Hybrid Master Equation for Jump-Diffusion Approximation of Biomolecular Reaction Networks

Derya Altıntan · Heinz Koeppl

Abstract Cellular reactions have a multi-scale nature in the sense that the abundance of molecular species and the magnitude of reaction rates can vary across orders of magnitude. This diversity naturally leads to hybrid models that combine continuous and discrete modeling regimes. In order to capture this multi-scale nature, we proposed jump-diffusion approximations in a previous study. The key idea was to partition reactions into fast and slow groups, and then to combine a Markov jump updating scheme for the slow group with a diffusion (Langevin) updating scheme for the fast group. In this study we show that the joint probability density function of the jump-diffusion approximation over the reaction counting process satisfies a hybrid master equation that combines terms from the chemical master equation and from the Fokker-Planck equation. Inspired by the method of conditional moments, we propose a efficient method to solve this master equation using the moments of reaction counters of the fast reactions given the reaction counters of the slow reactions. For each time point of interest, we then solve a set of maximum entropy problems in order to recover the conditional probability density from its moments. This finally allows us to reconstruct the complete joint probability density over all reaction counters and hence obtain an approximate solution of the hybrid master equation. Finally, we show the accuracy of the method applied to a canonical model of gene expression.

Keywords jump-diffusion approximation · chemical master equation · Fokker-Planck equation · maximum entropy approach

Derya Altıntan
Department of Mathematics, Selçuk University
E-mail: altintan@selcuk.edu.tr

Heinz Koeppl
Department of Electrical Engineering and Information Technology, Technische Universität Darmstadt
E-mail: heinz.koeppl@bcs.tu-darmstadt.de

Mathematics Subject Classification (2000) 60G05 · 60G07

1 Introduction

Reactions are inherently discrete and stochastic in nature [12, 15, 16]. Ignoring stochastic fluctuations and the integer nature of molecule counts can result in inappropriate models which especially fail for small reaction compartments as encountered in cell biology. Traditionally, the stochastic approach explains reaction dynamics using a discrete-state continuous-time Markov chains and describes the state of the system by the integer-valued number of molecules of each involved species. In this approach, the state vector of the system satisfies the random time change model (RTCM), which defines the reaction counting processes using time-warped Poisson processes [2]. Also, the probability mass function of these systems satisfies a set of differential equations referred to as the chemical master equation (CME) in the literature [18, 20]. When the number of molecules of the species in the system of interest is very high, the state vector of the system can be defined by real-valued concentrations instead of integer-valued particle numbers. Dynamics of such systems can approximately be capture through diffusion processes, and the state vector of the system satisfies an Itô stochastic differential equation (SDE) known as the chemical Langevin equation (CLE). Similarly, probability density function of these systems satisfies the Fokker-Planck equation (FPE) [18, 21]. In the thermodynamic limit, in which the number of molecules of species and the system volume both approach to infinity while the concentrations of species stay constant, the state of the system is given by the reaction rate equation (RRE) of the traditional deterministic modeling approach.

A major challenge of modeling the reaction networks using the CME is the curse of dimensionality. Each species of the system adds one dimension to the corresponding CME. Therefore, when the number of reachable states is high, it is infeasible to obtain the numerical solution of the CME. To avoid this drawback, different stochastic simulation algorithms (SSAs), such as the Doob-Gillespie algorithm and its variants, are proposed to obtain sample paths of the biochemical reaction system of interest [10, 19, 22]. The computational cost of these algorithms increases with the size or reactivity of the system and slow convergence of Monte Carlo estimates usually results in massive amounts of necessary simulation runs.

Moment approximations that analyze the dynamics of the reaction network under consideration using moments of the probability distribution satisfying the corresponding CME are often considered as an alternative [39]. In [13], the author propose the method of moments that computes the moments for any reaction network from the corresponding CME. In [31], a moment closure approximation that contains a finite set of ordinary differential equations (ODE) for the mean and the central moments by truncating the moment hierarchy at a certain order and using the Taylor series is introduced. Another moment closure method that approximates the moments with higher order, compared to the order of truncation, utilizing nonlinear closure functions of the lower order moments is introduced in [38]. A variational formulation is used in [6] to identify principled closure schemes and to relate moment closure to other approximation techniques.

Cellular reaction systems involve reactions with very different rates and species with very different abundances. Models only based on the traditional deterministic modeling approach fail to account for this nature. Therefore, different hybrid methods that couple the stochastic and deterministic modeling approaches were investigated. In general, hybrid methods separate reactions and/or species into different groups of reactions and/or species, and they use the diffusion or the deterministic modeling approach to describe the dynamics of fast reactions and/or species with high copy numbers, while a discrete Markov chain representation is utilized for slow reactions and/or species with low copy numbers. For example, in [8], the authors propose different hybrid models in four steps. In the first step, they partition the reactions/species into different classes, in the second step, they apply cycle averaging in reaction network, in the third step, they find super reactions which are fast enough to change the amount of continuous variables or have large stoichiometric vectors and finally, they propose their hybrid model based on Kramer's Moyal expansion. In [32], the authors express the probability density function (PDF) of a hybrid representation as a product of the probability distribution function of the discrete species and the conditional probability density function (CPDF) of the continuous species given discrete species. Then, they scale the population levels of the continuous species and use the Wentzel-Kramers-Brillouin approximation to approximate the CPDF. Finally, they derive a hybrid system involving the evolution equations for the PDF of the discrete species and for the expectation of the CPDF of the continuous species conditioned on the discrete species. In [37], another hybrid stochastic method which separates reactions into fast, slow subgroups is proposed. It uses the CLE to describe the dynamics of the fast reaction while the next reaction variant of the stochastic simulation algorithm [19] is applied for the slow ones. Different algorithms which take into account the fact that one or multiple slow reaction can fire within a time step of numerical integration of the corresponding Langevin equation were introduced. In [26], the authors proposed a hybrid method which splits the state vector into two class of stochastic variables. In this method, it is assumed that stochastic variables in the first class are normally distributed with a small variance while stochastic variables in the second class are Markovian. Then, Monte Carlo and Quasi Monte Carlo methods are used to approximate the expected values of the stochastic variables in the first class while SSA is used to obtain marginal probability of the variables in the second class. Different types of hybrid models can be seen in [7, 11, 14, 27].

In [25], the authors introduced the method of conditional moments (MCM) that can be considered as the combination of a hybrid method and a moment approximation method. The MCM separates species into two different classes involving species with high copy number of molecules and species with low copy number of molecules. Based on this decomposition, the joint probability density function satisfying the corresponding CME is also represented as a product of the marginal probabilities of species with low copy number of molecules and the conditional probabilities of species with high copy number of molecules conditioned on the remaining species with low copy numbers of molecules. To describe the dynamics of species with low copy number of molecules, the authors used marginal probabilities, while the conditional means and the centered conditional moments are used to model the dynamics of species with high copy numbers. In comparison to [25], in [3], the authors obtained

moments of the system of interest directly from the corresponding CME without using any partitioning of the species, and the maximum entropy approach is used to construct the corresponding probability distribution.

In [17], we developed a jump-diffusion approximation to model multi-scale behavior of cellular reactions. Based on an error bound, we separated reactions into fast and slow groups. We employed diffusion approximation for the fast reactions, while Markov jump process was kept for the slow ones. As a result, the state of the system was defined as the summation of the RTCM and the solution of the corresponding CLE. In this paper, based on this representation, we present the hybrid master equation (HME), which is the evolution equation for the joint probability density function of the jump diffusion approximation over the reaction counting process. The main contribution of the paper is to obtain evolution equation for the probability density function over the reaction counting process. Different than other studies [8, 25, 26, 32, 37], we define the probability density functions in terms of reaction counters and prove that evolution equation for this PDF which is called HME is the summation of the corresponding CME and the corresponding FPE [34]. Although it is possible to use different numerical methods to approximate the solution of the HME [35], we use the strategy proposed in [25].To solve the HME, we obtain the evolution equation for the marginal probability of slow reactions and the evolution equations for the conditional moments of the fast reactions given slow reactions [25]. Differently than [32], we don't use any scaling parameter for the probability density functions and don't approximate the conditional probability density function by using any series expansion. Also, we use not only expected values [26] of the reaction counters of the fast reactions given the reaction counters of slow reactions but also centered conditional moments of them. Using the maximum entropy approach, we construct the corresponding conditional probability at the time point of interest, which in turn gives the approximate solution of the corresponding HME. We note that the proposed algorithm approximates the joint probability density function over reaction counting process at a desired time point without counting firings of slow and fast reactions [37].

The rest of the paper is organized as follows: In Section 2, we describe the basic concepts of the stochastic modeling approach. In Section 3, we give a brief summary of the jump-diffusion approximation. We introduce the HME in Section 4. In Section 5, we construct an ODE system that will be used to obtain the approximate solution of the HME. In Section 6, we introduce the maximum entropy approach. In Section 7, we present numerical results and also explain the details of how we use the maximum entropy approach to construct the joint probability density function describing the HME. Section 8 concludes the paper.

Notation

Before we give the details of the mathematical derivations, we present the basic notations used through the present paper. We represent all random variables and their realizations by upper-case (i.e. A) and lower case (i.e. a) symbols, respectively, and we use bold symbols to represent the support of a random variable (i.e. $\mathbf{A}$). Also, e_j, $\bar{e}_j$ denote $(R-L) \times 1$, $L \times 1$, unit vectors with 1 in the $j-$th component and 0 in other coordinates.

2 Stochastic Modeling of Chemical Kinetics

In this study, we consider a well-mixed reaction system of M species, $S_1, S_2, \ldots, S_M$, interacting through $R \geq 1$ reaction channels $R_1, R_2, \ldots, R_R$ inside the reaction compartment with volume V. The $k-$th reaction channel of the system is described as follows:

$$r_{1k}S_1 + r_{2k}S_2 + r_{3k}S_3 + \ldots + r_{Mk}S_M \xrightarrow{\ell_k} p_{1k}S_1 + p_{2k}S_2 + p_{3k}S_3 + \ldots + p_{Mk}S_M,$$
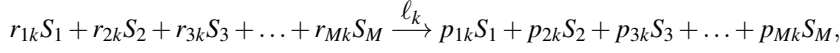
where $r_{jk}, p_{jk} \in \mathbb{N}$, $j = 1, 2, \ldots, M$, represent the number of molecules of species S_j consumed and produced with a single occurrence of the reaction R_k, respectively, and ℓ_k is the real-valued stochastic reaction rate constant. Let $X_i(t) \in \mathbb{N}_0$ denote the number of molecules of species S_i, $i = 1, 2, \ldots, M$, at time $t \geq 0$. Then, the state of the system at time t is $X(t) = (X_1(t), X_2(t), \ldots, X_M(t))^T \in \mathbb{N}_0^M$.

The classical stochastic modeling of biochemical networks assumes that the process of X is a continuous time Markov chain (CTMC). In this approach, the state vector, $X(t)$, is defined as a random variable of the Markov jump process. Each reaction channel R_k, $k = 1, 2, \ldots, R$, is specified by its stoichiometric vector (state-change vector) and its propensity function. The stoichiometric vector $v_k = (v_{1k}, v_{2k}, \ldots, v_{Mk}) \in \mathbb{Z}^M$ with $v_{jk} = p_{jk} - r_{jk}$, $j = 1, 2, \ldots, M$, represents the change in the state of the system after one occurrence of the reaction R_k. In other words, when the reaction R_k fires, the system state $X(t) = x$ jumps to a new state $x + v_k$. Given $X(t) = x$, the probability that one R_k reaction takes place in the time interval $[t, t + h)$ is $a_k(x)h + o(h)$ where $a_k(x) : \mathbb{N}_0^M \to \mathbb{R}_+$ represents the propensity function calculated by the law of mass action kinetics, i.e., $a_k(x) = \ell_k \prod_{i=1}^{M} \binom{x_i}{r_{ik}}$. Let $Z_k(t)$ denote the number of occurrence of the reaction R_k by the time t, then the state of the system at time t can be obtained as follows:

$$X(t) = X(0) + \sum_{k=1}^{R} Z_k(t) v_k.$$

If we represent the counting process $Z_k(t)$ in terms of the independent Poisson process denoted by ξ_k, such that $Z_k(t) = \xi_k \left(\int_0^t a_k(X(s))ds \right)$, then the state vector of the above CTMC satisfies the following RTCM [2]

$$X(t) = X(0) + \sum_{k=1}^{R} \xi_k \left(\int_0^t a_k(X(s))ds \right) v_k. \tag{2.1}$$

Let define the following probability mass function

$$p_t(x) = \mathrm{P}(X(t) = x).$$

Another way of analyzing this CTMC process is to consider the time evolution of the probability function $p_t(x)$. This probability mass function is the solution of the following Kolmogorov's forward equation, which is known as the CME [20]

$$\frac{\partial p_t(x)}{\partial t} = \sum_{k=1}^{R} [a_k(x - v_k)p_t(x - v_k) - a_k(x)p_t(x)]. \tag{2.2}$$

When the number of molecules in the system is very high, then the abundance of the species at time t can be represented by the real valued concentrations of the form $U(t) = V^{-1} X(t) \in \mathbb{R}^M_{\geq 0}$. In most cases, reaction channels in biochemical systems are bimolecular or monomolecular. If the $k-$th reaction channel R_k is bimolecular or monomolecular, then its propensity function satisfies the equality $a_k(x) = V \tilde{a}_k(u)$ where $\tilde{a}_k$ is the propensity function obtained involving the deterministic reaction rate constant $\tilde{\ell}_k$.

It is well known that the centered version of each independent Poisson process, ξ_k, in Equation (2.1) can be approximated through the independent Brownian motions $W_k(t)$ [2, 30]. Considering the fact that $(\xi_k(Vt) - Vt)/\sqrt{V}$ converges in distribution to the Brownian motion $W_k(t)$ for large V, we obtain the diffusion approximation of Equation (2.1) as given

$$U(t) = U(0) + \sum_{k=1}^{R} v_k \int_0^t \tilde{a}_k(U(s))ds + \frac{1}{\sqrt{V}} \sum_{k=1}^{R} v_k W_k \left(\int_0^t \tilde{a}_k(U(s))ds \right). \qquad (2.3)$$

The first and the second summand in the right hand-side of Equation (2.3) are called *drift* and *diffusion* terms, respectively. The time derivative of the state vector $U(t)$ satisfies an SDE, namely the CLE.

Let define the following probability density function

$$q_t(u)du = P(U(t) \in [u, u + du]).$$

Then, analog of the CME for this continuous process is represented by the following FPE [21]

$$\frac{\partial q_t(u)}{\partial t} = -\sum_{i=1}^{M} \frac{\partial}{\partial u_i} \left[\left(\sum_{k=1}^{R} v_{ik} \tilde{a}_k(u) \right) q_t(u) \right] + \frac{1}{2} \sum_{i,i'=1}^{M} \frac{\partial^2}{\partial u_i \partial u_{i'}} \left[\left(\sum_{k=1}^{R} v_{ik} v_{i'k} \tilde{a}_k(u) \right) q_t(u) \right].$$

Cellular processes consist of bimolecular reactions of very different speeds involving reactants of largely different abundances. Therefore, the models based only on the RTCM or only the diffusion approximation may be inappropriate to explain the dynamics of such multi-scale processes. In [17], we developed a jump-diffusion approximation to model such processes. In the following section, we will give a summary of this approximation.

3 Jump Diffusion Approximation

In jump-diffusion approximation [17], we partition the reactions into the fast subgroup, $\mathscr{C}$, and the slow subgroup, $\mathscr{D}$, and model the fast group using a diffusion process, while Markov chain representation is kept for the slow group.

In this approach instead of the CTMC process represented by X, we focus on the scaled abundances $\bar{X}_i^N = X_i/N^{\zeta_i}$, $i = 1, 2, \ldots, M$, and the scaled stochastic reaction rates $\kappa_j = \ell_j/N^{\eta_j}$, $j = 1, 2, \ldots, R$, such that $\bar{X}_i^N = O(1)$, $\kappa_j = O(1)$. Naturally, these scaled quantities will produce new scaled propensity functions as follows

$$a_k(X(t)) = N^{\eta_k + r_k \cdot \zeta} \bar{a}_k(\bar{X}^N(t)),$$

where $r_k = (r_{1k}, r_{2k}, \ldots, r_{Mk})$ and $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_M)$. We note that $\bar{a}_k(.)$ functions are also $O(1)$. Finally, scaling the time $t \to tN^\theta$ and defining $X^N(t) = \bar{X}^N(tN^\theta)$, we transform the state vector, $X(t)$, given by Equation (2.1) into the following scaled state vector

$$X^N(t) = X^N(0) + \sum_{k=1}^{R} \xi_k \left(\int_0^t \alpha_k(X^N(s))ds \right) \mu_k, \qquad (3.1)$$

where $\alpha_k(X^N(t)) = N^{\rho_k} \bar{a}_k(X^N(t))$, $\rho_k = \theta + \eta_k + r_k \cdot \zeta$ and $\mu_{ik} = v_{ik}/N^{\zeta_i}$, $i = 1, 2, \ldots, M$.

Modeling the fast reactions through diffusion approximation and modeling the slow reactions through Markov chains give the state vector of the jump diffusion approximation as follows:

$$Y(t) = Y(0) + \sum_{i \in \mathscr{D}} \xi_i \left(\int_0^t \alpha_i(Y(s))ds \right) \mu_i + \sum_{j \in \mathscr{C}} \int_0^t \alpha_j(Y(s))ds\,\mu_j \qquad (3.2)$$

$$+ \sum_{j \in \mathscr{C}} W_j \left(\int_0^t \alpha_j(Y(s))ds \right) \mu_j,$$

where $Y(0) = X^N(0)$, and W_j is a standard Brownian motion. If τ_1, τ_2 denote the successive firing times of reactions from the slow group, then for $\tau_1 < t < \tau_2$, only reactions from the fast group can fire. Therefore, in this time interval, the state vector of the system is given by

$$Y(t) = Y(\tau_1) + \sum_{j \in \mathscr{C}} \int_{\tau_1}^t \alpha_j(Y(s))ds\,\mu_j + \sum_{j \in \mathscr{C}} W_j \left(\int_{\tau_1}^t \alpha_j(Y(s))ds \right) \mu_j. \qquad (3.3)$$

The main contribution of this study is the derivation of an error bound for the mean $e(t) = \mathsf{E}\,|\,X^N(t) - Y(t)\,|$, which is used to partition the reaction set into fast and slow subgroups. Based on this error bound, we construct a dynamic partitioning algorithm that takes into account the fact that a fast reaction can return to a slow reaction or vice versa during the course of time.

By describing the state vector of the system as the summation of purely discrete and purely continuous components, we can introduce the HME, which defines the joint probability density function of the jump diffusion approximation over the reaction counting process. In the following section, we will obtain the HME.

4 Hybrid Master Equation

In jump diffusion approximation, we partition the reaction set into two subsets. As mentioned before, the first subset $\mathscr{C}$ involves reactions modeled by diffusion approximation, while the rest of the reactions constituting the slow set $\mathscr{D}$ are modeled by Markov chains. In the rest of the study, we will consider that there are L slow reactions, i.e., $|\mathscr{D}| = L$, and $R - L$ fast reactions in the system, i.e., $|\mathscr{C}| = R - L$.

Let $Z(t) = (Z_1(t), Z_2(t), \ldots, Z_R(t))^T$ be a vector of reaction counters such that $Z_i(t)$ denotes the number of occurrences of the reaction R_i, $i = 1, 2, \ldots, R$, during the time of the process until time $t > 0$. Similar to the idea of splitting the state

vector of the system into purely discrete and purely continuous parts, we also separate $Z(t) = (D(t), C(t))^T$ into purely discrete and continuous parts corresponding to the reaction counters of the slow, $D(t) \in \mathbb{N}^L$, and the fast reaction set, $C(t) \in \mathbb{R}_{\geq 0}^{R-L}$ such that $D_i(t) = Z_i(t), i \in \mathscr{D}$, and $C_j(t) = Z_j(t), j \in \mathscr{C}$. We also separate the stoichiometric vectors such that $\mu_i^D = \mu_i, i \in \mathscr{D}$, and $\mu_j^C = \mu_j, j \in \mathscr{C}$.

By using Equation (3.2), we will define reaction counters as follows:

$$D_i(t) = \xi_i\left(\int_0^t \alpha_i(Y(s))ds\right) = \xi_i\left(\int_0^t \tilde{\alpha}_i(D(s), C(s))ds\right), \qquad i \in \mathscr{D},$$

$$C_j(t) = \int_0^t \alpha_j(Y(s))ds + W_j\left(\int_0^t \alpha_j(Y(s))ds\right)$$
$$= \int_0^t \tilde{\alpha}_j(D(s), C(s))ds + W_j\left(\int_0^t \tilde{\alpha}_j(D(s), C(s))ds\right), \qquad j \in \mathscr{C},$$

where

$$\alpha_k(y) = \alpha_k\left(y(0) + \sum_{i \in \mathscr{D}} d_i \mu_i^D + \sum_{j \in \mathscr{C}} c_j \mu_j^C\right) = \tilde{\alpha}_k(d, c), k = 1, 2, \ldots, R. \qquad (4.1)$$

We note that if τ_1, τ_2 denote the successive firing times of reactions from the slow group, then for $\tau_1 < t < \tau_2$, $C(t)$ satisfies the following equation

$$C(t) = C(\tau_1) + \sum_{j \in \mathscr{C}} \left(\int_{\tau_1}^t \tilde{\alpha}_j(d, C(s))ds\right)e_j + \sum_{j \in \mathscr{C}} W_j\left(\int_{\tau_1}^t \tilde{\alpha}_j(d, C(s))ds\right)e_j, \quad (4.2)$$

where d denotes the number of slow reactions fired until time $\tau_1 > 0$.

The HME is the time derivative of the joint probability density function $p_t : \mathbb{N}^L \times \mathbb{R}_{\geq 0}^{R-L} \to \mathbb{R}_{\geq 0}$

$$p_t(d, c)dc = \mathrm{P}(D(t) = d, C(t) \in [c, c + dc]). \qquad (4.3)$$

Then, we can write

$$p_t(d, c) = p_t(c \mid d)p_t(d),$$

where

$$p_t(c \mid d)dc = \mathrm{P}(C(t) \in [c, c + dc] \mid D(t) = d)$$
$$p_t(d) = \mathrm{P}(D(t) = d).$$

To obtain the evolution equation for $p_t(d, c)$, which is called the HME, we need the following result whose details can be found in [34].

Result : Let $D(t) \in \mathbf{D} \subset \mathbb{N}^L$ be a discrete process and $C(t) \in \mathbb{R}_{\geq 0}^{R-L}$ be a continuous process. Define the joint probability density function as follows:

$$p_t(d, c) = p_t(c \mid d)p_t(d).$$

Then, the time derivative of this joint probability function, which is referred to as generalized Fokker-Planck equation (GFPE), has the following form

$$\frac{\partial}{\partial t} p_t(d,c) = \sum_{d' \in \mathbf{D}} a_{dd'} p_t(d',c) + \sum_{n_1,n_2,\ldots,n_{R-L}=1}^{\infty} \left(\prod_{i=1}^{R-L} \frac{(-1)^{n_i} \frac{\partial^{n_i}}{\partial c_i^{n_i}}}{n_i!} \right) [A_{n_1,n_2,\ldots,n_{R-L}} p_t(d,c)],$$

(4.4)

where

$$A_{n_1,n_2,\ldots,n_{R-L}} = \lim_{h \to 0} \frac{1}{h} \mathsf{E}[\prod_{i=1}^{R-L} \{C_i(t+h) - C_i(t)\}^{n_i} \mid D(t) = d, C(t), D(t+h) = d],$$

and

$$a_{dd'} = \lim_{h \to 0} \frac{1}{h} [\mathsf{P}(D(t+h) = d \mid D(t) = d', C(t)) - \delta_{dd'}].$$

(4.5)

where $\mathsf{E}[C(t) \mid d] = \int_{\mathbb{R}_{\geq 0}^{R-L}} c\, p_t(c \mid d)\, dc$. It is also proved that $A_{n_1,n_2,\ldots,n_{R-L}} = 0$ for all $\sum_{i=1}^{R-L} n_i \geq 3$. This gives us

$$\sum_{n_1,n_2,\ldots,n_{R-L}=1}^{\infty} \left(\prod_{i=1}^{R-L} \frac{(-1)^{n_i} \frac{\partial^{n_i}}{\partial c_i^{n_i}}}{n_i!} \right) [A_{n_1,n_2,\ldots,n_{R-L}} p_t(d,c)] = -\sum_{j=1}^{R-L} \frac{\partial}{\partial c_j} [B_j p_t(d,c)] \quad (4.6)$$

$$+ \frac{1}{2} \sum_{i,j=1}^{R-L} \frac{\partial^2}{\partial c_i \partial c_j} [B_{ij} p_t(d,c)],$$

where

$$B_j = \lim_{h \to 0} \frac{1}{h} \mathsf{E}[(C_j(t+h) - C_j(t)) \mid D(t) = d, C(t), D(t+dt) = d],$$

and

$$B_{ij} = \lim_{h \to 0} \frac{1}{h} \mathsf{E}[\{C_i(t+h) - C_i(t)\}\{C_j(t+h) - C_j(t)\} \mid D(t) = d, C(t), D(t+dt) = d].$$

Theorem 4.1 *Let $Z(t) = \{D(t), C(t)\}$ be a joint reaction counting process where $D(t)$ is a discrete random process with states $d \in \mathbf{D} \subset \mathbb{N}^L$, $L > 0$, and $C(t)$ is a continuous random process with states $c \in \mathbb{R}_{\geq 0}^{R-L}$, $R - L > 0$. Define Y as a multiscale process whose state vector is given in Equation (3.2). Then, the joint counting probability density function given in Equation (4.3) satisfies the following GFPE, which is referred to as the HME in the present paper.*

$$\frac{\partial}{\partial t} p_t(d,c) = \sum_{i \in \mathscr{D}} \left(\tilde{\alpha}_i(d - \bar{e}_i, c) p_t(d - \bar{e}_i, c) - \tilde{\alpha}_i(d,c) p_t(d,c) \right)$$

(4.7)

$$- \sum_{j \in \mathscr{C}} \frac{\partial}{\partial c_j} (\tilde{\alpha}_j(d,c) p_t(d,c)) + \frac{1}{2} \sum_{j \in \mathscr{C}} \frac{\partial^2}{\partial c_j^2} (\tilde{\alpha}_j(d,c) p_t(d,c)).$$

Proof By using Equation (4.4) and Equation (4.7), we obtain

$$\frac{\partial}{\partial t} p_t(d,c) = \sum_{d' \in \mathbf{D}} a_{dd'} p_t(d',c) - \sum_{j \in \mathscr{C}} \frac{\partial}{\partial c_j} (B_j p_t(d,c)) + \frac{1}{2} \sum_{i,j \in \mathscr{C}} \frac{\partial^2}{\partial c_i \partial c_j} (B_{ij} p_t(d,c)).$$

(4.8)

Now, let's focus on the first summand on the right hand-side of Equation (4.8), which can be rewritten as follows:

$$\sum_{d' \in \mathbf{D}} a_{dd'} p_t(d',c) = \sum_{\substack{d' \in \mathbf{D} \\ d \neq d'}} a_{dd'} p_t(d',c) + a_{dd} p_t(d,c).$$

(4.9)

Using Equation (4.5) gives

$$a_{dd} = \lim_{h \to 0} \frac{1}{h} [\mathrm{P}(D(t+h) = d \mid D(t) = d, C(t)) - 1],$$

which can be reformulated as follows

$$a_{dd} = \lim_{h \to 0} \frac{1}{h} [- \sum_{\substack{d' \in \mathbf{D} \\ d \neq d'}} \mathrm{P}(D(t+h) = d' \mid D(t) = d, C(t))].$$

By using this representation, we can rewrite Equation (4.9) in the following form

$$\sum_{d' \in \mathbf{D}} a_{dd'} p_t(d',c) = \sum_{\substack{d' \in \mathbf{D} \\ d \neq d'}} \Big(a_{dd'} p_t(d',c) - a_{d'd} p_t(d,c) \Big).$$

In our multi-scale process, we have L slow reactions, and one firing of the reaction R_j in this set updates d to $d + \bar{e}_j$. Starting from d, the system can jump to $d' = d + \bar{e}_j$, meaning that $a_{d+\bar{e}_j,d} = \tilde{\alpha}_j(d,c)$. In the same vein, to reach d, the system must supervene on $d - \bar{e}_j$, by definition $a_{d,d-\bar{e}_j} = \tilde{\alpha}_j(d - \bar{e}_j,c)$. As a result, we obtain the desired summand as follows:

$$\sum_{d' \in \mathbf{D}} a_{dd'} p_t(d',c) = \sum_{i \in \mathscr{D}} \Big(\tilde{\alpha}_i(d - \bar{e}_i,c) p_t(d - \bar{e}_i,c) - \tilde{\alpha}_i(d,c) p_t(d,c) \Big).$$

(4.10)

Now, we can concentrate on the second and the third summands of Equation (4.8). Jump diffusion approximation is based on the idea that between two successive firing times of the slow reactions, the fast reactions continue to fire. Hence, the state vector and also the reaction counting process of the fast reaction set will satisfy diffusion processes (see Equation 3.3,4.2). Therefore, B_j and B_{ij} values have the forms [21, 29]

$$B_j = \sum_{k \in \mathscr{C}} e_{jk} \tilde{\alpha}_k(d,c), \quad B_{ij} = \sum_{k \in \mathscr{C}} e_{ik} e_{jk} \tilde{\alpha}_k(d,c).$$

Substitution of B_j and B_{ij} values and Equation (4.10) into Equation (4.8) gives

$$\frac{\partial}{\partial t}p_t(d,c) = \sum_{i\in\mathscr{D}}\left(\tilde{\alpha}_i(d-\bar{e}_i,c)p_t(d-\bar{e}_i,c) - \tilde{\alpha}_i(d,c)p_t(d,c)\right)$$

$$- \sum_{j\in\mathscr{C}}\frac{\partial}{\partial c_j}\left(\sum_{k\in\mathscr{C}}e_{jk}\tilde{\alpha}_k(d,c)\right) + \frac{1}{2}\sum_{i,j\in\mathscr{C}}\frac{\partial^2}{\partial c_i\partial c_j}(\sum_{k\in\mathscr{C}}e_{ik}e_{jk}\tilde{\alpha}_k(d,c)p_t(d,c))$$

$$= \sum_{i\in\mathscr{D}}\left(\tilde{\alpha}_i(d-\bar{e}_i,c)p_t(d-\bar{e}_i,c) - \tilde{\alpha}_i(d,c)p_t(d,c)\right)$$

$$- \sum_{j\in\mathscr{C}}\frac{\partial}{\partial c_j}(\tilde{\alpha}_j(d,c)p_t(d,c)) + \frac{1}{2}\sum_{j\in\mathscr{C}}\frac{\partial^2}{\partial c_j^2}(\tilde{\alpha}_j(d,c)p_t(d,c)),$$

which completes the proof.

Based on the properties of the joint counting probability density function, we can write

$$p_t(d,c) = p_t(c\mid d)p_t(d).$$

Since we partition reaction counters into two subsets, we will also decompose the propensity functions. Using mass action kinetics to compute propensities is very popular, and for this large class we partition the propensity function of the reaction R_k, $\tilde{\alpha}_k(d,c)$, $k = 1,2,\ldots,R$, as follows:

$$\tilde{\alpha}_k(d,c) = \alpha_k(y(0) + \sum_{i\in\mathscr{D}}d_i\mu_i^D + \sum_{j\in\mathscr{C}}c_j\mu_j^C) = \kappa_k\prod_{s=1}^{M}(y_s(0) + \sum_{i\in\mathscr{D}}d_i\mu_{si}^D + \sum_{j\in\mathscr{C}}c_j\mu_{sj}^C)^{r_{sk}}$$

$$= \kappa_k\prod_{s=1}^{M}(\beta_s(d) + \gamma_s(c))^{r_{sk}} = \kappa_k\prod_{s=1}^{M}\sum_{n=0}^{r_{sk}}\binom{r_{sk}}{n}\beta_s^n(d)\gamma_s^{r_{sk}-n}(c)$$

where $\beta_s(d) = y_s(0) + \sum_{i\in\mathscr{D}}d_i\mu_{si}^D$, $\gamma_s(c) = \sum_{j\in\mathscr{C}}c_j\mu_{sj}^C$. With a slight abuse of notation but in order to avoid clutter, we will subsequently subsume the binomial coefficient into the definitions of β_s and γ_s. Based on this representation, the HME given in Equation (4.7) can be rewritten in the following form

$$\frac{\partial}{\partial t}p_t(d,c) = \sum_{i\in\mathscr{D}}\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\kappa_i\left(\beta_s^n(d-\bar{e}_i)\gamma_s^{r_{si}-n}(c)p_t(d-\bar{e}_i,c) - \beta_s^n(d)\gamma_s^{r_{si}-n}(c)p_t(d,c)\right)$$

$$- \sum_{j\in\mathscr{C}}\frac{\partial}{\partial c_j}\left(\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\gamma_s^{r_{sj}-n}(c)p_t(d,c)\right)$$

$$+ \frac{1}{2}\sum_{j\in\mathscr{C}}\frac{\partial^2}{\partial c_j^2}\left(\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\gamma_s^{r_{sj}-n}(c)p_t(d,c)\right). \tag{4.11}$$

Let $f(d): \mathbf{D}\to\mathbb{R}$ and $g(c): \mathbb{R}_{\geq 0}^{R-L}\to\mathbb{R}$ be any functions of d and c variables, respectively. To simplify the notation, we introduce one step operator in the following form

$$\mathscr{F}^{\bar{e}_i}\left(f(d)g(c)\right) = f(d+\bar{e}_i)g(c), i\in\mathscr{D}.$$

Based on this representation, we define

$$
\begin{aligned}
\Gamma(\beta(d),\gamma(c)) = & \sum_{i \in \mathscr{D}} (\mathscr{F}^{-\bar{e}_i} - I)\Big(\kappa_i \prod_{s=1}^{M} \sum_{n=0}^{r_{si}} \beta_s^n(d)\gamma_s^{r_{si}-n}(c)\Big) \\
& - \sum_{j \in \mathscr{C}} \frac{\partial}{\partial c_j}\Big(\kappa_j \prod_{s=1}^{M} \sum_{n=0}^{r_{sj}} \beta_s^n(d)\gamma_s^{r_{sj}-n}(c)\Big) \\
& + \frac{1}{2} \sum_{j \in \mathscr{C}} \frac{\partial^2}{\partial c_j^2}\Big(\kappa_j \prod_{s=1}^{M} \sum_{n=0}^{r_{sj}} \beta_s^n(d)\gamma_s^{r_{sj}-n}(c)\Big).
\end{aligned}
$$

Then, we can write Equation (4.11) in the following form

$$
\frac{\partial}{\partial t} p_t(d,c) = \Gamma(\beta(d),\gamma(c)) p_t(d,c). \tag{4.12}
$$

In the rest of the study, we will assume that $p_t(d,c)$ is zero at $c = 0$, $c = \infty$ [24,36]. In the folllowing section, we will explain how we obtain the solution of this HME.

5 Solution of the Hybrid Master Equation

To obtain the joint counting probability density function, $p_t(d,c)$, described by the HME given in Equation (4.12), we will approximate the process $C(t) \mid D(t)$ using its moments. Solving a maximum entropy problem for each conditional moment will produce the conditional probability function, $p_t(c \mid d)$. The multiplication of $p_t(c \mid d)$ with the marginal probabilities of the remaining discrete states, i.e., $p_t(d) = \int_{\mathbb{R}_{\geq 0}^{R-L}} p_t(d,c)\, dc$, will give us the desired joint probability density function $p_t(d,c)$.

In the rest of the study, time dependent conditional means and the centered conditional moments of the process $C(t) \mid D(t)$ will be denoted by

$$
\mathsf{E}_t[C_m \mid d] = \int_{\mathbb{R}_{\geq 0}^{R-L}} c_m p_t(c \mid d)\, dc, \quad m \in \mathscr{C},
$$

$$
\mathsf{E}_t[\tilde{C}^M \mid d] = \int_{\mathbb{R}_{\geq 0}^{R-L}} \prod_{j \in \mathscr{C}} \tilde{c}_j^{M_j} p_t(c \mid d) dc,
$$

where $\tilde{c} = c - \mathsf{E}_t[C \mid d]$, $M = (M_1, M_2, \ldots, M_{R-L})^T \in \mathbb{N}^{R-L}$. Now, based on the study [25], we want to construct a differential equation system to obtain $p_t(d)$, $\mathsf{E}_t[C_m \mid d]$, $\mathsf{E}_t[\tilde{C}^M \mid d]$. To construct this system, we will need the following Lemma [13,25].

Lemma 5.1 *Let $F_t : \mathbb{R}_{\geq 0}^{R-L} \longrightarrow \mathbb{R}$ be a polynomial function of c, and $p_t(d,c)$ satisfy differential Equation (4.12). Assume that sufficiently many moments of $p_t(d,c)$ with respect to c exist, and the joint counting probability density vanishes at $c = 0$ and $c = \infty$. Define the following conditional mean*

$$
\mathsf{E}_t[F_t(C) \mid d] = \int_{\mathbb{R}_{\geq 0}^{R-L}} F_t(c) p_t(c \mid d)\, dc.
$$

Then,

$$\frac{\partial}{\partial t}(\mathsf{E}_t[F_t(C)\mid d]p_t(d)) = \tilde{\Gamma}(\beta(d),F_t(c)\gamma(c))p_t(d) + \mathsf{E}_t[\frac{\partial}{\partial t}F_t(C)\mid d]p_t(d),$$

where

$$\tilde{\Gamma}(\beta(d),F_t(c)\gamma(c)) = \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i} - I)\Big(\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\beta_s^n(d)\mathsf{E}_t[F_t(C)\gamma_s^{r_{sj}-n}(C)\mid d]\Big)$$
$$+ \sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[\gamma_s^{r_{sj}-n}(C)\frac{\partial}{\partial c_j}F_t(C)\mid d]$$
$$+ \frac{1}{2}\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\mathsf{E}_t[\gamma_s^{r_{sj}-n}(C)\frac{\partial^2}{\partial c_j^2}F_t(C)\mid d].$$

Proof The proof of the Lemma can be found in Appendix A.1.

When $F_t(c) = 1$ in Lemma 5.1, we obtain the time derivative of the marginal probability $p_t(d)$, which is given in the following proposition.

Proposition 5.1

$$\frac{\partial}{\partial t}p_t(d) = \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i} - I)\Big(\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\beta_s^n(d)\mathsf{E}_t[\gamma_s^{r_{si}-n}(C)\mid d]p_t(d)\Big) \qquad (5.1)$$

The strategy of our method is to obtain $p_t(d)$ and $p_t(c\mid d)$ separately and construct the joint probability function using the equality $p_t(d,c) = p_t(c\mid d)p_t(d)$. To obtain the conditional probability $p_t(c\mid d)$, we will use evolution equations of the conditional means $\mathsf{E}_t[C_m\mid d]$, $m\in\mathscr{C}$, and the centered conditional moments $\mathsf{E}_t[\tilde{C}^M\mid d]$, $M\in\mathbb{N}^{R-L}$, which are the functions of $p_t(d)$, $\mathsf{E}_t[C_m\mid d]$, $\mathsf{E}_t[\tilde{C}^M\mid d]$. Equation (5.1) will be the first equation of our system. We note that the differential equation defining the marginal probability only depends on the slow reactions. To solve this differential equation, we need to reformulate the unknown conditional means $\mathsf{E}_t[\gamma_s^{r_{si}-n}(C)\mid d]$ through the known $\mathsf{E}_t[C_m\mid d]$, $\mathsf{E}_t[\tilde{C}^M\mid d]$. The details of this transformation can be found in Appendix A.2.

In the following proposition, we will obtain the time evolution equation for the conditional means $\mathsf{E}_t[C_m\mid d]$, $m\in\mathscr{C}$.

Proposition 5.2

$$p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[C_m\mid d] = \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i} - I)\Big(\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\beta_s^n(d)\mathsf{E}_t[C_m\gamma_s^{r_{si}-n}(C)\mid d]p_t(d)\Big)$$
$$+ \sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[\gamma_s^{r_{sj}-n}(C)\delta_{jm}\mid d]p_t(d)$$
$$- \mathsf{E}_t[C_m\mid d]\frac{\partial}{\partial t}p_t(d) \qquad (5.2)$$

where δ_{jm} is the Kronecker delta function.

Proof The proof of the proposition can be found in Appendix A.3.

In the following proposition, we will obtain $p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[\tilde{C}^M \mid d]$.

Proposition 5.3

$$p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[\tilde{C}^M \mid d] = \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i}-I)\Big(\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\beta_s^n(d)\mathsf{E}_t[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d]p_t(d)\Big)$$

$$+\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[M_j\gamma_s^{r_{sj}-n}(C)\tilde{C}^{M-e_j} \mid d]p_t(d) \qquad (5.3)$$

$$+\frac{1}{2}\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[M_j(M_j-1)\gamma_s^{r_{sj}-n}(C)\tilde{C}^{M-2e_j} \mid d]p_t(d)$$

$$-\sum_{j\in\mathscr{C}}M_j\mathsf{E}_t[\tilde{C}^{M-e_j} \mid d]p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[C_j \mid d]-\mathsf{E}_t[\tilde{C}^M \mid d]\frac{\partial}{\partial t}p_t(d).$$

Proof The proof of the theorem can be found in Appendix Section A.4.

Up to this section, we have obtained the time derivatives of the marginal probabilities as well as those of the conditional means and the centered conditional moments. These three equations will give us the following differential equation system.

Theorem 5.4 *Let* $p_t(d,c) = p_t(c \mid d)p_t(d)$ *satisfy Equation (4.7). Then, the time derivative of* $p_t(d)$, $\mathsf{E}_t[C_m \mid d]$, $m \in \mathscr{C}$ *and* $\mathsf{E}_t[\tilde{C}^M \mid d]$, $M = (M_1,M_2,\dots,M_{R-L})$, *satisfies the following system*

$$\frac{\partial}{\partial t}p_t(d) = \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i}-I)\Big(\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\beta_s^n(d)\mathsf{E}_t[\gamma_s^{r_{si}-n}(C) \mid d]p_t(d)\Big)$$

$$p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[C_m \mid d] = \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i}-I)\Big(\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\beta_s^n(d)\mathsf{E}_t[C_m\gamma_s^{r_{si}-n}(C) \mid d]p_t(d)\Big)$$

$$+\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[\gamma_s^{r_{sj}-n}(C)\delta_{jm}(C) \mid d]p_t(d)$$

$$-\mathsf{E}_t[C_m \mid d]\frac{\partial}{\partial t}p_t(d)$$

$$p_t(d)\frac{\partial}{\partial t}\mathsf{E}[\tilde{C}^M \mid d] = \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i}-I)\Big(\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\beta_s^n(d)\mathsf{E}_t[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d]p_t(d)\Big)$$

$$+\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[M_j\gamma_s^{r_{sj}-n}(C)\tilde{C}^{M-e_j} \mid d]p_t(d) \qquad (5.4)$$

$$+\frac{1}{2}\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[M_j(M_j-1)\gamma_s^{r_{sj}-n}(C)\tilde{C}^{M-2e_j} \mid d]p_t(d)$$

$$-\sum_{j\in\mathscr{C}}M_j\mathsf{E}_t[\tilde{C}^{M-e_j} \mid d]p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[C_j \mid d]-\mathsf{E}_t[\tilde{C}^M \mid d]\frac{\partial}{\partial t}p_t(d),$$

where δ_{jm} is kronecker delta function. Also, $\mathscr{F}^{\bar{e}_i}$ is a one step operator as follows:

$$\mathscr{F}^{\bar{e}_i}\Big(f(d)g(c)\Big) = f(d+\bar{e}_i)g(c), i \in \mathscr{D}.$$

where $f(d): \mathbf{D} \to \mathbb{R}$ and $g(c): \mathbb{R}_{\geq 0}^{R-L} \to \mathbb{R}$ be any functions of d and c variables, respectively.

In the following section, we will explain the details of the maximum entropy method which will be used to construct the conditional probability distribution $p_t(c \mid d)$.

6 Maximum Entropy

Assume that we want to obtain the solution of the HME under consideration at a specific time point $\tau > 0$. Solving the ODE system in (5.4) gives $p_\tau(d)$, $\mathsf{E}_\tau[C_m \mid d]$, $\mathsf{E}_\tau[\tilde{C}^M \mid d]$, $m \in \mathscr{C}$, $M = (M_1, M_2, \ldots, M_{R-L})^T \in \mathbb{N}^{R-L}$ values for the system of interest. Although the marginal probabilities, $p_\tau(d)$, can directly be obtained from the ODE system, we still do not know the corresponding conditional probability density function, $p_\tau(c \mid d)$, which will be used to construct the joint probability, $p_\tau(d,c)$, solving the corresponding HME.

To estimate the unknown conditional probability density functions using its moments, we will use the maximum entropy approach proposed by Jaynes [28]. Assume that we have a state space $\Omega = \mathbf{D} \times \mathbb{R}_{\geq 0}^{R-L}$ and our goal is to estimate the unknown probability density function $p_\tau: \Omega \to \mathbb{R}_{\geq 0}$. Let

$$\mathscr{S}_\tau^M = \int_{\mathbb{R}_{\geq 0}^{R-L}} \prod_{j \in \mathscr{C}} c_j^{M_j} p_\tau(c \mid d) dc, \quad M = (M_1, M_2, \ldots, M_{R-L}) \in \mathbb{N}^{R-L},$$

denote the moments of the joint probability density function at time point τ. It should be noted that when $M = e_m$, we obtain $\mathsf{E}_\tau[C_m \mid d]$. To guarantee that $p_\tau(c \mid d)$ is a probability function, we must impose the condition $\mathscr{S}_\tau^0 = 1$. Then, the approximation for the conditional probability density function $p_\tau(c \mid d)$ will be obtained solving the following constrained convex optimization problem

$$\text{Minimize} \quad \int_{\mathbb{R}_{\geq 0}^{R-L}} p_\tau(c \mid d) \ln(p_\tau(c \mid d)) dc$$

$$\text{Subject to} \quad \mathscr{S}_\tau^0 = \int_{\mathbb{R}_{\geq 0}^{R-L}} p_\tau(c \mid d) = 1$$

$$\mathscr{S}_\tau^{e_m} = \mathsf{E}_\tau[C_m \mid d] = \int_{\mathbb{R}_{\geq 0}^{R-L}} c_m p_\tau(c \mid d)$$

$$\mathscr{S}_\tau^M = \mathsf{E}_\tau[C^M \mid d] = \int_{\mathbb{R}_{\geq 0}^{R-L}} \prod_{j \in \mathscr{C}} c_j^{M_j} p_\tau(c \mid d)$$

Let N be the number of moment constraints and M^k, $k = 0, 1, \ldots, N$, denote different choices of vectors $M^k = (M_1^k, M_2^k, \ldots, M_{R-L}^k) \in \mathbb{N}^{R-L}$. To impose the conditions

given above, we will have $M^0 = 0$, $M^j = e_j$, $j = 1, 2, \ldots, R - L$. Then, the solution of this constrained optimization problem can be obtained maximizing the following Lagrange function

$$\mathscr{L}(p_\tau(c \mid d), \lambda(\tau)) = -\int_{\mathbb{R}^{R-L}_{\geq 0}} p_\tau(c \mid d) \ln(p_\tau(c \mid d)) dc$$

$$+ \sum_{k=0}^{N} \lambda_k(\tau) \Big(\int_{\mathbb{R}^{R-L}_{\geq 0}} \prod_{j \in \mathscr{C}} c_j^{M_j^k} p_\tau(c \mid d) dc - \mathscr{S}_\tau^{M^k} \Big),$$

where $\lambda_k \in \mathbb{R}$, $k = 1, 2, \ldots, N$ are referred to as Lagrange multipliers. Taking the derivative of $\mathscr{L}(p_\tau(c \mid d), \lambda(\tau))$ with respect to $p_\tau(c \mid d)$ will give the approximate solution of the conditional probability density for $p_\tau(c \mid d)$ in the following form

$$p_\tau^*(c \mid d) = \arg\max(\mathscr{L}(p_\tau(c \mid d), \lambda(\tau))) = \frac{1}{Z(N, \lambda(\tau))} \exp\Big(-\sum_{k=0}^{N} \lambda_k(\tau) \prod_{j \in \mathscr{C}} c_j^{M_j^k} \Big),$$

where $Z(N, \lambda(\tau))$ is a normalization constant [1,4,5]. Now, we can obtain the approximate solution of the joint probability density function which solves the HME under consideration by multiplying the obtained conditional probability function $p_\tau^*(c \mid d)$ with the marginal probability function $p_\tau(d)$.

7 Application

In the following we will apply our framework to a simple, illustrative reaction system. We will consider two reactions of the form

$$R_1 : \eta S_1 \xrightarrow{\kappa_1} \eta S_2, \quad R_2 : S_2 \xrightarrow{\kappa_2} S_1$$
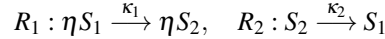
with $\eta \in \mathbb{N}$ and $\eta > 1$. Hence, conversion of S_1 into S_2 appear in infrequent bursts of size η, while the reverse reaction happens in frequent single conversions. The latter reaction may thus be amenable to a diffusion approximation. The state vector of the system at time $t \geq 0$ is defined by $Y(t) = (Y_1(t), Y_2(t))^T \in \mathbb{Z}^2_{\geq 0}$, where $Y_i(t)$ denote the number of molecules of species S_i, $i = 1, 2$.

The joint probability mass function, $p_t(d, c)$ over both reaction counters satisfies the following CME

$$\frac{\partial}{\partial t} p_t(d, c) = \tilde{\alpha}_1(d - 1, c) p_t(d - 1, c) - \tilde{\alpha}_1(d, c) p_t(d, c)$$

$$\tilde{\alpha}_2(d, c - 1) p_t(d, c - 1) - \tilde{\alpha}_2(d, c) p_t(d, c). \tag{7.1}$$

For the sake of simplicity and in order not to introduce another approximation step due to nonlinear propensity, we deviate from mass-action and assume linear propensities, i.e., affine in the counting variables

$$\tilde{\alpha}_1(d, c) = \kappa_1(\beta_1(d) + \gamma_1(c)), \tilde{\alpha}_2(d, c) = \kappa_2(\beta_2(d) + \gamma_2(c)),$$

with

$$\beta_1(d) = y_1(0) - \eta d, \gamma_1(c) = c, \beta_2(d) = y_2(0) + \eta d, \gamma_2(c) = -c.$$

According to the time scale separation mentioned above we now separate reactions and stoichiometric vectors as follows

$$\mathscr{D} = \{1\}, \mathscr{C} = \{2\}, \mu_1^D = (-\eta, \eta)^T, \mu_2^C = (1, -1)^T.$$

Then, the HME for the joint probability density function, $p_t(d,c)$, is defined as :

$$\frac{\partial}{\partial t} p_t(d,c) = \kappa_1(y_1(0) - \eta(d-1) + c) p_t(d-1,c) - \kappa_1(y_1(0) - \eta d + c) p_t(d,c) \quad (7.2)$$
$$- \frac{\partial}{\partial c} \left(\kappa_2(y_2(0) + \eta d - c) p_t(d,c) \right) + \frac{1}{2} \frac{\partial^2}{\partial c^2} \left(\kappa_2(y_2(0) + \eta d - c) p_t(d,c) \right)$$

The system of differential equation defining the marginal probabilities, the conditional means and the centered conditional moments has the following form (which we illustrate for the remainder of this section for $\eta = 2$)

$$\frac{\partial}{\partial t} p_t(d) = \left(\kappa_1(y_1(0) - 2(d-1)) + \kappa_1 \mathsf{E}_t[C \mid d-1] \right) p_t(d-1)$$
$$- \left(\kappa_1(y_1(0) - 2d) - \kappa_1 \mathsf{E}_t[C \mid d] \right) p_t(d) \quad (7.3)$$
$$p_t(d) \frac{\partial}{\partial t} \mathsf{E}_t[C \mid d] = \kappa_1(y_1(0) - 2(d-1)) \mathsf{E}_t[C \mid d-1] p_t(d-1)$$
$$- \kappa_1(y_1(0) - 2d) \mathsf{E}_t[C \mid d] p_t(d) + \kappa_1 \mathsf{E}_t[C^2 \mid d-1] p_t(d-1)$$
$$- \kappa_1 \mathsf{E}_t[C^2 \mid d] p_t(d) + \kappa_2(y_2(0) + 2d) p_t(d) - \kappa_2 \mathsf{E}_t[C \mid d] p_t(d)$$
$$- \mathsf{E}_t[C \mid d] \frac{\partial}{\partial t} p_t(d)$$
$$p_t(d) \frac{\partial}{\partial t} \mathsf{E}_t[\tilde{C}^2 \mid d] = \kappa_1(y_1(0) - 2(d-1)) \mathsf{E}_t[\tilde{C}^2 \mid d-1] p_t(d-1)$$
$$- \kappa_1(y_1(0) - 2d) \mathsf{E}_t[\tilde{C}^2 \mid d] p_t(d) + \kappa_1 \mathsf{E}_t[\tilde{C}^2 C \mid d-1] p_t(d-1)$$
$$- \kappa_1 \mathsf{E}_t[\tilde{C}^2 C \mid d] p_t(d) + \kappa_2(y_2(0) + 2d) p_t(d)$$
$$- \kappa_2 \mathsf{E}_t[C \mid d] p_t(d) - 2\kappa_2 \mathsf{E}_t[\tilde{C} C \mid d] p_t(d) - \mathsf{E}_t[\tilde{C}^2 \mid d] \frac{\partial}{\partial t} p_t(d).$$

Based on our previous discussions, we get the following system of differential equation which will be referred to as the moment equation system of the HME in the rest

of the study

$$
\begin{aligned}
\frac{\partial}{\partial t} p_t(d) = {} & \Big(\kappa_1 (y_1(0) - 2(d-1)) + \kappa_1 \mathsf{E}_t[C \mid d-1] \Big) p_t(d-1) \\
& - \Big(\kappa_1 (y_1(0) - 2d) - \kappa_1 \mathsf{E}_t[C \mid d] \Big) p_t(d)
\end{aligned}
\tag{7.4}
$$

$$
\begin{aligned}
p_t(d) \frac{\partial}{\partial t} \mathsf{E}_t[C \mid d] = {} & \kappa_1 (y_1(0) - 2(d-1)) \mathsf{E}_t[C \mid d-1] p_t(d-1) \\
& - \kappa_1 (y_1(0) - 2d) \mathsf{E}_t[C \mid d] p_t(d) \\
& + \kappa_1 \mathsf{E}_t[\tilde{\Psi}^2 \mid d-1] p_t(d-1) + \kappa_1 (\mathsf{E}_t[C \mid d-1])^2 p_t(d) \\
& - \kappa_1 \mathsf{E}_t[\tilde{C}^2 \mid d] p_t(d) - \kappa_1 (\mathsf{E}_t[C \mid d])^2 p_t(d) + \kappa_2 (y_2(0) + 2d) p_t(d) \\
& - \kappa_2 \mathsf{E}_t[C \mid d] p_t(d) - \mathsf{E}_t[C \mid d] \frac{\partial}{\partial t} p_t(d)
\end{aligned}
\tag{7.5}
$$

$$
\begin{aligned}
p_t(d) \frac{\partial}{\partial t} \mathsf{E}_t[\tilde{C}^2 \mid d] = {} & \Big(\mathsf{E}_t[\tilde{\Psi}^2 \mid d-1] + \{ \mathsf{E}_t[C \mid d-1] - \mathsf{E}_t[C \mid d] \}^2 \Big) \kappa_1 (y_1(0) - 2(d-1)) p_t(d-1) \\
& - \kappa_1 (y_1(0) - 2d) \mathsf{E}_t[\tilde{C}^2 \mid d] p_t(d) + \kappa_1 \mathsf{E}_t[C \mid d-1] \mathsf{E}_t[\tilde{\Psi}^2 \mid d-1] p_t(d-1) \\
& + 2\kappa_1 \{ \mathsf{E}_t[C \mid d-1] - \mathsf{E}_t[C \mid d] \} \mathsf{E}_t[\tilde{\Psi}^2 \mid d-1] p_t(d-1) \\
& + \kappa_1 \{ \mathsf{E}_t[C \mid d-1] - \mathsf{E}_t[C \mid d] \}^2 \mathsf{E}_t[C \mid d-1] p_t(d-1) \\
& - \kappa_1 \mathsf{E}_t[C \mid d] \mathsf{E}_t[\tilde{C}^2 \mid d] p_t(d) + \kappa_2 (y_2(0) + 2d) p_t(d) - \kappa_2 \mathsf{E}_t[C \mid d] p_t(d) \\
& - 2\kappa_2 \mathsf{E}_t[\tilde{C}^2 \mid d] p_t(d) - \mathsf{E}_t[\tilde{C}^2 \mid d] \frac{\partial}{\partial t} p_t(d),
\end{aligned}
\tag{7.6}
$$

where $\tilde{\Psi} = c - \mathsf{E}_t[C \mid d-1]$.

Substitution $\dfrac{\partial}{\partial t} p_t(d)$ into $p_t(d) \dfrac{\partial}{\partial t} \mathsf{E}_t[C \mid d]$, $p_t(d) \dfrac{\partial}{\partial t} \mathsf{E}_t[\tilde{C}^2 \mid d]$ will give a system of differential equations that is expressed only in terms of the marginal probabilities, the conditional means and the second centered moments. In our application, we close moment equations setting the third and the higher moments to zero. If $p_t(d) = 0$, then we will not be able to obtain $\mathsf{E}_t[C \mid d], \mathsf{E}_t[\tilde{C}^2 \mid d]$. To avoid this drawback, in [25], the authors proposed a successful initialization procedure.

Based on the fact that propensity functions must be non negative, we define the state space of the system as follows:

$$
\Omega = \mathbf{D} \times \mathbf{C} = \{ (d,c) \in \mathbf{D} \times \mathbb{R}_{\geq 0} : y_1(0) - 2d + c \geq 0, y_2(0) + 2d - c \geq 0, \mathbf{D} \subset \mathbb{N} \}.
$$

To obtain each conditional probability density function by solving the corresponding convex optimization problem on the state space of interest, we use the CVX toolbox of the MATLAB [23]. When the size of Ω is very high, the dimensionality of the optimization problem increases. Therefore, the CVX cannot produce accurate results.

To keep the dimension of the optimization problems small for the CVX, we construct state space iteratively using a similar strategy to the sliding window method [40]. In summary, our strategy is to solve the moment equation system of the HME using an appropriate discretization method. At each discretization step, we check the marginal probabilities. If they are higher than a given threshold, then we extend the state space of the variable d. This procedure continues until the time point of the interest is reached. Finally, depending on the state space of the variable d, we construct a state space for the variable c. Now, we can explain the details of the method.

In the first step of this construction, we define a feasible subset Ω^0 of Ω, $\Omega^0 = \mathbf{D}^0 \times \mathbf{C}^0$, in which the dimension of the optimization problem is acceptable for the

CVX. To avoid the problem of having $p_0(d) = 0$, we choose an initial Poisson distribution, $p_0(d,c)$, in the state space Ω^0 and compute the corresponding $p_0(d)$, $E_0[C \mid d]$, $E_0[\tilde{C}^2 \mid d]$, which will be considered as the initial conditions for the moment equation of the system of the HME.

Assume that we want to obtain the conditional counting probability density at time point $\tau > 0$. Then, we approximate the solution of the moment equation system of the HME on $[0, \tau]$ time interval using a numerical method. We choose a discretization time step Δ and define $t_j = j\Delta$, $j = 0, 1, \ldots, J$ such that $t_0 = 0$, $t_J = \tau$. As a result, we obtain subintervals $[t_j, t_{j+1}]$, $j = 0, 1, \ldots, J-1$. Let $p^j(d)$, $E^j[C \mid d]$, $E^j[\tilde{C}^2 \mid d]$ represent the approximate solution of the ODE system given in Equation (7.5) and $\mathbf{D}^j$ represents the state space of d at time point t_j. To construct $\mathbf{D}^1$, we will solve the moment equation system of the HME using initial conditions $p^0(d) \equiv p_0(d)$, $E^0[C \mid d] \equiv E_0[C \mid d]$, $E^0[\tilde{C}^2 \mid d] \equiv E_0[\tilde{C}^2 \mid d]$. Then, we will obtain $p^1(d)$, $E^1[C \mid d]$, $E^1[\tilde{C}^2 \mid d]$ values for each d variable in the state space

$$\mathbf{D}^0 = \{d : \min(\mathbf{D}^0) \leq d \leq \max(\mathbf{D}^0)\}.$$

To extend $\mathbf{D}^0$, we define a threshold $\varepsilon > 0$ and check the marginal probability $p^1_{max} \equiv p^1(max(\mathbf{D}^0))$. If $p^1_{max} > \varepsilon$, then we extend $\mathbf{D}^0$ as follows:

$$\mathbf{D}^1 = \{d : \min(\mathbf{D}^0) \leq d \leq \max(\mathbf{D}^0) + 1\}.$$

To approximate the solution of Equation (7.5) at time point t_2, we need to initialize the system on $\mathbf{D}^1$. Although we know $p^1(d)$, $E^1[C \mid d]$, $E^1[\tilde{C}^2 \mid d]$ for $d \in \mathbf{D}^0$, we have to impose initial conditions for $d = \max(\mathbf{D}^0) + 1$

$$p^1(\max(\mathbf{D}^0) + 1) = \frac{\sum\limits_{d \in \mathbf{D}^0} p^1(d)}{\mid \mathbf{D}^0 \mid + 1}, \quad E^1[C \mid \max(\mathbf{D}^0) + 1] = E^1[C \mid \max(\mathbf{D}^0)],$$

$$E^1[\tilde{C}^2 \mid \max(\mathbf{D}^0) + 1] = E^1[\tilde{C}^2 \mid \max(\mathbf{D}^0)],$$

where $\mid \mathbf{D}^0 \mid$ denotes the cardinality of the subset $\mathbf{D}^0$. We employ this procedure successively until the desired time point τ is reached. Let $\mathbf{D}^*$ denote the state space of d at time point τ. Here, we must choose $\varepsilon > 0$ such that, $\mathbf{D}^*$ must also be in the feasible region of the CVX. Now, we can construct the feasible state space for c denoted by $\mathbf{C}^*$. Since we know initial domain $\Omega^0 = \mathbf{D}^0 \times \mathbf{C}^0$, we only need to obtain (d,c) pairs for $d \in \mathbf{D}^* \setminus \mathbf{D}^0$. Then, for a given $\varepsilon > 0$, we construct the feasible region $\mathbf{C}^*$ for variable c as follows:

$$\mathbf{C}^* = \mathbf{C}^0 \cup \bar{\mathbf{C}} \quad \text{with} \quad \bar{\mathbf{C}} = \bigcup_{d \in \mathbf{D}^* \setminus \mathbf{D}^0} \mathbf{C}_d,$$

where

$$\mathbf{C}_d = \left\{ c : max(min(\mathbf{C}^0) - \varepsilon\sigma, 0) \leq c \leq max(\mathbf{C}^0) + \varepsilon\sigma) \wedge y_1(0) - 2d + c \geq 0 \wedge y_2(0) + 2d - c \geq 0 \wedge d \in \mathbf{D}^* \setminus \mathbf{D}^0 \right\},$$

where $\sigma = \sqrt{E^J[\tilde{C}^2 \mid d]}$. Here $max(\mathbf{C}^0)$ and $min(\mathbf{C}^0)$ denote the maximum and the minimum values of c of pairs $(c, max(\mathbf{D}^0)) \in \Omega^0$, respectively. As a result, we have

a feasible region $\Omega^* = \mathbf{D}^* \times \mathbf{C}^*$ for the CVX. Then, we can solve the corresponding convex optimization problems for each conditional counting probability density $p_\tau(c \mid d), d \in \mathbf{D}^*$ using the CVX. Finally, we can compute the approximate solution of $p_\tau(d,c)$. The resulting algorithm is presented in Algorithm 1.

Algorithm 1: Constructing feasible region for the CVX

Input: The state vector Y, the error bound ε, a discretization time step Δ, stoichiometric vectors μ_1^D, μ_2^C, the initial domain $\Omega^0 = \mathbf{D}^0 \times \mathbf{C}^0, \mathbf{D}^0, \mathbf{C}^0 \subset \mathbb{N}_0$, end of the simulation time $\tau > 0$.
Output: The conditional probability at time point τ, $p_\tau(c \mid d)$.

1 Set $t = 0$.
2 Calculate $p_0(d)$, $\mathsf{E}_0[C \mid d]$, $\mathsf{E}_0[\tilde{C}^2 \mid d]$ on domain Ω^0 by using a Poisson distribution.
3 Set $J = \tau/\Delta$ and define $t_j = j\Delta$, $j = 0, 1, \ldots, J$.
4 Set $p^0(d) \equiv p_0(d)$, $\mathsf{E}^0[C \mid d] \equiv \mathsf{E}_0[C \mid d]$, $\mathsf{E}^0[\tilde{C}^2 \mid d] \equiv \mathsf{E}_0[\tilde{C}^2 \mid d]$.
5 **for** $j = 1, 2, \ldots, J$ **do**
6   Solve moment equation of the HME by using any discretization based numerical method and obtain $p^j(d)$, $\mathsf{E}^j[C \mid d]$, $\mathsf{E}^j[\tilde{C}^2 \mid d]$ for $d \in \mathbf{D}^j$
7   **if** $p^j(max(\mathbf{D}^j)) > \varepsilon$ **then**
8   Set $\mathbf{D}^{j+1} = \mathbf{D}^j \cup \{max(\mathbf{D}^j) + 1\}$
9   Set $p^j(d) = p^j(d)$ for $d \in \mathbf{D}^j$
10   Define $p^j(max(\mathbf{D}^j) + 1) = \dfrac{\sum_{d \in \mathbf{D}} p^j(d)}{|\mathbf{D}^j| + 1}$, $\quad \mathsf{E}^j[C \mid max(\mathbf{D}^j) + 1] = \mathsf{E}^j[C \mid max(\mathbf{D}^j)]$,
        $\mathsf{E}^j[\tilde{C}^2 \mid max(\mathbf{D}^j) + 1] = \mathsf{E}^j[\tilde{C}^2 \mid max(\mathbf{D}^j)]$,
11 **end**
12 Set $\mathbf{D}^* = \mathbf{D}^J$
13 Set $\mathbf{C}^* = \mathbf{C}^0$
14 **for** $d \in \mathbf{D}^* \setminus \mathbf{D}^0$ **do**
15   Obtain

$$\mathbf{C}_d = \left\{ c : max(min(\mathbf{C}^0) - \varepsilon\sigma, 0) \leq c \leq max(\mathbf{C}^0) + \varepsilon\sigma) \wedge y_1(0) - 2d + c \geq 0 \wedge y_2(0) + 2d - c \geq 0 \wedge d \in \mathbf{D}^* \setminus \mathbf{D}^0 \right\}$$

16     $\mathbf{C}^* = \mathbf{C}^* \cup \mathbf{C}_d$
17 **end**
18 For each $d \in \mathbf{D}^*$ obtain $p_\tau(c \mid d)$ using the CVX.

In our numerical simulation study, the state of the system is initialized $y(0) = (50, 0)^T$ and the reaction rate constants of R_1, R_2 are given by $\kappa_1 = 0.2\mathrm{s}^{-1}$, $\kappa_2 = 0.4\mathrm{s}^{-1}$, respectively. We define

$$\Omega = \{(d,c) : 50 - 2d + c \geq 0, 2d - c \geq 0, d, c \in \{0, 1, 2, \ldots, 30\}\}.$$

The initial state space Ω^0 is

$$\Omega^0 = \{(d,c) \in \Omega : (d,c) \in \{0, 1, 2, \ldots, 8\}\}.$$

In Figures (7.1a) and (7.2a), one can see the state space Ω shown by the points with only green markers and Ω^0 shown by the points with black edged markers. The threshold for extending the region of the variable d is $\varepsilon = 10^{-6}$, and $\varepsilon = 2$. We obtain joint counting probability density function at time points $\tau = 0.5$ and $\tau = 1$.

We have used the Euler method with fixed time step $\Delta = 10^{-4}$. Figures (7.1a) and (7.2a) also show the Ω^* at time points $\tau = 0.5$ and $\tau = 1$, respectively. In both figures, the state space Ω^* is the union of the points denoted by markers with black and red edges. Figures (7.1b) and (7.2b) show the joint counting probability satisfying the CME given in Equation (7.1) at time points $\tau = 0.5$ and $\tau = 1$, respectively. To obtain these figures, we solve CME in this given restricted domain. To achieve this goal, we obtain the ODE system satisfying the CME and use Euler method with fixed time step $\Delta = 10^{-4}$.

Figures (7.1c) and (7.2c) indicate the approximate solution of the $p_\tau(d,c)$ satisfying Equation (7.2) obtained with Algorithm 1 at time points $\tau = 0.5$ and $\tau = 1$, respectively.
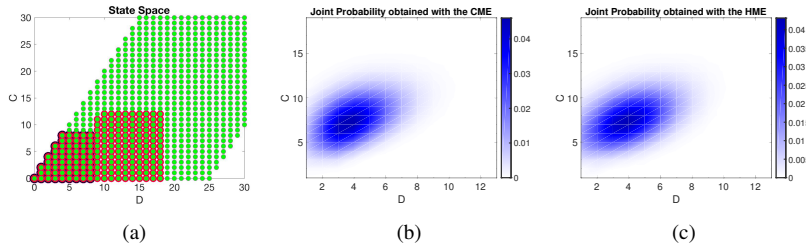

(a)   (b)   (c)

Fig. 7.1: The joint counting probability density function at time point $\tau = 0.5$. (a) The state space Ω is shown by the points only with green markers; Ω^0 is shown by the points with black edges; and Ω^* is the union of the points with black and red edges. (b) The joint counting probability density function satisfying the CME given in Equation (7.1) (c)The joint counting probability density function satisfying the HME given in Equation (7.2)

(a)   (b)   (c)

Fig. 7.2: The joint counting probability density function at time point $\tau = 1$. (a) The state space Ω is shown by the points only with green markers; Ω^0 is shown by the points with black edges; and Ω^* is the union of the points with black and red edges. (b) The joint counting probability density function satisfying the CME given in Equation (7.1) (c)The joint counting probability density function satisfying the HME given in Equation (7.2)
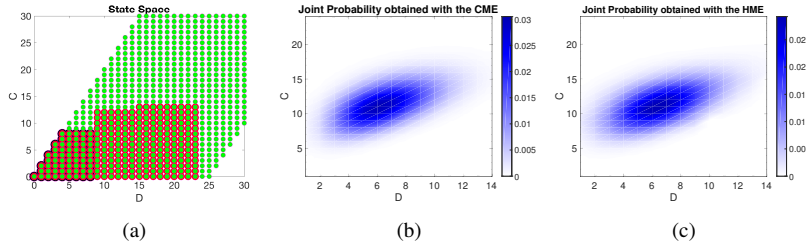
In Table 7.1, one can see two norm of the differences between the joint probabilities, marginal probabilities, conditional means and second central moments obtained from CME given in the Equation (7.1) and from the moment equation of the HME given in Equation (7.4). We should note that the error between the joint probability distributions decreases when the size of the domain of the HME increases. On the other hand, the errors in the marginal probabilities and centered means increases as the size of the domain of the HME increases, this is the result of only taking the evolution equations of the centered moments up to order 2.

Time	$\left\|p_{\tau,CME}(d,c) - p_{\tau,HME}(d,c)\right\|_2$	$\left\|p_{\tau,CME}(d) - p_{\tau,HME}(d)\right\|_2$	$\left\|\mathsf{E}_{\tau,CME}[C\,|\,d] - \mathsf{E}_{\tau,HME}[C\,|\,d]\right\|_2$	$\left\|\mathsf{E}_{\tau,CME}[\bar{C}^2\,|\,d] - \mathsf{E}_{\tau,HME}[\bar{C}^2\,|\,d]\right\|_2$
$\tau = 0.5$	0.0164	$6.1171e-05$	0.1800	1.4027
$\tau = 0.75$	0.0109	$1.0088e-04$	0.3883	1.7907
$\tau = 1$	0.0078	$1.4058e-04$	0.4050	1.9396

Table 7.1: The errors between the joint probability distributions, marginal distributions, centered conditional means and the second centered conditional moments of the CME satisfying Equation (7.1) and HME obtained by the Algorithm 1.

8 Conclusion

In this study, we present the hybrid master equation for jump-diffusion approximation, which models systems with multi-scale nature. The idea of jump diffusion approximation is to separate reactions into fast and slow groups based on an obtained error bound. Fast reactions are modeled using diffusion approximation, while Markov chain representation is employed for slow reactions. In this study, based on the study of Pawula [34], we prove that joint probability density of this hybrid model over reaction counting process satisfies the hybrid master equation, which is the summation of the corresponding chemical master equation and the Fokker-Planck equation. It can be said that while [17] presents a state vector representation for reaction networks with multi-scale nature, the current study complements it by obtaining evolution equation for the corresponding joint probability density over reaction counting process. We note that the main contribution of the paper is to prove that the joint probability density function of jump-diffusion approximation satisfies the HME that correspondingly involves terms from the chemical master equation modeling the reaction counters of the slow reactions and from the Fokker-Planck Equation modeling the reaction counters of fast reactions given the reaction counters of slow reactions. Although different solution procedures such as finite difference method or finite element method [35] can be used to approximate the solution of the corresponding HME, we propose an algorithm based on the method of conditional methods [25]. In our algorithm, we write the joint probability density function as the product of the conditional counting probability density of the fast reactions conditioned on the counting process of the slow reactions and the marginal probability of the counting

process of the slow reactions. To construct the conditional probability density functions at a specific time point, we used the maximum entropy approach. We use the CVX toolbox of the MATLAB to solve the constrained optimization problems. Based on restrictions of the CVX on the dimensionality of the optimization problems, we present a method which constructs feasible regions for the CVX. Although we consider that the main contribution of the paper is the derivation of the HME, there are two open questions in the proposed simulation study which obtains an approximation for the solution of the HME. The first question is about the application area of the proposed algorithm. In the present paper, we have applied the algorithm to a gene model which is a very small system. In our application, we have used zero or first order mass action rate laws to utilize the linearity of the conditional means. Although there is no restriction on propensity functions to obtain evolution equations for the conditional means and centered conditional moments, it is not easy to implement the proposed algorithm to systems involving nonlinear propensity functions. The second question is about the extension of the state space when the size of the system of interest is too large. In our application, we have one slow and one fast reaction. As a result, after solving the moment equation system of the HME on the feasible region which can be acceptable for the CVX, we can extend the state space of the discrete variable by adding 1 to the initial feasible state space of the discrete variable. Based on the new extended state space of the discrete variable, we can extend the state space of continuous variable. When we have more than one slow reaction in the system, the discrete variable will be a vector and we will not be able to extend the feasible region which in turn will cause problems for extension of the feasible region of the state space of the continuous variable. We will try to answer these two open questions in our future works.

A Appendix

A.1 Proof of Lemma 5.1

Proof Using Leibniz integral rule and the boundary conditions gives

$$\frac{\partial}{\partial t}\left(\mathsf{E}_t[F_t(C)\mid d]p_t(d)\right) = \frac{\partial}{\partial t}\left(\int_{\mathbb{R}_{\geq 0}^{R-L}} F_t(c)p_t(c\mid d)p_t(d)dc\right) = \int_{\mathbb{R}_{\geq 0}^{R-L}} F_t(c)\frac{\partial}{\partial t}p_t(d,c)\,dc$$
$$+ \int_{\mathbb{R}_{\geq 0}^{R-L}} p_t(d,c)\frac{\partial}{\partial t}F_t(c)\,dc$$

Inserting Equation (4.11) into the first integral yields

$$\frac{\partial}{\partial t}\left(E_t[F_t(C)\mid d]p_t(d)\right) = \int_{\mathbb{R}_{\geq 0}^{R-L}} F_t(c)\sum_{i\in\mathscr{D}}\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\left(\beta_s^n(d-\bar{e}_i)\gamma_s^{r_{si}-n}(c)p_t(d-\bar{e}_i,c) - \beta_s^n(d)\gamma_s^{r_{si}-n}(c)p_t(d,c)\right)dc$$
$$- \int_{\mathbb{R}_{\geq 0}^{R-L}} F_t(c)\sum_{j\in\mathscr{C}}\frac{\partial}{\partial c_j}\left(\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\gamma_s^{r_{sj}-n}(c)p_t(d,c)\right)dc$$
$$+ \frac{1}{2}\int_{\mathbb{R}_{\geq 0}^{R-L}} F_t(c)\sum_{j\in\mathscr{C}}\frac{\partial^2}{\partial c_j^2}\left(\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\gamma_s^{r_{sj}-n}(c)p_t(d,c)\right)dc$$
$$+ \int_{\mathbb{R}_{\geq 0}^{R-L}} p_t(d,c)\frac{\partial}{\partial t}F_t(c)dc.$$

Since $F_t(c)$ is a polynomial function and sufficiently many moments of $p_t(d,c)$ with respect to c exist, we can manipulate the integral as follows:

$$
\begin{aligned}
\frac{\partial}{\partial t}(E_t[F_t(C)\mid d]p_t(d)) &= \sum_{i\in\mathscr{D}}\kappa_i \int_{\mathbb{R}^{R-L}_{\geq 0}}\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\Big(\beta_s^n(d-\bar{e}_i)F_t(c)\gamma_s^{r_{si}-n}(c)p_t(d-\bar{e}_i,c)-\beta_s^n(d)F_t(c)\gamma_s^{r_{si}-n}(c)p_t(d,c)\Big)dc \\
&\quad -\sum_{j\in\mathscr{C}}\kappa_j \int_{\mathbb{R}^{R-L}_{\geq 0}}F_t(c)\frac{\partial}{\partial c_i}\Big(\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\gamma_s^{r_{sj}-n}(c)p_t(d,c)\Big)dc \\
&\quad +\frac{1}{2}\sum_{j\in\mathscr{C}}\kappa_j \int_{\mathbb{R}^{R-L}_{\geq 0}}F_t(c)\frac{\partial^2}{\partial c_j^2}\Big(\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\gamma_s^{r_{sj}-n}(c)p_t(d,c)\Big)dc \\
&\quad +\int_{\mathbb{R}^{R-L}_{\geq 0}}p_t(d,c)\frac{\partial}{\partial t}F_t(c)\,dc.
\end{aligned}
$$

Here, we want to draw the attention of the reader to the following mean which is used in our equations

$$
E_t[F_t(C)\gamma_s^{r_{sj}-n}(C)\beta_s(d)\mid d]=\int_{\mathbb{R}^{R-L}_{\geq 0}}F_t(c)\gamma_s^{r_{sj}-n}(c)\beta_s^n(d)p_t(c\mid d)\,dc=\beta_s(d)E_t[F_t(C)\gamma_s^{r_{sj}-n}(C)\beta_s^n(d)\mid d].
$$

Using this equality and the properties of the FPE will give us the following equation [9,33]:

$$
\begin{aligned}
\frac{\partial}{\partial t}(E_t[F_t(C)\mid d]p_t(d)) &= \sum_{i\in\mathscr{D}}\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}E_t[F_t(C)\gamma_s^{r_{si}-n}(C)\mid d-\bar{e}_i]\beta_s^n(d-\bar{e}_i)p_t(d-\bar{e}_i) \\
&\quad -\sum_{i\in\mathscr{D}}\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}E_t[F_t(C)\gamma_s^{r_{si}-n}(C)\mid d]\beta_s^n(d)p_t(d) \\
&\quad +\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)E_t[\gamma_s^{r_{sj}-n}(C)\frac{\partial}{\partial c_j}F_t(C)\mid d]p_t(d) \\
&\quad +\frac{1}{2}\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)E_t[\gamma_s^{r_{sj}-n}(C)\frac{\partial^2}{\partial c_j^2}F_t(C)\mid d]p_t(d) \\
&\quad +E_t[\frac{\partial}{\partial t}F_t(C)\mid d]p_t(d).
\end{aligned}
$$

Define

$$
\begin{aligned}
\tilde{\Gamma}(\beta(d),F_t(c)\gamma(c)) &= \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i}-I)\Big(\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\beta_s^n(d)E_t[F_t(C)\gamma_s^{r_{sj}-n}(C)\mid d]\Big) \\
&\quad +\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)E_t[\gamma_s^{r_{sj}-n}(C)\frac{\partial}{\partial c_j}F_t(C)\mid d] \\
&\quad +\frac{1}{2}\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}E_t[\gamma_s^{r_{sj}-n}(C)\frac{\partial^2}{\partial c_j^2}F_t(C)\mid d]
\end{aligned}
$$

Then, we obtain

$$
\frac{\partial}{\partial t}(E_t[F_t(C)\mid d]p_t(d))=\tilde{\Gamma}(\beta(d),F_t(c)\gamma(c))p_t(d)+E_t[\frac{\partial}{\partial t}F_t(C)\mid d]p_t(d)
$$

which completes the proof.

A.2 Mean of the propensity functions

To express $E_t[\gamma_s^{r_{si}-n}(C) \mid d]$ in terms of $E_t[C_m \mid d]$, $E_t[\tilde{C}^M \mid d]$, we will use the Taylor series expansion of $\gamma_s^{r_{si}-n}(C)$ around $E_t[C \mid d]$. The Taylor polynomial of degree q for $\gamma_s^{r_{si}-n}(C)$ around $E_t[C \mid d]$ has the following form

$$G_q(c) = \sum_{k_1+\ldots+k_{R-L}\leq q} \frac{\partial_1^{k_1}\ldots\partial_{R-L}^{k_{R-L}}}{k_1!\ldots k_{R-L}!} \gamma_s^{r_{si}-n}(E_t[C \mid d])(c_{k_1} - E_t[C_{k_1} \mid d])^{k_1}\ldots(c_{k_{R-L}} - E_t[C_{k_{R-L}} \mid d])^{k_{R-L}},$$
(A.1)

where $\partial_j^\ell = \dfrac{\partial^\ell}{\partial c_j^\ell}$, $\ell = k_1,\ldots,k_{R-L}$. In general cellular reactions are unimolecular or bimolecular. Therefore, the third and the higher order derivatives will be zero, meaning that the conditional mean of the function $\gamma_s^{r_{si}-n}(C)$ satisfies [25]

$$E_t[\gamma_s^{r_{si}-n}(C) \mid d] = \gamma_s^{r_{si}-n}(E_t[C \mid d]) + \frac{1}{2}\sum_{k,\ell=1}^{R-L} \frac{\partial^2}{\partial c_k \partial c_\ell} \gamma_s^{r_{si}-n}(E_t[C \mid d])E_t[\tilde{C}^{e_k+e_\ell} \mid d].$$
(A.2)

Here, we use the fact that $E_t[C_i - E_t[C_i \mid d] \mid d] = 0$, $i \in \mathscr{C}$. As a result, if we have a reaction with linear propensity, the second term in Equation (A.2) will also be zero and we will get $E_t[\gamma_s^{r_{si}-n}(C) \mid d] = \gamma_s^{r_{si}-n}(E_t[C \mid d])$.

A.3 Proof of Proposition 5.2

Proof We will use the product rule for derivatives as follows:

$$p_t(d)\frac{\partial}{\partial t}E_t[C_m \mid d] = \frac{\partial}{\partial t}\left(E_t[C_m \mid d]p_t(d)\right) - E_t[C_m \mid d]\frac{\partial}{\partial t}p_t(d).$$
(A.3)

By setting $F_t(c) = c_m$ in Lemma 5.1, we can obtain the first derivative on the right hand-side of Equation (A.3) as follows:

$$\frac{\partial}{\partial t}\left(E_t[C_m \mid d]p_t(d)\right) = \tilde{\Gamma}(\beta(d), c_m\gamma(c))p_t(d) + E_t[\frac{\partial}{\partial t}C_m \mid d]p_t(d).$$

By using equalities $\dfrac{\partial}{\partial t}C_m = 0$, $\dfrac{\partial^2}{\partial c_j^2}C_m = 0$, we get

$$p_t(d)\frac{\partial}{\partial t}E_t[C_m \mid d] = \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i} - I)\left(\kappa_i\prod_{s=1}^M\sum_{n=0}^{r_{si}}\beta_s^n(d)E_t[C_m\gamma_s^{r_{si}-n}(C) \mid d]p_t(d)\right)$$
$$+ \sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^M\sum_{n=0}^{r_{sj}}\beta_s^n(d)E_t[\gamma_s^{r_{sj}-n}(C)\delta_{jm} \mid d]p_t(d) - E_t[C_m \mid d]\frac{\partial}{\partial t}p_t(d),$$

where δ_{jm} is the Kronecker delta function.

In Equation (5.2), we have the conditional mean $E_t[\gamma_s^{r_{sj}-n}(C)\delta_{jm} \mid d]$. If $j \neq m$, then this term equals to zero; otherwise, we will have $E_t[\gamma_s^{r_{sj}-n}(C) \mid d]$. By using Equation (A.2), we can express $E_t[\gamma_s^{r_{sj}-n}(C) \mid d]$ also in terms of the conditional means and the centered conditional moments. Additionally, in Equation (5.2), we have $E_t[C_m\gamma_s^{r_{si}-n}(C) \mid d]$, which must also be reformulated in terms of the conditional means $E_t[C_m \mid d]$, $m \in \mathscr{C}$, and the centered conditional moments $E_t[\tilde{C}^M \mid d]$. To achieve this goal, we will add and subtract $E_t[C_m \mid d]$ term to and from $E_t[C_m\gamma_s^{r_{si}-n}(C) \mid d]$ as follows [25]:

$$E_t[C_m\gamma_s^{r_{si}-n}(C) \mid d] = E_t[(C_m - E_t[C_m \mid d] + E_t[C_m \mid d])\gamma_s^{r_{si}-n}(C) \mid d]$$
$$= E_t[\tilde{C}^{e_m}\gamma_s^{r_{si}-n}(C) \mid d] + E_t[C_m \mid d]E_t[\gamma_s^{r_{si}-n}(C) \mid d].$$

Similarly, adding and subtracting $E_t[C_m \mid d - \bar{e}_i]$ terms to and from $\mathsf{E}_t[C_m \gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i]$ will produce

$$
\begin{aligned}
\mathsf{E}_t[C_m \gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i] &= \mathsf{E}_t[(C_m - \mathsf{E}_t[C_m \mid d - \bar{e}_i] + \mathsf{E}_t[C_m \mid d - \bar{e}_i]) \gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i] \\
&= \mathsf{E}_t[\tilde{\Psi}^{e_m} \gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i] + \mathsf{E}_t[C_m \mid d - \bar{e}_i]\mathsf{E}_t[\gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i],
\end{aligned}
$$

where $\tilde{\Psi} = c - \mathsf{E}_t[C \mid d - \bar{e}_i]$ and for $M \in \mathbb{N}^{R-L}$, $\tilde{\Psi}^M = \prod_{j \in \mathscr{C}} \tilde{\Psi}_j^{M_j}$. Inserting the obtained $\mathsf{E}_t[C_m \gamma_s^{r_{si}-n}(C) \mid d]$ and $\mathsf{E}_t[C_m \gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i]$ values into Equation (5.2) will give us the following equation

$$
\begin{aligned}
p_t(d) \frac{\partial}{\partial t} \mathsf{E}_t[C_m \mid d] &= \sum_{i \in \mathscr{D}} \prod_{s=1}^{M} \sum_{n=0}^{r_{si}} \kappa_i \Big(\mathsf{E}_t[\tilde{\Psi}^{e_m} \gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i] + \mathsf{E}_t[C_m \mid d - \bar{e}_i]\mathsf{E}_t[\gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i] \Big) \beta_s^n(d - \bar{e}_i) p_t(d - \bar{e}_i) \\
&\quad - \sum_{i \in \mathscr{D}} \prod_{s=1}^{M} \sum_{n=0}^{r_{si}} \kappa_i \Big(\mathsf{E}_t[\tilde{C}^{e_m} \gamma_s^{r_{si}-n}(C) \mid d] + \mathsf{E}_t[C_m \mid d]\mathsf{E}_t[\gamma_s^{r_{si}-n}(C) \mid d] \Big) \beta_s^n(d) p_t(d) \\
&\quad + \sum_{j \in \mathscr{C}} \kappa_j \prod_{s=1}^{M} \sum_{n=0}^{r_{sj}} \Big(\mathsf{E}_t[\gamma_s^{r_{sj}-n}(C) \delta_{jm} \mid d] \beta_s^n(d) p_t(d) \Big) - \mathsf{E}_t[C_m \mid d] \frac{\partial}{\partial t} p_t(d).
\end{aligned}
\tag{A.4}
$$

As mentioned before, the Taylor series expansion of $\gamma_s^{r_{si}-n}(C)$ around $\mathsf{E}_t[C \mid d]$ will give us the possibility to reformulate $\mathsf{E}_t[\gamma_s^{r_{si}-n}(C) \mid d]$ using the conditional means and the centered conditional moments. Here, we must reformulate $\mathsf{E}_t[\tilde{C}^{e_m} \gamma_s^{r_{si}-n}(C) \mid d]$ and $\mathsf{E}_t[\tilde{\Psi}^{e_m} \gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i]$ using the corresponding conditional means and the centered moments. To do this, we will use the Taylor expansion given in Equation (A.1) as follows:

$$
\begin{aligned}
\mathsf{E}_t[\tilde{C}^M \gamma_s^{r_{si}-n}(C) \mid d] &= \gamma_s^{r_{si}-n}(\mathsf{E}_t[C \mid d])\mathsf{E}_t[\tilde{C}^M \mid d] + \sum_{k \in \mathscr{C}} \frac{\partial}{\partial c_k} \gamma_s^{r_{si}-n}(\mathsf{E}_t[C \mid d])\mathsf{E}_t[\tilde{C}^{M+e_k} \mid d] \\
&\quad + \frac{1}{2} \sum_{k \in \mathscr{C}} \frac{\partial^2}{\partial c_k \partial c_\ell} \gamma_s^{r_{si}-n}(\mathsf{E}_t[C \mid d])\mathsf{E}_t[\tilde{C}^{M+e_k+e_\ell} \mid d],
\end{aligned}
\tag{A.5}
$$

where $M = (M_1, M_2, \ldots, M_{R-L})^T \in \mathbb{N}^{R-L}$. It is clear that $\mathsf{E}_t[\tilde{\Psi}^M \gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i]$ can also be reformulated by using the Taylor expansion of $\gamma_s^{r_{si}-n}(C)$ around $\mathsf{E}_t[C \mid d - \bar{e}_i]$. Substitution of the new representations of $\mathsf{E}_t[\tilde{C}^M \gamma_s^{r_{si}-n}(C) \mid d]$ and $\mathsf{E}_t[\tilde{\Psi}^M \gamma_s^{r_{si}-n}(C) \mid d - \bar{e}_i]$, which only depend on the conditional means and the centered conditional moments conditioned on the corresponding discrete variable into Equation (5.2) will produce $p_t(d) \frac{\partial}{\partial t} \mathsf{E}_t[C_m \mid d]$ in terms of the marginal probabilities, the conditional means and the centered conditional moments.

A.4 Proof of Proposition 5.3

Proof Similar to our previous proofs, again we will use the product rule for derivatives as follows:

$$
p_t(d) \frac{\partial}{\partial t} \mathsf{E}_t[\tilde{C}^M \mid d] = \frac{\partial}{\partial t} \Big(\mathsf{E}_t[\tilde{C}^M \mid d] p_t(d) \Big) - \mathsf{E}_t[\tilde{C}^M \mid d] \frac{\partial}{\partial t} p_t(d).
$$

The first term in the right hand-side of the equation above can be obtained from Lemma 5.1 choosing $F(c) = \tilde{c}^M$. Then, we obtain

$$
\frac{\partial}{\partial t} \Big(\mathsf{E}_t[\tilde{C}^M \mid d] p_t(d) \Big) = \tilde{\Gamma}(\beta(d), \tilde{c}^M \gamma(c)) p_t(d) + \mathsf{E}_t[\frac{\partial}{\partial t} \tilde{C}^M \mid d] p_t(d).
$$

Since, we have

$$
\begin{aligned}
\frac{\partial}{\partial c_i} \tilde{C}^M &= \frac{\partial}{\partial c_i} \prod_{k \in \mathscr{C}} (c_k - \mathsf{E}_t[C_k \mid d])^{M_k} = M_i \tilde{C}^{M-e_i} \\
\frac{\partial^2}{\partial c_j^2} \tilde{C}^M &= \frac{\partial^2}{\partial c_j^2} \prod_{k \in \mathscr{C}} (c_k - \mathsf{E}_t[C_k])^{M_k} = M_j(M_j - 1) \tilde{C}^{M-2e_j} \\
\mathsf{E}_t[\frac{\partial}{\partial t} \tilde{C}^M \mid d] &= - \sum_{j \in \mathscr{C}} M_j \mathsf{E}_t[\tilde{C}^{M-e_j} \mid d] p_t(d) \frac{\partial}{\partial t} \mathsf{E}_t[C_j \mid d].
\end{aligned}
$$

We get

$$
p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[\tilde{C}^M \mid d] = \sum_{i\in\mathscr{D}}\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\Big(\beta_s^n(d-\bar{e}_i)\mathsf{E}_t[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d-\bar{e}_i]p_t(d-\bar{e}_i) - \beta_s^n(d)\mathsf{E}_t[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d]p_t(d)\Big)
$$

$$
+ \sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[M_j\gamma_s^{r_{sj}-n}(C)\tilde{C}^{M-e_j} \mid d]p_t(d)
$$

$$
+ \frac{1}{2}\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[M_j(M_j-1)\gamma_s^{r_{sj}-n}(C)\tilde{C}^{M-2e_j} \mid d]p_t(d)
$$

$$
- \sum_{j\in\mathscr{C}}M_j\mathsf{E}_t[\tilde{C}^{M-e_j} \mid d]p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[C_j \mid d] - \mathsf{E}_t[\tilde{C}^M \mid d]\frac{\partial}{\partial t}p_t(d).
$$

As a result, we obtain

$$
p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[\tilde{C}^M \mid d] = \sum_{i\in\mathscr{D}}(\mathscr{F}^{-\bar{e}_i}-I)\Big(\kappa_i\prod_{s=1}^{M}\sum_{n=0}^{r_{si}}\beta_s^n(d)\mathsf{E}_t[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d]p_t(d)\Big)
$$

$$
+ \sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[M_j\gamma_s^{r_{sj}-n}(C)\tilde{C}^{M-e_j} \mid d]p_t(d)
$$

$$
+ \frac{1}{2}\sum_{j\in\mathscr{C}}\kappa_j\prod_{s=1}^{M}\sum_{n=0}^{r_{sj}}\beta_s^n(d)\mathsf{E}_t[M_j(M_j-1)\gamma_s^{r_{sj}-n}(C)\tilde{C}^{M-2e_j} \mid d]p_t(d)
$$

$$
- \sum_{j\in\mathscr{C}}M_j\mathsf{E}_t[\tilde{C}^{M-e_j} \mid d]p_t(d)\frac{\partial}{\partial t}\mathsf{E}_t[C_j \mid d] - \mathsf{E}_t[\tilde{C}^M \mid d]\frac{\partial}{\partial t}p_t(d),
$$

which completes our proof.

To formulate the right hand-side of Equation (5.3) in terms of the marginal probabilities, the conditional means and the centered conditional moments, we must restate $\mathsf{E}_t[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d-e_i]$, $\mathsf{E}_t[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d]$ terms using these terms. $\mathsf{E}_t[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d]$, $\mathsf{E}_t[M_j\gamma_s^{r_{sj}-n}(C)\tilde{C}^{M-e_j} \mid d]$, $\mathsf{E}_t[M_iM_j\gamma_s^{r_{sk}-n}(C)\tilde{C}^{M-e_i-e_j} \mid d]$ can be expressed utilizing the corresponding Taylor series expansion given in Equation (A.5). To express $\mathsf{E}_t[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d-e_i]$ in terms of the marginal probabilities, the conditional means and the centered conditional moments, we will add and subtract $\mathsf{E}_t[C \mid d-e_i]$ to and from $\tilde{C}^M$ as follows:

$$
\tilde{c}^M = \Big(c-\mathsf{E}_t[C \mid d]\Big)^M = \Big(c-\mathsf{E}_t[C \mid d-\bar{e}_i]+\mathsf{E}_t[C \mid d-\bar{e}_i]-\mathsf{E}_t[C \mid d]\Big)^M
$$

$$
= \sum_{0\le k\le M}\binom{M}{k}\Big(\mathsf{E}_t[C \mid d-\bar{e}_i]-\mathsf{E}_t[C \mid d]\Big)^{M-k}\tilde{\Psi}^k.
$$

Then, we obtain

$$
\mathsf{E}[\tilde{C}^M\gamma_s^{r_{si}-n}(C) \mid d-\bar{e}_i] = \sum_{0\le k\le M}\binom{M}{k}\Big(\mathsf{E}_t[C \mid d-\bar{e}_i]-\mathsf{E}[C \mid d]\Big)^{M-k}\mathsf{E}[\tilde{\Psi}^k\gamma_s^{r_{si}-n}(C) \mid d-\bar{e}_i].
$$

Using the Taylor series expansion of $\gamma_s^{r_{si}-n}(C)$ around $\mathsf{E}_t[C \mid d-\bar{e}_i]$ gives us the corresponding Taylor series representation for $\mathsf{E}_t[\tilde{\Psi}^k\gamma_s^{r_{si}-n}(C) \mid d-\bar{e}_i]$. As a result, we can obtain the right-hand-side of Equation (5.3) in terms of the marginal probabilities, the conditional means and the centered moments.

References

1. Abramov, R.V.: The multidimensional maximum entropy moment problem: a review of numerical methods. Communications in Mathematical Sciences **8**(2), 377– 392 (2010)

2. Anderson, D.F., Kurtz, T.G.: Continuous time Markov chain models for chemical reaction networks. In: H. Koeppl, G. Setti, M.d. Bernardo, D. Densmore (eds.) Design and Analysis of Biomolecular Circuits. Springer-Verlag (2011)
3. Andreychenko, A., Mikeev, L., Wolf, V.: Model reconstruction for moment-based stochastic chemical kinetics. ACM Trans. Model. Comput. Simul. **25**(2) (2015)
4. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2004)
5. Bretthorst, G.L.: The maximum entropy method of moments and Bayesian probability theory. AIP Conf. Proc. **1553**, 3–15 (2013)
6. Bronstein, L., Koeppl, H.: A variational approach to moment-closure approximations for the kinetics of biomolecular reaction networks. The Journal of Chemical Physics **148**(1), 014,105 (2018)
7. Chevallier, A., Engblom, S.: Pathwise error bounds in multiscale variable splitting methods for spatial stochastic kinetics. SIAM J. Numer. Anal. **58**(1), 469–498 (2018)
8. Crudu, A., Debussche, A., Radulescu, O.: Hybrid stochastic simplifications for multiscale gene networks. BMC Systems Biology **3**(89) (2009)
9. Cseke, B., Schnoerr, D., Opper, M., Sanguinetti, G.: Expectation propagation for continuous time stochastic processes. Journal of Physics A: Mathematical and Theoretical **49**(49) (2016)
10. Doob, J.L.: Markoff chains–denumerable case. Transactions of the American Mathematical Society **58**(3), 455–473 (1945)
11. Duncan, A., Erban, R., Zygalakis, K.: Hybrid framework for the simulation of stochastic chemical kinetics. Journal of Computational Physics **326** (2016)
12. Eldar, A., Elowitz, M.B.: Functional roles for noise in genetic circuits. Nature **467**, 167–173 (2010)
13. Engblom, S.: Computing the moments of high dimensional solutions of the master equation. Applied Mathematics and Computation **180**, 498–515 (2006)
14. Engblom, S., Hellander, A., Lötstedt, P.: Multiscale Simulation of Stochastic Reaction-Diffusion Networks, pp. 55–79. Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology. Springer (2017)
15. Fedoroff, N., Fontana, W.: Small numbers of big molecules. Science **297** (2002)
16. Friedman, N., Cai, L., Xie, X.: Stochasticity in gene expression as observed by single-molecule experiments in live cells. Israel Journal of Chemistry **49**, 333–342 (2010)
17. Ganguly, A., Altıntan, D., Koeppl, H.: Jump-diffusion approximation of stochastic reaction dynamics: Error bounds and algorithms. Multiscale Model. Simul. **13**(4), 1390–1419 (2015)
18. Gardiner, C.W.: Handbook of stochastic methods, vol. 3. Springer Berlin (1985)
19. Gibson, M., Bruck, J.: Efficient exact stochastic simulation of chemical systems with many species and many channels. The Journal of Physical Chemistry A **104**(9), 1876–1889 (2000)
20. Gillespie, D.: A rigorous derivation of the chemical master equation. Physica A **188**, 404–425 (1992)
21. Gillespie, D.T.: The chemical Langevin equation. Journal of Chemical Physics **113**(1), 297–306 (2000)
22. Gillespie, D.T.: Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. **58**, 35–55 (2007)
23. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 2.1 (2014)
24. Grima, R., Thomas, P., Straube, A.V.: How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations? Journal of Chemical Physics **135**(8) (2011)
25. Hasenauer, J., Wolf, V., Kazeroonian, A., Theis, F.J.: Method of conditional moments (MCM) for the chemical master equation : A unified framework for the method of moments and hybrid stochastic-deterministic models. J. Math. Biol. **69**(3), 687–735 (2014)
26. Hellander, A., Lötstedt, P.: Hybrid method for the chemical master equation. Journal of Computational Physics **227**(1), 100–122 (2007)
27. Jahnke, T.: On reduced models for the chemical master equation. Multiscale Model. Simul. **9**(4), 1646–1676 (2011)
28. Jaynes, E.T.: Information theory and statistical mechanics. Physical Review **106**(4), 620 (1957)
29. Kampen, N.G.v.: The diffusion approximation for Markov process. In: I. Lamprecht, A.I. Zotin (eds.) Thermodynamics and Kinetics of Biological Processes, pp. 185–195. Walter de Gruyter and Co. (1982)
30. Kurtz, T.G.: Strong approximation theorems for density dependent Markov cahins. Stochast. Process. Appl. **6**(3), 177–191 (1978)
31. Lee, C.H., Kim, K.H., Kim, P.: A moment closure method for stochastic reaction networks. The Journal of Chemical Physics **130**, 134,107 (2009)
32. Menz, S., Latorre, J.C., Schütte, C., Huisinga, W.: Hybrid stochastic–deterministic solution of the chemical master equation. Multiscale Model. Simul. **10**(4), 1232–1262 (2012)

33. Otten, D.L., Vedula, P.: A quadrature based method of moments for nonlinear Fokker-Planck equations. Journal of Statistical Mechanics: Theory and Experiment **2011**(9) (2011)
34. Pawula, R.F.: Generalizations and extensions of the Fokker-Planck Kolmogorov equations. IEEE Transactions on Information Theory **13**(1) (1967)
35. Pichler, L., Masud, A., Bergman, L.A.: Numerical solution of the Fokker-Planck equation by finite difference and finite element methods - a comparative study. In: M. Papadrakakis, G. Stefanou, V. Papadopoulos (eds.) Computational Methods in Stochastic Dynamics, Computational Methods in Applied Sciences, pp. 69–85. Springer Netherlands (2013)
36. Risken, H., Haken, H.: The Fokker-Planck Equation: Methods of Solution and Applications Second Edition. Springer (1989)
37. Salis, H., Kaznessis, Y.: Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. The Journal of Chemical Physics **122**, 054,103 (2005)
38. Singh, A., Hespanha, J.P.: Approximate moment dynamics for chemically reacting systems. IEEE Trans. on Automat. Contr. **56**(2), 414–418 (2011)
39. Whittle, P.: On the use of the normal approximation in the treatment of stochastic processes. Journal of the Royal Statistical Society: Series B (Methodological) **19**(2), 268–281 (1957)
40. Wolf, V., Goel, R., Mateescu, M., Henzinger, T.A.: Solving the chemical master equation using sliding windows. BMC Systems Biology **4** (2010)