SAND2013-10187J

Hybrid Discrete/Continuum Algorithms for Stochastic Reaction Networks

Abstract

Direct solutions of the Chemical Master Equation (CME) governing Stochastic Reaction Networks (SRNs) are generally prohibitively expensive due to excessive numbers of possible discrete states in such systems. To enhance computational efficiency we develop a hybrid approach where the evolution of states with low molecule counts is treated with the discrete CME model while that of states with large molecule counts is modeled by the continuum Fokker-Planck equation. The Fokker-Planck equation is discretized using a 2^{nd} order finite volume approach with appropriate treatment of flux components to avoid negative probability values. The numerical construction at the interface between the discrete and continuum regions implements the transfer of probability reaction by reaction according to the stoichiometry of the system. The performance of this novel hybrid approach is explored for a two-species circadian model with computational efficiency gains of about one order of magnitude.

Keywords: Chemical Master Equation, Fokker-Planck Equation, Finite Volume, Flux Splitting, Hybrid Discrete-Continuum Models

1. Introduction

Stochastic noise is prevalent in a wide variety of systems, especially when the system behavior is affected or controlled by phenomena on a nanoscale, molecular level, where thermal noise introduces intrinsic variability in molecular interactions. Common examples of such phenomena are chemical, biochemical, or electrochemical reactions between small numbers of molecules, found in gene regulation, cell signaling, or interfacial electrochemistry, generally referred to in this paper as Stochastic Reaction Networks (SRNs). Given the relevance of these processes in applications ranging from bioremediation and bioenergy (bacterial behavior), biomedicine (immune system signaling), to electrical storage (electrodes), effective tools are needed for the simulation and analysis of SRNs.

Mathematically, SRNs are continuous time, discrete state Markov processes, governed by the Chemical Master Equation (CME) [1], which describes the time evolution of the probability of the system being in a particular state. Here the system state consists of the number of molecules of each species in the system. While the CME can be solved as a system of linear equations, its solution quickly becomes challenging as the number of possible system states increases exponentially with the number of species. For this reason, many studies rely on the Stochastic Simulation Algorithm (SSA)[2, 3], which can efficiently generate sample trajectories of the system state. Based on this fundamental simulation method, many advances have been made over the years [4–12] resulting in a wide range of powerful methods for sampling stochastic systems.

While SSA is a powerful tool for generating sample trajectories of the

system state, even in high dimensions, one of its drawbacks is that many samples are needed to construct an accurate probability density function of the system state. As the CME solution offers a full probability distribution of the system states, rather than just samples, it is able to better capture low probability events, which are often missed by sampling-based methods unless a prohibitively large number of samples is used. In recent years, various approaches have been developed to make the direct solution of the CME more computationally tractable. One approach is *finite state projection* [13–16], which reduces the size of the system of equations to be solved by only keeping track of those states that have a non-neglible probability associated with them. This projected state space is adaptively expanded and shrunk as the state probabilities vary in time. Another approach to reducing the number of equations is to group multiple states together via *aggregation* [17, 18]. For a given set of equations, advances in matrix exponentiation [18–21], or taking advantage of time scale separation [22] can speed up the CME solution. Ohter CME solution approaches rely on a spectral representation of the probabilities of the system states [17, 23–27].

The combination of these algorithms has led to dramatic speedups in the solution of the CME, enabling the study of small SRNs by solving their CME directly. For small system (a few species), this approach has been shown to be competitive with and complementary to the SSA sampling based approaches (see e.g. [19]).

Despite these advances in solving the CME, a direct solution is still intractable for systems with many species, especially if some of those species are present in large numbers of molecules, thereby significantly increasing

the size of the state space. For species that are present in large molecule counts, it is often appropriate to use a continuum representation for the system state [28], and rely on the Fokker-Planck Equation (FPE) [29, 30], which is the continuum equivalent to the discrete CME, or use deterministic rate equations. In this context, Sjöberg *et al.* [31] developed a hybrid approach that uses the FPE for species that are present in large numbers of molecules, while keeping the discrete CME for species that only have a few molecules. Other approaches use the discrete SSA approach for some species, and continuum, deterministic rate equations for others [32].

While the hybrid approaches offer large computational savings for systems where a clean separation is possible between the species that should be handled with a discrete formulation, and those that can be handled with a continuum approach, such a separation is not always feasible, and may vary as species molecule counts vary in time. Rather than relying on a static separation, where species are handled with either a discrete or continuum formulation, the current work varies the formulation over different areas in the species phase space. In the area where all species are present with few molecules, the CME is used for all species, since discreteness is important in that area. In areas corresponding to a large number of molecules for all species the FPE formulation is used. In regions where some species are present in small numbers, but others in large numbers, the formulation by Sjöberg *et al.* [31] is used.

By using the discrete CME formulation only where it is needed, the algorithm presented in this paper can allocate computational resources (*i.e.* grid density) more effectively. Also, as will be shown, the resulting set of

equations for the hybrid approach is very similar as the original set of linear equations corresponding to the CME. This means that the hybrid algorithm can be used in synergy with other approaches such as finite state projection and improved matrix exponentiation for further computational gains. For proof of concept, the approach was implemented for systems with two molecular species, and applied to the simulation of SRNs describing a simple metabolite model, and a circadian rhythm model. Canonical tests show second order convergence and good accuracy compared to full CME simulation, with an order of magnitude gain in computational efficiency.

This paper is organized as follows. Section 2 describes the CME approach to study SRNs and compares its direct solution with results based on the SSA. Section 3 introduces the Fokker-Planck approximation to the CME, and results based on a finite volume discretization of the FPE are compared to full CME results. The hybrid CME-FPE formulation is formulated in Section 4, numerical experiments are presented in Section 5, and conclusions are presented in Section 6.

2. Chemical Master Equation

In this section we describe the Chemical Master Equation (CME) approach to study the dynamics of Stochastic Reaction Networks. The CME models the evolution of the probability of the state of the system.

Consider a stochastic reaction network (SRN) for a system consisting of two species, X and Y. The 2-dimensional state space: $\boldsymbol{n} = (n_x, n_y) \in \mathbb{N}_0^2$, corresponds to molecule counts for species X and Y and $\mathbb{N}_0$ is the set of all non-negative integers. Denote the (jump, propensity) pairs by

$\{(\boldsymbol{\nu}_r, w_r(\boldsymbol{N}))\}_{r=1}^{R}$, where R is the number of all reactions. More specifically, $\boldsymbol{n} \to \boldsymbol{n} + \boldsymbol{\nu}_r$ with probability $w_r(\boldsymbol{n})$ per unit time.

The process is fully specified by the probabilities $p(\boldsymbol{n}; t)$ that evolve according to the CME:

$$\frac{d}{dt}p(\boldsymbol{n}; t) = \sum_{r=1}^{R} p(\boldsymbol{n} - \boldsymbol{\nu}_r; t)w_r(\boldsymbol{n} - \boldsymbol{\nu}_r) - \sum_{r=1}^{R} p(\boldsymbol{n}; t)w_r(\boldsymbol{n}). \tag{1}$$

Here, $p(\boldsymbol{n}; t)$ is the probability for the system to be in state $\boldsymbol{n}$ at time t. As a function of $\boldsymbol{n}$, $p(\boldsymbol{n}; t)$ is the probability mass function (PMF).

In a system with two species, X and Y, the right-hand side (RHS) for equation (1) corresponding to reaction r can be written as

$$p(n_x - \nu_{r,x}, n_y - \nu_{r,y}; t)w_r(n_x - \nu_{r,x}, n_y - \nu_{r,y}) - p(n_x, n_y; t)w_r(n_x, n_y) \tag{2}$$

where $\nu_{r,x}$, $\nu_{r,y}$ are the stoichiometric coefficients for species X, Y in reaction r, respectively, and $\boldsymbol{\nu}_r = (\nu_{r,x}, \nu_{r,y})$.

Figure 1 shows a schematic of a two dimensional CME system. In this figure, colored arrows correspond to reactions, with changes in the number of molecules being at most ± 1 for each species. Reactions leading to larger changes in the number of molecules are possible, however the corresponding sample "jumps" are not shown, to keep the sketch simple. Eq. (1) can be written in matrix form as

$$\frac{\partial p}{\partial t} = A^{(CME)} \cdot p \tag{3}$$

where p is an array of PMF values for all possible system states. For two-species systems, the components of $A^{(\mathrm{CME})}$ are given by

$$A^{(\mathrm{CME})}_{k,k} = -\sum_{r=1}^{R} w_r(n_x, n_y); \quad A^{(\mathrm{CME})}_{k,k+j} = \sum_{r=1}^{R} w_r(n_x - \nu_{r,x}, n_y - \nu_{r,y})|_{\nu_{r,y} \cdot N_x + \nu_{r,x} = j} \tag{4}$$

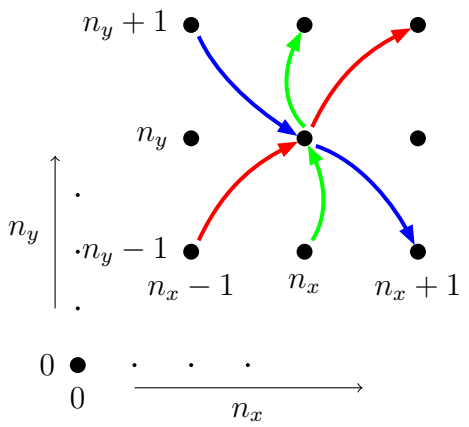

Figure 1: Schematic of the CME domain. The red arrows correspond to a reaction with $\nu = (1, 1)$, blue arrows to $\nu = (1, -1)$, and green arrows to $\nu = (0, 1)$.

Here, $k = n_y \cdot N_x + n_x$ is a linear index spanning the 2-D state space (n_x, n_y) in a row-major format, and N_x is the maximum number of molecules for species X. Similarly, element p_k is the PMF for state (n_x, n_y). The second summation in Eq. (4) is over all reactions for which $j = \nu_{r,y} \cdot N_x + \nu_{r,x}$.

In this section we compare two methodologies to solve the system of coupled ordinary differential equations (ODEs) defined by Eq. (3). In the first approach, we explore the stochastic simulation algorithm [2, 3] (SSA). SSA provides a mechanism for the time evolution of species numbers, effectively sampling the CME solution and allowing its statistical analysis. The SSA model is implemented in a simulation code using the StochKit library [33]. In the second approach, the CME system is integrated in time using a two-stage 2^{nd}-order total variation diminishing Runge-Kutta scheme [34].

Tests were conducted for two kinetic models, both involving two-species

systems. The first system involves five reactions modeling the creation of two metabolites, a joint reaction, and their destruction [35–37]:

$$\begin{aligned} \emptyset \xrightarrow{k_x} X, \quad \emptyset \xrightarrow{k_y} Y \\ X + Y \xrightarrow{k_2} \emptyset \\ X \xrightarrow{\mu} \emptyset, \quad Y \xrightarrow{\mu} \emptyset \end{aligned} \tag{5}$$

The propensities for the five reactions are given by $w = \{k_x, k_y, k_2 n_x n_y, \mu n_x, \mu n_y\}^T$, where n_x, n_y are the number of molecules of species X, Y, respectively. The matrix of stoichiometric coefficients is given by $\{(1,0),(0,1),(-1,-1),(-1,0),(0,-1)\}^T$. For the results presented in this section, we adopted the same values for the model parameters as in [37]: $k_x = k_y = 0.6$, $k_2 = \mu = 0.001$.

Figure 2 shows PMF contour lines at steady state, obtained using the two methodologies mentioned above: the SSA approach and direct integration of the CME. For all simulations, the initial condition was a 2D Gaussian "blob", symmetrical in both directions, centered at $(200, 200)$ and with standard deviations set to $\sigma_{n_x} = \sigma_{n_y} = 25$. The PMF field evolves towards the elongated steady-state solution seen in the figure. The SSA solution consists of the time evolution of an ensemble of system states. The initial condition for this ensemble are sampled from the same Gaussian distribution used as initial condition for the direct solution. The PMF values for the SSA simulations are obtained via Kernel Density Estimation (KDE) [38, 39]. For the metabolic configuration, the SSA approach using 10^7 realizations, shown in Fig. 2b, produces PMF values close to the ones obtained by the direct solution of the CME, shown in Fig. 2c. For this configuration, the SSA simulations required in order to obtain converged PMF values via KDE are about

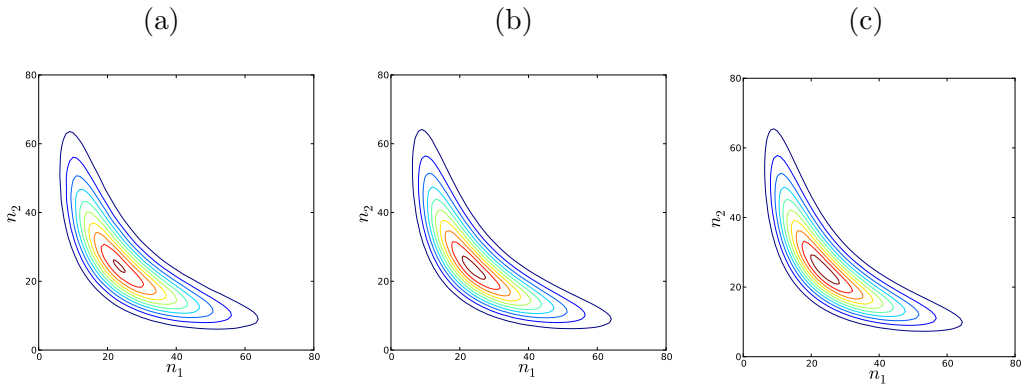

Figure 2: Steady state PMF contours for the metabolic model: (a) and (b) Kernel Density Estimate based on SSA solutions with 10^6 and 10^7 realizations; (c) direct integration of CME given by Eq. (3).

50% more expensive than the CME solution. Next we will compare the two approaches for a 2D system that exhibits more complicated dynamics.

This system models a circadian rhythm using two molecular species, the complex X and the repressor Y, and 4 reactions:

$$\begin{aligned} \emptyset &\xrightarrow{w_1} Y, \quad Y \xrightarrow{w_2} \emptyset \\ Y &\xrightarrow{w_3} X, \quad X \xrightarrow{w_4} Y \end{aligned} \tag{6}$$

The expressions for the propensities for the four reactions are provided below for completeness. Reference [35] provides more detail on this model.

$$
\begin{aligned}
w_1 &= \frac{\beta_R}{\delta_{MR}} \frac{\alpha_R \theta_R + \alpha'_R \gamma_R \tilde{A}(n_y)}{\theta_R + \gamma_R \tilde{A}(n_y)} \\
w_2 &= \delta_R n_y \\
w_3 &= \gamma_C \tilde{A}(n_y) n_y \\
w_4 &= \delta_A n_x
\end{aligned} \tag{7}
$$

where

$$
\tilde{A}(n_y) = \frac{1}{2}\left(\alpha'_A \rho(n_y) - K_d + \sqrt{\left(\alpha'_A \rho(n_y) - K_d\right)^2 + 4\alpha_A \rho(n_y) K_d}\right)
$$

and

$$
\rho(n_y) = \frac{\beta_A}{\delta_{MA}(\gamma_C n_y + \delta_A)}, \;\; K_d = \theta_A / \gamma_A.
$$

The coefficients used in the circadian rhythm models are shown in Table 1.

α_A	50	β_A	50	γ_A	1	δ_A	1	θ_A	50
α_R	0.01	β_R	5	γ_C	2	δ_R	0.2	θ_R	100
α'_A	500			γ_R	1	δ_{MA}	10		
α'_R	50					δ_{MR}	0.5		

Table 1: Coefficients for the circadian rhythm model [35].

Several numerical tests were performed with computational domain sizes starting from 2500^2 and scaled down to 1800×1400. These tests, results not shown, indicate that relevant system dynamics are contained in a rectangle $[0, 1400] \times [0, 1000]$, and that a 1800×1400 computational domain is sufficiently large to minimize the effect of boundaries on the simulation.

Figure 3 shows contour lines corresponding to the natural logarithm of PMF for the circadian rhythm at $t = 200$. This model exhibits limit-cycle behavior, with the peak PMF transitioning from the neighborhood of $(n_x, n_y) \approx (300, 1200)$ to regions with $n_y < 10$, then going through $(n_x, n_y) \approx (1000, 200)$ back to the first region. A logarithm scale is employed for the vertical axis to highlight the system dynamics for small molecule counts of species Y.

The circadian rhythm configuration is particularly difficult to simulate since relevant dynamics are observed for molecule counts spanning three orders of magnitude, and PMF values spanning five orders of magnitude. The SSA approach using 10^8 realizations, in Fig. 3b, while not fully converged, is about 100 times more expensive compared to the direct CME integration, shown in Fig. 3d.

In this section we presented results comparing the direct integration of the CME equation with results obtained using the SSA algorithm for the 2D systems. For these low-dimensional cases, the SSA algorithm, characterized by slow convergence properties of Monte-Carlo sampling methods, is less efficient compared to the direct integration of CME. For high-dimensional systems, however, the curse of dimensionality makes direct integration prohibitively expensive. The SSA approach is more affected by the system dynamics than its dimensionality and remains viable for generating sample trajectories of high-dimensional systems. However, when PMFs of the sampled states are needed, or when regions of low probability are of interest, the prohibitively large number of SSA samples to generate these values for such high dimensional systems presents a problem. Effective, non-sampling-based approaches

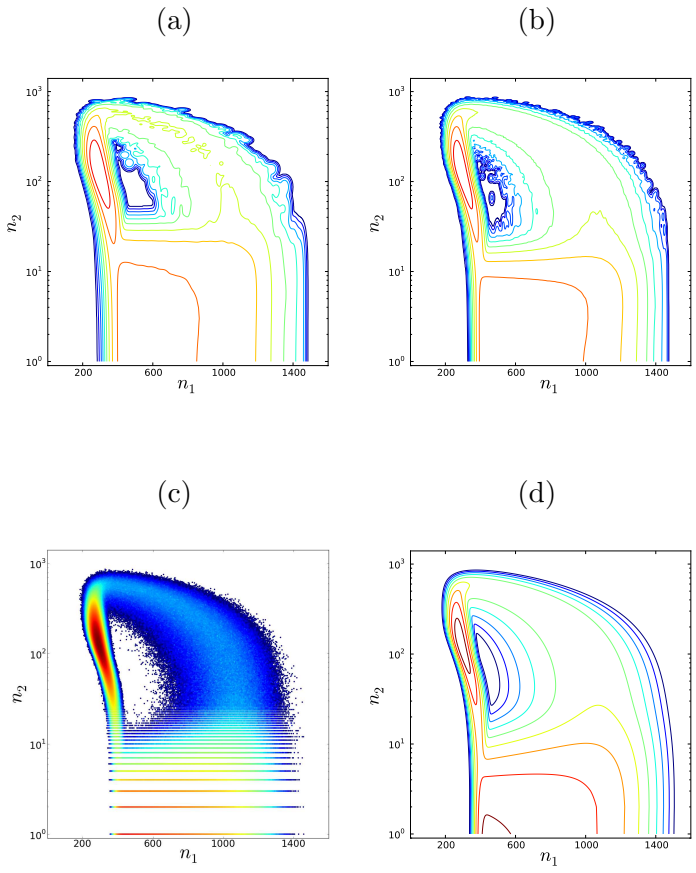

Figure 3: PMF contour plots for the circadian rhythm systems. (a) and (b): Kernel Density Estimate based on SSA solutions with 10^6 and 10^8 realizations; (c) SSA states colored according to the number of particles in each state; (d) direct integration of Eq. (3); The colormaps correspond to the natural logarithm of PMF in (a), (b), and (d) and natural logarithm of number of particles in (c). All frames employ a logarithmic scale for the molecule counts n_y of species Y.

would therefore be welcome in such situations. In the remainder of this paper we explore several approaches derived from the CME to improve the computational efficiency of solving for the state PMF.

3. Fokker-Planck Equation

In the limit of large numbers of molecules, the discrete states representing molecule counts can be replaced by the continuous concentrations $\boldsymbol{x} = \boldsymbol{n}/V$, where V is the system volume. The propensities, the jump vectors and the probability p are also rescaled, respectively,

$$\bar{w}_r(\boldsymbol{x}) = \frac{1}{V} w_r(\boldsymbol{n}), \qquad \boldsymbol{\xi}_r = \frac{\boldsymbol{\nu}_r}{V}, \qquad \bar{p}(\boldsymbol{x}; t) = V p(\boldsymbol{n}; t). \tag{8}$$

For the purposes of this text, the value of V is not relevant since it can be simplified out of the evolution equations. Therefore, we will choose volume units so that $V = 1$. Also, for notational simplicity, we will drop the overbars. In the discrete formulation, p represents the PMF, while in the continuum formulation p is the probability density function (PDF) which is the probability per unit volume. For a unified view, the PMF can be interpreted as a PDF that has been integrated over a hypercube of unit volume.

The Fokker-Planck equation (FPE) corresponding to the SRN introduced above is

$$\frac{\partial p(\boldsymbol{x}; t)}{\partial t} = \sum_{r=1}^{R} \nabla \cdot \vec{F}^{(r)} \tag{9}$$

where

$$\vec{F}^{(r)} = -\boldsymbol{\nu}_r \left(w_r(\boldsymbol{x}) p(\boldsymbol{x}; t) - \frac{1}{2} \boldsymbol{\nu}_r \cdot \nabla \left(w_r(\boldsymbol{x}) p(\boldsymbol{x}; t) \right) \right) \tag{10}$$

The flux $\vec{F}^{(r)}$ consists of advection and diffusion components

$$\vec{F}_a^{(r)} = -\boldsymbol{\nu}_r w_r(\boldsymbol{x}) p(\boldsymbol{x}; t), \quad \vec{F}_d^{(r)} = \frac{1}{2} \boldsymbol{\nu}_r \left(\boldsymbol{\nu}_r \cdot \nabla \left(w_r(\boldsymbol{x}) p(\boldsymbol{x}; t) \right) \right) \tag{11}$$

The FPE can be interpreted as a Taylor series expansion starting from the CME, and retaining the first order, advection, and the 2^{nd} order, diffusion, terms.

A finite volume approach is adopted for the discretization of Eqs. (9) and (10). This approach naturally preserves the total probability inside the computational domain. Further details are presented in the subsections below.

3.1. Finite Volume Discretization

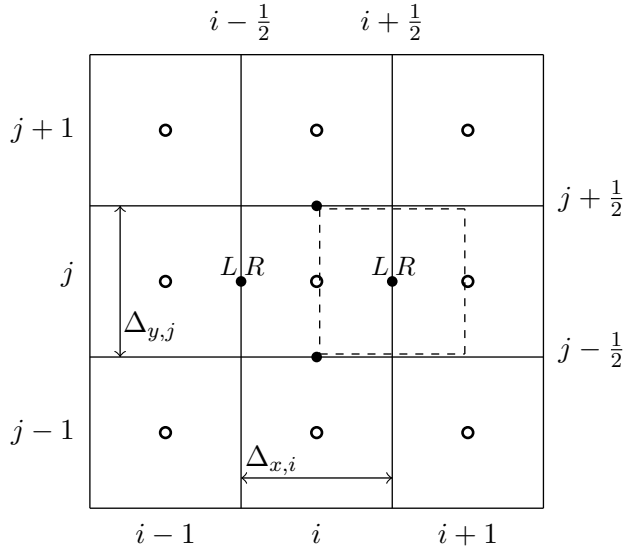

Figure 4: Schematic of the computational mesh for a 2D FPE domain. Cell centers are shown with open symbols while edge centers for cell (i, j) are shown with filled symbols.

Figure 4 shows a schematic of the computational grid in a 2D domain with $\boldsymbol{x} = (x, y)$. The horizontal axis coordinate x corresponds to species X while the vertical coordinate y to species Y. The PDF for the FPE solution

is represented by cell averages

$$p_{i,j} = \frac{1}{A_{\Omega_{ij}}} \int_{\Omega_{ij}} p \, d\omega \tag{12}$$

where $A_{\Omega_{ij}} = \Delta_{x,i}\Delta_{y,j}$. After integration over cell volumes, the FPE can be written in finite volume form as

$$\begin{aligned} \frac{\partial p_{i,j}}{\partial t} &= \frac{1}{A_{\Omega_{i,j}}} \sum_{r=1}^{R} \int_{\Omega_{i,j}} \nabla \cdot \vec{F}^{(r)} \mathrm{d}\omega \\ &= \sum_{r=1}^{R} \left(\frac{\bar{F}^{(r,x)}_{i+\frac{1}{2},j} - \bar{F}^{(r,x)}_{i-\frac{1}{2},j}}{\Delta_{x,i}} + \frac{\bar{F}^{(r,y)}_{i,j+\frac{1}{2}} - \bar{F}^{(r,y)}_{i,j-\frac{1}{2}}}{\Delta_{y,j}} \right) \end{aligned} \tag{13}$$

The fluxes in the *rhs* of Eq. (13) are edge-averaged fluxes along the edges of cell (i, j), *i.e.* for edge $(i + 1/2, j)$

$$\bar{F}^{(r,x)}_{i+\frac{1}{2}} = \frac{1}{\Delta_{y,j}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \left(\vec{F}^{(r)}(x_{i+\frac{1}{2}}, y) \cdot \vec{n} \right) dy$$

where $\vec{n}$ is the unit normal to edge $(i + 1/2, j)$ pointing outside of cell (i, j). The advection and diffusion components are discretized separately with specialized schemes that maintain positive probability values and are 2^{nd}-order accurate. For the advection component we adopt a MUSCL scheme [40] combined with a *local Lax-Friedrichs flux splitting* [41, 42] to approximate numerical fluxes at cell interfaces. This construction can be tailored to avoid Gibbs phenomena leading to negative PDFs near high-gradient regions. For ease of notation we present the formulation in the $x-$direction, dropping subscript j for brevity. The formulation extends naturally direction-by-direction to 2D or higher dimensions.

The numerical advection flux at edge $(i + 1/2)$ is written as

$$F^{(r,x)}_{a,i+\frac{1}{2}} = \frac{1}{2} \left(F^{(r,x)}_{a}(p^{L}_{i+\frac{1}{2}}) + F^{(r,x)}_{a}(p^{R}_{i+\frac{1}{2}}) \right) + a^{(r,x)}_{i+\frac{1}{2}} \left(p^{L}_{i+\frac{1}{2}} - p^{R}_{i+\frac{1}{2}} \right)$$

where $F_a^{(r)}$ is here the $x-$component of the advection flux. The $p^L_{i+1/2}$ and $p^R_{i+1/2}$ are "left" and "right" edge-averaged PDF values at $i+1/2$, extrapolated from cell-averaged values in cells i and $i+1$, respectively. Here $a^{(r,x)}_{i+1/2}$ is the peak value of $|\partial F_a^{(r,x)}/\partial p|$ in the vicinity of edge $(i+1/2)$

$$a^{(r,x)}_{i+1/2} = \max\left(|\nu_{r,x} w_{r,i}|, |\nu_{r,x} w_{r,i+1}|\right) \tag{14}$$

A linear extrapolation inside each finite volume cell is sufficient for an overall 2^{nd}-order discretization

$$p^L_{i+1/2} = p_i + 0.5\delta_x p_i, \quad p^R_{i+1/2} = p_{i+1} - 0.5\delta_x p_{i+1}$$

The construction of $\delta_x p_i$ requires a careful evaluation to avoid introducing new maxima in the solution. In this work we adopt a *minmod* limiter for these slopes. This limiter ensures the scheme is *Total Variation Diminishing* (TVD) [43]

$$\delta_x p_i = \Delta_{x,i} \cdot \text{minmod}\left(\frac{p_i - p_{i-1}}{\Delta_{i-1/2}}, \frac{p_{i+1} - p_i}{\Delta_{i+1/2}}, \frac{p_{i+1} - p_{i-1}}{\Delta_{i-1/2} + \Delta_{i+1/2}}\right)$$

where $\Delta_{x,i}$ is the width of cell i, and $\Delta_{i+1/2} = \frac{1}{2}(\Delta_{x,i} + \Delta_{x,i+1})$. The minmod function is defined as

$$\text{minmod}(a_1, a_2, \ldots) = \begin{cases} \text{sgn}(a_1) \min\limits_i(|a_i|) & \text{if all } a_i\text{'s have the same sign} \\ 0 & \text{otherwise} \end{cases}$$

with "sgn" denoting the sign function.

Numerical discretization of the diffusion term employs auxiliary control volumes and an example is shown with dashed lines in Fig. 4. This auxiliary control volume, Ω, corresponds to edge $(i+1/2, j)$. Further we employ a

similar notation as in [30], $q_r = w_r p$. The volume-averaged diffusion flux at $(i + 1/2, j)$ is computed as

$$\vec{F}^{(r)}_{d,i+1/2,j} = \frac{1}{2}\boldsymbol{\nu}_r \left(\boldsymbol{\nu}_r \cdot \frac{1}{A_\Omega} \int_\Omega \nabla q_r \, d\omega \right) = \frac{1}{2}\boldsymbol{\nu}_r \left(\boldsymbol{\nu}_r \cdot \frac{1}{A_\Omega} \int_{\partial\Omega} q_r \, \vec{n} \, d\sigma \right), \quad (15)$$

and the x-component of this flux is given by

$$\bar{F}^{(r,x)}_{d,i+1/2,j} = \frac{1}{4}\nu^2_{r,x} \left.\frac{\partial q_r}{\partial x}\right|_{i+\frac{1}{2},j} - \frac{1}{2}\nu_{r,x}\nu_{r,y} \left.\frac{\partial q_r}{\partial y}\right|_{i+\frac{1}{2},j} \quad (16)$$

Here $\vec{n}$ is the unit normal to the edges of control volume Ω, and the second integral in Eq. (15) is over all edges of this control volume. For uniform grids, the partial derivatives are given by

$$\left.\frac{\partial q_r}{\partial x}\right|_{i+\frac{1}{2},j} = \frac{q_{r,i+1,j} - q_{r,i,j}}{\Delta_x}, \quad \left.\frac{\partial q_r}{\partial y}\right|_{i+\frac{1}{2},j} = \frac{q_{r,i+\frac{1}{2},j+\frac{1}{2}} - q_{r,i+\frac{1}{2},j-\frac{1}{2}}}{\Delta_y} \quad (17)$$

with

$$q_{r,i+\frac{1}{2},j+\frac{1}{2}} = \frac{1}{4}\left(q_{r,i,j} + q_{r,i,j+1} + q_{r,i+1,j} + q_{r,i+1,j+1}\right)$$

For non-uniform grids, $q_{r,i+\frac{1}{2},j+\frac{1}{2}}$ is evaluated using weighted averages based on the sizes of cells adjacent to vertex $(i + \frac{1}{2}, j + \frac{1}{2})$.

3.2. Comparison of CME and FPE solutions

The FPE results using the discretizations presented in this section are compared to CME solutions to illustrate the range of applicability of the FPE approach. To this end we employ the metabolite system presented in the previous section in Eq. (5). In addition to the "nominal" setup we also consider cases based on modified propensity rates.

The red contour lines in Fig. 5a show the steady state field corresponding to the nominal parameters provided in Section 2. The blue contour lines

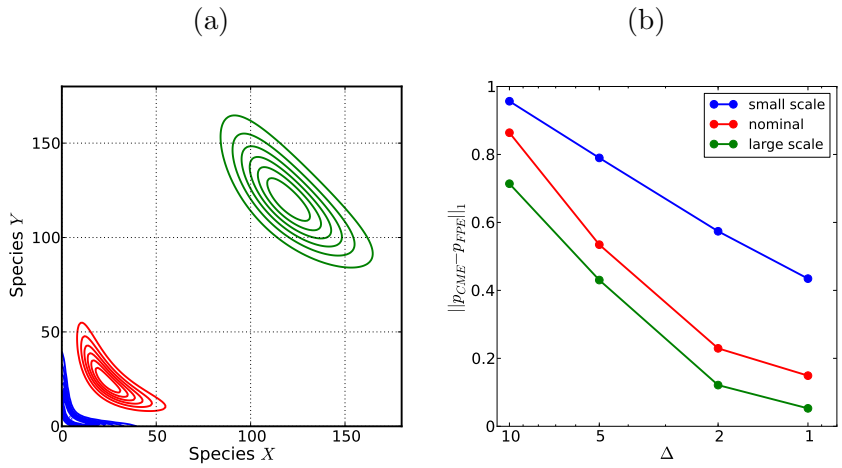

Figure 5: (a) Steady state CME solutions for the metabolite system; (b) L_1 error between CME solutions and FPE solutions with several grid size

show the steady state for a system where k_x and k_y are scaled down by a factor of 5 while k_2 is scaled up by the same factor. The steady state for this case corresponds to small molecule counts of both species. The green contour lines correspond to a system where the steady state occurs when both species are in large molecule counts. For this case k_x and k_y are scaled up by a factor of 5 while k_2 is scaled down by the same factor. The re-scaling of propensity rates is equivalent to re-scaling the volume V in Eq. (8). For large V, equivalent to the solution shown with green contours, the stochastic effects are diminished by averaging over a larger number of molecule counts, while for small V the effect is opposite and stochasic effects become more important.

Fig. 5b shows the L_1 error between FPE solution and the CME solution, for several grid sizes $\Delta_x = \Delta_y = \Delta$. While in principle one could generate

FPE solutions for smaller Δ values compared to the values shown in the figure, in practice the FPE approximation is only useful when the grid size spans more than one molecule count. For the system exhibiting dynamics at small molecule counts, shown in blue, the magnitude of the error between CME and FPE remains high rendering the FPE approach infeasible for this type of systems. At the other end of the spectrum chosen for this set of tests, shown in green, the error drops to small levels for $\Delta = 1 \ldots 2$.

Physical processes often operate at different scales with respect to copy numbers. For some processes the discrete CME approach is necessary to capture the reaction dynamics while for others the continuum FPE formulation is sufficient to resolve all dynamical scales. Results presented in this section suggest that a hybrid approach combining CME where necessary and FPE when the system dynamics occur at larger molecule counts can provide an accurate modeling approach, while improving computational efficiency.

4. Hybrid CME/FPE Formulation

We first introduce a formulation developed by Sjöberg *et al.* [31] that combines both discrete and continuum approximations. Let us assume that species X can be modeled by a continuum approximation while Y requires the discrete approach. We construct the hybrid approach starting from Eq. (2), splitting its *rhs* as

$$\underbrace{p(n_x - \nu_{r,x}, n_y - \nu_{r,y}) w_r(n_x - \nu_{r,x}, n_y - \nu_{r,y}) - p(n_x, n_y - \nu_{r,y}) w_r(n_x, n_y - \nu_{r,y})}_{\textbf{CME-x}}$$
$$+ \underbrace{[p(n_x, n_y - \nu_{r,y}) w_r(n_x, n_y - \nu_{r,y}) - p(n_x, n_y) w_r(i, j)]}_{\textbf{CME-y}}$$

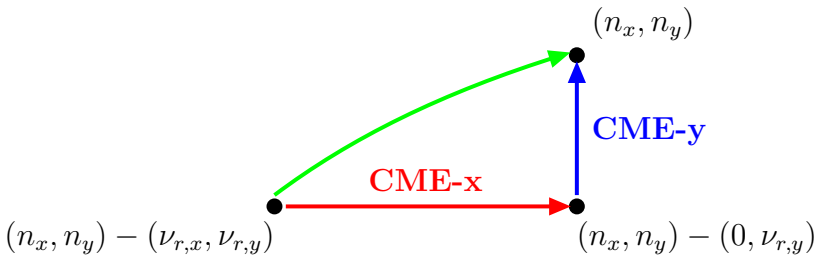

Figure 6: Schematic of CME decomposition into fractional "jumps" for a two-species system. The red arrow shows the fractional jump corresponding to species X and blue arrow corresponds to changes in molecule counts for species Y only.

In the expression above the time dependency is not shown, to simplify the notation. A schematic illustrating this decomposition is shown in Fig. 6. The first two sets of terms correspond to a fictitious system with reactions changing the molecule counts for X, while the other terms correspond to reactions modifying Y only. Based on the hypothesis that X can be modeled by a continuum approximation, we employ the FPE approximation to replace the first two terms in the expression above, while retaining the discrete terms for the dynamics in the Y direction.

$$\begin{aligned} \frac{\partial p}{\partial t} &= \sum_{r=1}^{R} \left. \frac{\partial F^{(r,x)}}{\partial x} \right|_{x,y-\nu_{r,y}} \\ &\quad + \sum_{r=1}^{R} \left[p(\boldsymbol{n} - (0,\nu_{r,y}); t) w_r(\boldsymbol{n} - (0,\nu_{r,y})) - p(\boldsymbol{n}; t) w_r(\boldsymbol{n}) \right] \end{aligned} \tag{18}$$

The first term in the *rhs* of Eq. (18) is discretized using the finite volume formulation for the FPE presented in Section 3.1, while the other terms represent discrete terms corresponding to the CME. This approach can be generalized to any number of species, separating the group of species for which the discrete effects are important from the group for which the continuum

formulation is sufficient.

Figure 7 shows a schematic of a hybrid CME-FPE (hCF) computational domain. In this figure the blue arrow corresponds to reactions with $\nu_{r,y} = 1$, and the green arrow to $\nu_{r,y} = -1$. In the hCF approach, for a two-species system, the PDF is now represented on one-dimensional grids or lines, and p becomes a probability density per unit length along lines of constant y. The 2-dimensional index (i, j) represents cell i of width $\Delta_{x,i}$ for species X and a molecule count j for species Y.

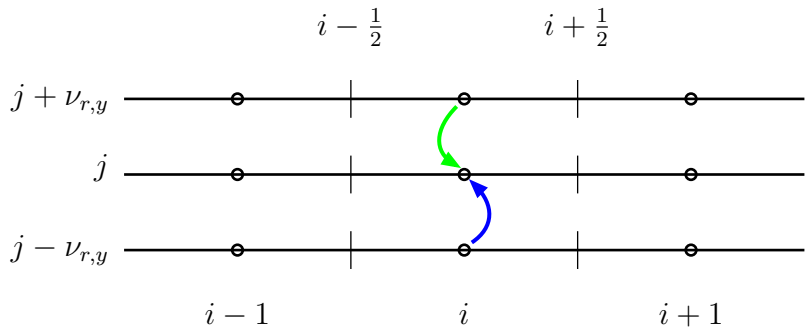

Figure 7: Schematic of the computational mesh for the hybrid CME-FPE domain.

The advection and diffusion components of flux $F^{r,x}$ in Eq. (18) are discretized similar to the expressions presented in Sec. 3.1. However, for this formulation, the flux in the $x-$direction corresponding to cell (i, j) is computed along the line $j - \nu_{r,y}$. Specifically, in the finite volume formulation, this contribution is given by

$$\frac{F^{(r,x)}_{i+\frac{1}{2},j-\nu_{r,y}} - F^{(r,x)}_{i-\frac{1}{2},j-\nu_{r,y}}}{\Delta_{x,i}} \tag{19}$$

where the convection and diffusion flux components on faces $(i \pm \frac{1}{2}, j - \nu_{r,y})$

are computed using the formulations presented in Section 3.1. The *rhs* of Eq. (18) is completed by the CME formulation along $(i, j - \nu_{r,y}) \rightarrow (i, j)$.

Since many models exibit both discrete, usually at small molecule counts, and continuum behaviour, we aim to design a hybrid approach that switches to the appropriate model, discrete CME, continuum FPE, or hCF depending on the system dynamics locally in the state space. Figure 8 shows an illustration of the region where all three types of sub-domains mentioned above join together to form a composite computational domain for a two-species system, X and Y. For small molecule counts, in the lower left corner, stochastic effects are potentially important for all species and we employ CME to model the system dynamics. For regions corresponding to larger molecule counts for one of the species, in sub-domains labeled hCF, we switch to a continuum FPE formulation for that species while still employing a CME formulation for the species with small molecule counts. Finally, at large molecule counts for all species, we switch to a continuum formulation. This approach, while allowing a more efficient allocation of computational resources, requires consistent treatment of probability fields across different formulations. The CME formulation models the transport of probability mass while the FPE approach models the transport of probability density. In the hCF scheme the probability field is also represented by a density, but corresponding to a unit of space of lower dimensionality compared to the full FPE scheme, *e.g.* per unit length rather than per unit volume in the 2D case.

The next sections introduce algorithmic constructions designed to ensure conservation of probability and preservation of order-of-accuracy for these transitions. We will present derivations that pertain to reactions with stoi-

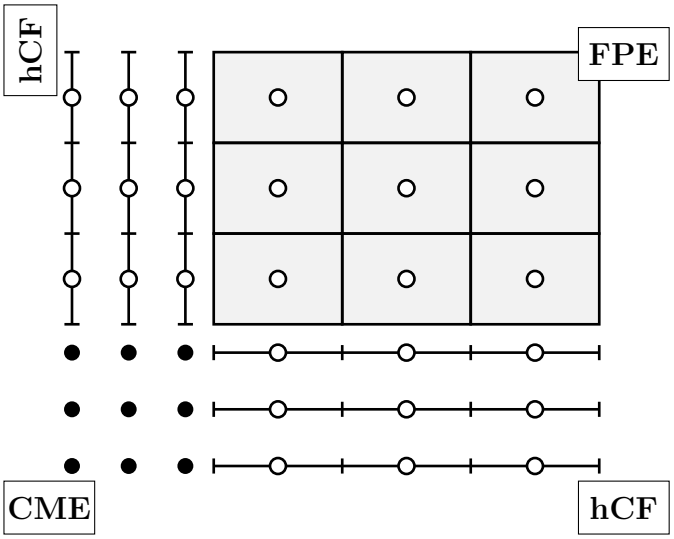

Figure 8: Schematic of a composite computational domain with CME, hCF, and FPE computational subdomains.

chiometric coefficients that are at most ± 1. Derivations for higher reaction orders are conceptually similar, however their practical implementation in a simulation code is more involved and is outside the scope of the current paper.

4.1. Coupling between CME and Hybrid CME-FPE Scheme

Figure 9 shows a schematic of the interface between CME and hCF computational domains. The blue arrow shows the transfer of probability corresponding to a reaction r with stoichiometric coefficients $\nu_{r,x} = 1$, while the green arrow corresponds to $\nu_{r,x} = -1$. For both reactions presented in the figure the value of $\nu_{r,y}$ is the same as $\nu_{r,x}$, however the derivations below are generic and applicable to cases when $\nu_{r,y} \neq \nu_{r,x}$. In this figure the index in the vertical direction is common between the two sub-domains. In the hor-

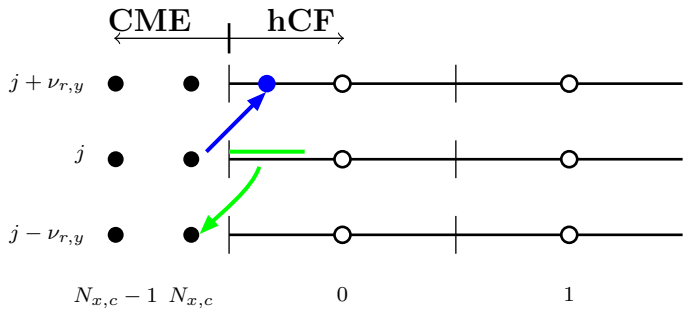

Figure 9: Schematic of the computational mesh at the interface between CME and hCF computational domains.

izontal direction, the CME domain extends up to a molecule count of $N_{x,c}$. The index counter in the horizontal direction is reset in the hCF domain, starting at 0.

In the first case, probability leaves state $(N_{x,c}, j)$ in the CME domain and "jumps" to cell $(0, j+\nu_{r,y})$ in the hCF domain. This transfer can be cast as a boundary condition for the Fokker-Planck flux on line j. The $(i-1/2)$ flux in Eq. (19) corresponding to line cell $i=0$ is given by

$$F^{(r,x)}_{-\frac{1}{2},j+\nu_{r,y}} = w_{r\,N_{x,c},j}\, p_{N_{x,c},j}$$

In the second case, with $\nu_{r,x}=-1$, probability leaves the hCF region towards the CME domain. This corresponds to

$$F^{(r,x)}_{-\frac{1}{2},j} = \int_{-\frac{\Delta_{x,0}}{2}}^{-\frac{\Delta_{x,0}}{2}+1} w_r(x)|_j\; p(x)|_j dx \tag{20}$$

Here, x represents the continuum coordinate along the j line, starting on the left boundary at $x=-\Delta_{x,0}/2$, and $p(x)|_j$ is the probability line density along

this line with cell averages $p_{0,j}$, $p_{1,j}$ and so on. The width of the integration domain is equal to 1 and represents the region from which probability can possibly be transferred to the CME states for a stoichiometric coefficient $\nu_{r,x} = -1$. For larger, in absolute sense, stoichiometric coefficients, the limits of this integral need to be adjusted accordingly.

In general, numerical quadrature can be employed to evaluate the integral in Eq. (20). For computational efficiency we adopt a linearized form

$$F^{(r)}_{-1/2,j} \approx \overline{w_r(x)}|_j \int_{-\frac{\Delta_{x,0}}{2}}^{-\frac{\Delta_{x,0}}{2}+1} p(x)|_j \; dx, \tag{21}$$

where $\overline{w_r(x)}|_j$ is the average propensity along the integration segment. This linearization can be employed for systems with propensity fields that have smooth dependence on the molecule counts. Numerical tests confirm this hypothesis for the two models presented in this paper. Further, in order to evaluate the integral in Eq. (21) we approximate $p(x)|_j$ by a 2^{nd} degree polynomial $P_2(x)$ constrained by the PMF at CME state $(N_{x,c}, j)$ and by the PDF along hCF cells $(0, j)$ and $(1, j)$

$$\begin{cases} \int_{\text{CME}_{N_{x,c}}} P_2(x)dx = p_{N_{x,c},j} \cdot 1 \\ \int_{\text{hCF}_0} P_2(x)dx = p_{0,j} \cdot \Delta_{x,0} \\ \int_{\text{hCF}_1} P_2(x)dx = p_{1,j} \cdot \Delta_{x,1} \end{cases} \tag{22}$$

Using the approximation in Eq. (21) and the solution of the linear system in Eq. (22), the boundary condition in Eq. (20) can be written as

$$F^{(r,x)}_{-1/2,j} = \overline{w_r(x)}|_j \left(c_1 p_{N_{x,c},j} + c_2 p_{0,j} + c_3 p_{1,j}\right) \tag{23}$$

In the computations presented in this paper, the cell sizes were uniform in each computational subdomain, *i.e.* $\Delta_{x,0} = \Delta_{x,1} = \Delta$. For this particular

case, the coefficients above are given by

$$c_1 = \frac{(\Delta-1)(2\Delta-1)}{(\Delta+1)(2\Delta+1)}, c_2 = \frac{7\Delta^2-1}{\Delta(\Delta+1)(2\Delta+1)}, c_3 = -\frac{\Delta-1}{\Delta(2\Delta+1)}$$

Numerical tests, results not shown, indicate that, for the metabolite and circadian rhythm models presented in this paper, a first order approximation for $p(x)|_j$, would have been sufficient to obtain a scheme that is overall 2^{nd}-order accurate. This approximation can be derived by assuming a first order polynomial for $p(x)|_j$ and employing the first two constraints in Eq. (22). The derivation above is presented for completeness to ensure a consistent 2^{nd}-order spatial discretization.

The boundary conditions for the CME sub-domain are set consistent to the boundary conditions derived above for the hybrid CME-FPE sub-domain. For the case with $\nu_{r,x} > 0$ the CME equations remain unchanged on the boundary since information propagates from the CME domain towards the hCF region. For the case with $\nu_{r,x} < 0$ the first term in Eq. (2) is replaced by the flux coming out of the hCF region, computed according to the expression in Eq. (23).

The approach presented in this section and throughout the rest of the paper leads to a smooth transition of information between heterogenous formulations while preserving the total probability mass in the computational domain. The interface conditions derived in this paper correspond to jump sizes normal to the boundary, $|\boldsymbol{\nu}_r| \leq 1$. For models with $|\boldsymbol{\nu}_{r,i}| > 1$, the numerical construction can follow similar arguments without any conceptual difficulties. However the algorithmic implementation for these cases is more involved and outside the scope of the current work.

4.2. Coupling between FPE and Hybrid CME-FPE Formulation

This section extends the above algorithmic construction to the 2D interface presented in Fig. 10. In this figure, the arrows show the transfer of probability from the hCF domain to the FPE domain, corresponding to a reaction with $\boldsymbol{\nu} = (1, 1)$. However, the numerical conditions derived below correspond to reactions with both $|\nu_{r,x}| \leq 1$ and $|\nu_{r,y}| \leq 1$.

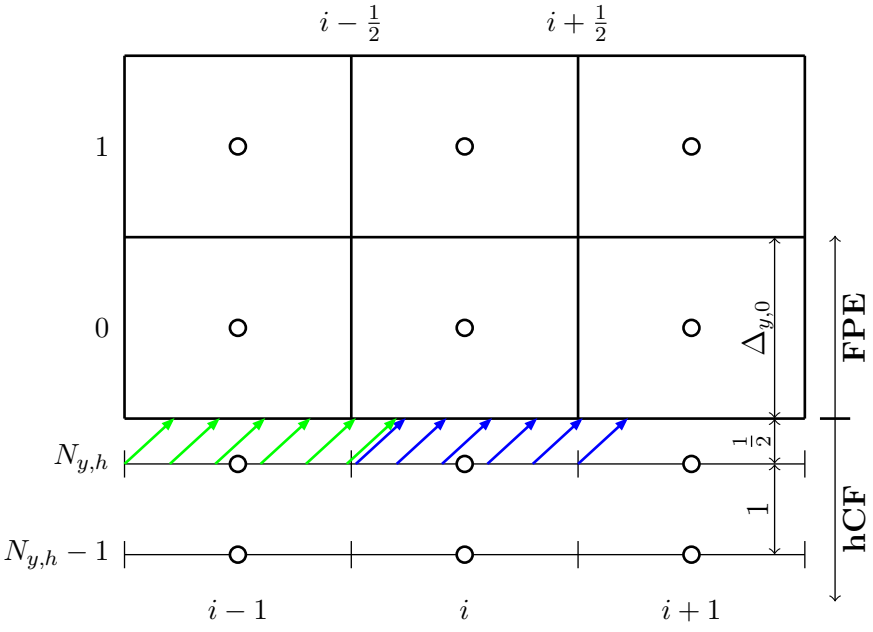

Figure 10: Schematic of the computational mesh at the interface between hybrid CME-FPE domain and the FPE domain. Lines $N_{y,h} - 1$ and $N_{y,h}$ are in the hybrid domain, while cells 0 and 1 are in the FPE domain.

We first derive the boundary conditions for the FPE domain. For the case shown in Fig. 10, with the probability flux going into the FPE domain, we use upwind considerations to set the boundary condition for the flux entering the FPE domain equal to the probability amount leaving the hCF domain.

Following the stoichiometry in the above schematic, the FPE flux on edge $(i, -1/2)$ for this case is given by

$$F^{(r,y)}_{i,-\frac{1}{2}} = \int_{x_i-\frac{\Delta_x+1}{2}}^{x_i+\frac{\Delta_x-1}{2}} w_r(x)|_{N_{y,h}}\, p(x)|_{N_{y,h}}\; dx \tag{24}$$

where $w_r(x)|_{N_{y,h}}$ and $p(x)|_{N_{y,h}}$ are the propensity and PDF along the top line in the hCF domain. The range for the integral above is determined by the jump values. For the case shown here, with $\boldsymbol{\nu} = (1, 1)$, the flux on the lower edge of FPE cell $(0, i)$ originates on segment $\left(x_i - \frac{\Delta_x}{2} - \frac{1}{2} : x_i + \frac{\Delta_x}{2} - \frac{1}{2}\right)$ on the last line in the hCF region. Similar arguments can be made for systems with jump sizes greater than one. To simplify the notation, we will drop subscript $N_{y,h}$ for the remainder of this section.

Similar to Section 4.1, we linearize the above expression and construct a 2^{nd}-order polynomial $P_2(x)$ to approximate $p(x)$. The coefficients of this polynomial are set so that line averages along adjacent cells $(i - 1 : i + 1)$ match the numerical solution

$$\begin{cases} \int_{hCF_{i-1}} P_2(x)dx = p_{i-1}\Delta_x \\ \int_{hCF_i} P_2(x)dx = p_i\Delta_x \\ \int_{hCF_{i+1}} P_2(x)dx = p_{i+1}\Delta_x \end{cases} \tag{25}$$

The above expressions assume a uniform grid in the $x-$direction. Given these constraints, the numerical flux in Eq. (24) is computed as

$$F^{(r,y)}_{i,-\frac{1}{2}} = \overline{w_r(x)} \left(\Delta_x p_i + \frac{(2\Delta_x + 1)p_{i-1} - 2p_i + (1 - 2\Delta_x)p_{i+1}}{8\Delta_x} \right) \tag{26}$$

Here, $\overline{w_r(x)}$ is the average propensity along segment $\left(x_i - \frac{\Delta_x+1}{2} : x_i + \frac{\Delta_x-1}{2}\right)$.

The numerical flux construction for $\boldsymbol{\nu} = (-1, 1)$ is similar to the above workflow, with averages now computed for segment $\left(x_i - \frac{\Delta_x - 1}{2} : x_i + \frac{\Delta_x + 1}{2}\right)$. For the case with $\boldsymbol{\nu} = (0, 1)$, the vertical flux is simply given by $w_{r,i} p_i \Delta_x$.

The expressions derived above correspond to $\nu_{r,y} > 0$. For these cases, no boundary conditions are necessary for the hCF domain. For reactions with $\nu_{r,y} = -1$, the probability flows from the FPE domain to the hCF domain. For this case edge-averaged fluxes on the edge of the domain, $F^{(r,y)}_{i,-\frac{1}{2}}$ are evaluated using the 2$^{\text{nd}}$-order formulations for the advection and diffusion components presented in Section 3.1. Continuum approximations for these fluxes are constructed as described above in Eqs. (25) and (26). Their contributions to hCF domain are then partitioned according to the value of $\nu_{r,x}$, i.e. acounting for the $\pm 1/2$ segments from adjacent cells as described above in this section.

4.3. Coupling between CME, FPE and Hybrid CME-FPE Formulations

In the previous sections we presented algorithms for the transition of probability between two formulations at a time: discrete CME and discrete-continuum hCF, or between hCF and the continuum FPE. In this section we present the numerical construction of interface conditions in the "corner" region, at the confluence between discrete, continuum, and hybrid discrete-continuum regions. Figure 11a shows a schematic corresponding to a reaction with $\boldsymbol{\nu} = (1, 1)$, while the illustration in Fig. 11b corresponds to a reaction with $\boldsymbol{\nu} = (1, -1)$. In this figure, the CME state (N_x, N_y) near the corner is shown with a filled circle while the FPE cell $(0, 0)$ is shown with light grey shading. The hCF lines $(0, N_y)$ and $(N_x, 0)$ are surrounded by dashed contours corresponding to "virtual" cells around the hCF grids.

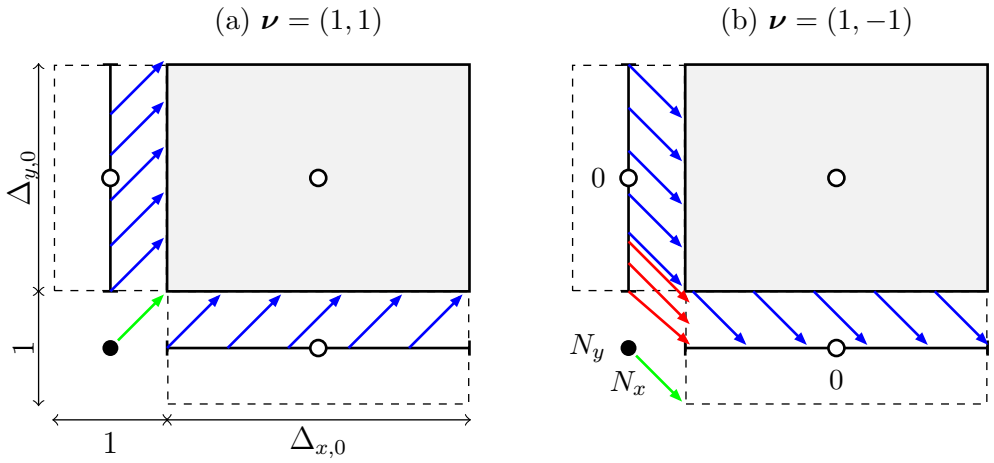

Figure 11: Schematic of the transfer of probability near the corner between CME, FPE, and hCF. Grid sizes near the corner are shown in frame (a) while grid indices are shown in frame (b).

For the reaction illustrated in Fig. 11a, the probability is transferred from the CME and hCF sub-domains towards the FPE sub-domain. We employ the same techniques as in Sections 4.1 and 4.2 to construct the interface fluxes for the FPE sub-domain. The flux on the bottom edge of the FPE cell shown in Fig. 11a is written as

$$F^{(r,y)}_{0,-\frac{1}{2}} = \frac{1}{2} w_{r,N_x,N_y}\, p_{N_x,N_y} + \int_{x_0-\frac{\Delta_x}{2}}^{x_0+\frac{\Delta_x-1}{2}} w_r(x)|_{N_y}\; p(x)|_{N_y}\, dx \tag{27}$$

The computation of integral in the *rhs* of Eq. (27) follows the steps outlined in Eqs. (20)-(23). Numerically the flux normal to the bottom edge of the FPE cell is given by

$$F^{(r,y)}_{0,-\frac{1}{2}} = \frac{1}{2} w_{r,N_x,N_y}\, p_{N_x,N_y} + \overline{w_r(x)}|_{N_y} \left(c_1 p_{N_x,N_y} + c_2 p_{0,N_y} + c_3 p_{1,N_y}\right) \tag{28}$$

and the coefficients above are given by

$$c_1 = \frac{(2\Delta_x - 1)}{8\,(\Delta_x + 1)},\; c_2 = \frac{(2\Delta_x - 1)\,(8\Delta_x^2 + 7\Delta_x + 1)}{16\Delta_x\,(\Delta_x + 1)},\; c_3 = -\frac{2\Delta_x - 1}{16\Delta_x}$$

The computation of the flux normal to the right edge of the FPE cell, $F^{(r,x)}_{-\frac{1}{2},0}$ employs a similar approach as above. For this case, with $\boldsymbol{\nu} = (1, 1)$, no boundary conditions are necessary for the CME and hCF sub-domains.

The construction of interface conditions for a reversed reaction, with $\boldsymbol{\nu} = (-1, -1)$ (schematic not shown) follows the same approach, now applied to 2^{nd}-order reconstructions for the edge-averaged fluxes normal to the FPE interfaces. The limits for the integrals along the FPE edges are set accordingly to the stoichiometry of each reaction, as explained in Section 4.1. Specifically, the probability flux exiting the FPE sub-domain and entering the lower hCF subdomain is given by.

$$F^{(r)}_{0,N_y} = \int_{x_0-\frac{\Delta_x-1}{2}}^{x_0+\frac{\Delta_x+1}{2}} F(x)^{r,y}|_{-1/2}\, dx \tag{29}$$

where $F(x)^{r,y}|_{-1/2}$ is the y-component of the combined FPE advection and diffusion fluxes along the lower edge of the bottom FPE cells. A 2^{nd}-order reconstruction for the FPE fluxes on the boundary is employed. This reconstruction is similar to the steps presented above. The transfer between FPE and the left hCF cell is computed with the same reconstruction approach along the left edges of the FPE cells.

The remaining segments on the lower and left edges of the corner FPE cells account for the transfer of probability between FPE and CME regions:

$$\int_{x_0-\frac{\Delta_x}{2}}^{x_0-\frac{\Delta_x-1}{2}} F(x)^{(r,y)}|_{-1/2}\; dx + \int_{y_0-\frac{\Delta_y}{2}}^{y_0-\frac{\Delta_y-1}{2}} F(y)^{(r,x)}|_{-1/2}\; dy \tag{30}$$

For the reaction illustrated in Fig. 11b, the red arrows illustrate the transfer of probability between the two hCF regions, in addition to the transfers between hCF and FPE sub-domains, shown with blue lines. For this case,

with $\boldsymbol{\nu} = (1, -1)$, the transfer from the vertical hCF line cell and the horizontal hCF cell is computed numerically as

$$\overline{w_r(y)} \int_{y_0 - \frac{\Delta_y}{2}}^{y_0 - \frac{\Delta_y - 1}{2}} p(y)|_{N_x} \, dy \tag{31}$$

Here, $\overline{w_r(y)}$ is the average propensity along segment $\left(y_0 - \frac{\Delta_y}{2} : y_0 - \frac{\Delta_y - 1}{2}\right)$. The integral in Eq. (31) is evaluated numerically by replacing $p(y)$ with a 2^{nd}-order polynomial reconstruction near the interface between the CME sub-domain and the vertical hCF sub-domain.

5. Numerical Experiments

The algorithms presented in Section 4 were numerically implemented in a software framework using mixed *python/C++* libraries. The data arrays and kinetic model and numerical experiment setup are implemented through *python* scripts. The computationally intensive flux evaluations are handled in *C++*. The data communication between the two languages is handled with *swig* tools [44] using array templates defined in *numpy.i* [45]. For all simulations presented below, the numerical time integration is based on a two-stage 2^{nd}-order total variation diminishing (*TVD*) Runge-Kutta scheme [34].

5.1. Convergence Tests

We performed several tests using synthetic one-reaction models to validate the numerical implementations presented above and to determine the overall order-of-accuracy of the numerical construction. Fig. 12 shows a schematic of a composite computational domain. In order to test the order of accuracy of the combined scheme we employ a 2D Gaussian "blob"

with $\sigma_x = \sigma_y = 50$. This initial condition is placed at several locations in the composite domain and evolved in time using select one-reaction models with $\boldsymbol{\nu} = (\pm 1, \pm 1)$. The arrows in Fig. 12 shows sample trajectories for the PMF/PDF fields. For example, line #4 corresponds to runs where the probability moves between the two hCF domains by crossing the FPE domain.

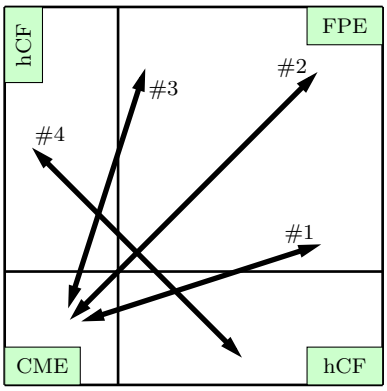

Figure 12: Schematic of the composite computational domain. The arrows show trajectories for sets of tests to check the order of accuracy of the CME-hCF-FPE formulations.

Figure 13 shows sample contour plots for run #2, using $\boldsymbol{\nu} = (1, 1)$. The black arrow in the figure shows the direction of propagation of the PMF, *i.e.* it propagates from the CME domain through the corner between all regions to the FPE domain. The first two frames correspond to the time when the peak PMF/PDF values are nearly centered at the corner. The detail view in Fig. 13b shows the dependency with the grid size once the solution enters domains emloying the FPE approximation to the CME. Similarly at later

times, shown in Figs. 13c and 13d, the FPE solution corresponding to $\Delta = 2$ (green contours) is closer to the full CME solution (red contours) compared to the coarse FPE solution (blue contours).

We verify the overall order of accuracy for the spatial discretization by comparing the solutions on successively refined grids. For the set of runs shown schematically in Fig. 12, we employ composite simulations with FPE grids $\Delta = \{2, 4, 8\}$. The L_1 error between composite simulation results and the CME solution reference is computed as

$$L_{1_\Delta} = \int_\Omega \left| p^{(\Delta)} - p^{CME} \right| d\omega \tag{32}$$

In order to compute the integral above, the CME results are interpolated to a coarser grid Δ using bicubic spline interpolations. We assume a power-law dependency of the L_1 error on the grid size, $L_{1_\Delta} \propto \Delta^\gamma$, with the exponent γ indicating the order-accuracy of the discretization. The order-of-accuracy can then be estimated as

$$\gamma = \log_2 \left(L_{1_\Delta} / L_{1_{\Delta/2}} \right) \tag{33}$$

For the set of simulations employed for testing the discretization order we found γ values in the range $1.82 \ldots 2.01$, confirming the theoretical 2^{nd}-order spatial accuracy of the numerical construction. Fig. 14 shows the error dependency on the grid size for the set of tests outlined in Fig. 12. Results for Set # 3 are not shown since the results nearly overlap with Set # 1. The theoretical 2^{nd}-order slope is shown with black line in Fig 14 for reference. Numerical simulations, results not shown, also confirm the 2^{nd}-order accuracy for the numerical implementation of the time integration scheme.

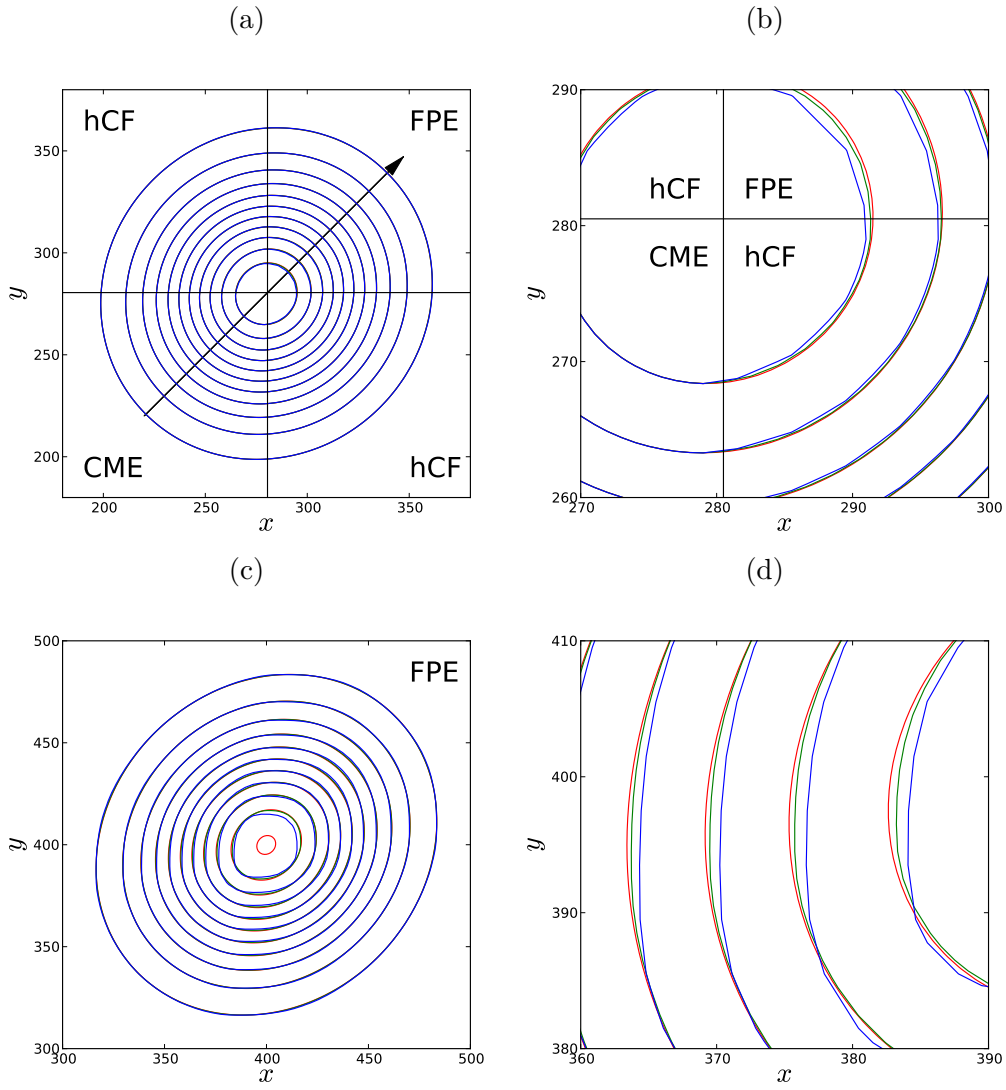

Figure 13: Contour plots of PMF/PDF fields corresponding to run #2 (see Fig. 12). Red contours correspond to full CME simulations, green contours to hybrid simulations with $\Delta = 2$, and blue contours to $\Delta = 4$. Frame (a) shows the solution as it passes through the corner between sub-domains and frame (b) shows a detail near the corner. Frame (c) shows the solution at a later time, fully contained in the FPE domain, with an enlarged detail shown in frame (d).

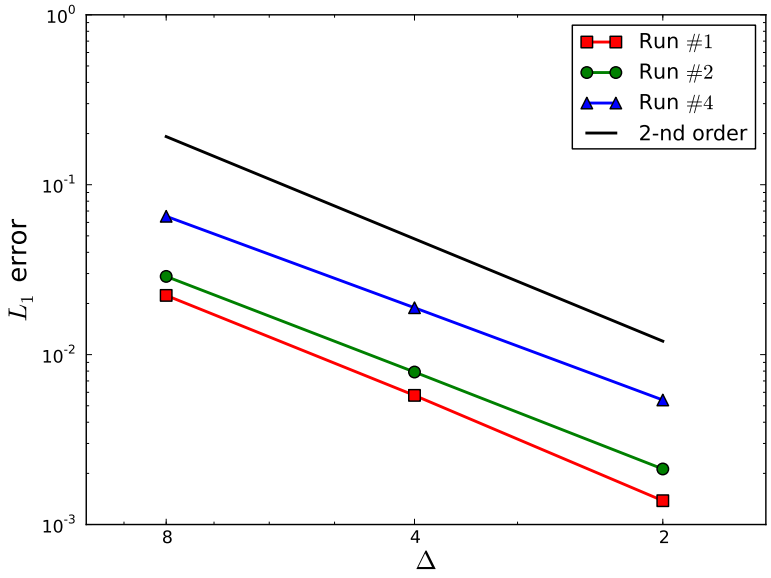

Figure 14: Spatial convergence test for the composite CME-FPE model. The test setup corresponds to paths shown in Fig. 12.

5.2. Circadian Rhythm Simulations

Next we consider composite runs for the circadian rhythm model, introduced in Sec. 2. The parameters for these simulations are provided in Table 2. Runs R_1 though R_3 use the same grid size for the FPE cells, $\Delta = 4$. These runs use progressively smaller CME regions, details shown on the second line in the table. Refer back to the sketch in Fig. 12 for the sub-domain placement in the global, composite computational domain. The extent of subdomains for run R_4 are similar to R_2. R_4 employs a smaller cell size, $\Delta = 2$ to explore the effect of grid refinement on the dynamics of this system.

Figure 15 shows the time evolution of the peak ot the PMF. For the circadian rhythm, the location of peak PMF (for CME) or PDF (for FPE) exhibits

Run	1	2	3	4
CME size	400×400	400×100	100×100	400×100
Δ	4	4	4	2

Table 2: Parameters for composite CME-hCF-FPE runs. For all runs the computational domain spans $[0, 1800] \times [0, 1400]$, and the FPE grids have the same grid spacing in each direction, $\Delta_x = \Delta_y = \Delta$.

a limit-cycle pattern, from a region where species Y has low molecule counts, to regions where both species have molecure counts in the hundreds. The results for runs R_1 through R_4 are compared to the full CME solution shown with black line. Since the initial conditions place the bulk of probability in the FPE sub-domain, at early times the peak PMF decreases compared to the CME solution due to the increased diffusivity given by larger grids. At later times the results for run R_1, which employs CME formulations for a region 400 wide at small molecule counts, are in very good agreement with the full CME solution. At the other end of the spectrum, it is clear that a setup with CME regions 100 wide at small molecule counts, combined with a grid size of 4 in the FPE regions, for R_3 is not sufficient to accurately capture the dynamics of the circadian rhythm model.

Figures 16 and 17 show contour lines of PMFs values at $t = 17$ and $t = 41$ for the composite runs listed in Table 2. These figures also show results for full CME and FPE simulations, respectively, for comparison purposes. The times for these snapshots were selected based on simulation dynamics. At $t = 17$ large PMF values can be simultaneously observed at small and large counts for species Y, while at $t = 41$ the PMF peaks at both small and

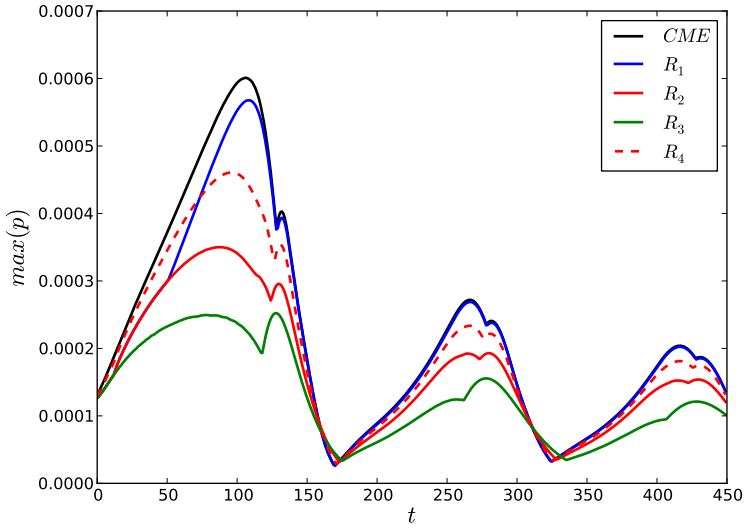

Figure 15: Time evolution of the peak PMF for composite simulations compared to the values from full CME simulation of the circadian rhythm. The PDFs from composite simulations are converted to PMFs for the purpose of this comparison.

intermediate counts of molecules for the same species. The contour values in Figs. 16 and 17 are computed on a logarithmic scale with red contours corresponding to large PMF values, of $o(10^{-4})$, and blue ones to small PMF values, of $o(10^{-12})$. The FPE simulation uses a grid size of 4. While the FPE simulation is about 10 times less expensive compared to the CME approach, the quality of the FPE results is very poor as this approach fails to capture the model dynamics at small molecule counts.

The composite approach results, shown on the second and third rows of these figures provide visually a very good agreement to the full CME solution. At early times, $t = 17$, all four composite runs capture the dynamics of low and high PMF values well. At later times, run R_3, which employs the

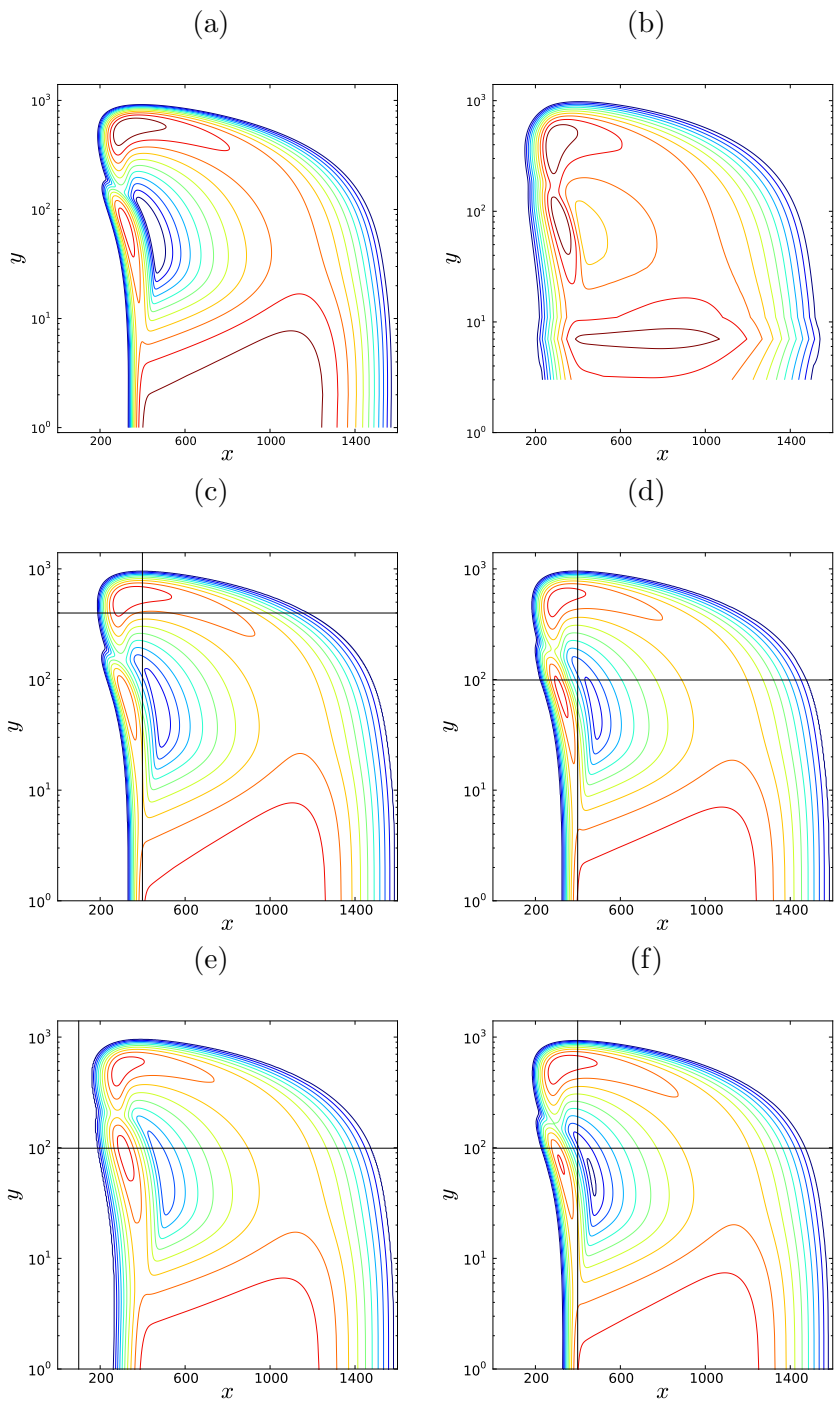

Figure 16: Comparison of (a) full CME and (b) FPE solutions, respectively, with composite simulations (c-f) for the circadian rhythm at $t = 17$. The initial conditions for these simulations are the same as for results presented in Fig. 3.

thinnest CME region at small molecule counts, exhibits a slightly diffused solution while runs R_1, R_2, and R_4 are still in a good visual agreement with the CME solution.

In order to quantify the error introduced by the mixture of CME and FPE formulations for the circadian rhythm model, we examine the L_1 discrepancies between the composite simulations results and the full CME approach. These results are shown in Fig. 18. The horizontal axis in this figure corresponds to the ratio between the average CPU time per time step for composite simulation and the corresponding CPU time for the CME simulation. The tags near select markers in the figure indicate the location of the CME/FPE interface for that data entry, *i.e.* "1" corresponds to a composite simulation with transition between CME and FPE formulations imposed at 100 molecule counts for both species, "3" to 300 molecule counts and "5" to 500 molecule counts. The transition for the remaining data sets are intuitive. For all grid sizes there is significant drop in the error compared to full CME results when the interface is shifted from 300 to 400 molecule counts. This suggests that, for the circadian rhythm model, the discrete dynamics are important for molecule counts less than 400. For simulations using grid sizes equal to 4 and 8, respectively, moving the interface towards larger molecule counts also leads to more expensive simulations. However, these simulations remain computationally efficient, and are about one order of magnitude less expensive compared to the CME simulations. It is interesting to note the evolution of the relative CPU time for the simulations using a grid size of 2. For these simulations, the CPU times are not necessarily correlated with the number of grid points. While further analysis of this behaviour is outside the

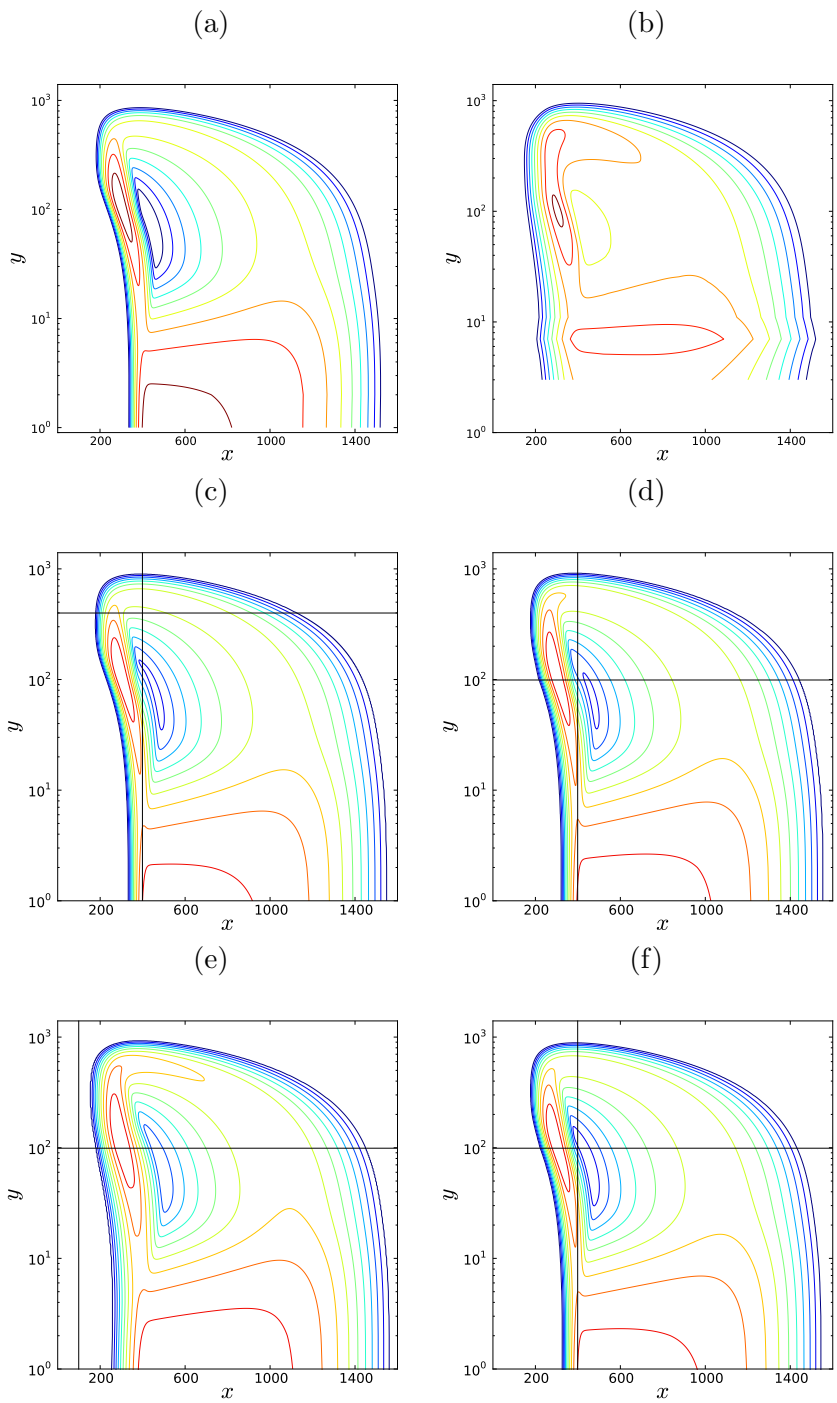

Figure 17: Comparison of (a) full CME and (b) FPE solutions, respectively, with composite simulations (c-f) for the circadian rhythm at $t = 41$. The initial conditions for these simulations are the same as for results presented in Fig. 3.

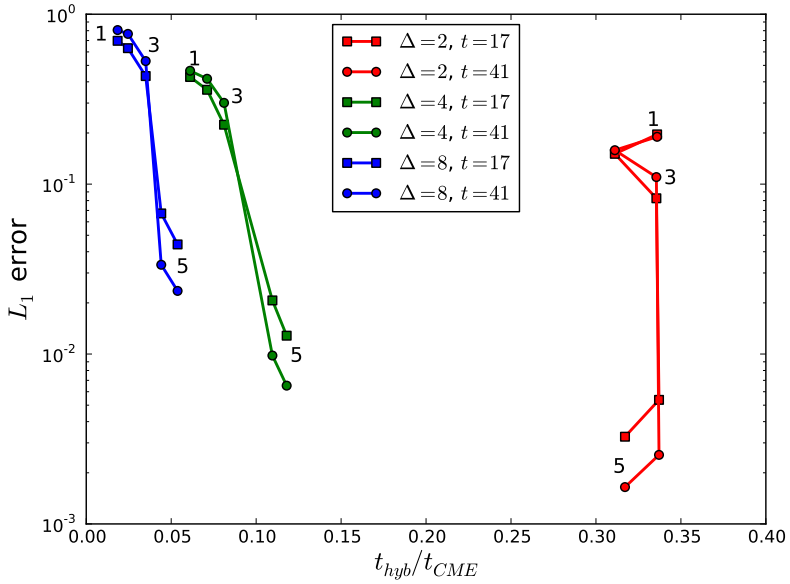

Figure 18: Comparison of L_1 error between PMF values obtained with CME and composite simulations with several grid sizes, at $t = 17$ and $t = 41$.

scope of this paper, we suspect that this is due to a less efficient memory and cache usage compared to the previous two simulations, which use a smaller grid.

For the circadian rhythm tests presented in this section the FPE only simulations are by design cheaper computationally compared to the already efficient composite simulations. However, the FPE solution fails to capture the key system features. The simulations employing the composite CME-hCF-FPE formulation are in very good agreement compared to the CME solution. The computational expense of the composite approach is about 10 times cheaper compared to the CME solution for an error of approximately 1%. This makes the composite approach promising for extensions to higher dimensional systems.

6. Conclusions

In this paper we introduce a hybrid modeling approach to study the behavior of Stochatic Reaction Networks. This approach employs a combination of the Chemical Master Equation (CME), which describes the evolution of the probability of the discrete system to be in a particular state, and the Fokker-Planck Equation (FPE) which is a continuum approximation to the CME in the limit of large molecule count.

In the hybrid formulation, the CME is used to advance the system state for species present in small molecule counts and for which discreteness is important. For species present in large molecule counts the FPE is used instead. The numerical construction for the transfer of probability between CME and FPE regions is designed according to the stoichiometry of each

reaction. The FPE is discretized using a finite volume formulation. The advection component of the FPE fluxes is discretized using a 2^{nd}-order MUSCL scheme with Lax-Friedrichs flux splitting and employs a minmod limiter to keep the probabilities positive even in regions with high gradients. The diffusion component is discretized with a 2^{nd}-order centered scheme. The overall order of the composite numerical scheme is verified to be approximately 2 by comparing sample solutions on successively refined grids.

The efficiency of the hybrid CME/FPE approach is tested on a circadian rhythm model consisting of two species and 4 reactions. The full CME and the hybrid solutions are found to be in very good agreement. For this model the hybrid approach is one order of magnitude cheaper compared to a full-domain CME computation. For comparison, obtaining the system state PMF with SSA sampling is about two orders of magnitude more expensive compared to the CME solution for this system. Also, an approximation with the FPE over the full domain, while computationally more efficient than the hybrid methodology, is not sufficiently accurate.

A key consideration regarding the performance of any approach for solving the CME is how the computational cost scales with the number of species. As discussed in the introduction, the computational cost of the full CME solution increases exponentially with the number of species. Sampling approaches like SSA are less dependent on the dimensionality of the system for generating individual trajectories in time of the system state. However, if PMF values of the system state are needed, the number of SSA samples required also increases prohibitively with the system dimensionality. The results presented in this paper suggest that a higher-dimensional implementation of the

current hybrid scheme can help alleviate this curse of dimensionality by reducing the number of grid points needed in each species dimension. Also, the hybrid approach in this paper can be combined with many other approaches for speeding up the CME simulation, such as finite state projection, adaptive mesh refinement in the FPE regime, stiffness reduction, or improved matrix exponentiation. Also, the hybrid scheme may be extended to treat some species with deterministic rate equations for further efficiency gains. Combined, all of these approaches can provide a direct way to evaluate of the PMF values for realistic SRNs with moderate dimensionality.

References

[1] D. Gillespie, A rigorous derivation of the chemical master equation, Phys. A 188 (1992) 404–425.

[2] D. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. 22 (1976) 403–434.

[3] D. Gillespie, Exact Stochastic Simulation of Coupled Chemical Reactions, Journal of Physical Chemistry 81 (1977) 2340–2361.

[4] M. Gibson, J. Bruck, Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels, J. Phys. Chem. A 104 (2000) 1876–1889.

[5] M. Rathinam, L. R. Petzold, Y. Cao, D. T. Gillespie, Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method., Journal of Chemical Physics 119 (2003) 12784 – 94.

[6] D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems., Journal of Chemical Physics 115 (2001) 1716 – 33.

[7] E. L. Haseltine, J. B. Rawlings, Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics., Journal of Chemical Physics 117 (2002) 6959 – 6969.

[8] Y. Cao, L. R. Petzold, M. Rathinam, D. T. Gillespie, The numerical stability of leaping methods for stochastic simulation of chemically reacting systems, Journal of Chemical Physics 121 (2004) 12169 – 12178.

[9] Y. Cao, D. Gillespie, L. Petzold, The slow-scale stochastic simulation algorithm, The Journal of Chemical Physics 122 (2005) 014116.

[10] Y. Cao, D. Gillespie, L. Petzold, Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems, Journal of Computational Physics 206 (2005) 395–411.

[11] W. E., D. Liu, E. Vanden-Eijnden, Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales, Journal of Computational Physics 221 (2007) 158–180.

[12] A. Sandu, A new look at the chemical master equation, Numerical Algorithms (2013).

[13] B. Munsky, M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, Journal of Chemical Physics 124 (2006) 44104–1–13.

[14] B. Munsky, M. Khammash, A multiple time interval finite state projection algorithm for the solution to the chemical master equation, Journal of Computational Physics 226 (2007) 818–835.

[15] B. Munsky, M. Khammash, The finite state projection approach for the analysis of stochastic noise in gene networks, IEEE Transactions on Automatic Control 53 (2008) 201–14.

[16] V. Sunkara, M. Hegland, An optimal Finite State Projection Method, Procedia Computer Science 1 (2010) 1579–1586.

[17] M. Hegland, C. Burden, L. Santoso, S. Macnamara, H. Booth, A solver for the stochastic master equation applied to gene regulatory networks, J. of Comp. and Appl. Math. 205 (2007) 708–724.

[18] S. MacNamara, A. Bersani, K. Burrage, R. Sidje, Stochastic chemical kinetics and the total quasi-steady-state assumption: application to the stochastic simulation algorithm and chemical master equation, Journal of Chemical Physics 129 (2008) 95–105.

[19] S. Macnamara, K. Burrage, R. Sidje, Multiscale modeling of chemical kinetics via the master equation, Multiscale Modeling and Simulation 6 (2008) 1146–1168.

[20] R. Sidje, K. Burrage, S. MacNamara, Inexact uniformization method for computing transient distributions of markov chains, SIAM Journal on Scientific Computing 29 (2007) 2562–80.

[21] J. Zhang, L. T. Watson, Y. Cao, A modified uniformization method for

the solution of the chemical master equation, Computers & Mathematics with Applications 59 (2010) 573–584.

[22] S. Peles, B. Munsky, M. Khammash, Reduction and solution of the chemical master equation using time scale separation and finite state projection, Journal of Chemical Physics 125 (2006) 204104–1–13.

[23] P. Deuflhard, W. Huisinga, T. Jahnke, M. Wulkow, Adaptive discrete Galerkin methods applied to the chemical master equation, SIAM J. Sci. Comp. 30 (2008) 2990–3011.

[24] S. Engblom, A discrete spectral method for the chemical master equation, Technical Report 36, Uppsala University, 2006.

[25] S. Engblom, Spectral approximation of solutions to the chemical master equation, Journal of Computational and Applied Mathematics 229 (2009) 208–221.

[26] J. Zhang, L. T. Watson, C. A. Beattie, Y. Cao, Radial Basis Function Collocation for the Chemical Master Equation, International Journal of Computational Methods 07 (2010) 477–498.

[27] T. Jahnke, T. Udrescu, Solving chemical master equations by adaptive wavelet compression, Journal of Computational Physics 229 (2010) 5724–5741.

[28] R. Khanin, D. J. Higham, Chemical Master Equation and Langevin regimes for a gene transcription model, Theoretical Computer Science 408 (2008) 31–40.

[29] C. W. Gardiner, Handbook of stochastic methods for physics, chemistry and the natural sciences, volume 13 of *Springer Series in Synergetics*, third ed., Springer-Verlag, Berlin, 2004.

[30] P. Sjöberg, P. Lötstedt, J. Elf, Fokker-Planck approximation of the master equation in molecular biology, Computing and Visualization in Science 12 (2009) 37–50.

[31] P. Sjöberg, PDE and Monte Carlo approaches to solving the master equation applied to gene regulation, Technical Report 2007-028, Uppsala University, http://www.it.uu.se/research/publications/reports/2007-028/2007-028-nc.pdf, 2007.

[32] A. Hellander, P. Lötstedt, Hybrid method for the chemical master equation, Journal of Comp. Physics 227 (2007) 100–122.

[33] K. Sanft, S. Wu, M. Roh, J. Fu, R. K. Lim, L. Petzold, Stochkit2: software for discrete stochastic simulation of biochemical systems with events, Bioinformatics 27 (2011) 2457–2458.

[34] C.-W. Shu, Essentially Non-Oscillatory and Weighted Essentially Non-Oscillatory Schemes for Hyperbolic Conservation Laws, Technical Report, ICASE Report No. 97-65; NASA/CR-97-206253, 1997.

[35] J. M. G. Vilar, H. Y. Kueh, N. Barkai, S. Leibler, Mechanisms of noise-resistance in genetic oscillators, Proc. Nat. Acad. Sci. 99 (2002) 5988–5992.

[36] J. Elf, J. Paulsson, O. G. Berg, M. Ehrenberg, Near-critical phenomena in intracellular metabolite pools, Biophysical Journal 84 (2003) 154–170.

[37] L. Ferm, P. Lötstedt, P. Sjöberg, Adaptive, conservative solution of the Fokker-Planck equation in molecular biology, Technical Report 2004-054, Department of Information Technology, Uppsala University, Uppsala, Sweden, 2004.

[38] D. Scott, Multivariate Density Estimation. Theory, Practice and Visualization, Wiley, New York, 1992.

[39] B. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.

[40] B. van Leer, Towards the ultimate conservative difference scheme, v. a second order sequel to godunov's method, J. Comp. Phys. 32 (1979) 101–136.

[41] P. D. Lax, Weak solutions of nonlinear hyperbolic equations and their numerical computation, Comm. Pure Appl. Math. 7 (1954) 159–193.

[42] R. LeVeque, Finite-Volume Methods for Hyperbolic Problems, Cambridge University Press, Cambridge, 2002.

[43] A. Harten, High resolution schemes for hyperbolic conservation laws, Journal of Computational Physics 49 (1983) 357 – 393.

[44] SWIG (Simplified Wrapper and Interface Generator), `http://swig.org/`, 2013.

[45] T. Oliphant, Python for scientific computing, Computing in Science & Engineering 9 (2007) 90.