Stochastic analysis of gene regulatory networks using finite state projections and singular perturbation

Brian Munsky, Slaven Peleš and Mustafa Khammash
Department of Mechanical Engineering
University of California
Santa Barbara, CA 93106-5070

***Abstract*— Considerable recent experimental evidence suggests that significant stochastic fluctuations are present in gene regulatory networks. The investigation of stochastic properties in genetic systems involves the formulation of a mathematical representation of molecular noise and devising efficient computational algorithms for computing the relevant statistics of the modeled processes. However, the complexity of gene regulatory networks poses serious computational difficulties and makes any quantitative prediction a difficult task. The recently proposed Finite State Projection (FSP) algorithm provides a promising approach to handling these problems, but thus far it has only been demonstrated for a certain class of problems. In this paper we show that the applicability of the finite projection approach can be enhanced by taking advantage of tools from the fields of modern control theory and dynamical systems. In particular, we present an approach that utilizes singular perturbation theory in conjunction with the Finite State Projection approach to improve the computation time and facilitate model reduction. We demonstrate the effectiveness of the resulting slow manifold FSP algorithm on a simple example arising in the cellular heat shock response mechanism.**

I. Introduction

Through evolution living organisms have developed complex robust control mechanisms with which they can regulate intracellular processes and adapt to changing environments. These mechanisms are triggered by small changes in sensitive control parameters, yet their functions are unaffected by relatively large variations in the system that may even include damage of whole parts of the mechanism itself. Understanding control processes that take place inside a cell is key to understanding fundamental biological functions of living organisms. However, despite dramatic new developments in experimental study of intracellular processes, the quantitative study of these systems is significantly hindered by computational complexity.

For many chemical processes reactants are present in large numbers and the reactions can be modeled at the macroscopic level with ordinary or stochastic differential equations (ODEs or SDEs). Inside a cell, however, matters can be entirely different. Here some chemicals such as proteins or RNA molecules may be present in only a few copies per cell. Since these trace chemicals may effect important changes in the cell's behavior, they cannot be ignored, yet they obviously cannot be treated as continuous concentrations. This consideration has given rise to a relatively new branch of computational research in chemistry: stochastic modeling of chemical reactions at the mesoscopic level.

For a discrete population models with N chemical species, the *configuration* of the system at any time is specified by the population of each species: $\mathbf{x} := [\ \xi_1 \quad \xi_2 \quad \dots \quad \xi_N\ ]^T$. The *configuration space* of the system is the N dimensional non-negative integer lattice that contains all possible configurations. Over time, the reacting system follows a trajectory, where each individual reaction is represented by a jump from the current configuration to another within the configuration space. With the assumptions that the system is well mixed, has constant volume and is kept at constant temperature, the problem becomes autonomous of time, and the waiting time between jumps can be modeled as an exponentially distributed random variable [1]. The probability of each configuration can then be described with the linear time invariant ordinary differential equation known as the Chemical Master Equation [1].

Until recently, it was believed that the CME was computationally intractable except in the simplest of circumstances. As a result, most discrete state stochastic models have been studied almost exclusively with the simple kinetic Monte Carlo simulation technique known as the stochastic simulation algorithm SSA [2]. Here random numbers are generated for every individual reaction event in order to determine (*i*) when the next reaction will occur, and (*ii*) which reaction it will be. However, for most systems, huge numbers of individual reactions may occur, and the SSA is often too computationally expensive. Two types of recent approximations have been made to improve efficiency of the SSA: time scale based system partitioning routines and τ leaping routines. In the first type, efficiency is gained by separating the system dynamics into slow and fast parts [3], [4]. If one is interested in short time scales, the slow processes are considered to be constant, and only the dynamics of the fast partition are considered. If one is interested in long term behaviors of the system, then the fast dynamics are averaged in order to concentrate upon the slow dynamics. In the second type of approximation to the SSA, some probabilistic aspects of the system are assumed to be constant over short discrete time intervals, and under this assumption one can develop a faster kinetic Monte Carlo algorithm that leaps from one time step to the next [5]–[7]. Both approximation types, system partitioning and τ leaping, have been very successful in increasing the scope of problems to which the SSA may be reasonably applied.

Moving in an entirely different direction, we recently

introduced the Finite State Projection (FSP) method as a new approach for directly approximating the solution to the CME [8]. The FSP provides a systematic method to perform bulk system reductions on the possibly infinite dimensional ODE in order to find a finite dimensional analytical approximation to the CME. Unlike Monte Carlo algorithms, the FSP does not rely on random number generation and provides precise lower and upper bounds on its approximation accuracy. While the earliest implementations of the FSP have shown much promise in dealing with a handful of biologically inspired models [8], [9], the method is still limited in problems where accuracy requirements demand consideration of an exorbitantly large state space. In these cases, model reduction techniques can be utilized to make the problem dimension more manageable. For example, drawing upon concepts of reachability and controlability from modern control theory, one can significantly reduce the order of the FSP [10]. Similarly, one can use the initial distribution and the CME to define a lower dimensional Krylov subspace in which to compute the system dynamics as has been done in [11]. Additionally, since the SSA benefits so strongly from system partitioning methods, it is natural to ask whether the same or similar tools may improve upon the FSP. In [12] we showed how time-scale separation can allow for huge reductions in the computation of a projected approximation of the CME. In this paper, we expand upon this work to explicitly illustrate how this reduction can be exploited in the development of a Slow-Manifold FSP algorithm that systematically determines the projection necessary to achieve a pre-specified error tolerance in the solution to the CME.

In the next section we provide some background on the basic FSP method. In section III we illustrate how linear perturbation theory can be used to reduce the order of the FSP problem and we formulate the Slow-Manifold FSP algorithm. In section IV we demonstrate this algorithm on an example from the field of systems biology. Finally, in section V, we conclude with remarks on the advantages of these approaches over the original FSP and outline a few directions for future work on the topic.

II. Background

We consider a well-mixed, fixed-temperature, and fixed volume system of N distinct reacting chemical species. At any point in time, we use an integer vector $\mathbf{x} := [\ \xi_1 \quad \xi_2 \quad \dots \quad \xi_N\ ]^T$ to describe the discrete populations of the N molecular species. Each $\mathbf{x}$ is a unique configuration of the system, and the nonnegative set $\mathbb{N}^N$ is the set of all possible configurations. We will *a priori* fix a sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ of elements in $\mathbb{N}^N$ and define $\mathbf{X} := [\mathbf{x}_1, \mathbf{x}_2, \dots\ ]^T$ as the *ordered configuration set*. Beginning at any $\mathbf{x}_i$, suppose that there can be a maximum of M possible reactions, where each reaction leads to a different configuration: $\mathbf{x}_i \rightarrow \mathbf{x}_j = \mathbf{x}_i + \nu_\mu$. For each μ^{th} reaction, ν_μ is the associated stoichiometry (directional transition on the configuration set), and $a_\mu(\mathbf{x}_i)$ is the associated propensity function. Define $p_i(t) = p(\mathbf{x}_i, t)$ as the probability that the system will have the i^{th} configuration at time t. At an incrementally small time later, $p_i(t+dt)$ is equal to the sum of (*i*) the probability that the system begins in $\mathbf{x}_i$ at t and remains there until $t+dt$, and (*ii*) the probability that the system is in $\mathbf{x}_j \neq \mathbf{x}_i$ at t and will transition to $\mathbf{x}_i$ during the time step, dt. This probability can be written as:

$$p_i(t+dt) = p_i(t)\left(1 - \sum_{\mu=1}^{M} a_\mu(\mathbf{x}_i)dt\right) + \sum_{\mu=1}^{M} p(\mathbf{x}_i - \nu_\mu, t)a_\mu(\mathbf{x}_i - \nu_\mu)dt. \quad (1)$$

From (1) it is easy to derive the differential equation known as the Chemical Master Equation, or CME [2], which based upon our enumeration of $\mathbf{X}$, can be written in matrix form:

$$\dot{\mathbf{P}}(t) = \mathbf{A}\mathbf{P}(t), \quad (2)$$

where $\mathbf{P}(t)$ is the probability distribution at time t. The generator matrix $\mathbf{A}$ is uniquely defined by the reaction stoichiometries and propensities and the choice of the enumeration of $\mathbf{X}$. $\mathbf{A}$ has the properties that it is independent of t; all of its diagonal elements are non-positive; all its off-diagonal elements are non-negative; and all its columns sum to zero. The solution to (2) beginning at $t = 0$ and ending at $t = t_f$ is the expression: $\mathbf{P}(t_f) = \Phi(0, t_f) \cdot \mathbf{P}(0)$. In the case where the configuration set has a small finite dimension, the Perron-Frobenius operator, $\Phi(0, t_f)$, is the exponential of $\mathbf{A}t_f$, and one can compute the solution: $\mathbf{P}(t_f) = \exp(\mathbf{A}t_f)\mathbf{P}(0)$. However, when $\mathbf{A}$ is infinite dimensional or extremely large, the corresponding analytic solution is unclear or vastly difficult to compute. In these cases, one may devise a systematic means of approximating the full system using finite dimensional sub-systems. This truncation approach lies behind the rationale for the Finite State Projection (FSP) method [8].

We must introduce some convenient notation. Let $J = \{j_1, j_2, j_3, \dots\}$ denote an index set. For any two index sets I and J, let $J \subseteq I$ denote that I contains every element from J, and let $J \bigcup I$ and $J \bigcap I$ be the union and intersection, respectively, of J and I. Let J' denote the complement of the set J, where unless stated otherwise the complement is taken with respect to the set of all non-negative integers. If $\mathbf{X}$ is an enumerated set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots\}$, then $\mathbf{X}_J$ denotes the subset $\{\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \mathbf{x}_{j_3}, \dots\}$. Furthermore, let $\mathbf{v}_J$ denote the subvector of $\mathbf{v}$ whose elements are chosen according to J, and let $\mathbf{A}_{IJ}$ denote the submatrix of $\mathbf{A}$ such that the rows have been chosen according to I and the columns have been chosen according to J. For example, if I and J are defined as $\{3, 1, 2\}$ and $\{1, 3\}$, respectively, then:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix}_{IJ} = \begin{bmatrix} g & k \\ a & c \\ d & f \end{bmatrix}.$$

For convenience let $\mathbf{A}_J$ denote the principle submatrix of $\mathbf{A}$, in which both rows and columns have been chosen according to J. Unless otherwise stated, all vector inequalities are to be interpreted componentwise. With this notation we can restate the following theorem from [8]:

Theorem 1 If $\mathbf{A} \in \mathbb{R}^{n\times n}$ has no negative off-diagonal elements, then $[\exp\mathbf{A}]_J \geq \exp[\mathbf{A}_J] \geq \mathbf{0}$.

Ref. [8] provides a detailed proof of this theorem. The matrix $\mathbf{A}$ in (2) has no negative off-diagonal terms and therefore satisfies the assumptions of Thm 1. Consider two finite configuration subsets $\mathbf{X}_{J_1}$ and $\mathbf{X}_{J_2}$, where $J_2 \supseteq J_1$. Since the full probability density vector, $\mathbf{P}(0)$ is nonnegative, Thm 1 assures that:

$$[\exp(\mathbf{A}_{J_2}t_f)]_{J_1}\mathbf{P}_{J_1}(0) \geq \exp(\mathbf{A}_{J_1}t_f)\mathbf{P}_{J_1}(0) \geq \mathbf{0}.$$

where $\mathbf{P}_{J_1}$ is the probability distribution of the elements of $\mathbf{X}$ indexed by J_1. This result guarantees that as one increases the finite configuration subset, the approximate solution increases monotonically for each element in the configuration subset.

In addition to being non-negative, $\mathbf{P}(t)$ sums to one. These properties and the nonnegativity of the off-diagonal elements of $\mathbf{A}$ allow one to state the following result [8].

Theorem 2 Consider a Markov process in which the probability distribution evolves according to the linear ODE, $\dot{\mathbf{P}}(t) = \mathbf{A}\mathbf{P}(t)$, where $\mathbf{A}$ has no negative off-diagonal entries. If for some finite index set J, $\varepsilon > 0$, and $t_f \geq 0$,

$$\mathbf{1}^T\exp(\mathbf{A}_J t_f)\mathbf{P}_J(0) \geq 1-\varepsilon, \tag{3}$$

then

$$\exp(\mathbf{A}_J t_f)\mathbf{P}_J(0) \leq \mathbf{P}_J(t_f), \text{ and} \tag{4}$$

$$||\mathbf{P}_J(t_f) - \exp(\mathbf{A}_J t_f)\mathbf{P}_J(0)||_1 \leq \varepsilon. \tag{5}$$

While Theorem 1 guarantees that as we add points to the finite configuration subset, the approximate solution monotonically increases, Theorem 2 provides a certificate of how close the approximation is to the true solution. Together the two theorems lead us to the FSP algorithm presented in [8]:

The Finite State Projection Algorithm

Inputs Reaction propensity functions and stoichiometry.
Initial probability density vector, $\mathbf{P}(0)$.
Final time of interest, t_f.
Total amount of acceptable error, $\varepsilon > 0$.
Step 0 Choose finite set of states, $\mathbf{X}_{J_o}$, for the FSP.
Initialize a counter, $i = 0$.
Step 1 Form $\mathbf{A}_{J_i}$ and compute $\Gamma_{J_i} = \mathbf{1}^T\exp(\mathbf{A}_{J_i}t_f)\mathbf{P}_{J_i}(0)$.
Step 2 If $\Gamma_{J_i} \geq 1-\varepsilon$, **Stop**.
$\exp(\mathbf{A}_{J_i}t_f)\mathbf{P}_{J_i}(0)$ is within ε of $\mathbf{P}_{J_i}(t_f)$.
Step 3 Add more states to find $\mathbf{X}_{J_{i+1}}$.
Increment i and return to **Step 1**.

For the basic FSP algorithm, if we wish to find a solution that is accurate to within ε at a time t_f, we must find a finite set of configurations such that the probability of ever leaving that set during the time interval $[0, t_f]$ is less than ε. For many problems, including the examples shown in [8], [9], this set of configurations may be small enough that we can easily compute a single matrix exponential to approximate the solution to the CME. However, in other situations that may not be the case. In the next section we apply some concepts of linear perturbation theory to reduce the order of the FSP problem and thereby significantly improve the efficiency and expand the applicability of the FSP algorithm.

III. The Slow-Manifold FSP Algorithm

In Step 1 of the FSP algorithm above we are faced with solving an N dimensional LTI system, $\dot{\mathbf{P}}(t) = \mathbf{A}\mathbf{P}(t)$, where $\mathbf{A}$ is made up from the propensity functions of the various reactions. In gene networks we can often identify m clusters of configurations within which transitions occur quite frequently, while transitions between the clusters are relatively rare. Let $\mathbf{H}$ be a generator matrix that is made up from the frequent transitions within these clusters, and let $\mathbf{G}$ be the generator made up of the remaining transitions from one cluster to another such that $\mathbf{A} = \mathbf{H} + \mathbf{G}$. With some permutation of the configuration space, $\mathbf{H}$ has a block diagonal structure, $\mathbf{H} = \text{diag}\{\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_m\}$, where each block $\mathbf{H}_i$ is itself a generator corresponding to a single cluster of fast interconnected configurations.

Each $\mathbf{H}_i$ is itself a generator and therefore has an eigenvalue equal to zero, which corresponds to a left eigenvector $\mathbf{u}_i = \mathbf{1}^T$ and a positive right eigenvector, $\mathbf{v}_i$, where $\mathbf{v}_i$ is scaled such that $\mathbf{u}_i\mathbf{v}_i = \mathbf{1}^T\mathbf{v}_i = 1$. We use these eigenvectors to define.

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & 0 & \ldots \\ 0 & \mathbf{u}_2 & \ldots \\ \vdots & \vdots & \ddots \end{bmatrix}, \text{ and } \mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & 0 & \ldots \\ 0 & \mathbf{v}_2 & \ldots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

Let $\mathbf{S} = \begin{bmatrix} \mathbf{V} & \mathbf{S}_R \end{bmatrix}$ be a square matrix in which each column is an right eigenvector of $\mathbf{H}$ that has been scaled to have an absolute sum of one. The inverse of $\mathbf{S}$ is given by $\mathbf{S}^{-1} = \begin{bmatrix} \mathbf{U} \\ \mathbf{S}_L \end{bmatrix}$ such that we have the following similarity transformation for $\mathbf{H}$:

$$\mathbf{S}^{-1}\mathbf{H}\mathbf{S} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda} \end{bmatrix}, \ \Lambda = \text{diag}(\lambda_{m+1}, \ldots, \lambda_N).$$

where the first m diagonal elements correspond to the zero eigenvalues of the $\mathbf{H}_i$ blocks, and we label non-zero eigenvalues of $\mathbf{H}$ so that $0 > \text{Re}\{\lambda_{m+1}\} \geq \text{Re}\{\lambda_{m+2}\}, \ldots \geq \text{Re}\{\lambda_N\}$. If we apply the coordinate transformation

$$\begin{bmatrix} \mathbf{y}_1(t) \\ \mathbf{y}_2(t) \end{bmatrix} = \mathbf{S}^{-1}\mathbf{P}(t) = \begin{bmatrix} \mathbf{U}\mathbf{P}(t) \\ \mathbf{S}_L\mathbf{P}(t) \end{bmatrix},$$

Our ODE becomes:

$$\begin{bmatrix} \dot{\mathbf{y}}_1(t) \\ \dot{\mathbf{y}}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{U}\mathbf{G}\mathbf{V} & \mathbf{U}\mathbf{G}\mathbf{S}_R \\ \mathbf{S}_L\mathbf{G}\mathbf{V} & \mathbf{\Lambda} + \mathbf{S}_L\mathbf{G}\mathbf{S}_R \end{bmatrix} \begin{bmatrix} \mathbf{y}_1(t) \\ \mathbf{y}_2(t) \end{bmatrix}. \tag{6}$$

If we can make a clear distinction between frequent and rare reactions, then we are assured that the magnitude of $\mathbf{G}\mathbf{V}$ is small with respect to $\mathbf{\Lambda}$. For convenience we define $\varepsilon = \frac{1}{|\text{Re}\{\lambda_{m+1}\}|}\,||\mathbf{G}\mathbf{V}||_1$ and $\mathbf{Q} = \frac{1}{\varepsilon}\mathbf{G}\mathbf{V}$, and we can make coordinate substitution $(\mathbf{y}_1, \mathbf{y}_2) = (\mathbf{y}, \varepsilon\mathbf{z})$ to get the standard singular perturbation model,

$$\begin{bmatrix} \dot{\mathbf{y}}(t) \\ \varepsilon\dot{\mathbf{z}}(t) \end{bmatrix} = \begin{bmatrix} \varepsilon\mathbf{U}\mathbf{Q} & \varepsilon\mathbf{U}\mathbf{G}\mathbf{S}_R \\ \varepsilon\mathbf{S}_L\mathbf{Q} & \varepsilon(\mathbf{\Lambda} + \mathbf{S}_L\mathbf{G}\mathbf{S}_R) \end{bmatrix} \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}(t) \end{bmatrix}. \tag{7}$$

We can apply linear perturbation theory to show that $|\mathbf{z}(t)| \leq \gamma(t) + O(\varepsilon)$, where $\gamma = |\mathbf{z}(0)| \exp(\lambda_{m+1} t)$ is the transient error of the fast manifold. For the slow manifold, we have

$$\mathbf{y}(t) = \exp(\mathbf{UGV}t)\mathbf{y}(0) + O(\varepsilon),$$

for all time $t \geq 0$[1].

The reverse transformation then yields

$$\begin{aligned}\mathbf{P} &= \mathbf{S}\begin{bmatrix}\mathbf{y}_1(t)\\ \mathbf{y}_2(t)\end{bmatrix} = \mathbf{S}\begin{bmatrix}\mathbf{y}(t)\\ \varepsilon\mathbf{z}(t)\end{bmatrix}\\ &= \mathbf{V}\mathbf{y}(t) + O(\varepsilon)\\ &= \mathbf{V}\exp(\mathbf{UGV}t)\mathbf{y}(0) + O(\varepsilon)\\ &= \mathbf{V}\exp(\mathbf{UGV}t)\mathbf{UP}(0) + O(\varepsilon). \qquad (8)\end{aligned}$$

It is important to note that due to truncation of Eqn 6, only contributions of the first m eigenvectors of $\mathbf{H}$ affect the approximate solution. Therefore, instead of calculating full eigensystems for each block $\mathbf{H}_i$, it suffices to find only eigenvectors associated with zero eigenvalues.

Applying this model reduction to the original FSP algorithm yields the following algorithm which we name the Slow-Manifold FSP algorithm:

The Slow-Manifold FSP Algorithm

Inputs Propensities and stoichiometries for all reactions.
Initial probability density vector, $\mathbf{P}(0)$.
Final time of interest, t_f.
Target FSP error, $\delta > 0$.

Step 0 Choose initial set of states, $\mathbf{X}_{J_o}$, for the FSP.
Initialize a counter, $k = 0$.

Step 1 Use fast reactions connecting states within $\mathbf{X}_{J_k}$ to form $\mathbf{H}_{J_k} = \text{diag}\{\mathbf{H}_1, \ldots, \mathbf{H}_m\}$.
Use remaining reactions to form $\mathbf{G}_{J_k}$.

Step 2 Find eigenvalues and vectors of each $\mathbf{H}_i$ and build matrices $\mathbf{U}$ and $\mathbf{V}$.
Estimate $\varepsilon = ||\mathbf{G}_{J_k}\mathbf{V}||_1 \,/\, |\lambda_{m+1}|$.
Compute $\gamma = |\mathbf{S}_L\mathbf{P}(0)|_1 \exp(\lambda_{m+1} t_f)$.

Step 3 Find $\mathbf{P}^{FSP}_{J_k}(t_f) = \mathbf{V}\exp(\mathbf{UG}_{J_k}\mathbf{V}t_f)\mathbf{UP}_{J_k}(0)$ and compute $\Gamma_{J_k} = \mathbf{1}^T\mathbf{P}^{FSP}_{J_k}(t_f)$.

Step 4 If $\Gamma_{J_k} \geq 1 - \delta$, **Stop**.
$\mathbf{P}^{FSP}_{J_k}(t_f)$ is within $\delta + \gamma + O(\varepsilon)$ of $\mathbf{P}_{J_k}(t_f)$.

Step 5 Add more states to find $\mathbf{X}_{J_{k+1}}$.
Increment k and return to **Step 1**.

Above we have used the non-traditional error estimate notation $\delta + \gamma + O(\varepsilon)$ to mean the following. If δ is largest, then the dominant error is most likely the result of the projection, and the slow manifold truncation error can be ignored. If γ is largest then the time t_f is too short for the transient dynamics to sufficiently diminish and additional eigenvectors must be included in the truncation. Finally, if ε is larger than δ and γ, then there is insufficient separation between the slow and fast dynamics and an alternative reduction scheme may be required.

[1]For a detailed derivation of these errors, see the appendix of [12].

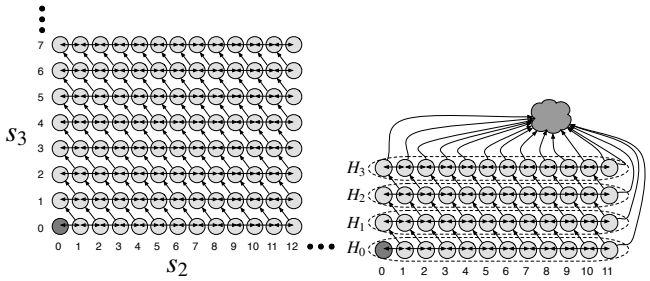

Fig. 1. (Left) Two dimensional lattice representing the configurations of the toy heat shock model. s_2 and s_3 are populations of free σ_{32} molecules and σ_{32}-RNAP compounds, respectively. Reactions $s_1 \rightleftharpoons s_2$, are represented by bidirectional horizontal arrows and reaction $s_2 \rightarrow s_3$ is represented with diagonal arrows. The total number of σ_{32} is constant, the system is uniquely defined by s_2 and s_3 alone. (Right) The same lattice after applying the FSP. Unlikely states have been aggregated into a single sink state. Each horizontal row of configurations is separated from the rest by the slow reaction 3 and the fast reactions are used to form the block generator $\mathbf{H}_i$.

In the following section we will illustrate this algorithm on model of the heat shock response in *E. coli*.

IV. Example

In order to deal with frequently changing situations, living organisms require mechanisms to deal with harsh environments. One such mechanism is the cellular heat shock response. At elevated temperatures proteins inside the cell begin to misfold and lose their geometry dependent functionality. In an attempt to stop this misfolding, the cell produces heat shock proteins including molecular chaperones and proteases. By helping to refold deformed proteins, molecular chaperones can restore the original functionality of those proteins. Proteases, on the other hand, help degrade damaged proteins before those damaged proteins impair the function of the cell. The core of the heat shock response mechanism in *E. coli* is the synthesis of the σ_{32}-RNAP complex [13]. Here we study a simplified model for σ_{32}-RNAP synthesis using the slow manifold finite state projection algorithm.

At normal physiological temperatures σ_{32} protein is found almost exclusively in a complex σ_{32}-DnaK. However, as the temperature increases this complex becomes less stable and the probability of finding free σ_{32} inside the cell increases. Once free, σ_{32} can combine with RNA polymerase to form a stable σ_{32}-RNAP complex, which initiates production of heat shock proteins. This simple regulatory mechanism is summarized by a set of three reactions, $s_1 \rightleftharpoons s_2 \rightarrow s_3$, where s_1, s_2 and s_3 correspond to the σ_{32}-DnaK complex, the σ_{32} heat shock regulator and the σ_{32}-RNAP complex, respectively. This model of the heat shock subsystem has been analyzed before using various computational methods including Monte Carlo implementations [3], [14].

In the biological system, the relative rates of the reactions are such that the reaction from s_2 to s_1 is by far the fastest, and σ_{32} molecules infrequently escape from DnaK long enough to form the σ_{32}-RNAP complex. The purpose of this mechanism is to strike a balance between fixing the damage produced by heat and saving the cell's resources, as a significant portion of cell energy is consumed when

producing heat shock proteins. The optimal response to the heat shock is not massive, but measured production of heat shock proteins, which leaves sufficient resources for other cellular functions. We use the following set of parameters values for the reaction rates: $c_1 = 10,\ c_2 = 4 \times 10^4,\ c_3 = 2, s_1(0) = 2000,\ s_2(0) = s_3(0) = 0$.

With only the reactions above, the total number of σ_{32}–free or in compounds–is constant, so that $s_1 + s_2 + s_3 = \text{const}$. With this constraint the reachable states of this three species problem can be represented on a two dimensional lattice as shown in Fig. 1(Left). For our initial conditions there are 2,001,000 reachable states, and the full chemical master equation is too large to be tackled directly, and we wish to find an approximate solution using the slow manifold FSP algorithm at time $t_f = 300s$. To do this we follow the steps of the algorithm as follows:

Step 0: We specify the problem parameters as given above. We choose a target 1-norm error of $\delta = 0.001$ for the distribution at time t_f, and we choose an initial FSP set $\mathbf{X}_{J_0}$ that includes all configurations such that $s_3 \leq 200$ and $s_2 \leq 11$.

Step 1: We form the fast generator $\mathbf{H}_{J_k}$. For the configurations in $\mathbf{X}_{J_0}$, reaction 1 has propensities ranging from 18,000 to 20,000, reaction 2 has propensities ranging up to 110,000, and reaction 3 has a propensity that ranges no higher than 22. Thus there is a clear separation between the fast transitions (reactions 1 and 2) and the slow transitions (reaction 3). Furthermore, since s_3 changes only during the slow reaction, each level of $s_3 = i$ defines a cluster of fast interconnected configurations; these are shown as horizontal rows in Fig. 1. We will use the index set I_i to denote the set of configurations in the i^{th} cluster. The generator $\mathbf{H}_i$ is formed from all of the reactions that begin and end within $\mathbf{X}_{I_i}$. In this case there are 12 points in each i^{th} cluster corresponding to $s_3 = i$, and $s_2 = \{0, 1, 2, \ldots, 11\}$; if we were to change the number of points in our projection, the size and number of clusters would change accordingly. The full fast generator is defined $\mathbf{H}_J = \text{diag}(\mathbf{H}_{I_0}, \mathbf{H}_{I_1}, \ldots, \mathbf{H}_{I_{200}})$. For the reaction rates given above, the first nonzero eigenvalue of $\mathbf{H}_J$ can be computed to be $\lambda_{m+1} \approx -4 \times 10^4$.

The rare transition generator $\mathbf{G}_J$ is then formed from the remaining transitions coming from two components: (*i*) reaction 3 corresponding to a transition from set $\mathbf{X}_{I_i}$ to $\mathbf{X}_{I_{i+1}}$, and (*ii*) all transitions taking the system from the set $\mathbf{X}_J$ to its complement $\mathbf{X}_{J'}$.

Step 2: We build the matrices $\mathbf{U}$ and $\mathbf{V}$ from the left and right eigenvectors corresponding to the zero eigenvalues of $\mathbf{H}_J$. We compute the norm $||\mathbf{G}_J\mathbf{V}||_1 \approx 2.0$, and use this to estimate how well the time scale is separated between the frequent and rare events:

$$\varepsilon = \frac{||\mathbf{G}_J\mathbf{V}||_1}{|\lambda_{m+1}|} = \frac{2}{4 \times 10^4} = 0.5 \times 10^{-4}.$$

We also compute the transient error due to the initial conditions, $\gamma = |\mathbf{S}_L\mathbf{P}(0)|_1 \exp(\lambda_{m+1}t_f) \approx \exp(-1.2 \times 10^7)$, which is far smaller that δ, and can therefore be neglected.

Step 3: We apply perturbation theory and approximate the FSP solution as

$$\mathbf{P}^{FSP}_{J_k}(t_f) \approx \mathbf{V}\exp(\mathbf{U}\mathbf{G}_{J_k}\mathbf{V}t_f)\mathbf{U}\mathbf{P}_{J_k}(0),$$

which has error $\gamma + O(\varepsilon)$ compared to the unreduced FSP solution. In order to determine whether this FSP solution is sufficient, we compute $\Gamma_{J_k} = \mathbf{1}^T\mathbf{P}^{FSP}_{J_k}(t_f) = 7.3 \times 10^{-8}$. It is not.

Step 4: Since the $\Gamma_{J_0} < 1 - \delta$, we need to add more configurations and return to Step 1. If we define $\mathbf{X}_{J_1}$ to include all configurations such that $s_3 \leq 250$ and $s_2 \leq 11$, we find $\Gamma_{J_1} = 0.034$, which is little better. If we further expand the configuration set to include up to 300 or 350 molecules of s_3, we compute $\Gamma_{J_2} = 0.921$ and $\Gamma_{J_3} = 0.99997$ respectively. The FSP solution on the set $\mathbf{X}_{J_3}$ exceeds the precision of our stopping criteria, and we know that $\mathbf{P}^{FSP}_{J_3}(t_f)$ approximates the true solution within $\delta = 0.001$. Furthermore, since $\varepsilon = 0.0005$ and $\gamma \ll \delta$, the error introduced by the time scale separation is of the same order.

For this problem we are interested in determining how the population of σ_{32}-RNAP compounds grows in time if the temperature is constant and slightly above normal physiological level. This number is proportional to the number of heat shock proteins produced in the cell. With the original FSP algorithm, we have computed the probability distribution for s_3 at $t_f = 300s$, and Fig. 2 illustrates this probability distribution. The FSP solution is shown with solid lines, and the dots represent the distribution of s_3 as computed using our current slow manifold FSP algorithm. The difference between the two results is indistinguishable. However, if we are to compare the relative computational effort of the two algorithms, we find a significant improvement. Table I(top) shows the computational time and accuracy for each step of the FSP and the slow manifold FSP algorithm. From the table we find that the time scale reduction reduces the computational effort at each step by a factor of almost 1000 compared to the original FSP algorithm. For further comparison I(bottom) shows the total computational effort of the FSP and the slow manifold FSP, as well as two Monte Carlo methods: the original SSA and an SSA approximation very similar to the Slow Scale SSA [3] in which we have applied the same time scale reduction approach. The 1-norm difference between the original FSP and the slow manifold FSP was found to be 6.6×10^{-4}, which is indeed on the same order as $\varepsilon = 5 \times 10^{-4}$ and is smaller than the required error tolerance $\delta = 1 \times 10^{-3}$. However, by making this small sacrifice in accuracy, the slow manifold FSP converges in a fraction of the time of any of the other methods.

V. Conclusion

Until recently, it was thought that the chemical master equation could not be solved analytically except for the most trivial systems. Previous work on the Finite State Projection demonstrated that for many biological systems, bulk system reductions could bring models closer into the fold of solvable problems. Here we have shown that the

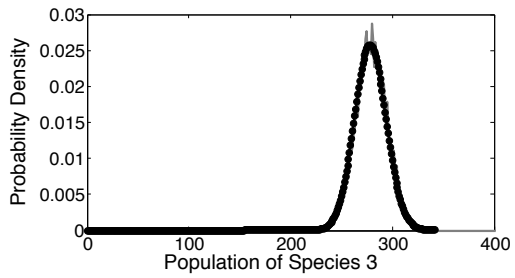

Fig. 2. Probability distribution for s_3 calculated at $t_f = 300s$. The slow manifold FSP solution (dots) approximates well the FSP solution (smooth solid lines). The jagged grey line represents the corresponding slow scale SSA solution after 10^4 realizations. See also Table I.

TABLE I

(TOP) A COMPARISON OF THE FSP AND THE SLOW-MANIFOLD FSP (FSP-SM) ALGORITHMS FOR CONFIGURATION SETS $\mathbf{X}_{J_1} \subset \mathbf{X}_{J_2} \subset \mathbf{X}_{J_3}$. (BOTTOM) TOTAL COMPUTATIONAL EFFORT AND ACCURACY OF THE FSP, FSP-SM, SSA AND A SSA-SM ALGORITHM FOR THE TOY HEAT SHOCK MODEL AT $t_f = 300s$.

Configuration set	$\|\|\mathbf{Error}\|\|_1$	**Comp. Time (s)**[a]	
		FSP	**FSP-SM**
$\mathbf{X}_{J_1}$: $s_3 \leq 250, s_2 \leq 11$	0.97	253	0.39
$\mathbf{X}_{J_2}$: $s_3 \leq 300, s_2 \leq 11$	0.08	439	0.54
$\mathbf{X}_{J_3}$: $s_3 \leq 350, s_2 \leq 11$	2×10^{-5}	647	0.67

Method	**# Simulations**	**Time (s)**	$\|\|\mathbf{Error}\|\|_1$
FSP	–[b]	1472	$\leq 2 \times 10^{-5}$
SSA	10^3	> 20000	≈ 0.25
Slow-Manifold FSP	–	1.88	$\approx 6.6 \times 10^{-4}$
Slow Scale SSA	10^3	82	≈ 0.24
Slow Scale SSA	10^4	826	≈ 0.066
Slow Scale SSA	10^5	8130	≈ 0.027

[a]All computations have been performed in Matlab 7.2 on a 2.0 MHz PowerPC Dual G5.

[b]The FSP and Slow-Manifold FSP algorithms are run only once with a specified allowable total error of 10^{-3}.

Finite State Projection method can be further enhanced when solving the chemical master equation for systems involving multiple time scales. In this work we have presented a Slow Manifold version of the FSP algorithm. Based upon singular perturbation theory tools developed in previous work [12], this algorithm provides a powerful computational tool for studying intracelular processes and gene regulatory networks.

By casting the stochastic chemical kinetics problem in a familiar linear system context, the FSP provides valuable insight into the bewildering complexity that intracellular processes exhibit. Model reductions not only speed up computations, but they also enable us to distinguish which aspects of a system are important and which are not. It is plausible that the main features of intracellular dynamics can be captured in a relatively small subset of the state space, as the results obtained by FSP reduction on the heat shock model suggest. Depending on the observation time of interest, some reactions can be neglected, while some will contribute only through their averages. Preliminary success with our approach gives us a hope that relatively simple models for intracellular processes can be discovered in the process of reducing the FSP solution.

Of course, one can easily envision that additional model reductions may be possible to even further enhance the power of the FSP. Indeed other reductions based upon control theory [10] and Krylov subspace methods [11] are already becoming apparent. Also, in our computations we have used off the shelf numerical routines for eigensystem calculations and matrix exponentiation. Further improvements in computational speed can be achieved if these routines are optimized for matrices which define master equations and their special properties. Parallelization could also provide significant benefit in the reduction process as many the eigensystem identification of each block can be done independently of the rest. We intend to investigate these possibilities in the future.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant NSF-ITR CCF-0326576 and the Institute for Collaborative Biotechnologies through Grant DAAD19-03-D-0004 from the U.S. Army Research Office.

REFERENCES

[1] van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. Elsevier, 2001.

[2] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 81, no. 25, pp. 2340–2360, May 1977.

[3] Y. Cao, D. Gillespie, and L. Petzold, "The slow-scale stochastic simulation algorithm," *J. Chem. Phys.*, vol. 122, no. 014116, Jan. 2005.

[4] H. Salis and Y. Kaznessis, "Accurate hybrid stochastic simulation of a system of coupled chemical or biological reactions," *J. Chem. Phys.*, vol. 112, no. 054103, 2005.

[5] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *J. Chem. Phys.*, vol. 115, no. 4, pp. 1716–1733, Jul. 2001.

[6] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, "Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method," *J. Chem. Phys.*, vol. 119, no. 24, pp. 12 784–12 794, Dec. 2003.

[7] T. Tian and K. Burrage, "Binomial leap methods for simulating stochastic chemical kinetics," *J. Chem. Phys.*, vol. 121, no. 21, pp. 10 356–10 364, Dec. 2004.

[8] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *J. Chem. Phys.*, vol. 124, no. 044104, 2006.

[9] B. Munsky, A. Hernday, D. Low, and M. Khammash, "Stochastic modeling of the pap-pili epigenetic switch," *Proc. FOSBE*, pp. 145–148, August 2005.

[10] B. Munsky and M. Khammash, "A reduced model solution for the chemical master equation arising in stochastic analyses of biological networks," *Proc. 45th IEEE CDC*, Dec. 2006.

[11] K. Burrage, M. Hegland, S. Macnamara, and R. Sidje, "A krylov-based finite state projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems," *Proc. of The A.A.Markov 150th Anniversary Meeting*, 2006.

[12] S. Peles, B. Munsky, and M. Khammash, "Reduction and solution of the chemical master equation using time-scale separation and finite state projection," *J. Chem. Phys.*, vol. 125, no. 204104, Nov. 2006.

[13] H. El Samad, H. Kurata, J. Doyle, C. Gross, and K. M., "Surviving heat shock: Control strategies for robustness and performance," *PNAS*, vol. 102, no. 8, p. 27362741, 2005.

[14] H. El Samad, M. Khammash, L. Petzold, and D. Gillespie, "Stochastic modeling of gene regulatory networks," *Int. J. Robust Nonlin.*, vol. 15, pp. 691–711, 2005.