PARALLEL IN TIME SIMULATION OF MULTISCALE STOCHASTIC CHEMICAL KINETICS[*]

STEFAN ENGBLOM[†]

Abstract. A version of the time-parallel algorithm parareal is analyzed and applied to stochastic models in chemical kinetics. A fast predictor at the macroscopic scale (evaluated in serial) is available in the form of the usual reaction rate equation. A stochastic simulation algorithm is used to obtain an exact realization of the process at the mesoscopic scale (in parallel). The underlying stochastic description is a jump process driven by the Poisson measure. A convergence result in this arguably difficult setting is established, suggesting that a homogenization of the solution is advantageous. We devise a simple but highly general such technique. Three numerical experiments on models representative to the field of computational systems biology illustrate the method. For nonstiff problems, it is shown that the method is able to quickly converge even when stochastic effects are present. For stiff problems, we are instead able to obtain fast convergence to a homogenized solution. Overall, the method builds an attractive bridge between, on the one hand, macroscopic deterministic scales and, on the other hand, mesoscopic stochastic ones. This construction is clearly possible to apply also to stochastic models within other fields.

1. Introduction. It has been shown in several studies [17, 47, 51, 54] that stochastic descriptions of biochemical networks are necessary for understanding and explaining the mechanisms inside living cells. Such networks often contain species present in copy numbers down to a few hundreds [30], hinting why discreteness and stochasticity becomes important. Examples of when random effects are pronounced include the regularity and drift of circadian oscillations [61], spontaneous separation in bistable systems [16], and the creation of new steady-states [60].

An accurate such stochastic model can be obtained directly from microphysical considerations as a consequence of the Markov property [28]. Essentially, in a diffusion-controlled system, reactive collisions between molecules are rare events, implying a rapid loss of dependence on the system's history.

Sample realizations of any given chemical network can be obtained by randomly firing reactions and correspondingly updating the state of the system; the most well known such algorithm is Gillespie's *stochastic simulation algorithm* (SSA) [27]. Although conceptually simple, detailed simulation of complex reaction networks, as found, for example, in living cells remains a computationally very intensive problem.

The *parareal* algorithm for the solution of evolution equations on a time interval $[0, T]$ was suggested in the note [44] as a method for the *para*llel solution of problems in

---

[†]School of Computer Science and Communication, Royal Institute of Technology, S-100 44 Stockholm, Sweden (steng@csc.kth.se).

real time. An improved version, better tuned to nonlinear problems, was subsequently applied to problems in control theory [46] and molecular dynamics [4], to mention just a few.

The method is built around a predictor-corrector step in which a *coarse* solver is used as a preconditioner to a *fine* solver. It thus incorporates both multigrid and domain decomposition ideas, but is unique in that it is a purely parallel algorithm; it has no value when executed on a serial machine.

The idea suggested in this paper is to use the macroscopic approximation of the chemical system, the *reaction rate equation*, as the coarse solver while a stochastic simulation is used as the fine scale solver. This setup is available also for other types of stochastic evolution equations but has not received much attention. The reason is probably that any deterministic model is a quite poor formal approximation. On the other hand, any macroscopic trend in the solution is likely to be quickly obtained. Also, with a deterministic coarse solver, highly efficient implicit integrators are available, which greatly simplifies the implementation.

The present paper is related to references [6, 45]. In the former, the analysis of parareal when applied to ordinary differential equations (ODEs) and stochastic ODEs (SDEs) driven by Wiener processes is treated. In the latter, the parareal algorithm is applied to linear mass-balance (deterministic) chemical kinetics with disparate rates by using a coarse solver in the form of a reduced model.

The paper is divided into sections as follows: in section 2 the mathematical model of stochastic kinetics is described in some detail, and we briefly review the parareal algorithm in section 3. We present a convergence analysis in section 4; the main result indicates that strong convergence might be slow for systems with large Lipschitz constants. A way around this is suggested in the form of a simple but very general homogenization procedure in parallel. This "parareal homogenization" could well be an interesting approach to other stiff stochastic problems. We perform three quite different numerical experiments in section 5, where the practical implementation of the algorithm is also discussed. Finally, in section 6 we conclude the paper by discussing the outcome of the experiments and suggesting some future work and generalizations.

2. Stochastic chemical kinetics. In this section we discuss the usual mesoscopic stochastic model for chemical reactions (the master equation). As it is needed in the analysis in section 4, we shall also give an equivalent and mathematically more appealing representation in terms of a certain stochastic jump process. For monographs on applications of the master equation, see [23, 36]. The mathematical treatment of jump processes can be found in [3, 13, 20, 26, 34].

2.1. Chemical reactions, the master equation, and a hierarchy of methods. We consider in this paper a general homogeneous chemical network consisting of D different species reacting according to R prescribed reactions. At any given time t, the *state* of the system is an integer vector $x \in \mathbf{Z}_+^D = \{0, 1, 2, \dots\}^D$ counting the number of molecules of each kind. In a stochastic description, each reaction is a change of the state according to a certain transition rule, and the intensity of this process is governed by a *reaction propensity* $w_r : \mathbf{Z}_+^D \to \mathbf{R}_+$. This is the transition probability per unit of time for moving from the state x to $x - \mathbb{N}_r$:

$$(2.1) \qquad x \xrightarrow{w_r(x)} x - \mathbb{N}_r,$$

where, by convention, $\mathbb{N}_r \in \mathbf{Z}^D$ is the transition step and is the rth column in the *stoichiometric matrix* $\mathbb{N} \in \mathbf{Z}^{D \times R}$.

For instance, in a given well-stirred volume Ω the *elementary* reactions can be written using the states $x = [a, b]^T$,

$$(2.2) \qquad\qquad \emptyset \xrightarrow{k_1\Omega} A,$$

$$(2.3) \qquad\qquad A \xrightarrow{k_2 a} \emptyset,$$

$$(2.4) \qquad\qquad A + A \xrightarrow{k_3 a(a-1)/\Omega} \emptyset,$$

$$(2.5) \qquad\qquad A + B \xrightarrow{k_4 ab/\Omega} \emptyset.$$

The propensities are generally scaled such that

$$(2.6) \qquad\qquad w_r(x) = \Omega u_r(x/\Omega)$$

for some function u_r which does not depend on Ω. Intensities of this form are called *density dependent* and arise naturally in a number of situations [20, Chap. 11].

Arguably the most common stochastic description of chemically reacting systems is the *chemical master equation* (CME). To state it, let $p(x, t)$ be the probability that a certain number x of molecules is present at time t. The CME [23, 36] is then given by

$$(2.7) \qquad \frac{\partial p(x, t)}{\partial t} = \sum_{\substack{r=1 \\ x+\mathbb{N}_r^- \geq 0}}^{R} w_r(x + \mathbb{N}_r)p(x + \mathbb{N}_r, t) - \sum_{\substack{r=1 \\ x-\mathbb{N}_r^+ \geq 0}}^{R} w_r(x)p(x, t),$$

where the transition steps are decomposed into positive and negative parts as $\mathbb{N}_r = \mathbb{N}_r^+ + \mathbb{N}_r^-$.

In a pioneering paper from 1976, Gillespie [27] showed that exact samples from the master equation can be obtained via the SSA. This follows essentially by identifying it with the forward Kolmogorov equation associated to a certain continuous-time Markov chain to be introduced shortly. Importantly, Gillespie [28] later also showed that the CME is an exact physical description when the system is well stirred and in thermal equilibrium.

For chemical networks where the number of reactions per unit of time is large, detailed simulation with SSA becomes inefficient. As a remedy, Gillespie [29] proposed the *tau-leap method* with the ability to simultaneously "leap" over many reactions at once. The method has since been developed [2, 52, 59] and analyzed [32, 33, 42, 56] by several authors.

A related issue is *stiffness*: many interesting models become prohibitively expensive to solve by explicitly simulating the various involved scales. Several different model reduction techniques have been proposed for this situation [10, 15, 31]. Also, an implicit version of the tau-leap method has been developed [55], but this method converges in a very weak sense only [12, 43, 56].

An alternative approximation is the *Langevin equation* [20, Chap. 7], which is a *continuous* SDE driven by Wiener processes. On ignoring the noise term and keeping only the drift, the *reaction rate* equation is obtained. This is an ODE in D variables approximately evolving the expectation value of the mesoscopic model; exact moment equations are obtained for linear kinetics only. The hierarchy of solution methods (SSA → tau-leap → Langevin SDE → ODE) can thus be thought of as a transition between the mesoscopic model and the macroscopic one [7, 29, 38, 39, 40].

2.2. Sample path representation. A representation equivalent to but perhaps more direct than the master equation (2.7) was proposed fairly recently in [53] and further detailed in [42]. The idea is to construct a sample path representation in the form of jump SDEs driven by Poisson processes. As we shall see and as pointed out in [42, 53], this representation is particularly useful for numerical analysis.

We therefore introduce the stochastic variable $X(t) \in \mathbf{Z}_+^D$ counting at time $t \geq 0$ the number of molecules of each type. Reactions are understood to occur instantly, and hence the process is right continuous only; the notation $X(t-)$ is used to denote the value of the process prior to any reactions occurring at time t. We assume the existence of a probability space[1] $(\Sigma, \mathbf{F}, \mathbf{P})$ with the filtration $\mathbf{F}_{t \geq 0}$ containing R-dimensional standard Poisson processes. The transition probabilities in (2.1) define *counting processes* $\pi_r(t)$ [13, Chap. 2.5] according to

$$(2.8) \qquad E\left[\pi_r(t + dt) - \pi_r(t)|\mathbf{F}_t\right] = w_r(X(t-)) \, dt + o(dt).$$

In turn, the counting processes determine the process $X(t)$ [20, Chap. 6.4]:

$$(2.9) \qquad X_t = X_0 - \sum_{r=1}^{R} \mathbb{N}_r \pi_r(t).$$

A representation of $\pi_r(t)$ in terms of a unit-rate Poisson process Π_r in an *operational* or *scaled* time can also be given [13, Chap. 3.1, 4.4]:

$$(2.10) \qquad \pi_r(t) = \Pi_r \left(\int_0^t w_r(X(s-)) \, ds \right).$$

This *random time change representation* has appeared earlier in the literature [40, 41], but has also received more recent attention in the current context of chemical kinetics [1, 2].

We now wish to move beyond (2.9) and obtain a more transparent representation in the form of a jump SDE driven by the proper measure. Processes with nonlinearly dependent jump intensities are quite difficult to handle, since the noise enters with a complicated dependence on the state. This is in contrast to the more familiar Itô SDE driven linearly by Wiener processes.

For the present purposes, the Poisson random measure [35, Chap. II.1c] $\mu(dt \times dz; \sigma)$, with $\sigma \in \Sigma$ defines an increasing sequence of arrival times $\tau_i \in \mathbf{R}_+$ with corresponding "marks" z_i uniformly distributed in $[0, \bar{W}]$ [13, Chap. 1.4], where $\bar{W}$ is to be defined shortly. The deterministic intensity of $\mu(dt \times dz)$ is the Lebesgue measure $m(dt \times dz) = dt \times dz$ [26, Part II, Chap. 2, sect. 4].

The propensities have been general nonlinear functions up to this moment, but we need to impose conditions so as to bound the total intensity of all reaction channels. It is not difficult to see that a *finite* intensity implies a bounded solution in the mean square sense for finite times (see Lemma 4.2 below). For convenience, we take the approach in [42] and formally specialize the investigation to closed systems. This implies the existence of a bounded set S containing at any time t the state of the system and is reasonable from physical considerations; any real application must necessarily involve a bounded total number of molecules.

---

[1]Note that, throughout the paper, the letter Ω is *not* used for the probability sample space, but it consistently denotes the system volume.

Summarizing, we have the following.

Assumption 2.1. The state of the chemical network satisfies $X(t) \in S \subset \mathbf{Z}_+^D$. When confined to this set, all propensities are Lipschitz continuous in their argument with respect to the Euclidean norm $|w_r(x) - w_r(y)| \leq L_r \|x - y\|$. It follows that all propensities are bounded such that

$$(2.11) \qquad \bar{W}_r := \sum_{s=1}^{r} \max_{x \in S} w_s(x)$$

exists and is bounded for all r. We define also the numbers $L = \sum_r L_r$, $\bar{W} = \bar{W}_R$, and $W = \max_{x \in S} \sum_r w_r(x)$. Note that $W \leq \bar{W}$.

It is known that estimates based on Lipschitz constants generally become overly pessimistic for diffusive nonlinear problems with transient scales. For examples of relevance to the present work, see the analysis of splitting-based time-parallel methods in [9, Chap. 7.9] and for results in the setting of jump processes, see [33]. We therefore introduce a *one-sided* Lipschitz constant as follows: define

$$(2.12) \qquad f(x) = -\sum_{r=1}^{R} \mathbb{N}_r w_r(x),$$

and let Λ be the smallest constant such that for all $x, y \in S$,

$$(2.13) \qquad (f(x) - f(y), x - y) \leq \Lambda \|x - y\|^2$$

holds in the usual Euclidean inner product. By Assumption 2.1, we always have that $\Lambda \leq L\|N\|$, but we are mainly interested in the cases where Λ is even smaller since this yields better estimates.

The frequency of each reaction is controlled through a set of indicator functions $\hat{w}_r : S \times [0, \bar{W}] \to \{0, 1\}$ defined as follows:

$$(2.14) \qquad \hat{w}_r(x;\, z) = \begin{cases} 1 & \text{if } 0 \leq z - \bar{W}_{r-1} < w_r(x), \\ 0 & \text{otherwise.} \end{cases}$$

We can now represent the counting process (2.10) in terms of the Poisson random measure [35, Chap. II.1d] via a *thinning* of the measure using an acceptance-rejection strategy [13, Chap. 4.3]:

$$(2.15) \qquad \pi_r(t) = \int_0^t \int_0^{\bar{W}} \hat{w}_r(X(t-);\, z)\, \mu(dt \times dz).$$

Note that this thinning is not the same as in [42, 53] where instead $z \in [0, W]$; the present construction yields sharper estimates when comparing trajectories formed from different initial data.

The representation (2.15) combined with (2.9) leads us to the sample path representation in terms of a Skorohod jump SDE [34, Chap. IV.9]:

$$(2.16) \qquad dX_t = -\sum_{r=1}^{R} \mathbb{N}_r \int_0^{\bar{W}} \hat{w}_r(X(t-);\, z)\, \mu(dt \times dz).$$

Often one is interested in separating (2.16) into its "drift" and "jump" terms [26, Part II, Chap. 2, sect. 4]

$$dX_t = -\sum_{r=1}^{R} \mathbb{N}_r w_r(X(t-))\,dt - \sum_{r=1}^{R} \mathbb{N}_r \int_0^{\bar{W}} \hat{w}_r(X(t-);z)(\mu-m)(dt \times dz)$$

(2.17) $$=: dX_D(t) + dX_J(t).$$

We will use the initial value conventions $X_D(0) = X(0)$ and $X_J(0) = 0$ throughout the paper. Note that the second term in (2.17) driven by the compensated measure $(\mu-m)$ is a martingale of mean zero. On taking expectation values and approximating, we obtain $dEX_t \approx d\tilde{x}(t)$, where

(2.18) $$d\tilde{x}(t) = -\sum_{r=1}^{R} \mathbb{N}_r w_r(\tilde{x}(t))\,dt = f(\tilde{x}(t))\,dt$$

and $\tilde{x}(0) = EX(0)$. This approximation constitutes, up to a scaling by the system volume Ω, the usual macroscopic reaction rate equation.

3. The parareal algorithm. We continue by giving a short account of the parareal algorithm. For a quick introduction with additional references, the survey [57] can be recommended. See also [5, 22, 58] for abstract convergence results, stability analysis, and more.

Consider a general time-dependent problem written in operator form:

(3.1) $$\dot{u} = -A(u), t \in [0,T], \text{ with } u(0) = u_0.$$

The parareal algorithm is obtained from two solution operators. Define first the *fine* propagator $\mathcal{F}_t y \equiv \tilde{u}(t)$, where $\tilde{u}$ accurately solves (3.1) subject to the initial condition $\tilde{u}(0) = y$. Second, the parareal algorithm employs a *coarse* propagator $\mathcal{C}$ that does the same thing but faster and presumably less accurately. For instance, a combination often used in analysis is an explicit Euler method using a small time-step for $\mathcal{F}$ and an implicit Euler method with a larger time-step for $\mathcal{C}$.

Time is now discretized in $N = T/\Delta t$ chunks of time, and any solver $S \in \{\mathcal{F}_{\Delta t}, \mathcal{C}_{\Delta t}\}$ can be used to compute a numerical solution:

(3.2) $$\begin{bmatrix} I & & & \\ -S & I & & \\ & \ddots & \ddots & \\ & & -S & I \end{bmatrix} \begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_N \end{bmatrix} = \begin{bmatrix} u_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{u_0},$$

or simply $B(S)\mathbf{v} = \mathbf{u_0}$ in matrix notation. The parareal algorithm emerges as the fix-point iteration obtained by using $B(\mathcal{C}_{\Delta t})^{-1}$ as an approximate inverse to $B(\mathcal{F}_{\Delta t})$:

(3.3) $$\mathbf{v}_{k+1} = \mathbf{v}_k - B(\mathcal{C}_{\Delta t})^{-1}(B(\mathcal{F}_{\Delta t})\mathbf{v}_k - \mathbf{u_0}).$$

Let $v_{k,n}$ denote the nth element in the kth iterate $\mathbf{v}_k$, and let $v_{0,0} = u_0$ and $v_{0,n} = \mathcal{C}_{\Delta t} v_{0,n-1}$ for $n \in [1,N]$ to start up the algorithm. One readily verifies the recursion

(3.4) $$v_{k,n} = \mathcal{F}_{\Delta t} v_{k-1,n-1} - [\mathcal{C}_{\Delta t} v_{k-1,n-1} - \mathcal{C}_{\Delta t} v_{k,n-1}],$$

where the point of the algorithm is that the evaluation of $\mathcal{F}$ is now trivially parallel. At any iteration k, the preconditioned residual is given by

$$(3.5) \qquad \mathbf{p}_k = \mathbf{v}_k - \mathbf{v}_{k+1}$$

and is useful as an estimate of the error.

A nice survey of the parareal method is found in [22] where the connection to several earlier methods is made, including time-parallel multigrid and waveform relaxation methods. The process can be viewed as a Newton method applied to a certain nonlinear problem using a Jacobian approximated by finite differences. This provides an explanation to the superlinear convergence often observed on bounded intervals. It is known that the parareal algorithm converges poorly for problems where the operator has eigenvalues with large imaginary parts [58]. Another interesting result is that the parareal scheme can be made unconditionally stable by using a suitable relaxation of the coarse solver [5].

4. Analysis. In this section we analyze the parareal algorithm obtained by using the rate equation (2.18) and the jump SDE (2.16) as the coarse and fine solver, respectively. In section 4.1 we give some results that cover local properties of solutions to these equations. The actual convergence analysis is found in section 4.2, where convergence in mean square and convergence of the first moment is investigated. As noted in [6], weak convergence is not so interesting for the parareal algorithm, since weak estimates typically come into play when many trajectories are simultaneously computed in a Monte Carlo simulation; parallelism can then be trivially achieved with optimal efficiency. Our interest in the convergence of the first moment mainly stems from the possibility of homogenizing the model on the fly by a suitable averaging filter. This idea is discussed in section 4.3.

Because the process $X(t)$ is integer valued, we mention here for clarity the trivial inequality

$$(4.1) \qquad |X| \leq X^2 \Longrightarrow E|X| \leq EX^2.$$

When X is large, (4.1) is an overestimate, and we will switch to the standard inequality

$$(4.2) \qquad E|X| \leq (EX^2)^{1/2}.$$

In what follows we will use various elementary inequalities and energy-type estimates based on the one-sided Lipschitz constant defined in (2.13). Without serious loss of generality and for the sake of keeping a transparent notation, we therefore treat the one-dimensional case only. In particular, note that the Poisson random measure in (2.16) is in fact scalar.

4.1. Preliminaries. We start by citing the following convenient lemma.

LEMMA 4.1 (Lemma 3.1 and 3.2 in [42]; see also Lemma 2.3.2 in [3]). *Define* $\Delta X_t = X(t) - X(t-)$. *Then for any fixed time $s > 0$, $\Delta X_s = 0$ (a.s.). Furthermore, if w is a continuous function and $t_1 < t_2$, then*

$$(4.3) \qquad \int_{t_1}^{t_2} \Delta w(X_t)\, dt = 0.$$

We continue by establishing two lemmas concerning the stability of the jump process $X(t)$.

LEMMA 4.2. *Let $X(t \geq 0)$ evolve according to (2.16), and define the jump term $X_J(t)$ as in (2.17) with $X_J(0) = 0$. Then*

$$(4.4) \qquad EX_J(t)^2 \leq \|\mathbb{N}\|^2 Wt.$$

Furthermore,

$$(4.5) \qquad E[X_t - X_0]^2 \leq 2\|\mathbb{N}\|^2 Wt + 2\|\mathbb{N}\|^2 W^2 t^2.$$

Proof. The left side of (4.4) can be written

$$E\left(\int_0^t \int_0^{\bar{W}} \sum_{r=1}^R -\mathbb{N}_r \hat{w}_r(X(s-); z)(\mu - m)(ds \times dz)\right)^2$$

$$= E\int_0^t \int_0^{\bar{W}} \left(\sum_{r=1}^R -\mathbb{N}_r \hat{w}_r(X(s-); z)\right)^2 m(ds \times dz)$$

$$\leq E\int_0^t \int_0^{\bar{W}} \|\mathbb{N}\|^2 \sum_{r=1}^R \hat{w}_r(X(s-); z)^2 m(ds \times dz),$$

where the Itô isometry for jump processes was used [26, Part II, Chap. 2, sect. 5], [3, Lemma 4.2.2]. Using that $\hat{w}_r^2 = \hat{w}_r$ we complete the proof of (4.4). For (4.5) we use the drift/jump split (2.17) and the inequality $(a + b)^2 \leq 2a^2 + 2b^2$ to obtain

$$E[X_t - X_0]^2 \leq 2EX_J(t)^2 + 2E\left(\int_0^t \sum_{r=1}^R -\mathbb{N}_r w_r(X(s-)) \, ds\right)^2. \qquad \square$$

For the proof of our next lemma, we shall need the case $\varepsilon = 0$ of the following nonlinear inequality, remindful of Grönwall's lemma.

PROPOSITION 4.3. *Let u and b be nonnegative continuous functions in some positive interval including zero, and let $\varepsilon \geq 0$ be a constant. Then*

$$(4.6) \qquad u(t) \leq \left(\varepsilon^2 + \int_0^t b(s)u(s) \, ds\right)^{1/2} \quad \text{implies} \quad u(t) \leq \varepsilon + \frac{1}{2}\int_0^t b(s) \, ds.$$

Proof. Define

$$D_\varepsilon(t) \equiv \left(\varepsilon^2 + \int_0^t b(s)u(s) \, ds\right)^{1/2} - \varepsilon - \frac{1}{2}\int_0^t b(s) \, ds.$$

Then $D_\varepsilon(0) = 0$ and, for $\varepsilon > 0$, assuming the first inequality in (4.6), we readily obtain $D'_\varepsilon(t) \leq 0$ for $t > 0$ so that D_ε is nonpositive. This clearly implies (4.6). The result remains true when $\varepsilon \to 0+$ by using continuity from the right. $\square$

LEMMA 4.4. *Let the drift term be defined as in (2.17) with $X_D(0) = X(0)$, and let $\tilde{x}(t)$ be the reaction rate approximation (2.18) to the expected value of $X(t)$ with $\tilde{x}(0) = X(0)$. Then*

$$(4.7) \qquad E[X_D(t) - \tilde{x}(t)]^2 \leq L^2 \|\mathbb{N}\|^4 Wt^3 \left(\frac{\exp(\Lambda t) - 1}{\Lambda t}\right)^2.$$

Although we will mainly use (4.7) in the much weaker form $E[X_D(t) - \tilde{x}(t)]^2 = \mathcal{O}\left(t^3\right)$, the actual values of the various constants are of interest. For instance, less sharp estimates with Λ replaced with L are easier to obtain but place severe restrictions on time-steps.

Proof. Using the one-sided Lipschitz bound (2.13) we get

$$\frac{d}{dt}[X_D(t) - \tilde{x}(t)]^2 = 2(f(X_t) - f(\tilde{x}_t), X_D(t) - \tilde{x}(t))$$
$$\leq 2\Lambda[X_D(t) - \tilde{x}(t)]^2 + 2(f(X(t)) - f(X_D(t)), X_D(t) - \tilde{x}(t))$$
$$\leq 2\Lambda[X_D(t) - \tilde{x}(t)]^2 + 2L\|N\|\|X_J(t)\|\|X_D(t) - \tilde{x}(t)|.$$

Hence on taking expectation and integrating using the Cauchy–Schwartz inequality and the first part of Lemma 4.2 we get

$$\exp(-2\Lambda t)E[X_D(t) - \tilde{x}(t)]^2$$
$$\leq \int_0^t 2L\|N\| \exp(-2\Lambda s)\|N\|W^{1/2}s^{1/2} \left(E[X_D(s) - \tilde{x}(s)]^2\right)^{1/2} ds.$$

Now (4.7) follows by applying the nonlinear Grönwall-type inequality of Proposition 4.3 with $u(t)^2 = \exp(-2\Lambda t)E[X_D(t) - \tilde{x}(t)]^2$ and some straightforward calculations. $\square$

Remark. In a closed system we can identify the magnitude of X with the system volume Ω; $\max_{x \in S} \|x\| = \mathcal{O}(\Omega)$. For density-dependent propensities (see (2.6)), this implies that $W \sim \Omega$, while L does not scale with Ω. Lemma 4.2 and 4.4 then yield

$$(4.8) \qquad\qquad \left(E[X(t) - \tilde{x}(t)]^2\right)^{1/2} \leq C_t\Omega^{1/2}$$

for some bounded constant C_t. By Chebyshev's inequality, we thus have convergence in probability $\Omega^{-1}X(t) \to \Omega^{-1}\tilde{x}(t)$ as $\Omega \to \infty$. For more precise reasoning of this type, see [20, 38, 39, 40].

We proceed with a result that is crucial to the convergence properties of the parareal algorithm.

THEOREM 4.5. *Let $X(t \geq 0)$ and $Y(t \geq 0)$ evolve according to (2.16) using the same underlying set of Poisson processes but with different initial values X_0 and Y_0. Define the jump terms $X_J(t)$ and $Y_J(t)$ as in Lemma* 4.2. *Then*

$$(4.9) \qquad E[Y_J(\Delta t) - X_J(\Delta t)]^2 \leq L\|\mathbb{N}\|^2\Delta t|Y_0 - X_0|(1 + \mathcal{O}(\Delta t)).$$

Proof. By the Itô isometry, we can bound the left side of (4.9) by

$$E \int_0^{\Delta t} \int_0^{\bar{W}} \|\mathbb{N}\|^2 \sum_{r=1}^R [\hat{w}_r(Y(t-); z) - \hat{w}_r(X(t-); z)]^2 m(dt \times dz)$$
$$\leq E \|\mathbb{N}\|^2 \int_0^{\Delta t} \sum_{r=1}^R |w_r(Y(t-)) - w_r(X(t-))| dt,$$

where the usefulness of the thinning proposed in section 2.2 is evident. From the

Lipschitz assumption and Lemma 4.1 we get the bound

$$\leq L\|\mathbb{N}\|^2 \int_0^{\Delta t} E\left|Y(t-) - X(t-)\right| dt = L\|\mathbb{N}\|^2 \int_0^{\Delta t} E\left|Y(t) - X(t)\right| dt$$

$$\leq L\|\mathbb{N}\|^2 \Delta t\, |Y_0 - X_0| + L\|\mathbb{N}\|^2 \int_0^{\Delta t} E\left|Y(t) - Y_0\right| + E\left|X(t) - X_0\right| dt$$

$$\leq L\|\mathbb{N}\|^2 \Delta t |Y_0 - X_0| + 4L\|\mathbb{N}\|^4 \int_0^{\Delta t} Wt + W^2 t^2\, dt,$$

where the integer inequality (4.1) and the second part of Lemma 4.2 were used in the last line. □

We now set out to prove a result which is similar in spirit to Theorem 4.5, but which is used to study the error in the first moment alone. For this purpose, we add the assumption that f in (2.12) is two times continuously differentiable with a uniformly bounded second derivative. We note that this restriction is only a very minor one in light of the assumptions already made; see also the elementary propensities in (2.2)–(2.5). A Taylor's expansion then implies that

$$(4.10) \qquad \|Ef(X) - f(EX)\| \leq \lambda E\|X - EX\|^2,$$

where λ is some constant.

THEOREM 4.6. *Define $X_D(t)$ and $\tilde{x}(t)$ as in Lemma 4.4. Then*

$$(4.11) \qquad |EX_D(\Delta t) - \tilde{x}(\Delta t)| \leq \|N\|^2 \lambda W \Delta t^2 \frac{\exp(\Lambda\Delta t) - 1}{\Lambda\Delta t} + \mathcal{O}\left(\Delta t^3\right).$$

Note first that (4.11) is $1/2$ order higher with respect to Δt than that obtained more directly from (4.7). Second, for density dependent propensities it holds that $\lambda \propto \Omega^{-1}$, so that the multiplicative constant in (4.11) is *independent of both the system size Ω and of the Lipschitz constant L.*

Proof. The proof parallels that of Lemma 4.4. Since $EX_J(t) = 0$, we get with $\varphi(t) \equiv [EX_D(t) - \tilde{x}(t)]^2$ that

$$\frac{d}{dt}\varphi^2 = 2(Ef(X_t) - f(\tilde{x}_t), \varphi) \leq 2\Lambda\varphi^2 + 2(Ef(X_t) - f(EX_t), \varphi).$$

From (4.10) and Lemmas 4.2 and 4.4 we readily get

$$|Ef(X(t)) - f(EX(t))| \leq \lambda E[X_t - EX_t]^2$$
$$\lambda(1+t)EX_J(t)^2 + \lambda(1+1/t)E[X_D(t) - \tilde{x}(t)]^2 \leq \|N\|^2 \lambda Wt + \mathcal{O}\left(t^2\right).$$

Hence on integrating we get

$$\exp(-2\Lambda t)\varphi(t)^2 \leq \int_0^t \left(2\|N\|^2 \lambda W s + \mathcal{O}\left(s^2\right)\right)\exp(-2\Lambda s)\varphi(s)\, ds,$$

where Proposition 4.3 applies with $u(t)^2 = \exp(-2\Lambda t)\varphi(t)^2$. □

4.2. Convergence of the parareal algorithm. For the application of the parareal algorithm, we let $\mathcal{F}y$ be the stochastic evolution in a time-step Δt of the jump process $X(t)$ obeying (2.16) with initial data y. Following (2.17), we write $\mathcal{F} = \mathcal{F}_D + \mathcal{F}_J$ for the drift- and jump-term, respectively. The coarse solver $\mathcal{C}y$ is

instead the evolution in a time-step Δt using the deterministic rate equation (2.18) starting from y.

We denote by $X = [X_0, \ldots, X_N]$ with $X_n = X(t_n)$ the exact solution at times $t_n \equiv n\Delta t$. $\tilde{X}_{k,n} \approx X_n$ is the numerical approximative solution obtained after the kth iteration of the parareal algorithm:

$$(4.12) \qquad \tilde{X}_{k,0} = X_0, \qquad k \geq 0,$$

$$(4.13) \qquad \tilde{X}_{0,n} = \mathcal{C}\tilde{X}_{0,n-1}, \qquad n \geq 1,$$

$$(4.14) \qquad \tilde{X}_{k,n} = \mathcal{F}\tilde{X}_{k-1,n-1} - \mathcal{C}\tilde{X}_{k-1,n-1} + \mathcal{C}\tilde{X}_{k,n-1},$$

where $(k, n) \geq 1$ in (4.14).

We now wish to analyze the root–mean-square (RMS) error defined by

$$(4.15) \qquad e_{k,n}^2 = E[\tilde{X}_{k,n} - X_n]^2.$$

This measure satisfies by (4.14) the recursion

$$
\begin{aligned}
e_{k,n}^2 &= E[T_1 + T_2 + T_3]^2 \\
(4.16) \qquad &= E\left[T_1^2 + T_2^2 + T_3^2 + 2T_1T_2 + 2T_1T_3 + 2T_2T_3\right],
\end{aligned}
$$

where

$$(4.17) \qquad T_1 \equiv \mathcal{F}_J\tilde{X}_{k-1,n-1} - \mathcal{F}_J X_{n-1},$$

$$(4.18) \qquad T_2 \equiv \mathcal{F}_D\tilde{X}_{k-1,n-1} - \mathcal{C}\tilde{X}_{k-1,n-1} - \mathcal{F}_D X_{n-1} + \mathcal{C}X_{n-1},$$

$$(4.19) \qquad T_3 \equiv \mathcal{C}\tilde{X}_{k,n-1} - \mathcal{C}X_{n-1}.$$

By Theorem 4.5 using (4.1) and Lemma 4.4 we get

$$(4.20) \qquad ET_1^2 \leq L\|\mathbb{N}\|^2 \Delta t e_{k-1,n-1}^2 (1 + \mathcal{O}(\Delta t)),$$

$$(4.21) \qquad ET_2^2 \leq \mathcal{O}\left(\Delta t^3\right).$$

We also have the bound

$$(4.22) \qquad ET_3^2 \leq \exp(2\Lambda\Delta t)e_{k,n-1}^2.$$

As for the cross-terms, we have that $ET_1T_3 = 0$ since T_1 is a martingale of zero mean and T_3 is a deterministic (nonanticipating) function. The remaining terms can be easily estimated as

$$(4.23) \qquad 2ET_1T_2 \leq \Delta t ET_1^2 + \Delta t^{-1} ET_2^2 \leq \mathcal{O}\left(\Delta t^2\right) e_{k,n-1}^2,$$

$$(4.24) \qquad 2ET_2T_3 \leq \Delta t^{-1} ET_2^2 + \Delta t ET_3^2 \leq \mathcal{O}\left(\Delta t^2\right) + \mathcal{O}\left(\Delta t\right) e_{k,n-1}^2.$$

Summarizing, we have from (4.16) upon ignoring higher order terms in Δt,

$$
\begin{aligned}
e_{k,n}^2 &\leq L\|\mathbb{N}\|^2 \Delta t e_{k-1,n-1}^2 + \exp(2\Lambda\Delta t)e_{k,n-1}^2 \\
(4.25) \qquad &= \Delta t S_{\mathcal{F}}^2 e_{k-1,n-1}^2 + \exp(2\Delta t S_{\mathcal{C}})e_{k,n-1}^2,
\end{aligned}
$$

where $S_{\mathcal{F}} \equiv L^{1/2} \|\mathbb{N}\|$ and $S_{\mathcal{C}} \equiv \Lambda$ are the stability constants associated with the fine and coarse solver, respectively. Define $M = \max_n e_{0,n}$ in view of (4.8). We readily recognize the binomial recurrence in (4.25) so that

$$(4.26) \qquad e_{k,n}^2 \le M^2 \binom{n}{k} \Delta t^k S_{\mathcal{F}}^{2k} \exp(2\Delta t S_{\mathcal{C}}(n-k)).$$

If we now look at iteration $k = \mathcal{O}(1)$ and interval $n = N = T/\Delta t$, (4.26) can be simplified into

$$(4.27) \qquad e_{k,n} \le C_{1,T} S_{\mathcal{F}}^k,$$

where $C_{1,T}$ is a bounded constant for any given total time T. Equation (4.27) shows mean square convergence for contractive problems where $S_{\mathcal{F}} < 1$ only; it is unclear what happens for systems with larger Lipschitz constants. A key to understanding how the analysis can be refined lies in the fact that the integer inequality (4.1) had to be applied before using Theorem 4.5 and arriving at (4.20). For an initial large error, the inequality (4.2) is sharper but introduces a nonlinear dependence:

$$(4.28) \qquad e_{k,n}^2 \le \Delta t S_{\mathcal{F}}^2 e_{k-1,n-1} + \exp(2\Delta t S_{\mathcal{C}}) e_{k,n-1}^2.$$

This motivates our interest in the next proposition.

PROPOSITION 4.7. *Consider for $(k,n) \ge 0$ a sequence $a_{k,n}$ of nonnegative numbers. Let $a_{k,0} = 0$ for $k \ge 0$, and suppose that $a_{0,n} \le C$ for $n > 0$ and some nonnegative constant C. Then for $k \le n$ the inequality*

$$a_{k,n} \le A a_{k-1,n-1}^{1/2} + B a_{k,n-1},$$

where $A \ge 0$ and $B \ge 1$ can be "solved" in terms of the implied and simpler inequality

$$(4.29) \qquad a_{k,n} \le A^{2^{2-2^{1-k}}} B^{n-k} (n-k)^{2^{2-2^{1-k}}} C^{2^{-k}}.$$

A proof by induction is easily constructed once the functional form in (4.29) has been obtained.

To apply Proposition 4.7, we put $A = \Delta t S_{\mathcal{F}}^2$, $B = \exp(2\Delta t S_{\mathcal{C}})$, and identify $e_{k,n} = a_{k,n}^{1/2}$ so that $C = M^2 = \max_n e_{0,n}^2$ is the largest initial mean-square error. The result is for $k \ll n = N$ that

$$(4.30) \qquad e_{k,n} \le C_{2,T} M^{2^{-k}}$$

for some constant $C_{2,T}$. During the first few iterations, this bound is sharper than (4.27), which is only valid when the error is already small. For instance, if as in (4.8) we have that $M \sim \Omega^{1/2}$, then $e_{k,N} \sim [\Omega^{1/2}, \Omega^{1/4}, \Omega^{1/8}, \dots]$ for the first few iterations $k = [0, 1, 2, \dots]$. Since we have that the magnitude of the solution scales as Ω (cf. the remark after Lemma 4.4), this implies that for large enough system size Ω, the *relative* error decays quite quickly, at least initially. Note also that the dependence on the Lipschitz constant is much less critical in (4.30) than in (4.27), since $C_{2,T} \le T^2 L^2 \exp(\Lambda T)$.

As it provides us with additional insight, we also comment on the *error in the first moment*. Define for this purpose $\bar{e}_{k,n} = |E[\tilde{X}_{k,n} - X_n]|$. Proceeding as before and using Theorem 4.6 twice, it is not difficult to see that this error satisfies

$$(4.31) \qquad \bar{e}_{k,n} \le \mathcal{O}\left(\Delta t^2\right) \bar{e}_{k-1,n-1} + \exp(\Lambda \Delta t) \bar{e}_{k,n-1} \lesssim \binom{n}{k} \Delta t^{2k} \le C_{3,T} \Delta t^k.$$

The estimates (4.27), (4.30), and (4.31) show that *for Δt sufficiently small, the contractive parts converge in mean square while the noncontractive parts converge in the sense of* (4.31) *only*. This interpretation opens up for modifying the scheme in such a way that the fast scales are actively filtered away.

4.3. Homogenization. The previous analysis suggests that high accuracy may be difficult to reach with the proposed parareal algorithm for problems with a large Lipschitz constant L. Unless the one-sided Lipschitz constant Λ is much smaller or even negative, the RMS convergence *asymptotically* is dictated by (4.27) where $S_{\mathcal{F}} \propto \sqrt{L}$, despite the fact that the transient convergence in (4.30) is more robust with respect to L.

On the other hand, for certain problems involving very rapid scales, pathwise convergence may not be so interesting to obtain. For instance, this is often true in the "stiff noise"-case, where the trajectory has the appearance of a trend (which may or may not be transient) superpositioned with a rapidly varying noise. In such cases one is often rather interested in convergence to a *homogenized* solution, which by itself is a weak approximation to the original system. This situation occurs in the current context when rapid reaction channels almost balance each other so that the interesting dynamics occurs on a much slower scale.

Unlike the deterministic case where implicit solvers generally evolve stiff problems efficiently, it has been proposed that model reduction techniques are more appropriate in the stochastic setting [43]. For instance, the implicit tau-leap method has only been shown to be weakly convergent under the rather strong assumption of linear propensities [12, 56].

Provided that the coarse solver is sufficiently dissipative, the biggest challenge in directly using the current parareal algorithm for stiff problems is to detect when the numerical solution is of sufficient quality; since convergence is at best weak, any jumps on the order of the noise associated with the fast scale could in principle be accepted.

An easy way around this is to replace the fine propagator $\mathcal{F}$ with a homogenized version $\mathcal{F}^h$. A simple example is

$$(4.32) \qquad \mathcal{F}^h X_0 := \frac{1}{\delta t} \int_{\Delta t - \delta t}^{\Delta t} Y(t)\, dt, \qquad \text{where } Y(t) = \mathcal{F}_t X_0,$$

i.e., an average of the exact trajectory over an interval δt. This interval should be large enough to contain several fast reactions but short enough to be essentially independent on the slow scales. One can easily think of much more advanced filters to create a suitable homogenization, but we settled for the immediate (4.32).

Let us remark that an essentially "exact" trajectory is available within the homogenized solution, since the nonaveraged solution is always computed first. This "exact" trajectory is just as noisy as the true solution but allows for jumps related to the fast scale in between successive time intervals $[t_n, t_{n+1}]$. An intuitive interpretation is that the homogenized trajectory is an exact sample from a nearby model containing additional reactions that are scheduled at deterministic time-steps Δt. In principle, the intensity of this unknown process could be estimated and be put in relation to the rest of the propensities.

5. Numerical examples. After discussing the implementation of the proposed method, we will consider three numerical examples. The first is a typical example of when stochastic models are necessary, since mean-field equations give incorrect results. Nevertheless, when they are used as a preconditioner in the parareal algorithm,

convergence to the true solution is quite fast. The second example is representative of situations involving multiple scales with fast transients. Finally, we obtain a trajectory from the *reaction-diffusion* master equation, which is a good representative of very large networks. Here, the macroscopic model is just the familiar reaction-diffusion PDE.

5.1. Implementation. For the purpose of performing experiments, we have implemented a serial version of the parareal algorithm (4.12)–(4.14) in Matlab. The coarse solver $\mathcal{C}$ is simply the reaction rate equation (2.18) solved by any suitable ODE-solver; we typically used Matlab's `ode23s` but occasionally got away with just a single step of the backward Euler method.

As the fine solver $\mathcal{F}$, we used an implementation of the *next reaction method* (NRM) [25], which is statistically equivalent to the SSA but specifically tuned to large networks. The reason for preferring this particular algorithm is that *consistent Poisson processes may easily be sampled*. This feature is achieved by simply using the same sequence of random numbers for each reaction channel. Although there are other efficient SSAs available, such as the *optimized direct method* [11] or the *sorting direct method* [48], it appears that they are more difficult to adapt in this respect. To appreciate where this technicality enters in the analysis, note that in Theorem 4.5, Y_t and X_t are assumed to be trajectories conditioned on the same Poisson process. If they are not related in this respect, the associated constants are much larger.

In order to estimate the quality of the numerical solution, we used the relative Euclidean norm of the preconditioned residual as given by (3.5):
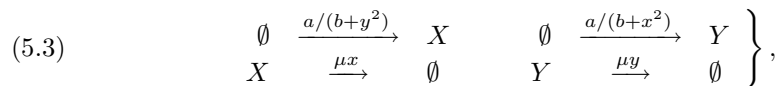
$$(5.1) \qquad e_{k+1} \sim \max_n D^{-1/2} \|(v_{k,n} - v_{k+1,n})/(1 + v_{k+1,n})\|$$

(elementwise division). This was then taken as an approximation to the true relative error

$$(5.2) \qquad e_{k+1} = \max_n D^{-1/2} \|(v_{k+1,n} - u_n)/(1 + u_n)\|.$$

We generally did not round any fractional results obtained from the coarse solver. Formally, this introduces a complication in the analysis as the solution space becomes a continuum (see [42, Rem. 3.1]), but we have not found any benefits in forcing the solution to stay in $\mathbf{Z}_+^D$. However, for fractional states, the propensities should explicitly be set to zero whenever executing the reaction would result in a negative copy number.

5.2. Stochastic toggle switch. A *toggle switch* found within the regulatory network of *Escherichia coli* has been modeled by two mutually cooperatively repressing gene products X and Y [24]. The model is

$$(5.3) \qquad \left.\begin{array}{ccc} \emptyset & \xrightarrow{a/(b+y^2)} & X \\ X & \xrightarrow{\mu x} & \emptyset \end{array} \qquad \begin{array}{ccc} \emptyset & \xrightarrow{a/(b+x^2)} & Y \\ Y & \xrightarrow{\mu y} & \emptyset \end{array}\right\},$$

with parameters $a = 3000$, $b = 11000$, and $\mu = 10^{-3}$. In (5.3), note that if the number of X-molecules is larger than the number of Y-molecules, then the production of Y-molecules is inhibited, and the system finds a stable state with $x > y$. However, the intrinsic noise can make the roles of X and Y suddenly switch to a state where instead, the production of X-molecules is inhibited. In this way, the system (5.3) constitutes an interesting example where the deterministic dynamics clearly differs from that obtained by a stochastic simulation.
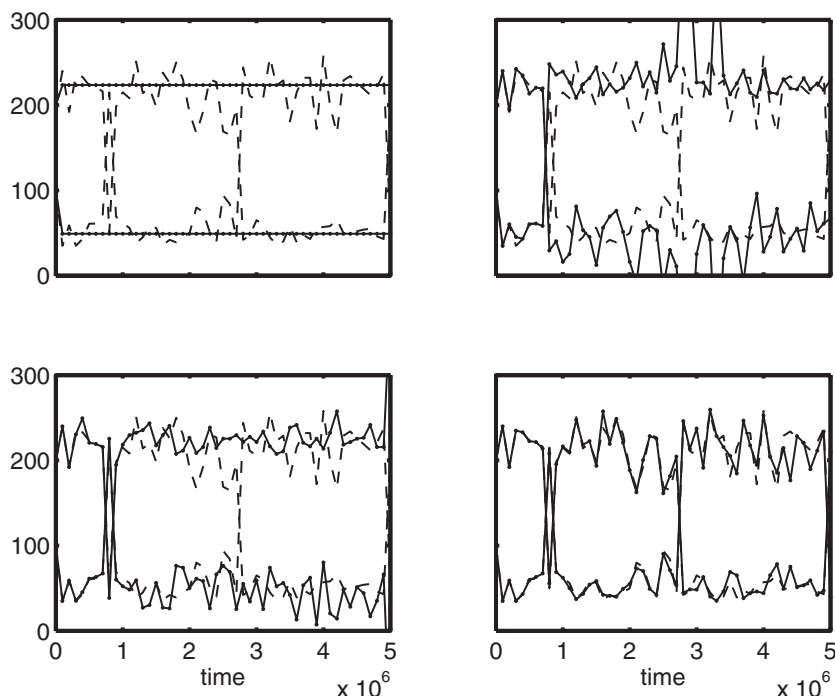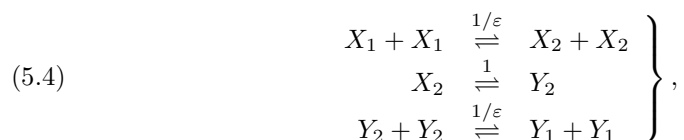
Fig. 1. *Dashed: exact trajectory of the two components of the toggle switch* (5.3) *versus time. Solid: solution obtained in parallel after* 0, 1, 2, *and* 4 *iterations. Top left: the initial solution from the reaction rate equation immediately settles at one of the stable states. Top right and bottom left: the first and second switches are correctly obtained after, respectively, one and two iterations. Bottom right: all* 4 *spontaneous changes of state have been correctly detected. Note the rather large level of noise.*

In Figure 1 the exact trajectory up to final time $T = 5 \times 10^6$ is displayed together with the approximation obtained after a few iterations by the parareal algorithm on a *very* coarse grid with $\Delta t = T/50 = 10^5$. Interestingly, the next correct place to switch is found for each new iteration so that all such events are correctly located within the first four iterations. Although the numerical solution sometimes overshoots, *some* information is evidently being correctly propagated through the system.

The relative errors and residuals are plotted in Figure 2. It is seen that the convergence initially is very rapid but then reaches a plateau where the error stays within a small fraction of the order of the noise level. This behavior is in line with the analysis in section 4: far away from the true trajectory the convergence is dictated by the nonlinear bound (4.30). Once the *absolute* error has been brought down to close to $\mathcal{O}(1)$, however, the presence of the Lipschitz constant in (4.27) implies a slower and less regular convergence. On the other hand, the accuracy obtained is quite reasonable and is at the order of a 1% perturbation of the propensities in (5.3).

5.3. Homogenization of disparate rates. As a specific model containing two different scales, we consider *fast dimerization* combined with *slow isomerization*

$$
(5.4) \qquad \left. \begin{array}{rcl}
X_1 + X_1 & \overset{1/\varepsilon}{\rightleftharpoons} & X_2 + X_2 \\
X_2 & \overset{1}{\rightleftharpoons} & Y_2 \\
Y_2 + Y_2 & \overset{1/\varepsilon}{\rightleftharpoons} & Y_1 + Y_1
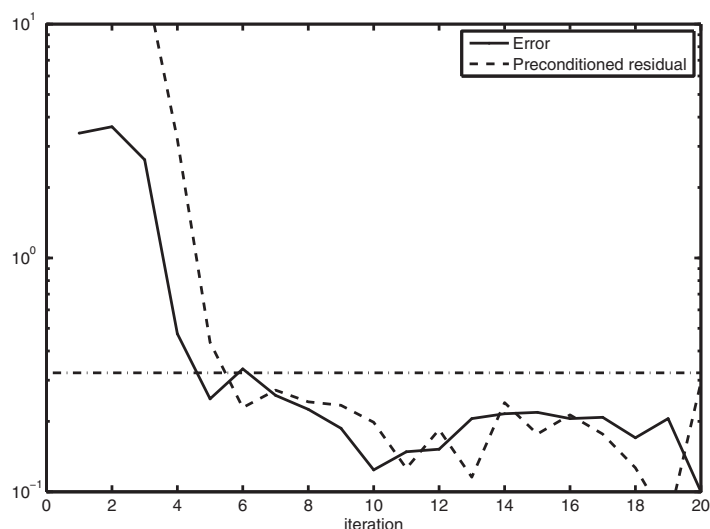\end{array} \right\},
$$

FIG. 2. *Maximum relative errors and residuals (see (5.1) and (5.2)) obtained for the toggle switch during the first 20 iterations. The accuracy is quickly improved during the first 5 iterations and then settles more slowly. The horizontal line (dash-dot) is the induced difference when the propensities are perturbed by ±1%.*

where the small parameter ε controls the difference in scales. For this example, we took $\varepsilon = 10^{-3}$ and initial data $[x_1, x_2, y_1, y_2](0) = [15, 5, 30, 10]$ with final time $T = 10$ so that the fast channels fire about 10^4 times more often than the slow ones. Note also that the fast channels have quadratic propensities (e.g., $w_1 = x_1(x_1 - 1)/\varepsilon$), and hence the rate equation is inexact.

For the parareal discretization, we used $\Delta t = T/50$ as before and an extremely simple coarse solver in the form of a single step with the linearized backward Euler method. The homogenization procedure described in section 4.3 was used with the fine solver defined in (4.32) using $\delta t = \Delta t/2$. The resulting combination was very effective indeed in obtaining a homogenized solution; see Figures 3 and 4.

5.4. Stochastic reaction-diffusion. There are many interesting chemical systems where the spatial distribution of the species must be taken into account. The usual thermodynamical equilibrium assumption ("well stirredness") no longer holds, as the transport of the molecules through the solvent is slow compared to average reaction times or when some reactions are strongly localized. The length scale over which the system can be regarded as homogeneous is now much shorter.

Examples of when both the stochasticity of the reactions and the spatial distribution are necessary to explain experimental data are found in [14, 16, 21, 37, 49]. For such systems, the diffusion at a molecular level can be treated as a special set of linear propensities. This yields the reaction-diffusion master equation (RDME) [23, Chap. 8], [36, Chap. XIV], which evolves the probability density of the system in the same manner as the CME (2.7). However, the dimensionality of the state-space is much higher and computing even a single trajectory can be a very computationally intensive problem. Note also that, as the rate equation now forms the reaction-diffusion PDE, the master operator defines a continuous spectrum of scales so that a clear separation into slow/fast ones is no longer possible.
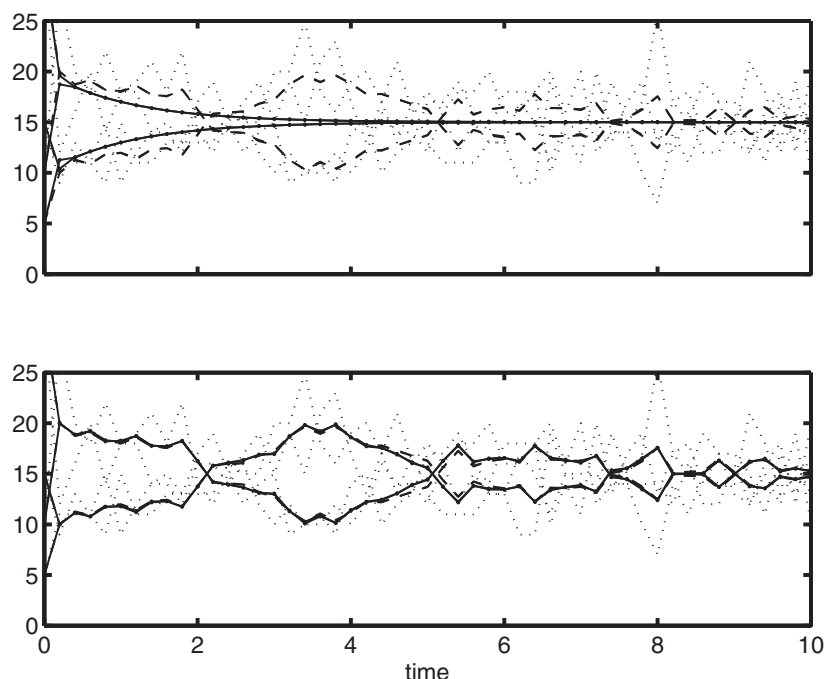
FIG. 3. *Dashed: homogenized solution for* (5.4) *versus time. Solid: solution after* 0 *(top) and* 1 *(bottom) iteration of the parareal algorithm. For comparison, the corresponding nonhomogenized trajectory is also plotted (dotted). For this example, a single step of the backward Euler method was a sufficient accurate coarse scale solver which therefore is extremely cheap to evaluate.*
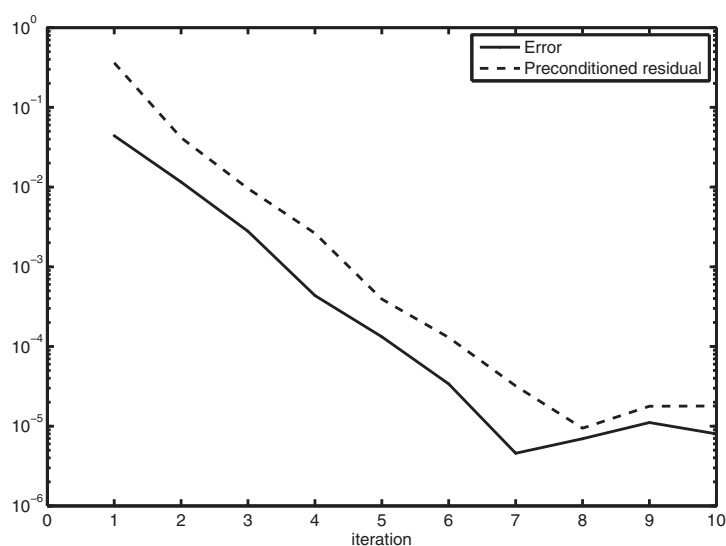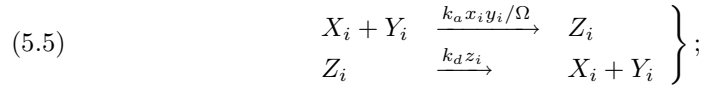

FIG. 4. *Convergence history for* (5.4) *homogenized according to* (4.32).

We shall consider a small example in the present section as follows. For $i = 1 \ldots 5$, denote by the triplet $(x, y, z)_i$ the number of X-, Y- and Z-molecules in cell i. The cells each have volume Ω and for simplicity are connected in an array 1-2-3-4-5. Within
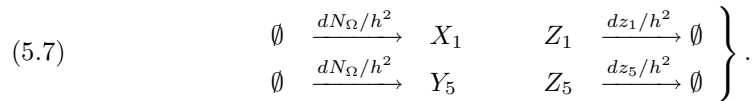
each cell i we specify the *reactions*

$$(5.5) \qquad \left. \begin{array}{lll} X_i + Y_i & \xrightarrow{k_a x_i y_i / \Omega} & Z_i \\ Z_i & \xrightarrow{k_d z_i} & X_i + Y_i \end{array} \right\};$$

compare (2.2)–(2.5). Furthermore, the molecules can *diffuse* to any neighboring cell $j = i \pm 1 \in \{1, \dots, 5\}$ according to

$$(5.6) \qquad \left. \begin{array}{lll} X_i & \xrightarrow{dx_i / h^2} & X_j \\ Y_i & \xrightarrow{dy_i / h^2} & Y_j \\ Z_i & \xrightarrow{dz_i / h^2} & Z_j \end{array} \right\},$$

where $h \equiv \Omega^{1/3}$ is the length-scale. Finally, there are also inflow and outflow at the boundaries

$$(5.7) \qquad \left. \begin{array}{llll} \emptyset & \xrightarrow{dN_\Omega / h^2} & X_1 \qquad Z_1 & \xrightarrow{dz_1 / h^2} \emptyset \\ \emptyset & \xrightarrow{dN_\Omega / h^2} & Y_5 \qquad Z_5 & \xrightarrow{dz_5 / h^2} \emptyset \end{array} \right\}.$$

The scaling of these propensities can be understood as a consequence of the fact that the mean first exit time τ of Brownian diffusion from a ball of radius r is given by $\tau = r^2 / 3d$ [50, Chap. 7.4].

In order to fully describe the system, we have chosen constants to approximately coincide with those in [16] for *Escherichia coli*; see also [8]. The model is parameterized by the integer N_Ω defining the total volume through $V = 10^{-15} \times N_\Omega / 25 = 5\Omega$ (l). The rate constants are given by $k_a = 10^8$ $(M^{-1}s^{-1})$, $k_d = 10$ (s^{-1}), and $d = 10^{-10}$ $(m^2 s^{-1})$. Since we have that $N_\Omega \propto \Omega$, for convenience, we loosely refer to both quantities as the "system size." The model is normalized around $N_\Omega = 25$ molecules per cell, and as initial data we simply took $(x_i, y_i, z_i) = (N_\Omega, N_\Omega, 0)$ in all cells.

For the parareal algorithm, we again took $N = 50$ intervals with total time $T = 1$. In the semidiscrete case, one can show that the reaction rate equation for the diffusion part (5.6) is equivalent to a mass-lumped finite element method for the macroscopic diffusion equation; this observation has been used to generalize the mesoscopic model to more complicated geometries [19]. We note in passing that this setting opens up for extremely efficient coarse-grain solvers based on multigrid techniques when the spatial resolution increases.

In Figure 5 the reduction of the error for each parareal step is displayed. There is a visible trend with better convergence as the system size increases and in line with the analysis in section 4.2, it can also be seen that the error reduction is the largest initially. Despite the rather high level of noise, the solution obtained after a few iterations can hardly be distinguished from the exact one (cf. Figure 6). In Figure 7 we have used instead the homogenized fine scale solver (4.32), and the convergence improves a lot. Again, smoothing the output from the fine solver by integrating over a short period of time has the effect of scaling down the effective Lipschitz constant, and convergence to the homogenized solution becomes fast.
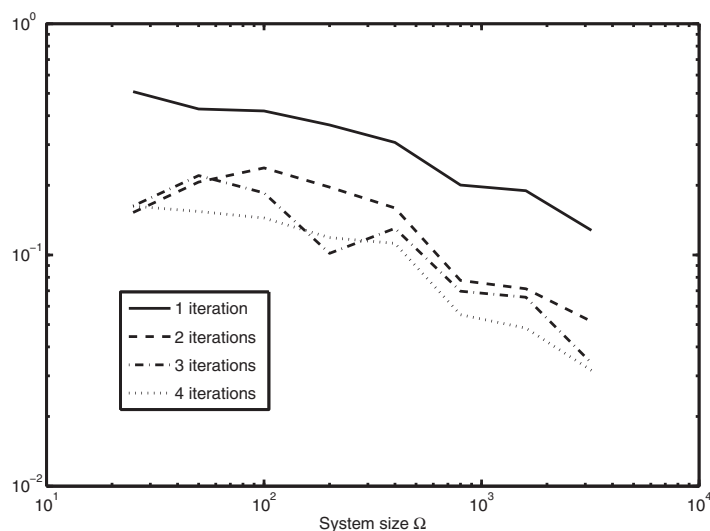
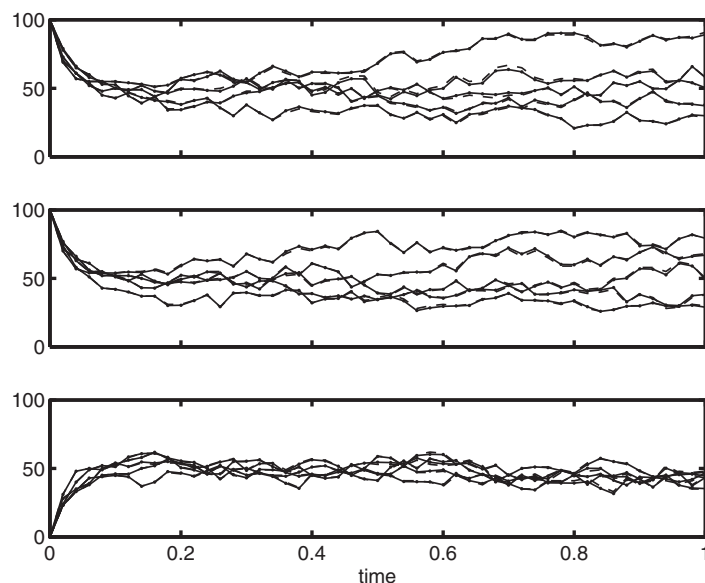FIG. 5. *Reduction of the relative error per the first few parareal iterations when the system size* Ω *increases.*

FIG. 6. *Sample solution for the reaction-diffusion system* (5.5), (5.6), *and* (5.7). *The system-size parameter is for this example* $N_\Omega = 100$ *molecules per cell. Solid: solution obtained after 4 parareal iterations over* 50 *simulated processors. Dashed and barely visible: exact trajectory. From top to bottom are the trajectories for the X-, the Y-, and the Z-molecules, respectively. Towards the end of the simulation there is a strong negative correlation between X- and Y-molecules, and there is also a spatial correlation due to the unsymmetric inflow* (5.7).

6. Conclusions. The master equation, or equivalently, a special type of jump SDEs accurately describes well stirred chemical reactions at the mesoscopic level. Diffusion can be incorporated by a special set of reactions implying that non-homogeneous problems also can be modeled using the same type of process. In the macroscopic limit the usual rate equation (or the reaction-diffusion PDE) emerges.
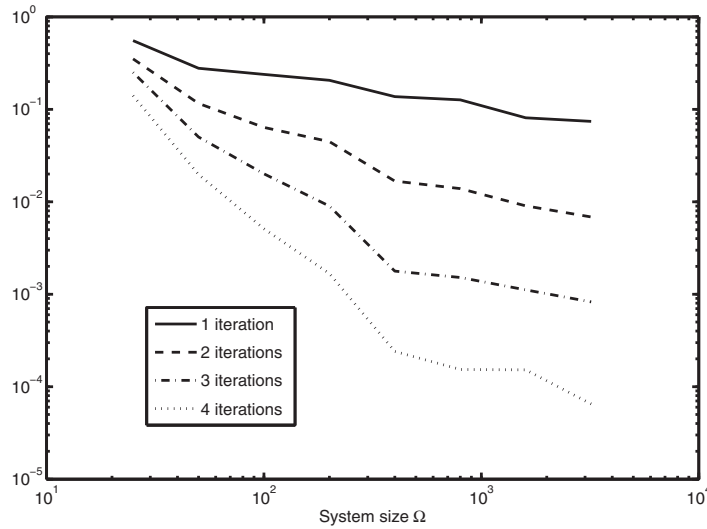
FIG. 7. *As in Figure* 5 *but using a homogenized fine solver* (4.32) *with* $\delta t = \Delta t/4$. *The convergence to the homogenized solution is faster and more regular.*

The parareal algorithm can be applied by using the rate equation as a predictor for the jump process. Convergence is then asymptotically dictated by the Lipschitz constant of the system, but it also improves as the system's size is increased. In the macroscopic limit, convergence in probability is obtained.

For models involving rapid transients, the best one can generally hope for is convergence in moment, but it is then not clear how the error should be monitored. A remedy is to homogenize the model on the fly by averaging the process in time. This homogenization is very general and should be applicable to a wide range of stochastic time-dependent problems. We remark that obtaining this homogenized solution is impractical on a serial computer, since the full solution must be obtained first. Furthermore, the proposed homogenization can be put in contrast to other multiscale methods [10, 15]. Typically, one computes a solution to a homogenized *equation* obtained by determining some kind of averaged coefficients. With the parareal algorithm, one rather directly obtains a homogenized *solution.*

The range of applicability of the method can be increased by adding a white-noise term to the governing equations, thus encompassing highly general Lévy-type SDEs [3]. In this general setting, the suggested homogenization provides a way of applying the parareal algorithm to stiff problems.

We would also like to relate our results to those in reference [6] mentioned in the introduction. There it is shown that the parareal/Euler forward combination for a *nonstiff* Wiener SDE yields a mean square estimate $\mathcal{O}\left(\Delta t^{k/2}\right)$ for the kth parareal iteration assuming a practically exact fine solver. This holds true in the present context as well, provided that we use the appropriate version of the forward Euler method, namely, the tau-leap method [33, 42]. *However*, the resulting method is highly nontrivial, if possible at all, to implement efficiently. The reason is that the same realizations of Poisson processes have to be used for both the coarse and the fine solver. The list of events simulated by the fine solver must somehow be reused and searched through leading to a very expensive coarse solver. It is an open question if one can somehow circumvent this issue. We remark, however, that the multilevel

improvement of the parareal algorithm as analyzed in [6] applies also in the current setting. The idea here is to let the coarse solver again be an instance of the parareal algorithm and with this improvement, parallel efficiency can be close to one.

One of the most promising applications for the proposed method seems to be for reaction-diffusion models such as the one investigated in section 5.4. Here the macroscopic model can typically be acceptable in many, but not all, subvolumes. Some of the species are typically present in fairly large copy numbers where noise is less pronounced. A few iterations of the suggested parareal algorithm can then be understood as a kind of deterministic/stochastic hybrid method in line with the method in [7, 31]. A feature with the parareal setup is that one never needs to explicitly determine what parts should be treated deterministically.

REFERENCES

[1] D. F. ANDERSON, *A modified next reaction method for simulating chemical systems with time dependent propensities and delays*, J. Chem. Phys., 127 (2007), p. 214107, doi:10.1063/1.2799998.

[2] D. F. ANDERSON, *Incorporating postleap checks in tau-leaping*, J. Chem. Phys., 128 (2008), pp. 054103–054111, doi:10.1063/1.2819665.

[3] D. APPLEBAUM, *Lévy Processes and Stochastic Calculus*, Cambridge Stud. Adv. Math. 93, Cambridge University Press, Cambridge, 2004.

[4] L. BAFFICO, S. BERNARD, Y. MADAY, G. TURINICI, AND G. ZÉRAH, *Parallel-in-time molecular-dynamics simulations*, Phys. Rev. E (3), 66 (2002), pp. 057701–057704, doi:10.1103/PhysRevE.66.057701.

[5] G. BAL, *On the convergence and the stability of the parareal algorithm to solve partial differential equations*, in Domain Decomposition Methods in Science and Engineering, Lecture Notes Comput. Sci. Eng. 40, Y. Maday, ed., Springer, Berlin, 2005, pp. 425–432, doi:10.1007/3-540-26825-1_43.

[6] G. BAL, *Parallelization in time of (stochastic) ordinary differential equations*, submitted, http://www.columbia.edu/~gb2030.

[7] K. BALL, T. G. KURTZ, L. POPOVIC, AND G. REMPALA, *Asymptotic analysis of multiscale approximations to reaction networks*, Ann. Appl. Probab., 16 (2006), pp. 1925–1961, doi:10.1214/105051606000000420.

[8] O. G. BERG AND P. H. VON HIPPEL, *Diffusion-controlled macromolecular interactions*, Ann. Rev. Biophys. Chem., 14 (1985), pp. 131–160, doi:10.1146/annurev.bb.14.060185.001023.

[9] K. BURRAGE, *Parallel and Sequential Methods for Ordinary Differential Equations*, in Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 1995.

[10] Y. CAO, D. T. GILLESPIE, AND L. R. PETZOLD, *The slow-scale stochastic simulation algorithm*, J. Chem. Phys., 122 (2005), p. 014116, doi:10.1063/1.1824902.

[11] Y. CAO, H. LI, AND L. PETZOLD, *Efficient formulation of the stochastic simulation algorithm for chemically reacting systems*, J. Chem. Phys., 121 (2004), pp. 4059–4067, doi:10.1063/1.1778376.

[12] Y. Cao, L. R. Petzold, M. Rathinam, and D. T. Gillespie, *The numerical stability of leaping methods for stochastic simulation of chemically reacting systems*, J. Chem. Phys., 121 (2004), pp. 12169–12178, doi:10.1063/1.1823412.

[13] D. R. Cox and V. Isham, *Point Processes*, Monogr. Appl. Probab. Stat., Vol. 12, Chapman and Hall, London, 1980.

[14] M. Dobrzyński, J. V. Rodríguez, J. A. Kaandorp, and J. G. Blom, *Computational methods for diffusion-influenced biochemical reactions*, Bioinformatics, 23 (2007), pp. 1969–1977, doi:10.1093/bioinformatics/btm278.

[15] W. E, D. Liu, and E. Vanden-Eijnden, *Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates*, J. Chem. Phys., 123 (2005), p. 194107, doi:10.1063/1.2109987.

[16] J. Elf and M. Ehrenberg, *Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases*, Syst. Biol., 1 (2004), pp. 230–236, doi:10.1049/sb:20045021.

[17] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Stochastic gene expression in a single cell*, Science, 297 (2002), pp. 1183–1186, doi:10.1126/science.1070919.

[18] S. Engblom, *Time-parallel simulation of stochastic chemical kinetics*, in Numerical Analysis and Applied Mathematics 1048, T. E. Simos, ed., American Institute of Physics, College Park, MD, 2008, pp. 174–177, doi:10.1063/1.2990884.

[19] S. Engblom, L. Ferm, A. Hellander, and P. Lötstedt, *Simulation of stochastic reaction-diffusion processes on unstructured meshes*, SIAM J. Sci. Comput., 31 (2009), pp. 1774–1797, doi:10.1137/080721388.

[20] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley Ser. Probab. Math. Stat., John Wiley & Sons, New York, 1986.

[21] D. Fange and J. Elf, *Noise-induced Min phenotypes in E. coli*, PLoS Comput. Biol., 2 (2006), pp. 637–648, doi:10.1371/journal.pcbi.0020080.

[22] M. J. Gander and S. Vandewalle, *Analysis of the parareal time-parallel time-integration method*, SIAM J. Sci. Comput., 29 (2007), pp. 556–578, doi:10.1137/05064607X.

[23] C. W. Gardiner, *Handbook of Stochastic Methods*, 3rd ed., Springer Ser. Synergetics, Springer, Berlin, 2004.

[24] T. S. Gardner, C. R. Cantor, and J. J. Collins, *Construction of a genetic toggle switch in Escherichia coli*, Nature, 403 (2000), pp. 339–342, doi:10.1038/35002131.

[25] M. A. Gibson and J. Bruck, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem., 104 (2000), pp. 1876–1889, doi:10.1021/jp993732q.

[26] I. I. Gihman and A. V. Skorohod, *Stochastic Differential Equations*, Ergebnisse der Mathematik und ihrer Grenzgebiete 72, Springer, Berlin, 1972.

[27] D. T. Gillespie, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22 (1976), pp. 403–434, doi:10.1016/0021-9991(76)90041-3.

[28] D. T. Gillespie, *A rigorous derivation of the chemical master equation*, Phys. A, 188 (1992), pp. 404–425, doi:10.1016/0378-4371(92)90283-V.

[29] D. T. Gillespie, *Approximate accelerated stochastic simulation of chemically reacting systems*, J. Chem. Phys., 115 (2001), pp. 1716–1733, doi:10.1063/1.1378322.

[30] P. Guptasarama, *Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of Escherichia coli?*, Bioessays, 17 (1995), pp. 987–997, doi:10.1002/bies.950171112.

[31] E. L. Haseltine and J. B. Rawlings, *Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics*, J. Chem. Phys., 117 (2002), pp. 6959–6969, doi:10.1063/1.1505860.

[32] E. Hausenblas, *Error analysis for approximation of stochastic differential equations driven by Poisson random measures*, SIAM J. Numer. Anal., 40 (2002), pp. 87–113, doi:10.1137/S0036142999360275.

[33] D. J. Higham and P. E. Kloeden, *Numerical methods for nonlinear stochastic differential equations with jumps*, Numer. Math., 101 (2005), pp. 101–119, doi:10.1007/s00211-005-0611-8.

[34] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland Math. Library 24, North-Holland/Kodansha, Amsterdam/Tokyo, 1981.

[35] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, Grundlehren der mathematischen Wissenschaften 288, Springer, Berlin, 1987.

[36] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 5th ed., Elsevier, Amsterdam, 2004.

[37] K. KRUSE AND J. ELF, *Kinetics in spatially extended systems*, in System Modeling in Cellular Biology, From Concepts to Nuts and Bolts, Z. Szallasi, J. Stelling, and V. Periwal, eds., MIT Press, Cambridge, MA, 2006, pp. 177–198.

[38] T. G. KURTZ, *Solutions of ordinary differential equations as limits of pure jump Markov processes*, J. Appl. Probab., 7 (1970), pp. 49–58.

[39] T. G. KURTZ, *Limit theorems for sequences of jump Markov processes approximating ordinary differential processes*, J. Appl. Probab., 8 (1971), pp. 344–356.

[40] T. G. KURTZ, *Strong approximation theorems for density dependent Markov chains*, Stochastic Process. Appl., 6 (1978), pp. 223–240, doi:10.1016/0304-4149(78)90020-0.

[41] T. G. KURTZ, *Representations of Markov processes as multiparameter time changes*, Ann. Probab., 8 (1980), pp. 682–715, doi:10.1214/aop/1176994660.

[42] T. LI, *Analysis of explicit tau-leaping schemes for simulating chemically reacting systems*, Multiscale Model. Simul., 6 (2007), pp. 417–436, doi:10.1137/06066792X.

[43] T. LI, A. ABDULLE, AND W. E., *Effectiveness of implicit methods for stiff stochastic differential equations*, Commun. Comput. Phys., 3 (2008), pp. 295–307.

[44] J.-L. LIONS, Y. MADAY, AND G. TURINICI, *Résolution d'EDP par un schéma en temps "pararéel"*, C. R. Acad. Sci. Paris Sér. I Math., 332 (2000), pp. 661–668, doi:10.1016/S0764-4442(00)01793-6.

[45] Y. MADAY, *Parareal in time algorithm for kinetic systems based on model reduction*, in High-Dimensional Partial Differential Equations in Science and Engineering, CRM Proc. Lecture Notes 41, A. Bandrauk, M. C. Delfour, and C. Le Bris, eds., AMS/CRM, Providence, RI, 2007, pp. 183–194.

[46] Y. MADAY AND G. TURINICI, *A parareal in time procedure for the control of partial differential equations*, C. R. Acad. Sci. Paris Sér. I Math., 335 (2002), pp. 387–392, doi:10.1016/S1631-073X(02)02467-6.

[47] H. H. MCADAMS AND A. ARKIN, *It's a noisy business. Genetic regulation at the nanomolar scale*, Trends Gen., 15 (1999), pp. 65–69, doi:10.1016/S0168-9525(98)01659-X.

[48] J. M. MCCOLLUM, G. D. PETERSON, C. D. COX, M. L. SIMPSON, AND N. F. SAMATOVA, *The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior*, Comput. Biol. Chem., 30 (2006), pp. 39–49, doi:10.1016/j.compbiolchem.2005.10.007.

[49] R. METZLER, *The future is noisy: The role of spatial fluctuations in genetic switching*, Phys. Rev. Lett., 87 (2001), pp. 068103–068107, doi:10.1103/PhysRevLett.87.068103.

[50] B. ØKSENDAL, *Stochastic Differential Equations*, 6th ed., Springer, Berlin, 2005.

[51] J. PAULSSON, O. G. BERG, AND M. EHRENBERG, *Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 7148–7153, doi:10.1073/pnas.110057697.

[52] M. F. PETTIGREW AND H. RESAT, *Multinomial tau-leaping method for stochastic kinetic simulations*, J. Chem. Phys., 126 (2007), p. 084101, 2007, doi:10.1063/1.2432326.

[53] S. PLYASUNOV, *On hybrid simulation schemes for stochastic reaction dynamics*, http://arxiv.org/abs/math/0504477.

[54] J. M. RASER AND E. K. O'SHEA, *Noise in gene expression: Origins, consequences and control*, Science, 309 (2005), pp. 2010–2013, doi:10.1126/science.1105891.

[55] M. RATHINAM, L. PETZOLD, Y. CAO, AND D. T. GILLESPIE, *Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method*, J. Chem. Phys., 119 (2003), pp. 12784–12794, doi:10.1063/1.1627296.

[56] M. RATHINAM, L. R. PETZOLD, Y. CAO, AND D. T. GILLESPIE, *Consistency and stability of tau-leaping schemes for chemical reaction systems*, Multiscale Model. Simul., 4 (2005), pp. 867–895, doi:10.1137/040603206.

[57] G. A. STAFF, *The parareal algorithm*, www.idi.ntnu.no/~elster/notur-cluster03.

[58] G. A. STAFF AND E. M. RØNQUIST, *Stability of the parareal algorithm*, in Domain Decomposition Methods in Science and Engineering, Lecture Notes Comput. Sci. Eng. 40, Y. Maday, ed., Springer, Berlin, 2005, pp. 449–456, doi:10.1007/3-540-26825-1_46.

[59] T. TIAN AND K. BURRAGE, *Binomial leap methods for simulating stochastic chemical kinetics*, J. Chem. Phys., 121 (2004), pp. 10356–10364, doi:10.1063/1.1810475.

[60] Y. TOGASHI AND K. KANEKO, *Molecular discreteness in reaction-diffusion systems yields steady states not seen in the continuum limit*, Phys. Rev. E (3), 70 (2004), pp. 020901–1–020901–4, doi:10.1103/PhysRevE.70.020901.

[61] J. M. G. VILAR, H. Y. KUEH, N. BARKAI, AND S. LEIBLER, *Mechanism of noise-resistance in genetic oscillators*, Proc. Natl. Acad. Sci., 99 (2002), pp. 5988–5992, doi:10.1073/pnas.092133899.