bioinf. **8**, 210 (2007).
[45]P. J. E. Goss and J. Peccoud, Proc. Natl. Acad. Sci. U.S.A. **95**, 6750 (1998).
[46]M. Feinberg, Chem. Eng. Sci. **42**, 2229 (1987).
[47]V. N. Reddy, M. L. Mavrovouniotis, and M. N. Liebman, Proc. Int. Conf. Intell. Syst. Mol. Biol. **1**, 328 (1993).
[48]I. Zevedei-Oancea and S. Schuster, In Silico Biol. **3**, 323 (2003).
[49]G. Craciun and M. Feinberg, SIAM J. Appl. Math. **66**, 1321 (2006).
[50]D. Angeli, P. de Leenheer, and E. Sontag, Math. Biosci. **210**, 598 (2007).
[51]D. Angeli, P. de Leenheer, and E. Sontag, J. Math. Biol. **61**, 581 (2010).
[52]S. H. Lam and D. A. Goussis, Int. J. Chem. Kinet. **26**, 461 (1994).
[53]M. S. Okino and M. L. Mavrovouniotis, Chem. Rev. **98**, 391 (1998).
[54]Y. Cao, D. T. Gillespie, and L. R. Petzold, J. Comput. Phys. **206**, 395 (2005).
[55]D. F. Anderson, G. Craciun, and T. G. Kurtz, Bull. Math. Biol. **72**, 19471970 (2010).
[56]K. R. Sanft, S. Wu, M. Roh, J. Fu, R. K. Lim, and L. R. Petzold, Bioinformatics **27**, 2457 (2011).
[57]E. R. Gansner, *Drawing Graphs with Graphviz* (2011), see http://www.graphviz.org/pdf/libguide.pdf.
[58]K. R. Sanft, S. Wu, D. T. Gillespie, and L. R. Petzold, "Automatic slow-scale stochastic simulation algorithm with dynamic partitioning and efficient partial equilibrium approximation," J. Comput. Phys. (submitted).
[59]L. A. Harris, A. M. Piccirilli, E. R. Majusiak, and P. Clancy, Phys. Rev. E **79**, 051906 (2009).