

## MIT Open Access Articles

*The Enhanced Finite State Projection algorithm, using conditional moment closure and time-scale separation*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Kwon, Ukjin, Naghnaeian, Mohammad and Del Vecchio, Domitilla. 2020. "The Enhanced Finite State Projection algorithm, using conditional moment closure and time-scale separation." Proceedings of the IEEE Conference on Decision and Control, 2020-December.

**As Published:** 10.1109/CDC42340.2020.9303796

**Publisher:** Institute of Electrical and Electronics Engineers (IEEE)

**Persistent URL:** <https://hdl.handle.net/1721.1/138545>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# The Enhanced Finite State Projection algorithm, using conditional moment closure and time-scale separation\*

Ukjin Kwon<sup>1</sup>, Mohammad Naghnaeian<sup>2</sup> and Domitilla Del Vecchio<sup>3</sup>

**Abstract**—The Chemical Master Equation (CME) is commonly used to describe the stochastic behavior of biomolecular systems. However, in general, the CME’s dimension is very large or infinite, so analytical or even numerical solutions may be difficult to achieve. The truncation methods such as the Finite State Projection (FSP) algorithm alleviate this issue to some extent but not completely. To further resolve such a computational burden, we propose the Enhanced Finite State Projection (EFSP) algorithm, in which the ubiquitous time-scale separation is utilized to reduce the dimension of the CME. Our approach combines the original FSP algorithm and the model reduction technique that we developed, to approximate an infinite dimensional CME with a finite dimensional CME that contains the slow species only. Unlike other time-scale separation methods, which rely on the fast-species counts’ stationary conditional probability distributions, our model reduction technique relies on only the first few conditional moments of the fast-species counts. This is possible because we apply conditional moment closure to close the fast-species counts’ dynamics. In addition, each iteration of the EFSP algorithm relies on the solution of the approximated CME that contains the slow species only, unlike the original FSP algorithm relies on the solution of the full CME. These two properties provide a significant computation advantage. The benefit of our algorithm is illustrated through a protein binding reaction example.

## I. INTRODUCTION

To analyze the behavior of biomolecular systems, deterministic or stochastic methods can be used [1]. At the single-cell level, the randomness of molecular events can have substantial repercussions on an emergent system’s behavior [2]. Therefore, deterministic models, in many cases, fail to capture the inherent randomness of biomolecular systems, so stochastic approaches are often needed. The Chemical Master Equation (CME) gives the temporal description of the progression of a system’s state probability distribution [1]. However, when the number of molecular counts is large or unbounded, the dimension of the CME is large or countably infinite. Therefore, analytical or computational solutions of the CME are very difficult to obtain in general.

To obtain sample paths that result from the CME, the Stochastic Simulation Algorithm (SSA) [3] is used. However, when the number of reactions increases or especially in the

presence of multiple time scale, this algorithm can become computationally expensive, resulting in long simulation time. To address the simulation time issue, [4] [5] ran the SSA algorithm only with the slow reactions. These approaches require an approximation of the fast-species counts as a function of the slow-species counts, which need considerable extra computation. For example, [4] approximated the stationary conditional probability distribution of the fast-species counts as functions of the slow-species counts. Obtaining the stationary conditional probability distribution of the fast-species counts is equivalent to obtaining the fast-species counts’ stationary conditional moments of all different orders [6], and the number of conditional moments grow exponentially with the number of the fast-species counts.

Another way to address the CME’s computational issue is to use the Finite State Projection (FSP) algorithm developed by [7]. When the number of molecular counts is unbounded, the system’s state space becomes an infinite set. The FSP algorithm finds an upper bound to the molecular count of each species and truncates the system’s state space as a finite set, so that the truncated finite dimensional system’s trajectories of probabilities are as close as desired to those of the original infinite dimensional CME. In several examples, the FSP algorithm outperforms SSA algorithms in terms of computational efficiency as well as accuracy [7]. However, when multiple time scales exist, the FSP algorithm also confronts a computational issue. This is because the algorithm equally treats the transition rates between states, even though they typically vary over several orders of magnitude [8].

To handle this problem, Peleš [8] combined the FSP algorithm and time-scale separation. In particular, Peleš applied the FSP algorithm to the approximated CME that contains the slow species only. To obtain the approximated CME with the slow species only, the author applied the time-scale separation technique developed by [9] [6], and approximated the stationary conditional probability distribution of the fast-species counts as functions of the slow-species counts as [4] did. However, the size of the vector of these stationary distributions grows exponentially with the number of the fast-species counts.

In this paper, we propose the Enhanced Finite State Projection (EFSP) algorithm, which combines the original FSP algorithm and the model reduction technique that we developed [10] to approximate an infinite dimensional CME with a finite dimensional CME, which contains the slow species only. Instead of considering the fast-species counts’ stationary conditional probability distributions, our model reduction technique considers the first  $n$  conditional moments

\*This work was supported by Korea Foundation for Advanced Studies and the Air Force Office of Scientific Research under grant FA9550-14-1-0060

<sup>1</sup>U. Kwon is with the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, Cambridge, MA, USA [ujkwon@mit.edu](mailto:ujkwon@mit.edu)

<sup>2</sup>M. Naghnaeian is with the Mechanical Engineering Department, Clemson University, South Carolina, USA [mnaghna@clemson.edu](mailto:mnaghna@clemson.edu)

<sup>3</sup>D. Del Vecchio is with the Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge, MA, USA [ddv@mit.edu](mailto:ddv@mit.edu)

of the fast-species counts, where  $n$  is an arbitrary (small) number, which gives a trade off between approximation accuracy and computational complexity. We assume that the number of fast-species counts is bounded, which is reasonable in many biomolecular systems of practical interest [1], but allow an unbounded number of slow-species counts.

For each iteration of the EFSP algorithm, we use the truncated state space of the slow species, which is a finite set, as an input. Then, we write Ordinary Differential Equations (ODEs) for both the marginal probability distribution of the slow-species counts, with the truncated state space at that iteration, and for the first  $n$  conditional moments of the fast-species counts. Next, we apply our model reduction technique [10] and obtain a low dimensional CME with the slow species only. We expand the truncated state space of the slow species until the approximation error between the original CME and the approximated CME is sufficiently small, by using the approximated CME that contains the slow species only. Therefore, for each iteration, unlike the original FSP algorithm, which relies on the solution of the full CME, the EFSP algorithm only relies on the solution of the approximated CME that contains the slow species only. In addition, different from [8], our method does not require the slow-species counts' stationary conditional probability distribution. These two differences of the EFSP algorithm provide a significant computation advantage.

## II. PRELIMINARIES

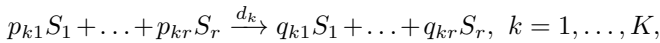
In this section, we define notations that are used throughout this paper.  $\mathbb{Z}_{\geq 0}$  and  $\mathbb{R}_{\geq 0}$  are the sets of nonnegative integers and real numbers, respectively. For any positive integer  $n$ ,  $\mathbb{Z}_{\geq 0}^n$  ( $\mathbb{R}_{\geq 0}^n$ ) implies the set of  $n$ -dimensional vectors with each entry in  $\mathbb{Z}_{\geq 0}$  ( $\mathbb{R}_{\geq 0}$ ). Given a nonnegative integer  $w$  and an  $n$ -dimensional vector  $Z = [z_1, z_2, \dots, z_n]^T$ , we define  $\Psi_w(Z)$  to be the vector made up of entries of the form  $z_1^{k_1} z_2^{k_2} \dots z_n^{k_n}$  where  $k_i \in \mathbb{Z}_{\geq 0}$ , for  $i = 1, 2, \dots, n$ , and  $\sum_{i=1}^n k_i = w$ . For instance, when  $Z = [z_1, z_2, z_3]$ ,

$$\begin{aligned}\Psi_1(Z) &= [z_1, z_2, z_3]^T, \\ \Psi_2(Z) &= [z_1^2, z_1 z_2, z_1 z_3, z_2^2, z_2 z_3, z_3^2]^T.\end{aligned}\quad (1)$$

The  $l_\infty$  and  $l_1$  norms of a vector  $Z = [z_1, z_2, \dots, z_n]^T$  are defined as  $\|Z\|_\infty = \max_i |z_i|$  and  $\|Z\|_1 = \sum_{i=1}^n |z_i|$ , respectively. For the  $l_\infty$  norm, we eliminate the subscript  $\infty$  and simply note  $\|Z\|$ . We call a vector  $P \in \mathbb{R}_{\geq 0}^p$  a probability vector when  $\|P\|_1 = 1$ . The  $l_\infty$  induced norm of matrix  $M$  is defined as  $\|M\| = \max_i \sum_{j=1}^n |m_{ij}|$ . The  $l_1$  to  $l_\infty$  induced norm of matrix  $M$  is defined as  $\|M\|_{l_1-l_\infty} = \max_{i,j} |m_{ij}|$ . We define  $\mathcal{R}[M]_i$  as the  $i^{th}$  row of  $M = [m_{ij}] \in \mathbb{R}^{m \times n}$ , which implies that

$$\mathcal{R}[M]_i = [m_{i1} \quad m_{i2} \quad \dots \quad m_{in}], \text{ for } i = 1, 2, \dots, m.$$

Now, we consider a biomolecular system with  $r$  species,  $S_1, \dots, S_r$ , and  $K$  reactions of the form:



where  $q_{kl} - p_{kl}$  is the change in the number of molecules of  $S_l$  by the  $k^{th}$  reaction and  $d_k$  stands for the  $k^{th}$  reaction

rate constant. Let  $s_i$ , for  $i = 1, 2, \dots, r$ , be the molecular count for each species as a discrete random variable and let  $\mathbf{s} = [s_1, s_2, \dots, s_r]^T$  be the state of the system. Then, for any  $\mathbf{s} \in \mathbb{Z}_{\geq 0}^r$ , the CME takes the form

$$\frac{\partial P(\mathbf{s}, t)}{\partial t} = \sum_{k=1}^K [-a_k(\mathbf{s})P(\mathbf{s}, t) + a_k(\mathbf{s} - \gamma_k)P(\mathbf{s} - \gamma_k, t)] \quad (2)$$

where  $\gamma_k$  is a stoichiometry vector and  $a_k(\mathbf{s})$  is a propensity function. When we let  $p_k = [p_{k1}, \dots, p_{kr}]^T$  and  $q_k = [q_{k1}, \dots, q_{kr}]^T$ , for  $k = 1, \dots, K$ , then  $\gamma_k = q_k - p_k$ .  $a_k(\mathbf{s})$  is proportional to  $d_k$  and  $a_k(\mathbf{s})dt$  is the probability that the  $k^{th}$  reaction takes place in an infinitesimal time step  $dt$  [11].

Let  $\Omega_s$  be the state space of all species, which implies  $\mathbf{s} \in \Omega_s$ . Since  $\Omega_s$  is a subset of a countable set  $\mathbb{Z}_{\geq 0}^r$ , it is also countable. Let  $\{\mathbf{s}_i\} = \{\mathbf{s}_1, \mathbf{s}_2, \dots\}$  be an enumeration of  $\Omega_s$ , and define  $S = [\mathbf{s}_1, \mathbf{s}_2, \dots]^T$ . Then according to [7], when we let  $P(S, t) = [P(\mathbf{s}_1, t), P(\mathbf{s}_2, t), \dots]^T$ , which is the probability density state vector at time  $t$ , (2) can be written as a single linear expression:

$$\begin{aligned}\frac{d}{dt}P(S, t) &= MP(S, t), \text{ given } P(S, t_0), \text{ where} \\ M_{ij} &= \begin{cases} -\sum_{k=1}^K a_k(\mathbf{s}_j) & \text{if } i = j, \\ a_k(\mathbf{s}_j) & \text{if } \mathbf{s}_j = \mathbf{s}_i - \gamma_k, \\ 0 & \text{Otherwise.} \end{cases}\end{aligned}\quad (3)$$

In this work, we consider biomolecular systems in which the chemical reactions take place on two time-scales. Let  $K_s$  be the number of slow reactions and  $K_f$  be the number of fast reactions where  $K_s + K_f = K$ . We are using a small positive parameter  $\epsilon (\ll 1)$ , which quantifies a time-scale separation between the slow and fast reactions. Then, we can separate propensity functions as slow reactions' propensity functions  $a_k(\mathbf{s})$ , for  $k = 1, 2, \dots, K_s$ , and fast reactions' propensity functions  $a_k(\mathbf{s})$ , for  $k = K_s + 1, K_s + 2, \dots, K_s + K_f = K$ . Based on the slow and fast reactions, we can define the slow and fast species as follow. Upon firing the fast reactions, slow species counts never change. On the other hand, for each fast species, there exists at least one fast reaction that changes the fast-species counts. According to [12], there exists a proper linear coordinate transformation that identifies the slow and fast species in the system. Let  $X_1, \dots, X_l$  be the slow species and  $Y_1, \dots, Y_q$  be the fast species, where  $l + q = r$ . Let  $x_i$ , for  $i = 1, 2, \dots, l$ ,  $y_j$ , for  $j = 1, 2, \dots, q$ , be the molecular count for each slow and fast species, respectively. Let  $\mathbf{x} = [x_1, x_2, \dots, x_l]^T$  and  $\mathbf{y} = [y_1, y_2, \dots, y_q]^T$  be the state of the system for each slow and fast species, respectively. Then  $\mathbf{s} \in \mathbb{Z}_{\geq 0}^r$  can be represented as  $\mathbf{s} = (\mathbf{x}, \mathbf{y})$ , where  $\mathbf{x} \in \mathbb{Z}_{\geq 0}^l$  stands for the slow-species counts and  $\mathbf{y} \in \mathbb{Z}_{\geq 0}^q$  stands for the fast-species counts. In addition, for  $k = 1, 2, \dots, K$ , propensity function  $a_k(\mathbf{s})$  can be written as  $a_k(\mathbf{x}, \mathbf{y})$ . Now we make the following assumptions:

*Assumption 2.1:* There exist nonnegative integers  $y_{tot}^j$  such that

$$0 \leq y_j \leq y_{tot}^j, \quad \text{for } j = 1, 2, \dots, q.$$

*Assumption 2.2:* All of the propensity functions are polynomial in  $\mathbf{s}$  [11]. In addition, the order of each polynomial is less than or equal to 2.

Assumption 2.1 requires that the number of fast species is bounded, which is reasonable in many biomolecular systems of practical interest [1]. Gillespie derived that propensity functions are polynomial under suitable conditions such as well-mixedness [11]. Assumption 2.2 states that the order of each polynomial for the propensity function is at most two because reactions are either uni-molecular or bi-molecular. This is a standard assumption considering that  $n$ -molecular reactions ( $n > 2$ ) have low probability compared to a sequence of bi-molecular reactions [1]. In addition, propensity functions can be written as follows [10] [13]:

$$\begin{aligned} a_k(\mathbf{x}, \mathbf{y}) &= b_k(\mathbf{x}) + c_k(\mathbf{x})\Psi_1(\mathbf{y}) + d_k\Psi_2(\mathbf{y}), k = 1, 2, \dots, K_s, \\ a_k(\mathbf{x}, \mathbf{y}) &= \frac{1}{\epsilon}(b_k(\mathbf{x}) + c_k(\mathbf{x})\Psi_1(\mathbf{y}) + d_k\Psi_2(\mathbf{y})), \\ k &= K_s + 1, K_s + 2, \dots, K, \end{aligned} \quad (4)$$

where  $\Psi_1(\mathbf{y})$  and  $\Psi_2(\mathbf{y})$  are defined in (1).  $b_k(\mathbf{x})$  is a polynomial in  $\mathbf{x}$  with order less than or equal to 2.  $c_k(\mathbf{x})$  is a matrix with appropriate dimensions, and each component of  $c_k(\mathbf{x})$  is a polynomial in  $\mathbf{x}$  with order less than or equal to 1.  $d_k$  is a constant matrix with appropriate dimension. The propensity functions of the slow reactions are an order of  $\epsilon$  of the propensity functions of the fast reactions.

We let  $\Omega_x$  and  $\Omega_y$ , be the state space of the slow species and the fast species, respectively. Since each slow-species counts is unbounded,  $\Omega_x = \mathbb{Z}_{\geq 0}^l$ . On the other hand, because of Assumption 2.1,  $\Omega_y \subset \mathbb{Z}_{\geq 0}^q$  and it is a finite set. Then  $\mathbf{x}$  and  $\mathbf{y}$  will be vectors of random variables taking values in the sets  $\Omega_x$  and  $\Omega_y$ , respectively. Then,  $\Omega_s = \Omega_x \times \Omega_y$ , which is a subset of  $\mathbb{Z}_{\geq 0}^r$ .  $\Omega_y$  is a countable set because it is a finite set, and  $\Omega_x$  is also a countable set because it is a finite product of countable sets. Let  $\{\mathbf{x}_i\}$  and  $\{\mathbf{y}_j\}$  be an enumeration of  $\Omega_x$  and  $\Omega_y$ , respectively.

Then for any  $\mathbf{x} \in \Omega_x$  and  $\mathbf{y} \in \Omega_y$ , we can rewrite the CME as

$$\begin{aligned} \frac{d}{dt}P(\mathbf{x}, \mathbf{y}, t) &= \sum_{k=1}^{K_s} [-a_k(\mathbf{x}, \mathbf{y})P(\mathbf{x}, \mathbf{y}, t) \\ &+ a_k(\mathbf{x} - \gamma_{x,k}, \mathbf{y} - \gamma_{y,k})P(\mathbf{x} - \gamma_{x,k}, \mathbf{y} - \gamma_{y,k}, t)] \\ &+ \sum_{k=K_s+1}^K [-a_k(\mathbf{x}, \mathbf{y})P(\mathbf{x}, \mathbf{y}, t) \\ &+ a_k(\mathbf{x}, \mathbf{y} - \gamma_{y,k})P(\mathbf{x}, \mathbf{y} - \gamma_{y,k}, t)], \end{aligned} \quad (5)$$

where, for  $k = 1, 2, \dots, K$ ,  $a_k(\mathbf{x}, \mathbf{y})$  are propensity functions for the slow and fast reactions defined in (4), and  $\gamma_{x,k}$  and  $\gamma_{y,k}$  are corresponding stoichiometry vectors, for the slow species and the fast species, respectively [5]. For the fast reactions ( $k = K_s + 1, \dots, K$ ),  $\gamma_{x,k}$  is 0, because the slow-species counts are not changed by the fast reactions.

### III. BASIC SETUP

In this section, we define ODEs for the slow-species counts' marginal probability distribution,  $P(\mathbf{x}, t)$ , and for the fast-species counts' first  $n$  conditional moments,  $Y_n(\mathbf{x}, t)$ , based

on (5). To proceed, we define  $P(\mathbf{y}, t|\mathbf{x})$  as the conditional probability distribution of the fast-species counts given the slow-species counts. Then, these two distributions,  $P(\mathbf{x}, t)$  and  $P(\mathbf{y}, t|\mathbf{x})$ , jointly specify the full distribution  $P(\mathbf{s}, t) = P(\mathbf{x}, \mathbf{y}, t)$  via  $P(\mathbf{x}, \mathbf{y}, t) = P(\mathbf{x}, t)P(\mathbf{y}, t|\mathbf{x})$ , by Bayes' theorem. Then we define

$$\begin{aligned} \mu_w(\mathbf{x}, t) &= \mathbb{E}[\Psi_w(\mathbf{y})|\mathbf{x}] = \sum_{\mathbf{y} \in \Omega_y} \Psi_w(\mathbf{y})P(\mathbf{y}, t|\mathbf{x}), \\ Y_n(\mathbf{x}, t) &= [\mu_1(\mathbf{x}, t)^T, \mu_2(\mathbf{x}, t)^T, \dots, \mu_n(\mathbf{x}, t)^T]^T, \end{aligned} \quad (6)$$

for any  $\mathbf{x} \in \Omega_x$ ,  $w \in \mathbb{Z}_{\geq 0}$ ,  $1 \leq n \leq |\Omega_y|$ , where  $\mu_w(\mathbf{x}, t)$  and  $Y_n(\mathbf{x}, t)$  denote the fast-species counts'  $w^{\text{th}}$  and first  $n$  conditional moments, respectively. For  $w = 1, 2, \dots, n$ , let  $f_w$  be a matrix whose multiplication with  $Y_n(\mathbf{x}, t)$  isolates  $\mu_w(\mathbf{x}, t)$ , i.e.,

$$\mu_w(\mathbf{x}, t) = f_w Y_n(\mathbf{x}, t). \quad (7)$$

Now we can derive ODEs from (5) for the slow-species counts' marginal probability distribution,  $P(\mathbf{x}, t)$ , as in (8):

*Proposition 3.1:* [10] For the CME in (5) with Assumptions 2.1 and 2.2, given  $\mathbf{x} \in \Omega_x$ , we have

$$\begin{aligned} \frac{d}{dt}P(\mathbf{x}, t) &= \sum_{k=1}^{K_s} [-\mathbb{E}[a_k(\mathbf{x}, \mathbf{y})|\mathbf{x}]P(\mathbf{x}, t) \\ &+ \mathbb{E}[a_k(\mathbf{x} - \gamma_{x,k}, \mathbf{y})|\mathbf{x} - \gamma_{x,k}]P(\mathbf{x} - \gamma_{x,k}, t)]. \end{aligned} \quad (8)$$

By using (4), we can further express the conditional expectation of propensity function  $a_k(\mathbf{x}, \mathbf{y})$  as

$$\begin{aligned} \mathbb{E}[a_k(\mathbf{x}, \mathbf{y})|\mathbf{x}] &= \sum_{\mathbf{y}} a_k(\mathbf{x}, \mathbf{y})P(\mathbf{y}, t|\mathbf{x}) \\ &= b_k(\mathbf{x}) + c_k(\mathbf{x})\mu_1(\mathbf{x}, t) + d_k\mu_2(\mathbf{x}, t), \end{aligned} \quad (9)$$

for given  $\mathbf{x}$  and  $1 \leq k \leq K_s$ . When we let  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots]^T$ , according to Munsky [7], the slow-species counts' marginal probability distribution in (8) can be written as a single linear expression:

$$\frac{d}{dt}P(X, t) = A(Y_2(\mathbf{x}, t))P(X, t), \text{ given } P(X, t_0), \quad (10)$$

$$\text{where, } P(X, t) = [P(\mathbf{x}_1, t), P(\mathbf{x}_2, t), \dots, P(\mathbf{x}_i, t), \dots]^T \quad (11)$$

is the slow-species counts' marginal probability distribution vector at time  $t$  and

$$A_{ij} = \begin{cases} -\sum_{k=1}^{K_s} \mathbb{E}[a_k(\mathbf{x}_j, \mathbf{y})|\mathbf{x}_j] & \text{if } i = j, \\ \mathbb{E}[a_k(\mathbf{x}_j, \mathbf{y})|\mathbf{x}_j] & \text{if } \mathbf{x}_j = \mathbf{x}_i - \gamma_{x,k}, 1 \leq k \leq K_s, \\ 0 & \text{Otherwise.} \end{cases} \quad (12)$$

In (10),  $P(X, t)$  is an infinite dimensional vector, because  $\Omega_x$ , the state space of the slow species is infinite. We can also derive ODEs from (5) for the fast-species counts' first  $n$  conditional moments,  $Y_n(\mathbf{x}, t)$ , as in (13):

*Proposition 3.2:* [10] For the CME in (5) with Assumptions 2.1 and 2.2, for  $\mathbf{x} \in \Omega_x$  and  $1 \leq n \leq |\Omega_y|$ , we have

$$\frac{d}{dt}Y_n(\mathbf{x}, t) = C(\mathbf{x})Y_n(\mathbf{x}, t) + c_1(\mathbf{x}) + c_2\mu_{n+1}(\mathbf{x}, t) + \epsilon G(t). \quad (13)$$

To proceed further, let us define some notations. Let  $I_s = \{i_1, i_2, \dots, i_a\}$  and  $I_x = \{i_1, i_2, \dots, i_b\}$  denote a finite ordered index set for all species and slow species, respectively.

Then, for given  $I_s$  and  $I_x$ , we can define a truncated finite state space,  $\Omega_s^{I_s}$  and  $\Omega_x^{I_x}$ , as follows:

$$\Omega_s^{I_s} = \{\mathbf{s}_{i_1}, \mathbf{s}_{i_2}, \dots, \mathbf{s}_{i_a}\}, \Omega_x^{I_x} = \{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_b}\}.$$

For any matrix  $A$  and a given ordered index set  $I$ , let  $A_I$  denote the principal submatrix of  $A$ , in which both rows and columns have been chosen and ordered according to  $I$ . For example, when  $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ ,  $I = \{2, 3\}$ , then  $A_I = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}$ . In addition, for any vector  $X$ ,  $X_I$  is the vector of those elements of  $X$  indexed by  $I$ . For example, when  $X = [0.3, 0.7, 0.1]^T$ ,  $I = \{3, 2\}$ , then  $X_I = [0.1, 0.7]^T$ .

Based on (3), and a given finite ordered index set  $I_s$ , we define  $P^{I_s}(S, t)$  as solutions of

$$\frac{d}{dt} P^{I_s}(S, t) = M_I P^{I_s}(S, t), P^{I_s}(S, t_0) = P_{I_s}(S, t_0), \quad (14)$$

where  $P^{I_s}(S, t)$  is the approximated finite dimensional probability distribution. Also, based on (10), and a given finite ordered index set  $I_x$ , we define  $P^{I_x}(X, t)$  as solutions of

$$\frac{d}{dt} P^{I_x}(X, t) = [A(Y_2(\mathbf{x}, t))]_{I_x} P^{I_x}(X, t), P^{I_x}(X, t_0) = P_{I_x}(X, t_0), \quad (15)$$

where  $P^{I_x}(X, t)$  is the approximated finite dimensional marginal probability distribution of the slow-species counts.

Although, we are not assuming that the slow species' count is finite, we iteratively truncate the space space of the slow species to  $I_x$  and compute the error due to such a truncation. If the error is larger than a desired value, we enlarge  $I_x$  and repeat the procedure. This process is streamlined in Section 7, where we propose the EFSP algorithm. Therefore, for now, we assume that  $I_x$  is a given finite set. Then, for the given  $I_x$ , let  $\Sigma_{true}$  be

$$\Sigma_{true} : \begin{cases} \frac{d}{dt} P^{I_x}(X, t) = [A(Y_2(\mathbf{x}, t))]_{I_x} P^{I_x}(X, t), \\ \epsilon \frac{d}{dt} Y_n(\mathbf{x}, t) = C(\mathbf{x}) Y_n(\mathbf{x}, t) + c_1(\mathbf{x}) \\ + c_2 \mu_{n+1}(\mathbf{x}, t) + \epsilon G(t), \end{cases} \quad (16)$$

for any  $\mathbf{x} \in \Omega_x^{I_x}$ . When  $n = |\Omega_y|$ ,  $\Sigma_{true}$  is closed, because  $\mu_{n+1}(\mathbf{x}, t)$  can be represented as an affine function of  $Y_n(\mathbf{x}, t)$  [5]. However, in general, when  $1 \leq n < |\Omega_y|$ , the dynamics of the fast-species counts' conditional moments are not closed, because  $\mu_{n+1}(\mathbf{x}, t)$  is not a function of  $Y_n(\mathbf{x}, t)$  anymore. To solve this problem, we apply a robust conditional moment closure method to approximate  $\mu_{n+1}(\mathbf{x}, t)$  as a function of  $Y_n(\mathbf{x}, t)$  to close the dynamics. The next section proposes the robust moment closure technique developed by [13], and it can be applied to the conditional moments.

#### IV. ROBUST CONDITIONAL MOMENT CLOSURE

The Robust Moment Closure (RMC) was originally developed by [13] for the moments dynamics. Here we modify it so to make it applicable to the dynamics of the conditional moments. For any  $\mathbf{x} \in \Omega_x^{I_x}$ , we define the fast-species counts' conditional probability distribution as

$$P_{Y|X}(\mathbf{x}, t) = [P(\mathbf{y}_1, t|\mathbf{x}), \dots, P(\mathbf{y}_j, t|\mathbf{x}), \dots, P(\mathbf{y}_{|\Omega_y|}, t|\mathbf{x})]^T.$$

Then for each  $\mu_{n+1}(\mathbf{x}, t)$  and  $Y_n(\mathbf{x}, t)$ , there exists unique  $H_n$  and  $V_n$  that satisfy

$$\mu_{n+1}(\mathbf{x}, t) = H_n P_{Y|X}(\mathbf{x}, t), Y_n(\mathbf{x}, t) = V_n P_{Y|X}(\mathbf{x}, t).$$

For example, when  $q = 1$ ,

$$H_n = [0 \quad 1^{n+1} \quad 2^{n+1} \quad \dots \quad (|\Omega_y|)^{n+1}], \quad (17)$$

$$V_n = \begin{bmatrix} 0 & 1 & 2 & \dots & |\Omega_y| \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1^n & 2^n & \dots & |\Omega_y|^n \end{bmatrix}. \quad (18)$$

Our objective is to approximate  $\mu_{n+1}(\mathbf{x}, t)$  as a function of  $Y_n(\mathbf{x}, t)$ , so that (15) becomes closed, denoted as

$$\mu_{n+1}(\mathbf{x}, t) \approx \phi(Y_n(\mathbf{x}, t)),$$

where  $\phi(\cdot)$  can be a nonlinear function. Since the conditional probability distribution,  $P_{Y|X}(\mathbf{x}, t)$ , is not known,  $\phi(\cdot)$  should be chosen such that the worst-case error between  $\mu_{n+1}(\mathbf{x}, t)$  and  $\phi(Y_n(\mathbf{x}, t))$  is minimized. Therefore, the following min-max problem:

$$\inf_{\phi} \sup_{P_{Y|X}(\mathbf{x}, t) \in \mathbb{P}} \|\mu_{n+1}(\mathbf{x}, t) - \phi(Y_n(\mathbf{x}, t))\|, \quad (19)$$

should be solved. According to [13], without a priori information on the conditional probability distribution,  $P_{Y|X}(\mathbf{x}, t)$ , a solution for (19) is obtained when  $\phi(Y_n(\mathbf{x}, t))$  is an affine function of  $Y_n(\mathbf{x}, t)$ , which can be written as

$$\phi(Y_n(\mathbf{x}, t)) = K Y_n(\mathbf{x}, t) + K_0.$$

In addition, we can obtain  $K$  and  $K_0$  by solving the linear program

$$\begin{aligned} \min_{K_0, K} \quad & \gamma \\ \text{s.t.} \quad & -\gamma \mathbf{1}^T \leq \mathcal{R}[H_n - (K V_n + K_0 \mathbf{1}^T)]_i \leq \gamma \mathbf{1}^T \end{aligned} \quad (20)$$

for  $i = 1, 2, \dots, p$ , where  $p$  is the number of rows in  $H_n$ . Let the linear program in (20)'s object value be  $\rho_n$ . When  $n$  is fixed,  $\rho_n$  is a constant. Then

$$\begin{aligned} & \|\mu_{n+1}(\mathbf{x}, t) - \phi(Y_n(\mathbf{x}, t))\| = \\ & \|H_n P_{Y|X}(\mathbf{x}, t) - (K V_n P_{Y|X}(\mathbf{x}, t) + K_0)\|, \end{aligned}$$

which is the approximation error between  $\mu_{n+1}(\mathbf{x}, t)$  and  $\phi(Y_n(\mathbf{x}, t))$ , is bounded by  $\rho_n$  for any  $P_{Y|X}(\mathbf{x}, t)$ .

On the right-hand side of  $\Sigma_{true}$ , we substitute  $K Y_n(\mathbf{x}, t) + K_0$  for  $\mu_{n+1}(\mathbf{x}, t)$ , and obtain

$$\Sigma_{closed} : \begin{cases} \frac{d}{dt} \tilde{P}^{I_x}(X, t) = [A(\tilde{Y}_2^\epsilon(\mathbf{x}, t))]_{I_x} \tilde{P}^{I_x}(X, t) = \tilde{A}(t) \tilde{P}^{I_x}(X, t) \\ \epsilon \frac{d}{dt} \tilde{Y}_n^\epsilon(\mathbf{x}, t) = C(\mathbf{x}) \tilde{Y}_n^\epsilon(\mathbf{x}, t) + c_1(\mathbf{x}) \\ + c_2 (K \tilde{Y}_n^\epsilon(\mathbf{x}, t) + K_0) + \epsilon G(t). \end{cases} \quad (21)$$

Let us define  $\tilde{\mu}_w^\epsilon(\mathbf{x}, t) = f_w \tilde{Y}_n^\epsilon(\mathbf{x}, t)$ .

*Remark 4.1:* If  $C(\mathbf{x}) + c_2 K$  is a stable matrix, the approximation error between  $\mu_w(\mathbf{x}, t)$  and  $\tilde{\mu}_w^\epsilon(\mathbf{x}, t)$  is bounded. To ensure the stability by construction, we can augment (20) with a linear matrix inequality and conduct an iterative algorithm. This procedure is in [10]. Here we assume that  $C(\mathbf{x}) + c_2 K$  is a stable matrix.

## V. TIME-SCALE SEPARATION

We can check that (21) is written in standard singular perturbation form [9]. Let  $\tilde{Y}_n^0(\mathbf{x})$  be the solution of

$$C(\mathbf{x})\tilde{Y}_n^0(\mathbf{x}) + c_1(\mathbf{x}) + c_2(K\tilde{Y}_n^0(\mathbf{x}) + K_0) = 0, \quad (22)$$

which can be obtained by letting  $\epsilon = 0$  in (21). Since  $C(\mathbf{x}) + c_2K$  is a stable matrix,  $Y_n^\epsilon(\mathbf{x}, t)$  converges exponentially fast to  $\tilde{Y}_n^0(\mathbf{x})$  [9]. By replacing  $\tilde{Y}_2^\epsilon(\mathbf{x}, t)$  with  $\tilde{Y}_2^0(\mathbf{x})$  on the right-hand side of (21), we can obtain

$$\Sigma_{reduced} : \left\{ \frac{d}{dt} \bar{P}^{I_x}(X, t) = [A(\tilde{Y}_2^0(\mathbf{x}))]_{I_x} \bar{P}^{I_x}(X, t) = \bar{A} \bar{P}^{I_x}(X, t) \right\}, \quad (23)$$

$\Sigma_{reduced}$  is composed of the slow species only, because we approximate the conditional moments of the fast-species counts,  $\tilde{Y}_2^\epsilon(\mathbf{x}, t)$ , as functions of the slow-species counts,  $\tilde{Y}_2^0(\mathbf{x})$ .  $\Sigma_{reduced}$  is a positive system if and only if  $\bar{A}$  is a Metzler matrix [14]. However, this is not guaranteed in general. Because of (9) and (12), any off-diagonal element of  $\bar{A}$  has the form

$$b_k(\mathbf{x}) + c_k(\mathbf{x})\tilde{\mu}_1^0(\mathbf{x}) + d_k\tilde{\mu}_2^0(\mathbf{x}), \quad (24)$$

where  $\tilde{\mu}_w^0(\mathbf{x}) = f_w\tilde{Y}_n^0(\mathbf{x})$ , for  $w = 1, 2$ . Therefore,  $\bar{A}$  is a Metzler matrix if and only if (24) is non-negative for all  $\mathbf{x} \in \Omega_x^{I_x}$ . For a given  $\mathbf{x} \in \Omega_x^{I_x}$  and  $k = 1, 2, \dots, K_s$ , we define a linear program

$$\begin{aligned} \min_{h_1(\mathbf{x}), h_2(\mathbf{x})} \quad & \|h_1(\mathbf{x}) - \tilde{\mu}_1^0(\mathbf{x})\| + \|h_2(\mathbf{x}) - \tilde{\mu}_2^0(\mathbf{x})\| \\ \text{s.t.} \quad & b_k(\mathbf{x}) + c_k(\mathbf{x})h_1(\mathbf{x}) + d_k h_2(\mathbf{x}) \geq 0. \end{aligned} \quad (25)$$

Let the optimal solution to (25) be  $h_1(\mathbf{x}) = \hat{\mu}_1(\mathbf{x})$  and  $h_2(\mathbf{x}) = \hat{\mu}_2(\mathbf{x})$  and the object value be  $\lambda^x$ . By replacing  $\tilde{\mu}_1^0(\mathbf{x})$  and  $\tilde{\mu}_2^0(\mathbf{x})$  with  $\hat{\mu}_1(\mathbf{x})$  and  $\hat{\mu}_2(\mathbf{x})$  in  $\Sigma_{reduced}$ , we obtain

$$\Sigma_{final} : \left\{ \frac{d}{dt} \hat{P}^{I_x}(X, t) = [A(\hat{Y}_2(\mathbf{x}))]_{I_x} \hat{P}^{I_x}(X, t) = \hat{A} \hat{P}^{I_x}(X, t) \right\}, \quad (26)$$

where  $\hat{Y}_2(\mathbf{x}) = [\hat{\mu}_1(\mathbf{x})^T, \hat{\mu}_2(\mathbf{x})^T]^T$ . In (26),  $\Sigma_{final}$  is a positive system because  $\hat{A}$  is a Metzler matrix, which is guaranteed by (25). Furthermore,  $\hat{A}$  is a stable matrix and both  $\|\hat{\mu}_1(\mathbf{x}) - \tilde{\mu}_1^0(\mathbf{x})\|$  and  $\|\hat{\mu}_2(\mathbf{x}) - \tilde{\mu}_2^0(\mathbf{x})\|$  are bounded by  $\lambda^x$ .

*Remark 5.1:* [10] When  $n = |\Omega_y|$ , both  $\rho_n$  and  $\lambda^x$  are 0. Now the approximation errors for both the fast-species counts' conditional moments and the slow-species counts' marginal probability distribution should be quantified.

## VI. ERROR QUANTIFICATION

### A. Conditional Moments of the Fast-Species Counts

Here, we first consider the fast-species counts' conditional moments. The following theorem derives the approximation error between the fast-species counts' conditional moments of  $\Sigma_{true}$ ,  $\mu_w(\mathbf{x}, t)$ , and those of  $\Sigma_{final}$ ,  $\hat{\mu}_w(\mathbf{x})$ .

*Theorem 6.1:* [10] Given  $t_f > t_0 > 0$  and  $\mathbf{x} \in \Omega_x^{I_x}$ , for a sufficiently small  $\epsilon$ , the approximation error between  $\mu_w(\mathbf{x}, t)$  and  $\hat{\mu}_w(\mathbf{x})$  satisfies

$$\sup_{t \in [t_0, t_f]} \|\mu_w(\mathbf{x}, t) - \hat{\mu}_w(\mathbf{x})\| \leq \Delta_{w, \epsilon}^x + \lambda^x + O(\epsilon),$$

for  $w = 1, 2$ , where,

$$\Delta_{w, \epsilon}^x = \int_{t_0}^{t_f} \|f_w \exp\{\frac{1}{\epsilon}(C(\mathbf{x}) + c_2K)(t_f - \tau)\}\| d\tau \frac{\rho_n}{\epsilon} \|c_2\|.$$

Furthermore, there exist  $\Delta_\epsilon > 0$  and  $\epsilon^* > 0$  such that  $\sup_{t \in [t_0, t_f]} \|\mu_w(\mathbf{x}, t) - \hat{\mu}_w(\mathbf{x})\| \leq \Delta_\epsilon + O(\epsilon)$  for all  $\mathbf{x} \in \Omega_x^{I_x}$ ,  $\epsilon \in (0, \epsilon^*)$  and  $w = 1$  or  $2$ .

### B. Marginal Probability Distribution of the Slow-Species Counts

Next, we quantify the approximation error of the slow-species counts' marginal probability distribution by using Theorem 6.1. In  $\Sigma_{final}$ , we construct a positive system by making  $\hat{A}$  Metzler and stable matrix. Now, let us regard  $\Sigma_{final}$  as our nominal system and (15) as the perturbed system. Then we can express the perturbed system as follows:

$$\frac{d}{dt} P^{I_x}(X, t) = (\hat{A} + \Delta_1(t)) P^{I_x}(X, t), \quad (27)$$

where

$$\Delta_1(t) = [A(Y_2(\mathbf{x}, t))]_{I_x} - \hat{A}.$$

By using Theorem 6.1, we can prove that  $l_1 - l_\infty$  norm of  $\Delta_1(t)$  is bounded.

*Lemma 6.2:* [10] There is a constant  $k_1$  such that

$$\|\Delta_1(t)\|_{l_1 - l_\infty} \leq k_1 \Delta_\epsilon + O(\epsilon).$$

Now we can quantify the approximation error between the slow-species counts' marginal probability distribution of the original CME,  $P^{I_x}(X, t)$ , and those of  $\Sigma_{final}$ ,  $\hat{P}^{I_x}(X, t)$ , as follows:

*Theorem 6.3:* [10] Given  $t_f > t_0 > 0$ , the approximation error between  $P^{I_x}(X, t)$  and  $\hat{P}^{I_x}(X, t)$  satisfies

$$\begin{aligned} & \sup_{t \in [t_0, t_f]} \|P^{I_x}(X, t) - \hat{P}^{I_x}(X, t)\| \\ & \leq k_1 \int_{t_0}^{t_f} \left\| \exp\left\{\hat{A}\right\}(t_f - \tau) \right\| d\tau \Delta_\epsilon + O(\epsilon). \end{aligned}$$

*Corollary 6.4:* [10] As  $\epsilon \rightarrow 0$ , the right-hand side of the inequality in Theorem 6.3 goes to  $k\Delta_0$ , where

$$\begin{aligned} k &= k_1 \int_{t_0}^{t_f} \left\| \exp\left\{\hat{A}(t_f - \tau)\right\} \right\| d\tau, \\ \Delta_0 &= \lim_{\epsilon \rightarrow 0} \Delta_\epsilon = \sup_{w \in \{1, 2\}, \mathbf{x} \in \Omega_x^{I_x}} (\Delta_{w, 0}^x + \lambda^x), \text{ and } \Delta_{w, 0}^x = \\ & \lim_{\epsilon \rightarrow 0} \Delta_{w, \epsilon}^x = \int_0^\infty \|f_w \exp\{(C(\mathbf{x}) + c_2K)t\}\| dt \rho_n \|c_2\|. \end{aligned}$$

*Remark 6.5:* When  $n = |\Omega_y|$ , the right-hand side of the inequality in Theorem 6.3 goes to  $O(\epsilon)$ . This is because when  $n = |\Omega_y|$ , both  $\rho_n$  and  $\lambda^x$  go to 0 by Remark 5.1, so  $\Delta_\epsilon$  goes to 0.

Corollary 6.4 and Remark 6.5 show that the approximation error between the slow-species counts' marginal probability distribution of the original system and those of the approximated system decreases as  $\epsilon$  decreases. Now we should discuss about how to find  $I_x$ , that approximate the infinite dimensional CME as a finite dimensional CME that contains the slow species only.

## VII. THE FSP AND THE EFSP ALGORITHMS

In this section, we first illustrate the FSP algorithm developed by [7] in general, for (2) and (3). Then we propose the EFSP algorithm which can be applied to (10), where two time-scale exists.

### A. The FSP algorithm in general

The FSP algorithm provides a systematic method to find a finite ordered index set  $I_s$ , so that the approximated probability distribution,  $P^{I_s}(S, t)$  in (14), is sufficiently close to the original infinite-dimensional probability distribution,  $P(S, t)$ . In particular, the solution of (14) for  $t \in [t_0, t_f]$  is

$$P^{I_s}(S, t) = \exp(M_{I_s} t) P^{I_s}(S, t_0).$$

According to [7], for any pair of index sets,  $I_{s,1} \subset I_{s,2}$ ,

$$[\exp(M_{I_{s,2}})]_{I_{s,1}} \geq \exp(M_{I_{s,1}}) \geq 0. \quad (28)$$

Since the probability density vector  $P(S, t)$  is always non-negative, (28) guarantees that

$$[\exp(M_{I_{s,2}} t_f)]_{I_{s,1}} P^{I_{s,1}}(S, t_0) \geq \exp(M_{I_{s,1}} t_f) P^{I_{s,1}}(S, t_0). \quad (29)$$

This result assures that when we gradually expand ordered index set  $I_{s,j}$  (as  $I_{s,1} \subset I_{s,2} \dots \subset I_{s,j} \dots$ ), the approximation monotonically improves.

In addition, for given  $\delta > 0$ ,  $t_f \geq 0$  and  $I_s$ , if

$$\left\| \exp(M_{I_s} t_f) P^I(S, t_0) \right\|_{I_{s,1}} = \mathbf{1}^T [\exp(M_{I_s} t_f) P^{I_s}(S, t_0)] \geq 1 - \delta \quad (30)$$

$$\begin{aligned} &\text{is satisfied then } \exp(M_{I_s} t_f) P^{I_s}(S, t_0) \\ &\leq P_{I_s}(S, t_f) \leq \exp\{(M_{I_s} t_f)\} P^{I_s}(S, t_0) + \delta \mathbf{1} \end{aligned} \quad (31)$$

is guaranteed [7]. (30) and (31) imply that the approximate solution,  $P^{I_s}(S, t_f) = \exp(M_{I_s} t_f) P^{I_s}(S, t_0)$  never exceeds the actual solution,  $P_{I_s}(S, t_f)$ , and  $\|P_{I_s}(S, t_f) - P^{I_s}(S, t_f)\| \leq \delta$ , when (30) is satisfied.

We can depict the underlying idea of the FSP algorithm, by representing all possible states as nodes on an infinite  $r$ -dimensional integer lattice, where  $r$  is the number of species in the biomolecular system and each node corresponds to distinct state,  $s_i$ . Fig. 1 (Top) of [7] shows a lattice for  $r = 2$ . Here, infinite lattice is projected onto the finite subset enclosed by the gray square, that corresponds to an index set  $I_s$ . This projected state space is shown in Fig. 1 (Bottom) of [7], where  $I_s$  represents the truncated state space, and  $I'_s$  represents the complement of  $I_s$ , where  $I'_s$  is aggregated to a single point. (14) illustrates the truncated CME with the index set  $I_s$ . (14) reflects transitions between states within  $I_s$  as well as reactions that starts from  $I_s$  and end in  $I'_s$ . However, the equation ignores reactions that begin in  $I'_s$  and end in  $I_s$  or  $I'_s$ . (29) shows that as the index set  $I_s$  increases, more trajectories are maintained and the probability of remaining in  $I_s$  increases. (31) shows that the probability that the original infinite dimensional system is currently in  $I_s$  must be larger than or equal to the probability that the system has stayed in  $I_s$  for all times,  $t \in [t_0, t_f]$ .

Now we can apply the FSP algorithm to (3) to find a finite ordered index set  $I_s$  that truncates the original infinite state space to a finite state space such that the error

$$\|P^{I_s}(S, t) - P_{I_s}(S, t)\| \leq \delta.$$

#### \* The Finite State Projection Algorithm

Step 0.

Choose the final time of interest,  $t_f$ .

Specify the acceptable error,  $\delta > 0$ .

Choose an initial finite set of states,  $I_0$  for the FSP.

Initialize a counter,  $j = 0$ .

Step 1.

Get  $\Gamma_{I_{s,j}} = \mathbf{1}^T P^{I_{s,j}}(S, t_f) = \mathbf{1}^T \exp(M_{I_{s,j}} t_f) P^{I_{s,j}}(S, t_0)$ .

Step 2.

If  $\Gamma_{I_{s,j}} \geq 1 - \delta$ :  $I_s = I_{s,j}$ , Stop.

$P^{I_s}(S, t_f)$  approximates  $P_{I_s}(S, t_f)$  within error  $\delta$ .

Else: Go to Step 3.

Step 3.

Add more states to  $I_{s,j}$  and obtain  $I_{s,j+1}$ .

Increment  $j$  as  $j + 1$  and return to Step 1.

In Step 0,  $I_0$  can be determined based on the initial probability distribution  $P(S, t_0)$ . In Step 3 of the FSP algorithm, a method to expand  $I_{s,j}$  to  $I_{s,j+1}$  is not explicitly stated. There may be many methods to expand the state space, and [7] illustrates one way to perform the expansion, called  $N$ -step reachability. Let  $I_{s,0}$  be the initial state and define  $I_{s,j}$  inductively. Let  $I_{s,j+1}$  contain all states in  $I_{s,j}$  combined with all states which can be reached from  $I_{s,j}$  in 1 reaction. Then,  $I_{s,j}$  denotes the set of all states which can be reached from the initial state in  $j$  or fewer reactions. This is how the algorithm expands the state space in Step 3. Munsky showed that for sufficiently large  $j$ ,  $\Gamma_{I_{s,j}} = \mathbf{1}^T \exp(M_{I_{s,j}} t_f) P^{I_{s,j}}(S, t_0) \geq 1 - \delta$  is satisfied [7]. In addition, for  $I$  that we find in the FSP algorithm, we have that

$$\|P_{I_s}(S, t) - P^{I_s}(S, t)\| \leq \delta,$$

for  $t \in [t_0, t_f]$  is guaranteed.

However, when multiple time scales exist, the FSP algorithm suffers from computational issues for two reasons. First, when the algorithm expands  $I_{s,j}$  in Step 3, it equally treats the transition rates between states, even though transitions by the fast reactions are much more probable than transitions by the slow reactions. Second, to compute  $\Gamma_{I_{s,j}}$  in Step 1, the algorithm has to solve the full CME at  $t = t_f$ , to obtain  $P^{I_{s,j}}(S, t_f)$ , which contains both slow and fast species. To handle these problems, we propose the EFSP algorithm, when two time-scale exists. In Step 3 of the EFSP algorithm, we aggregate the fast species and apply the  $N$ -step reachability to the slow species, by using the fact that transitions by the fast reactions are much more probable than transitions by the slow reactions. In Step 1 of the EFSP algorithm, we use the model reduction technique to approximate the original CME as  $\Sigma_{final}$ , which contains the slow species only, and solve the approximated CME at  $t = t_f$ .

### B. The EFSP algorithm where two time-scale exists

The EFSP algorithm provides a systematic method to find a finite ordered index set  $I_x$ , so that the approximated slow-species counts' marginal probability distribution,  $\hat{P}^{I_x}(X, t)$  in (26), is sufficiently close to the original slow-species

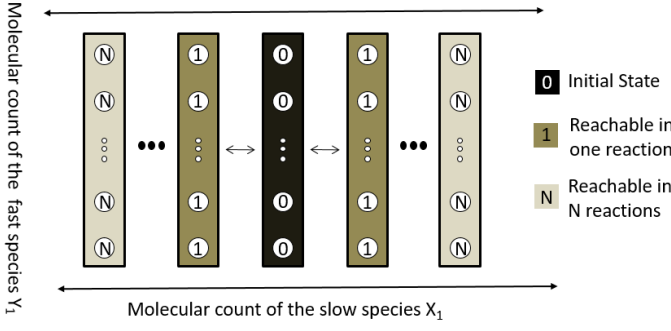


Fig. 1: Here, we visualize a two dimensional lattice, when one slow species ( $X_1$ ) and one fast species ( $Y_1$ ) exist. Based on the fact that moving vertically is much more probable then moving horizontally, we aggregate the fast species and apply the  $N$ -step reachability to the slow species.

counts' marginal probability distribution,  $P(X, t)$ , which is an infinite dimensional vector.

\* *The Enhanced Finite State Projection Algorithm*

Step 0.

Choose the final time of interest,  $t_f$ .

Specify the acceptable error,  $\delta > 0$ .

Choose an initial finite set of states,  $I_{x,0}$  for the FSP.

Initialize a counter,  $i = 0$ .

Step 1.

Get  $\Gamma_{I_{x,i}} = \mathbb{1}^T \hat{P}^{I_{x,i}}(X, t_f) = \mathbb{1}^T \exp(\hat{A}_{I_{x,i}} t_f) \hat{P}^{I_{x,i}}(X, t_0)$ .

Step 2.

If  $\Gamma_{I_{x,i}} \geq 1 - \delta$ :  $I_x = I_{x,i}$ , Stop.

$\hat{P}^{I_{x,i}}(X, t_f)$  approximates  $P_{I_x}(X, t_f)$  within error  $\delta$ .

Else: Go to Step 3.

Step 3.

Add more elements to  $I_{x,i}$  and obtain  $I_{x,i+1}$ .

Increment  $i$  as  $i + 1$  and return to Step 1.

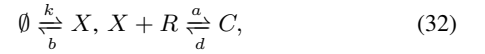
In Step 3 of the EFSP algorithm, a method to expand  $I_{x,i}$  to  $I_{x,i+1}$  is not explicitly stated. Fig. 1 shows a two-dimensional integer lattice when only one species for both slow ( $X_1$ ) and fast species ( $Y_1$ ) exist. Based on the fact that moving horizontally (by the fast reactions) is much more probable than moving vertically (by the slow reactions), we aggregate all state, that have the same slow-species counts, as a single state, and apply the  $N$ -step reachability procedure to the slow species only. It solves the first computational issue of the original FSP algorithm.

In Step 1, unlike the original FSP algorithm, which has to calculate original CME that contains both fast and slow species, the EFSP algorithm relies on the approximated CME,  $\Sigma_{final}$ , that contains the slow species only. It solves the second computational issue of the original FSP algorithm. To obtain  $\hat{P}^{I_{x,i}}(X, t_f)$  at  $i^{th}$  iteration, we have to approximate  $Y_n(\mathbf{x}, t)$  as  $\hat{Y}_n(\mathbf{x})$  for all  $\mathbf{x} \in \Omega_x^{I_{x,i}}$ . However,  $\hat{Y}_n(\mathbf{x})$  for  $\mathbf{x} \in \Omega_x^{I_{x,i-1}}$  is already calculated at  $(i-1)^{th}$  iteration, so we additionally need to calculate  $\hat{Y}_n(\mathbf{x})$  only for  $\mathbf{x} \in \Omega_x^{I_{x,i}} \setminus \Omega_x^{I_{x,i-1}}$ . In addition, unlike other time-scale separation methods, which rely on the fast-species counts'

stationary conditional probability distributions, our model reduction technique relies on only the first few conditional moments of the fast-species counts, which provides significant computational advantage. Approximation error occurs by the model reduction and the truncation of the state space. Corollary 6.4 and Remark 6.5 assure that the approximation error by the model reduction is small enough for sufficiently small  $\epsilon$ . In addition, [7] shows that for sufficiently large projection space, approximation error by truncating the state space is sufficiently small. Therefore, even though there is no guarantee that the EFSP algorithm converges, we expect that this algorithm will converge for sufficiently small  $\epsilon$  and large truncated state space. To show the utility of our algorithm, we consider a protein binding example.

### VIII. EXAMPLE

In this section, we consider a protein binding reaction to show the utility of our method. We can write the chemical reactions as follows [1]:



where  $X$ ,  $R$ ,  $C$  is protein, promoter and complex, respectively, and  $k$ ,  $b$ ,  $a$ ,  $d$  is protein's production and decay rate, promoter's binding and unbinding reaction rate, respectively. In (32), there exists a constant  $R_t$  that satisfies  $R_t = R + C$ , because the total concentration of the promoter is conserved. We know that  $aR_t$ ,  $d \gg b, k$ , because the protein and the promoter's binding and unbinding reactions are much faster than the protein's decay and production reactions. Therefore, we can divide reactions in two groups:  $\emptyset \xrightleftharpoons[b]{k} X$  are two slow

reactions and  $X + R \xrightleftharpoons[d]{a} C$  are two fast reactions. When we define  $X_1 = X + C$ , we can find out that  $X_1$  is a slow species and  $C$  is a fast species which is bounded by  $R_t$ , because count of  $X_1$  is never changed and count of  $C$  is changed by the fast reactions. When we define  $\epsilon = \frac{b}{d}$ , which satisfies  $0 < \epsilon \ll 1$  and let  $k = b$ ,  $aR_t = \frac{d}{2}$ ,  $\mathbf{s} = [x_1, c]^T$ , we can derive the following propensity functions and corresponding stoichiometries for both fast and slow reactions as

$$\begin{aligned} a_1(x_1, c) &= b, \gamma_1 = [+1, 0]^T, \\ a_2(x_1, c) &= b(x_1 - c), \gamma_2 = [-1, 0]^T, \\ a_3(x_1, c) &= \frac{1}{\epsilon} \frac{b}{2VR_t} (x_1 - c)(R_t - c), \gamma_3 = [0, +1]^T, \\ a_4(x_1, c) &= \frac{1}{\epsilon} bc, \gamma_4 = [0, -1]^T, \end{aligned}$$

where,  $V$  is the volume.

The state space of the fast species is  $\Omega_y = \{0, 1, 2, \dots, R_t\}$ , which is a finite set, and the state space of the slow species is  $\Omega_x = \{0, 1, 2, \dots\}$  which is an infinite set. Based on the above propensity functions and stoichiometries, for  $x_1 \in \Omega_x$  and  $c \in \Omega_y$ , we can derive ODEs for the slow-species counts' marginal probability distribution and the fast-species counts' first 2 conditional moments as follows:

$$\begin{aligned} \frac{d}{dt} P(x_1, t) &= -bx_1 P(x_1, t) + b(x_1 + 1)P(x_1 + 1, t) \\ &\quad - b\mu_1(x_1 + 1, t)P(x_1 + 1, t) + b\mu_1(x_1, t)P(x_1, t) \\ &\quad - bP(x_1, t) + bP(x_1 - 1, t), \end{aligned} \quad (33)$$



$$\begin{aligned} \epsilon \frac{d}{dt} Y_2(x_1, t) = & \left[ \begin{array}{cc} -b - \frac{b(x_1 + R_t)}{2V R_t} & \frac{b}{2V R_t} \\ b + \frac{b(2bx_1 R_t - bx_1 - bR_t)}{2V R_t} & -2b + \frac{b - 2bx_1 - 2bR_t}{2V R_t} \end{array} \right] Y_2(x_1, t) \\ & + \left[ \frac{bx_1}{2V} \right] + \left[ \frac{0}{V R_t} \right] \mu_3(x_1, t) + O(\epsilon) \\ = & C(x_1) Y_2(x_1, t) + c_1(x_1) + c_2 \mu_3(x_1, t) + O(\epsilon), \end{aligned}$$

Here,  $\mu_3(x_1, t)$  is not a function of  $Y_2(x_1, t)$ . Therefore, to close the dynamics, we need to approximate  $\mu_3(x_1, t)$  as an affine function of  $Y_2(x_1, t)$  as follows:

$$\mu_3(x_1, t) \approx K_{32} \mu_2(x_1, t) + K_{31} \mu_1(x_1, t) + K_{30}.$$

By solving the linear program in (20), we can obtain

$$K_{32} = 15, K_{31} = -56, K_{30} = 30 \text{ and } \rho_2 = 30.$$

When we substitute  $\mu_3(x_1, t)$  with  $\phi(Y_2(x_1, t))$  in (33), we can check that ODEs for the fast-species counts first 2 conditional moments are closed. To approximate  $Y_2(x_1, t)$  as functions of the slow-species counts, we let  $\epsilon = 0$  in  $\Sigma_{closed}$ , and obtain

$$\begin{aligned} C(x_1) \begin{bmatrix} \tilde{\mu}_1^0(x_1) \\ \tilde{\mu}_2^0(x_1) \end{bmatrix} + c_1(x_1) \\ + c_2(K_{32} \tilde{\mu}_2^0(x_1) + K_{31} \tilde{\mu}_1^0(x_1) + K_{30}) = 0. \end{aligned} \quad (34)$$

By solving (34), we can obtain  $\tilde{\mu}_2^0(x_1) = K_{21} \tilde{\mu}_1^0(x_1) + K_{20}$ ,

$$\begin{aligned} \text{where } K_{20} &= \frac{ax_1 R_t + 2aK_{30}}{2aR_t + 2ax_1 - 2aK_{32} - a + 2dV}, \\ K_{21} &= \frac{2ax_1 R_t - aR_t - ax_1 + 2aK_{31} + dV}{2aR_t + 2ax_1 - 2aK_{32} - a + 2dV}, \\ \tilde{\mu}_1^0(x_1) &= \frac{aK_{20} + ax_1 R_t}{aR_t + ax_1 - aK_{21} + dV}. \end{aligned} \quad (35)$$

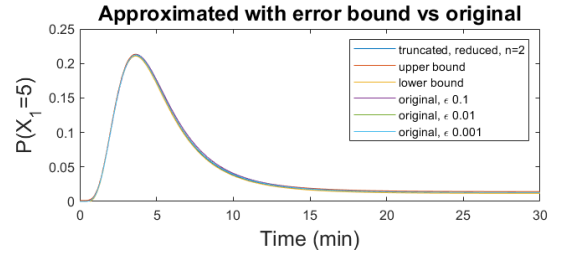
We can obtain the CME with the slow species only, by substituting  $\mu_1(x_1, t)$  as  $\tilde{\mu}_1^0(x_1)$  in (33). Now we should apply the EFSP algorithm to the original CME to find  $I_x$ .

Let  $b = 0.4[\text{min}^{-1}]$ ,  $R_t = 10[\text{molecules}]$ ,  $V = 1[\mu\text{m}^3]$ ,  $t_0 = 0[\text{min}]$ ,  $t_f = 30[\text{min}]$ ,  $X_1(t_0) = 10[\text{molecules}]$ ,  $C(t_0) = 2[\text{molecules}]$ ,  $\Omega_{x,0}^{I_x} = \{0, 1, \dots, X_1(t_0) + C(t_0) = 12\}$  and the EFSP error  $\delta = 10^{-3}$ . When we apply the EFSP algorithm,  $\Gamma_{I_x,0} = \mathbb{1}^T \hat{P}^{I_x,0}(X, t_f) = 0.8859 < 1 - \delta$ . After some iterations, when  $\Omega_{x,0}^{I_x} = \{0, 1, 2, \dots, 15\}$ ,  $\Gamma_{I_x} = \mathbb{1}^T \hat{P}^{I_x}(X, 0) = 0.99989 > 1 - \delta$ . This implies we succeed to truncate  $\Omega_x$  as  $\Omega_x^{I_x} = \{0, 1, \dots, 15\}$ , which is a finite set.

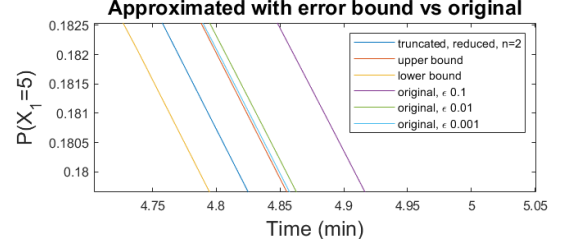
Fig. 2(a) compares  $P(X_1 = 5)$  of the original CME with  $10^{-6}$  error bound (which can be obtained by original FSP algorithm) with  $\epsilon = 0.1, 0.01, 0.001$  and those of  $\Sigma_{final}$  with  $n = 2$ ,  $\Omega_x^{I_x} = \{0, 1, 2, \dots, 15\}$  and  $\delta = 10^{-3}$ . Fig. 2(b) is the extended view of Fig. 2(a). The simulation result shows that  $P(X_1 = 5)$  of the original CME with  $\epsilon = 0.001$  is almost the same as those of  $\Sigma_{final}$ , which is obtained by the EFSP algorithm. Therefore, we can conclude that as  $\epsilon \rightarrow 0$ , our approach gives a valid result.

## IX. CONCLUSION

In this paper, we proposed the EFSP algorithm, by combining the FSP algorithm and our model reduction technique. The EFSP algorithm can be useful for the analysis and design of biomolecular systems.



(a) Comparing  $P(X_1 = 5)$  of the original CME with  $\epsilon = 0.1$  to  $0.001$  and those of  $\Sigma_{final}$  with  $n = 2$ ,  $\Omega_x^{I_x} = \{0, 1, \dots, 15\}$ ,  $\delta = 0.001$ .



(b) Extended view of the above graph.

Fig. 2: For this simulation,  $\epsilon = 0.1$  to  $0.001$ ,  $n = 2$ ,  $b = 0.4[\text{min}^{-1}]$ ,  $V = 1[\mu\text{m}^3]$ ,  $R_t = 10[\text{molecules}]$  are used.

## REFERENCES

- [1] D. Del Vecchio and R. M. Murray, *Biomolecular feedback systems*. Princeton University Press Princeton, NJ, 2015.
- [2] J. L. Spudich and D. E. Koshland, "Non-genetic individuality: chance in the single cell," *Nature*, vol. 262, no. 5568, p. 467, 1976.
- [3] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of physical chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.
- [4] C. V. Rao and A. P. Arkin, "Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm," *The Journal of chemical physics*, vol. 118, no. 11, pp. 4999–5010, 2003.
- [5] C. A. Gómez-Urbe, G. C. Verghese, and A. R. Tzafri, "Enhanced identification and exploitation of time scales for model reduction in stochastic chemical kinetics," *The Journal of chemical physics*, vol. 129, no. 24, p. 244112, 2008.
- [6] G. G. Yin and Q. Zhang, *Continuous-time Markov chains and applications: a singular perturbation approach*, vol. 37. Springer, 2012.
- [7] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *The Journal of chemical physics*, vol. 124, no. 4, p. 044104, 2006.
- [8] S. Peleš, B. Munsky, and M. Khammash, "Reduction and solution of the chemical master equation using time scale separation and finite state projection," *The Journal of chemical physics*, vol. 125, no. 20, p. 204104, 2006.
- [9] H. K. Khalil, "Nonlinear systems," *Upper Saddle River*, 2002.
- [10] U. Kwon, M. Naghnaeian, and D. Del Vecchio, "Approximation of the chemical master equation using conditional moment closure and time-scale separation," in *2019 American Control Conference (ACC)*, pp. 585–592, IEEE, 2019.
- [11] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A: Statistical Mechanics and its Applications*, vol. 188, no. 1-3, pp. 404–425, 1992.
- [12] M.-N. Contou-Carrere, V. Sotiropoulos, Y. N. Kaznessis, and P. Daoutidis, "Model reduction of multi-scale chemical langevin equations," *Systems & Control Letters*, vol. 60, no. 1, pp. 75–86, 2011.
- [13] M. Naghnaeian and D. Del Vecchio, "Robust moment closure method for the chemical master equation," in *2017 IEEE Conference on Control Technology and Applications (CCTA)*, pp. 967–972, IEEE, 2017.
- [14] W. Mitkowski, "Dynamical properties of metzler systems," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 2008.