



<http://www.diva-portal.org>

This is the published version of a paper published in *SIAM Journal on Numerical Analysis*.

Citation for the original published paper (version of record):

Engblom, S. (2015)

Strong convergence for split-step methods in stochastic jump kinetics.

SIAM Journal on Numerical Analysis, 53: 2655-2676

<http://dx.doi.org/10.1137/141000841>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-239548>

STRONG CONVERGENCE FOR SPLIT-STEP METHODS IN STOCHASTIC JUMP KINETICS*

STEFAN ENGBLOM†

Abstract. Mesoscopic models in the reaction-diffusion framework have gained recognition as a viable approach to describing chemical processes in cell biology. The resulting computational problem is a continuous-time Markov chain on a discrete and typically very large state space. Due to the many temporal and spatial scales involved many different types of computationally more effective multiscale models have been proposed, typically coupling different types of descriptions within the Markov chain framework. In this work we look at the strong convergence properties of the basic first order Strang, or Lie–Trotter, split-step method, which is formed by decoupling the dynamics in finite time steps. Thanks to its simplicity and flexibility, this approach has been tried in many different combinations. We develop explicit sufficient conditions for pathwise well-posedness and convergence of the method, including error estimates, and we illustrate our findings with numerical examples. In doing so, we also suggest a certain partition of unity representation for the split-step method, which in turn implies a concrete simulation algorithm under which trajectories may be compared in a pathwise sense.

Key words. operator splitting, partition of unity, Lie–Trotter formula, continuous-time Markov chain, jump process, rate equation

AMS subject classifications. Primary, 65C40, 60H35; Secondary, 60J28, 92C45

DOI. 10.1137/141000841

1. Introduction. Since their introduction by Gillespie [21, 22], stochastic models of chemical reactions have become ubiquitous tools in describing the kinetics of living cells. Since complete molecular dynamics–type descriptions of most biochemical processes are either impractical or out of reach for complexity reasons, stochastic models have remained popular as a viable alternative. Formulated in a way which resembles the macroscopic viewpoint, but with randomness taking certain microscopic effects into account, *mesoscopic* stochastic models attempt to strike a balance between computational feasibility and accuracy. In fact, a common theme in several studies is the discrepancy between deterministic and stochastic descriptions [6, 31, 38].

Due to the presence of multiple scales in species abundance and in reaction rates, the computational problem of simulating well-stirred or spatially extended models has caught a lot of attention. For example, these features are the driving motivation behind the development of *hybrid methods* [5, 25] and the various kinds of *model reduction* techniques that have been proposed [12, 14, 17, 24]. Similarly, more efficient time discretization “tau-leap” methods were proposed early on [23], and has since then been modified and analyzed in various ways [1, 30, 34].

As a means to facilitate multiscale and multiphysics coupling in the method’s development in general, *split-step* methods have a long story. Originally developed via (finite-dimensional) operator splitting [37], in the present case these methods became particularly important in the more computationally demanding spatial stochastic reaction-diffusion setting [16]. An analysis in the sense of convergence in distribu-

*Received by the editors December 19, 2014; accepted for publication (in revised form) October 7, 2015; published electronically December 10, 2015. The research of the author was supported by the Swedish Research Council within the UPMARC Linnaeus center of Excellence.

<http://www.siam.org/journals/sinum/53-6/100084.html>

†Division of Scientific Computing, Department of Information Technology, Uppsala University, SE-751 05 Uppsala, Sweden (stefane@it.uu.se).

tion in the master equation setting was presented in [27]. A practical method based on splitting to simulate fractional diffusion was reported in [8], and an adaptive reaction-diffusion simulator was suggested in [26]. Finally, in [4], the splitting technique was used to bring out parallelism from an otherwise strictly serial description.

In this work we look at the strong pathwise convergence of the basic first order (in the operator sense) split-step method. A key issue here is to devise a meaningful coupling between different trajectories conditioned on different split-step discretizations. We solve this by using a partition of unity representation which as a by-product also implies a practical algorithm. Sufficient conditions for strong convergence of order $1/2$ that apply notably also to *open* chemical systems are described and this is also confirmed in our numerical experiments. An interesting observation is that, although still only formally of strong order $1/2$, the second order (again in the operator sense) Strang splitting performs considerably better than the first order splitting.

In section 2 we recapitulate the description of chemical processes as continuous-time Markov chains, in the nonspatial as well as in the spatially extended case. Our main theoretical findings, including explicit conditions for strong convergence, are reported in section 3. Numerical illustrations are presented in section 4, and a concluding discussion is found in section 5.

2. Stochastic jump kinetics. We summarize in this section the mathematical background required in the description of biochemical processes. A recapitulation of the traditional well-stirred setting is found in section 2.1. Pathwise descriptions are found in section 2.2, where some fundamental tools from stochastic analysis are also reviewed. Finally, in section 2.3 the required extensions to encompass also spatially extended models are indicated.

2.1. Well-stirred Markovian reactions. In a memory-less Markovian chemical system, at any instant t , the *state* is an integer vector $X(t) \in \mathbf{Z}_+^D$ counting the number of molecules of each of D species. The *reactions* are prescribed transitions of the state according to an intensity law, or *reaction propensity*:

$$(2.1) \quad w_r : \mathbf{Z}_+^D \rightarrow \mathbf{R}_+,$$

$$(2.2) \quad \mathbf{P}[X(t+dt) = x - \mathbb{N}_r | X(t) = x] = w_r(x) dt + o(dt).$$

The system is thus fully described by the pair $[\mathbb{N}, w(x)]$, that is, the *stoichiometric matrix* $\mathbb{N} \in \mathbf{Z}^{D \times R}$, and $w(x) \equiv [w_1(x), \dots, w_R(x)]^T$, the column vector of reaction propensities. An important remark is that useful physical descriptions are always *conservative*, for all propensities it holds that $w_r(x) = 0$ whenever $x - \mathbb{N}_r \notin \mathbf{Z}_+^D$ [10, Chap. 8.2.2, Definition 2.4].

The *chemical master equation* [28, Chap. V], or *Kolmogorov's forward differential system* [10, Chap. 8.3] governs the law of the state $X(t)$ conditioned on some initial state. Put $p(x, t) = \mathbf{P}(X(t) = x | X(0) = x_0)$. Then

$$(2.3) \quad \frac{\partial p(x, t)}{\partial t} = \sum_{r=1}^R w_r(x + \mathbb{N}_r) p(x + \mathbb{N}_r, t) - w_r(x) p(x, t) =: \mathbb{M}^T p(x, t),$$

where \mathbb{M} is the *infinitesimal generator* of the process.

2.2. Pathwise representations. The use of pathwise representations in the analysis of Markov processes on discrete state spaces in continuous time was pioneered by Kurtz in a series of paper (see [18] and the references therein). We thus postulate the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where the filtration $\mathcal{F}_{t \geq 0}$ contains R -dimensional

Poisson processes. The transition law (2.2) implies a certain counting process which can be constructed from a standard unit-rate Poisson process Π_r . The state $X(t)$ can then be written

$$(2.4) \quad X_t = X_0 - \sum_{r=1}^R \mathbb{N}_r \Pi_r \left(\int_0^t w_r(X_{s-}) ds \right).$$

This is Kurtz’s *random time change representation* [18, Chap. 6.2] which gives rise to the notion of *operational time* in the argument to each of the R independent Poisson processes. Note that, in (2.4), by $X(t-)$ is meant the state before any transitions at time t .

It is sometimes convenient to use an equivalent construction in terms of a random counting measure [9, Chap. VIII]. We denote by $\mu_r(dt) = \mu_r(w_r(X_{t-}) dt; \omega)$ for $\omega \in \Omega$ the random measure associated with the counting process whose intensity at any instant t is $w_r(X_{t-})$. Thus with deterministic intensity $E[\mu_r(dt)] = E[w_r(X_{t-}) dt]$, this defines an increasing sequence of exponentially distributed counts $\tau_i \in \mathbf{R}_+$. With $\boldsymbol{\mu} = [\mu_1, \dots, \mu_R]^T$ we can now write (2.4) in the compact differential form

$$(2.5) \quad dX_t = -\mathbb{N}\boldsymbol{\mu}(dt).$$

For realistic chemical systems the number of molecules must somehow be bound a priori. We encapsulate this property by requiring the existence of a certain weighted norm

$$(2.6) \quad \|x\|_{\mathbf{l}} := \mathbf{l}^T x, \quad x \in \mathbf{Z}_+^D,$$

normalized such that $\min_i \mathbf{l}_i = 1$. Equipped with this norm we formulate, following [15] closely (see also [11, 33]) the assumption below.

Assumption 2.1. For arguments $x, y \in \mathbf{Z}_+^D$ we assume that

- (i) $-\mathbf{l}^T \mathbb{N}w(x) \leq A + \alpha \|x\|_{\mathbf{l}}$,
- (ii) $(-\mathbf{l}^T \mathbb{N})^2 w(x)/2 \leq B + \beta_1 \|x\|_{\mathbf{l}} + \beta_2 \|x\|_{\mathbf{l}}^2$,
- (iii) $|w_r(x) - w_r(y)| \leq L_r(P) \|x - y\|$, for $r = 1, \dots, R$, and $\|x\|_{\mathbf{l}} \vee \|y\|_{\mathbf{l}} \leq P$.

With the exception of α , all parameters $\{A, B, \beta_1, \beta_2, L\}$ are assumed to be non-negative.

In order to state an a priori result concerning the regularity of the solutions to (2.5), following [36, section 3.1.2], we define the following family of spaces of pathwise locally bounded processes:

$$(2.7) \quad S_{\mathcal{F}}^{p,\text{loc}}(\mathbf{Z}_+^D) = \left\{ X(t, \omega) : \begin{array}{l} X_t \in \mathbf{Z}_+^D \text{ is } \mathcal{F}_t\text{-adapted such that} \\ E \sup_{t \in [0, T]} \|X_t\|_{\mathbf{l}}^p < \infty \text{ for } \forall T < \infty \end{array} \right\}.$$

THEOREM 2.2 (Theorem 4.7 in [15]). *Let X_t be a solution to (2.5) under Assumption 2.1(i) and (ii) with $\beta_2 = 0$. Then if $\|X_0\|_{\mathbf{l}}^p < \infty$, $\{X_t\}_{t \geq 0} \in S_{\mathcal{F}}^{p,\text{loc}}(\mathbf{Z}_+^D)$. If $\beta_2 > 0$ then the conclusion remains under the additional requirement that $\|X_0\|_{\mathbf{l}}^{p+1} < \infty$.*

Below we will frequently use the stopping time

$$(2.8) \quad \tau_P := \inf_{t \geq 0} \{\|X_t\|_{\mathbf{l}} > P\}$$

and put $\hat{t} = t \wedge \tau_P$ for some finite t defining an interval of interest. As an example, a differential form of Itô’s change of variables formula can be derived formally by simply

summing over jump times [3, Chap. 4.4.2],

$$(2.9) \quad df(X_t) = \sum_{r=1}^R f(X_{t-} - \mathbb{N}_r) - f(X_{t-}) \mu_r(dt).$$

More carefully, *Dynkin's formula* for the stopped process is then given by [10, Chap. 9.2.2],

$$(2.10) \quad \mathbb{E} f(X_{\hat{t}}) - \mathbb{E} f(X_0) = \int_0^{\hat{t}} \sum_{r=1}^R \mathbb{E} [(f(X_s - \mathbb{N}_r) - f(X_s))w_r(X_s)] ds.$$

In order to efficiently work with the Poisson representation (2.4) in the sense of mean square, the following two lemmas which follow [13] very closely will be critical.

LEMMA 2.3. *Let Π be a unit-rate Poisson process and T a bounded stopping time, both adapted to \mathcal{F}_t . Then*

$$(2.11) \quad \mathbb{E}[\Pi(T)] = \mathbb{E}[T],$$

$$(2.12) \quad \mathbb{E}[\Pi^2(T)] = 2 \mathbb{E}[\Pi(T)T] - \mathbb{E}[T^2] + \mathbb{E}[T].$$

Proof. Let $\tilde{\Pi}(t) := \Pi(t) - t$ be the compensated process. This is a martingale and Doob's optional sampling theorem implies $\mathbb{E}[\tilde{\Pi}(T)] = 0$ [32, Theorem 17, Chap. I.2], which is (2.11). Similarly $Z(t) := \tilde{\Pi}^2(t) - t$ is a martingale [32, Theorem 24, Chap. I.3] and the sampling theorem yields $\mathbb{E}[Z(T)] = 0$, or,

$$0 = \mathbb{E}[\Pi^2(T) - 2\Pi(T)T + T^2 - T],$$

which is (2.12). \square

LEMMA 2.4. *Let Π be a unit-rate Poisson process and T_1, T_2 bounded stopping times, all adapted to \mathcal{F}_t . Then*

$$(2.13) \quad \mathbb{E}[|\Pi(T_2) - \Pi(T_1)|] = \mathbb{E}[|T_2 - T_1|],$$

$$(2.14) \quad \mathbb{E}[(\Pi(T_2) - \Pi(T_1))^2] = 2 \mathbb{E}[(\Pi(T_2) - \Pi(T_1))(T_1 \vee T_2)] - \mathbb{E}[|T_2^2 - T_1^2|] + \mathbb{E}[|T_2 - T_1|].$$

The formulation (2.13) was recently used in [19] to provide for a related analysis in the sense of convergence in mean.

Proof. Assume first that $T_2 \geq T_1$. By Lemma 2.3 (2.11),

$$\mathbb{E}[\Pi(T_2) - \Pi(T_1)] = \mathbb{E}[T_2 - T_1].$$

For general stopping times, say S_1, S_2 , (2.13) follows upon substituting $T_1 := S_1 \wedge S_2$ and $T_2 := S_1 \vee S_2$ into this equality.

Next put $X := \mathbb{E}[(\Pi(T_2) - \Pi(T_1))^2]$ and assume anew that $T_2 \geq T_1$. We find

$$\begin{aligned} X &= \mathbb{E}[\Pi(T_2)^2 + \Pi(T_1)^2 - 2\Pi(T_1)\Pi(T_2)] \\ &= \mathbb{E}[\Pi(T_2)^2 + \Pi(T_1)^2] - 2 \mathbb{E}[\Pi(T_1) \mathbb{E}[\Pi(T_2)|\mathcal{F}_{T_1}]]. \end{aligned}$$

To evaluate the iterated expectation, note that

$$\mathbb{E}[\tilde{\Pi}(T_2)|\mathcal{F}_{T_1}] = \tilde{\Pi}(T_1) \implies \mathbb{E}[\Pi(T_2)|\mathcal{F}_{T_1}] = \Pi(T_1) - T_1 + \mathbb{E}[T_2|\mathcal{F}_{T_1}].$$

Hence

$$E[\Pi(T_1) E[\Pi(T_2)|\mathcal{F}_{T_1}]] = E[\Pi(T_1)^2] - E[\Pi(T_1)T_1] + E[\Pi(T_1)T_2],$$

and so

$$X = E[\Pi(T_2)^2 - \Pi(T_1)^2] + 2 E[\Pi(T_1)T_1] - 2 E[\Pi(T_1)T_2].$$

Applying Lemma 2.3 (2.12) twice then yields

$$X = 2 E[(\Pi(T_2) - \Pi(T_1))T_2] - E[T_2^2 - T_1^2] + E[T_2 - T_1].$$

As before the substitutions $T_1 := S_1 \wedge S_2$ and $T_2 := S_1 \vee S_2$ imply (2.14) for general stopping times S_1, S_2 . \square

Lemma 2.4 will be applied as follows. Assuming first the a priori bound $T_1 \vee T_2 \leq B$ we get from (2.13)–(2.14) that

$$(2.15) \quad E[(\Pi(T_2) - \Pi(T_1))^2] \leq (2B + 1) E[|T_2 - T_1|].$$

Let \mathcal{F}_t be the filtration adapted to $\{\tilde{\Pi}_r\}_{r=1}^R$. Then for any fixed time t , $T_r(t) := \int_0^t w_r(X(s)) ds$ is a stopping time adapted to [2, Lemma 3.1]

$$\tilde{\mathcal{F}}_u^r := \sigma\{\Pi_r(s), s \in [0, u]; \Pi_{k \neq r}(s), s \in [0, \infty]\}.$$

Intuitively, since $X(t) = \sum_r \Pi_r(T_r(t))N_r$, the event $\{T_r(t) < u\}$ depends on Π_r during $[0, u]$ and on $\{\Pi_k, k = 1 \dots R, k \neq r\}$ during $[0, \infty)$. However, since the $\{\Pi_r\}_{r=1}^R$ are all independent, $\tilde{\Pi}_r$ is still a martingale with respect to $\tilde{\mathcal{F}}_u^r$ (and not only with respect to $\mathcal{F}_u^r = \sigma\{\Pi_r(s), s \in [0, u]\}$). Hence we can apply the stopping time theorems to $T_r(t)$ and the previous lemmas apply.

For an approximating process, say $\tilde{X} \approx X$, assuming a suitable random time representation in the form of (2.4) is available, these results will remain valid for \tilde{X} as well. To conclude, given the bound

$$(2.16) \quad \int_0^t w_r(X(s)) ds \vee \int_0^t \tilde{w}_r(\tilde{X}(s)) ds \leq B,$$

we get from (2.15) that

$$(2.17) \quad E \left[\left(\Pi_r \left(\int_0^t w_r(X(s)) ds \right) - \Pi_r \left(\int_0^t \tilde{w}_r(\tilde{X}(s)) ds \right) \right)^2 \right] \leq (2B + 1) E \left[\left| \int_0^t w_r(X(s)) ds - \int_0^t \tilde{w}_r(\tilde{X}(s)) ds \right| \right].$$

2.3. Incorporating spatial dependence. Well-stirred modeling of chemical kinetics relies on *homogeneity*, that is, that the probability of finding a molecule is equal throughout the volume. There are many situations of interest where this assumption is violated, for instance, when slow molecular transport allows concentration gradients to build up. A way to approach this situation is through compartmentalization techniques [28, Chap. XIV], which lead to models with a very large number of generalized reaction channels. Since split-step methods are a particularly promising computational technique here, we briefly review this framework.

The basic premise is that, although the full volume V is not well-stirred, it can be subdivided into smaller cells V_j such that their individual volume $|V_j|$ is small enough that diffusion suffices to make each cell practically well-stirred.

The state of the system is thus now an array $\mathbf{X} \in \mathbf{Z}_+^{D \times K}$ consisting of D chemically active species \mathbf{X}_{ij} , $i = 1, \dots, D$, in K cells, $j = 1, \dots, K$. This state is changed by chemical reactions occurring in each cell (vertically in \mathbf{X}) and by diffusion where molecules move to adjacent cells (horizontally in \mathbf{X}).

Since each cell is assumed to be well-stirred, (2.5) governs the reaction dynamics

$$(2.18) \quad d\mathbf{X}_t = -\mathbb{N}\boldsymbol{\mu}(dt),$$

where $\boldsymbol{\mu}$ is now R -by- K with $E[\boldsymbol{\mu}_{rj}] = E[w_{rj}(\mathbf{X}_{(\cdot,j)}(t-)) dt]$. Transport of a molecule from V_k to V_j can also be thought of as a special kind of reaction,

$$(2.19) \quad \mathbf{X}_{ik} \xrightarrow{q_{kji}\mathbf{X}_{ik}} \mathbf{X}_{ij},$$

where the rate constant q_{kji} is nonzero only for connected cells. In practice, for any given spatial discretization, numerical methods may be used to define the diffusion rates consistently [16]. We obtain from (2.19) the mesoscopic diffusion model

$$(2.20) \quad d\mathbf{X}_t = \mathbb{S} \otimes (-\boldsymbol{\nu}^T + \boldsymbol{\nu})(dt),$$

where $\mathbb{S} \in \mathbf{Z}^{1 \times K}$ of all 1's, where $\boldsymbol{\nu}$ is K -by- K -by- D with $E[\boldsymbol{\nu}_{kji}] = E[q_{kji}\mathbf{X}_{ik}(t-)] dt$, and where the array operations are suitably defined. In (2.20), note how diffusion *exit* events are paired with *entry* events via the terms $-\boldsymbol{\nu}^T$ and $\boldsymbol{\nu}$, respectively.

By superposition of (2.18) and (2.20) we arrive at the reaction-diffusion model

$$(2.21) \quad d\mathbf{X}_t = -\mathbb{N}\boldsymbol{\mu}(dt) + \mathbb{S} \otimes (-\boldsymbol{\nu}^T + \boldsymbol{\nu})(dt).$$

As was already noted in [16], part of the interest in split-step methods comes from simulating reactions and diffusions by different methods. For simplicity, in the rest of the paper we shall take the well-stirred case (2.5) as our target of study. In doing so we keep in mind that the reaction-diffusion (or reaction-transport) case (2.21) does fall under the same general class of descriptions.

3. Analysis of split-step methods. In this section we present our main theoretical findings. The splitting we choose to analyze is defined in the master equation setting in section 3.1. In order to couple trajectories and obtain pathwise comparisons, the splitting is redefined in the operational time framework in section 3.2, where some a priori estimates are also derived. After developing a few further preliminary results in section 3.3, the theory is put together in section 3.4, where our main convergence result is presented.

Throughout this section we let C denote a positive constant which may be different at each occurrence.

3.1. Operator splitting and the master equation. While we take another approach below, traditionally, split-step methods are constructed via operator splitting of the master equation (2.3). Assume the split into two sets of reaction pathways can be written as

$$(3.1) \quad \mathbb{N} = \begin{bmatrix} \mathbb{N}^{(1)} & \mathbb{N}^{(2)} \end{bmatrix},$$

$$(3.2) \quad w(x) = \begin{bmatrix} w^{(1)}(x); w^{(2)}(x) \end{bmatrix},$$

where $\mathbb{N}^{(i)}$ is D -by- R_i , $i \in \{1, 2\}$, $R_1 + R_2 = R$, and where the propensity column vectors have the corresponding dimensions. The simplest possible split-step method, and the one we choose to analyze in this paper, can then be written in integral form

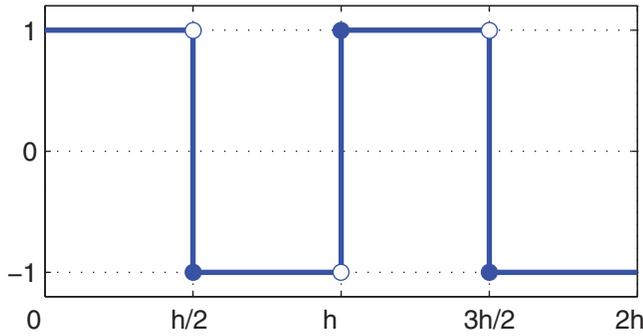


FIG. 1. Definition of the piecewise constant càdlàg kernel function $\sigma_h(x)$.

as (compare (2.3))

$$(3.3) \quad \tilde{p}_h(x, t + h) = p_h(x, t) + \int_t^{t+h} \sum_{r \in \mathcal{R}_1} w_r(x + \mathbb{N}_r) \tilde{p}_h(x + \mathbb{N}_r, s) - w_r(x) \tilde{p}_h(x, s) ds,$$

$$(3.4) \quad p_h(x, t + h) = \tilde{p}_h(x, t + h) + \int_t^{t+h} \sum_{r \in \mathcal{R}_2} w_r(x + \mathbb{N}_r) p_h(x + \mathbb{N}_r, s) - w_r(x) p_h(x, s) ds,$$

where $\mathcal{R}_1 = \{1, \dots, R_1\}$, $\mathcal{R}_2 = \{R_1 + 1, \dots, R\}$. Loosely speaking, (3.3) evolves the dynamics of the first set of reactions in an auxiliary variable \tilde{p}_h from time t to $t + h$, and (3.4) similarly evolves the second.

3.2. Splitting in operational time. To obtain a concrete pathwise formulation which is more amenable to analysis we first define the kernel step function

$$(3.5) \quad \sigma_h(t) = 1 - 2 \left(\lfloor t/(h/2) \rfloor \bmod 2 \right),$$

for convenience also visualized in Figure 1. This is a piecewise constant càdlàg function which may be used to introduce “switching” events into the process that does not affect the state but turns on or off selected parts of the dynamics. More precisely, the split-step method (3.3)–(3.4) for (2.4) can be written in the operational time form

$$(3.6) \quad Y_t = Y_0 - \sum_{r \in \mathcal{R}_1} \mathbb{N}_r \Pi_r \left(\int_0^t (1 + \sigma_h(s)) w_r(Y_{s-}) ds \right) - \sum_{r \in \mathcal{R}_2} \mathbb{N}_r \Pi_r \left(\int_0^t (1 - \sigma_h(s)) w_r(Y_{s-}) ds \right).$$

For convenience here and below we shall suppress the dependency on the split-step length h ; we simply write $Y(t)$ (or Y_t) instead of $Y_h(t)$.

Equation (3.6) is a partition of unity representation in that, at any instant in (3.6), one of the sets of reactions is turned off while the other operates at twice the intensity. Since the length of each interval where the same set of reactions is active is $h/2$, effectively the unit time for those channels is evolved in steps of length h , in agreement with (3.3)–(3.4). In section 4.3.3 below we show that the same type of representation may be used when analyzing also the second order Strang split method.

The main advantage with (3.6) over (3.3)–(3.4) is that the former may be compared pathwise to (2.4). Indeed, the convergence results in section 3.4 concern the behavior of $E \|(X - Y)(t)\|^2$ as the split step $h \rightarrow 0$, where $X(t), Y(t)$ are solutions to (2.4) and (3.6), respectively. The approach to coupling processes via the random time change representation was first used by Kurtz [29] and practically implies the *common reaction path* method for simulating coupled processes [7, 35] (see also section 4.1 below).

Assumption 3.1. In the following, our working assumptions will be that Assumption 2.1 holds for *both* subsystems $[\mathbb{N}^{(i)}, w^{(i)}(x)]$, $i \in \{1, 2\}$, in (3.1)–(3.2) and with *the same* weight vector \mathbf{l} . We separate the constants of the two subsystems by using superscripts, as in $A^{(i)}$, $i \in \{1, 2\}$, and additionally define $A^{(0)} := A^{(1)} \vee A^{(2)}$.

The assumption that the weight vector \mathbf{l} is the same for both subsystems as well as for the original description (2.4) is mainly for convenience as it avoids switching back and forth between equivalent norms. A further comment is that, in view of a finite time step h it makes sense to require that both subsystems are well-posed in the sense of Theorem 2.2. However, one can rightly ask if this is really necessary as $h \rightarrow 0$; for h small enough finite-time explosions are likely not going to be a problem. On balance we chose to settle with the current *sufficient* conditions, as a complete theory likely must contain several special cases (for an illustration, see the numerical example in section 4.5).

THEOREM 3.2 (moment bound). *Let $Y(t)$ satisfy (3.6) under Assumption 3.1. Then for any integer $p \geq 1$,*

$$(3.7) \quad E \|Y_t\|_{\mathbf{l}}^p \leq (\|Y_0\|_{\mathbf{l}}^p + 1) \exp(Ct) - 1,$$

with $C > 0$ a constant which depends on p and on the relevant constants of the assumptions, but not on the split step h .

It will be convenient to quote the following basic inequality.

LEMMA 3.3 (Lemma 4.6 in [15]). *Let $H(x) \equiv (x + y)^p - x^p$ with $x \in \mathbf{R}_+$ and $y \in \mathbf{R}$. Then for integer $p \geq 1$ we have the bounds*

$$(3.8) \quad H(x) \leq pyx^{p-1} + 2^{p-4}p(p-1)y^2 [x^{p-2} + |y|^{p-2}],$$

$$(3.9) \quad |H(x)| \leq p|y|2^{p-2} [x^{p-1} + |y|^{p-1}].$$

Proof of Theorem 3.2. The starting point is Dynkin’s formula (2.10) under the stopping time on $\|Y_t\|_{\mathbf{l}}$ defined in (2.8). With $f(x) = \|x\|_{\mathbf{l}}^p = [\mathbf{l}^T x]^p$ we get

$$(3.10) \quad E \|Y_{\hat{t}}\|_{\mathbf{l}}^p = \|Y_0\|_{\mathbf{l}}^p + E \int_0^{\hat{t}} (1 + \sigma_h(s)) \underbrace{\sum_{r \in \mathcal{R}_1} w_r(Y_s) \left[[\mathbf{l}^T (Y_s - \mathbb{N}_r)]^p - [\mathbf{l}^T Y_s]^p \right]}_{=: G_1(Y_s)} ds + E \int_0^{\hat{t}} (1 - \sigma_h(s)) \underbrace{\sum_{r \in \mathcal{R}_2} w_r(Y_s) \left[[\mathbf{l}^T (Y_s - \mathbb{N}_r)]^p - [\mathbf{l}^T Y_s]^p \right]}_{=: G_2(Y_s)} ds.$$

Using the assumptions on the first subsystem $[\mathbb{N}^{(1)}, w^{(1)}(x)]$ and the first part of

Lemma 3.3 we obtain the bound

$$\begin{aligned} G_1(y) &\leq p(A^{(1)} + \alpha^{(1)} \|y\|_{\mathbf{l}}) \|y\|_{\mathbf{l}}^{p-1} \\ &\quad + 2^{p-3}p(p-1)(B^{(1)} + \beta_1^{(1)} \|y\|_{\mathbf{l}} + \beta_2^{(1)} \|y\|_{\mathbf{l}}^2)(\|y\|_{\mathbf{l}}^{p-2} + \delta^{p-2}) \\ &\leq \frac{C}{2} (1 + \|y\|_{\mathbf{l}}^p), \end{aligned}$$

say, in which $\delta := \|\mathbf{l}^T \mathbb{N}^{(1)}\|_{\infty}$ and where Young’s inequality was used several times to arrive at the second bound. A similar bound is readily found for G_2 so summing up from (3.10) we get

$$(3.11) \quad \mathbb{E} \|Y_{\hat{t}}\|_{\mathbf{l}}^p \leq \|Y_0\|_{\mathbf{l}}^p + \int_0^{\hat{t}} C(1 + \mathbb{E} \|Y_s\|_{\mathbf{l}}^p) ds,$$

where $\hat{s} = s \wedge \tau_P$. By Grönwall’s inequality this implies the bound

$$(3.12) \quad \mathbb{E} \|Y_{\hat{t}}\|_{\mathbf{l}}^p \leq (\|Y_0\|_{\mathbf{l}}^p + 1) \exp(Ct) - 1,$$

which is independent of P . We therefore arrive at (3.7) by letting $P \rightarrow \infty$ and using Fatou’s lemma. \square

Recall that the quadratic variation of a process $(X_t)_{t \geq 0}$ in \mathbf{R}^D can be defined by (convergence in probability)

$$(3.13) \quad [X]_t = \lim_{\|M\| \rightarrow 0} \sum_{k \in M} \|X_{t_{k+1}} - X_{t_k}\|^2,$$

where the mesh $M = \{0 = t_0 < t_1 < \dots < t_n = t\}$ and where $\|M\| = \max_k |t_{k+1} - t_k|$. Similarly, we define for later use also the total variation

$$(3.14) \quad V_{[0,t]}(X) = \lim_{\|M\| \rightarrow 0} \sum_{k \in M} \|X_{t_{k+1}} - X_{t_k}\|.$$

LEMMA 3.4. *Let $Y(t)$ satisfy (3.6) under Assumption 3.1. Then the quadratic variation of $\|Y_t\|_{\mathbf{l}}^p$ is bounded by*

$$(3.15) \quad \mathbb{E} [\|Y\|_{\mathbf{l}}^p]_t^{1/2} \leq \mathbb{E} \int_0^t C(1 + \|Y_s\|_{\mathbf{l}}^p + \beta_2^{(0)} \|Y_s\|_{\mathbf{l}}^{p+1}) ds,$$

where $C > 0$ again is independent of the split step h and where $\beta_2^{(0)} := \beta_2^{(1)} \vee \beta_2^{(2)}$.

Proof. Instead of as in (3.10), for brevity we shall use the following compact notation for sums involved in the two subsystems which form the split-step method:

$$(3.16) \quad \sum_{r \in \mathcal{R}_1, \mathcal{R}_2} (1 \pm \sigma_h(s))F(r) := \sum_{r \in \mathcal{R}_1} (1 + \sigma_h(s))F(r) + \sum_{r \in \mathcal{R}_2} (1 - \sigma_h(s))F(r).$$

Keeping this in mind we have

$$(3.17) \quad \mathbb{E} [\|Y\|_{\mathbf{l}}^p]_{\hat{t}}^{1/2} = \mathbb{E} \left[\left(\int_0^{\hat{t}} \sum_{r \in \mathcal{R}_1, \mathcal{R}_2} \left([\mathbf{l}^T (Y_s - \mathbb{N}_r)]^p - [\mathbf{l}^T Y_s]^p \right)^2 \mu_r(ds) \right)^{1/2} \right].$$

Writing $\mu_r(dt) = (1 \pm \sigma_h(t))w_r(Y_{t-}) dt + \tilde{\mu}_r(dt)$, and from the inequality $\|\cdot\| \leq \|\cdot\|_1$ we get after using that the random measure compensated with the deterministic intensity is a local martingale,

$$(3.18) \quad \leq \mathbb{E} \left[\int_0^{\hat{t}} \sum_{r \in \mathcal{R}_1, \mathcal{R}_2} (1 \pm \sigma_h(s))w_r(Y_s) \left| \left[\mathbf{l}^T(Y_s - \mathbb{N}_r) \right]^p - \left[\mathbf{l}^T Y_s \right]^p \right| ds \right],$$

or, after using Lemma 3.3 (3.9),

$$(3.19) \quad \leq \mathbb{E} \left[\int_0^{\hat{t}} \sum_r (1 \pm \sigma_h(s)) p |\mathbf{l}^T \mathbb{N}_r| w_r(Y_s) 2^{p-2} \left[\|Y_s\|_{\mathbf{l}}^{p-1} + |\mathbf{l}^T \mathbb{N}_r|^{p-1} \right] ds \right] \\ \leq \mathbb{E} \left[\int_0^{\hat{t}} C(B^{(0)} + \beta_1^{(0)} \|Y_s\|_{\mathbf{l}} + \beta_2^{(0)} \|Y_s\|_{\mathbf{l}}^2) \left[\|Y_s\|_{\mathbf{l}}^{p-1} + \delta^{p-1} \right] ds \right]$$

by Assumption 2.1(ii), where $\delta = \|\mathbf{l}^T \mathbb{N}\|_{\infty}$. Using Theorem 3.2 and letting $P \rightarrow \infty$ we arrive at the stated bound. \square

THEOREM 3.5. *Let Y_t satisfy (3.6) under Assumption 3.1 with $0 = \beta_2^{(0)} := \beta_2^{(1)} \vee \beta_2^{(2)}$. Then if $\|Y_0\|_{\mathbf{l}}^p < \infty$, $\{Y_t\}_{t \geq 0} \in S_{\mathcal{F}}^{p, \text{loc}}(\mathbf{Z}_+^D)$ for all $h > 0$. If $\beta_2^{(0)} > 0$, then the conclusion remains under the additional requirement that $\|Y_0\|_{\mathbf{l}}^{p+1} < \infty$.*

Note that the somewhat technical details concerning the case $\beta_2^{(0)} \neq 0$ are shared with the solution of (2.5) itself; see Theorem 2.2.

Proof. This result follows essentially by combining Theorem 3.2 and Lemma 3.4. We get, under the same stopping time as before,

$$\|Y_{\hat{t}}\|_{\mathbf{l}}^p = \|Y_0\|_{\mathbf{l}}^p + \int_0^{\hat{t}} (1 + \sigma_h(s))G_1(Y_s) + (1 - \sigma_h(s))G_2(Y_s) ds + M_{\hat{t}}$$

with G_1 and G_2 defined in (3.10). The quadratic variation of the local martingale $M_{\hat{t}}$ can be estimated via Lemma 3.4,

$$(3.20) \quad \mathbb{E} [M_{\hat{t}}]^{1/2} \leq \mathbb{E} \int_0^{\hat{t}} C(1 + \|Y_s\|_{\mathbf{l}}^p + \beta_2^{(0)} \|Y_s\|_{\mathbf{l}}^{p+1}) ds.$$

The case $\beta_2^{(0)} = 0$. Using the bound (3.11) for the drift part as obtained in the proof of Theorem 3.2 we get

$$\|Y_{\hat{t}}\|_{\mathbf{l}}^p \leq \|Y_0\|_{\mathbf{l}}^p + \int_0^{\hat{t}} C(1 + \|Y_s\|_{\mathbf{l}}^p) ds + |M_{\hat{t}}|;$$

combining this with (3.20) we obtain from Burkholder’s inequality [32, Chap. IV.4] that

$$\mathbb{E} \sup_{s \in [0, \hat{t}]} \|Y_s\|_{\mathbf{l}}^p \leq \|Y_0\|_{\mathbf{l}}^p + \int_0^{\hat{t}} C(1 + \mathbb{E} \sup_{s' \in [0, s]} \|Y_{s'}\|_{\mathbf{l}}^p) ds.$$

It follows that $\mathbb{E} \sup_{s \in [0, \hat{t}]} \|Y_s\|_{\mathbf{l}}^p$ is bounded in terms of the initial data and time t . Using Fatou’s lemma the result follows by letting $P \rightarrow \infty$.

The case $\beta_2^{(0)} > 0$. By Theorem 3.2 we still have from (3.20) that

$$\mathbb{E} [M_{\hat{t}}]^{1/2} \leq \int_0^{\hat{t}} C(1 + \mathbb{E} \|Y_s\|_{\mathbf{l}}^{p+1}) ds \leq (e^{C\hat{t}} - 1)(\|Y_0\|_{\mathbf{l}}^{p+1} + 1),$$

where we similarly obtain a bound in terms of the initial data $\|Y_0\|_{\mathbf{l}}^{p+1}$. \square

3.3. Auxiliary lemmas. It is clear by now that the qualities of the kernel function $\sigma_h(\cdot)$ will play a role in the behavior of the split-step method. This motivates the following brief discussion.

LEMMA 3.6. *Let $f : \mathbf{R} \rightarrow \mathbf{R}$ be a càdlàg piecewise constant function. Then*

$$(3.21) \quad \left| \int_0^t \sigma_h(s)f(s) ds \right| \leq \frac{h}{2}|f(t)| + \frac{h}{2}V_{[0,t]}(f),$$

where the total absolute variation may be exchanged with the square root of the quadratic variation $[f]_t^{1/2}$. Furthermore, if t is a multiple of h , then the first term on the right side of (3.21) vanishes.

Proof. Define

$$(3.22) \quad \Sigma_h(t) \equiv \int_0^t \sigma_h(s) ds,$$

and observe that $|\Sigma_h(\cdot)| \leq h/2$. Denote the left side of (3.21) by J . Then with $(t_k)_{k=0}^N$, the points of discontinuity of f in $(0, t)$, but augmented with the two boundary points $\{0, t\}$, we obtain from summation by parts that

$$J = \left| \sum_{k=0}^{N-1} f(t_k) \Delta \Sigma_h(t_k) \right| = \left| f(t)\Sigma_h(t) - \sum_{k=0}^{N-1} \Delta f(t_k) \Sigma_h(t_{k+1}) \right|.$$

The stated result now follows from the triangle inequality and, for the case of the quadratic variation, from the Cauchy–Schwarz inequality. \square

LEMMA 3.7. *Let $G : \mathbf{R}^D \rightarrow \mathbf{R}$ be a globally Lipschitz continuous function with Lipschitz constant L and let $f : \mathbf{R} \rightarrow \mathbf{R}^D$ be a piecewise constant càdlàg function. Then*

$$(3.23) \quad \left| \int_0^t \sigma_h(s)G(f(s)) ds \right| \leq \frac{h}{2}|G(f(t))| + \frac{h}{2}LV_{[0,t]}(f)$$

with the same additional variants and simplifications as listed in Lemma 3.6.

Proof. This follows because $g(t) := G(f(t))$ satisfies the requirements for f in Lemma 3.6; clearly $|\Delta g(t_k)| = |\Delta G(f(t_k))| \leq L\|\Delta f(t_k)\|$. \square

3.4. Strong convergence. We are now ready to formulate and prove our main result of the paper.

THEOREM 3.8 (strong convergence). *Let $X(t)$ and $Y(t)$ be solutions to (2.4) and (3.6) with $X_0 = Y_0$ and under Assumptions 2.1 and 3.1, respectively. Then*

$$(3.24) \quad \mathbb{E} \|(Y - X)(t)\|^2 \leq h C_t$$

for C_t some constant dependent on the final time t .

In the formulation above, the actual estimate that goes into (3.24) is elaborated upon in (3.28) below. Also, for brevity and inspired by most actual use, we only consider the case of deterministic initial data.

Proof. Under the stated assumptions both processes are well-behaved (Theorems 2.2 and 3.5), so the strategy of proof is to define a suitable stopping time τ_P such that the probability that $t \geq \tau_P$ can be made arbitrarily small. We put

$$\tau_P := \inf_{t \geq 0} \{\|X_t\|_{\mathbf{L}} \vee \|Y_t\|_{\mathbf{L}} > P\},$$

and as before $\hat{t} = t \wedge \tau_P$. Subtracting (2.4) from (3.6) we get

$$(3.25) \quad (Y - X)(\hat{t}) = - \sum_{r \in \mathcal{R}_1, \mathcal{R}_2} \mathbb{N}_r \left[\Pi_r \left(\int_0^{\hat{t}} (1 \pm \sigma_h(s)) w_r(Y_{s-}) ds \right) - \Pi_r \left(\int_0^{\hat{t}} w_r(X_{s-}) ds \right) \right],$$

where the sign is to be chosen in accordance with (3.16).

Under the stopping time and using the local Lipschitz condition we first produce the basic estimates

$$\int_0^{\hat{t}} w_r(X_{s-}) ds \leq \int_0^{\hat{t}} w_r(0) + PL_r ds \leq C_0^{(r)}(P)\hat{t},$$

$$\int_0^{\hat{t}} (1 \pm \sigma_h(s)) w_r(Y_{s-}) ds \leq 2C_0^{(r)}(P)\hat{t}$$

with, for brevity, $L_r \equiv L_r(P)$, and using that $|1 \pm \sigma_h| = 2$.

Taking the square Euclidean norm and expectation value of (3.25) we find by Lemma 2.4 in the form of (2.17) using the above basic bounds that

$$(3.26) \quad \mathbb{E} \|(Y - X)(\hat{t})\|^2 \leq \|\mathbb{N}\|^2 \sum_{r=1}^R (4C_0^{(r)}\hat{t} + 1) \mathbb{E}[A_r]$$

in terms of

$$A_r = \left| \int_0^{\hat{t}} w_r(Y_{s-}) - w_r(X_{s-}) ds \pm \int_0^{\hat{t}} \sigma_h(s) w_r(Y_{s-}) ds \right|.$$

Using Assumption 2.1(iii) anew we get

$$A_r \leq \int_0^{\hat{t}} L_r \|(Y - X)(s-)\| ds + \underbrace{\left| \int_0^{\hat{t}} \sigma_h(s) w_r(Y_{s-}) ds \right|}_{=: B_r}.$$

Also, by Lemma 3.7 we obtain

$$B_r \leq \frac{h}{2} |w_r(Y_{\hat{t}-})| + \frac{h}{2} L_r [Y]_{\hat{t}-}^{1/2}.$$

For the quadratic variation we note that, since $\|\cdot\| \leq \|\cdot\|_1 \leq \|\cdot\|_L$, we have that $[Y]_t \leq \|[Y]_L\|_t$, where the case $p = 1$ of Lemma 3.4 applies. Estimating using Theorem 3.2 we get after some work,

$$(3.27) \quad \mathbb{E}[B_r] \leq C_1^{(r)}(P)h + C_2^{(r)}(P)h\hat{t} \exp(C\hat{t}),$$

where the constants do not depend on h , and where $C_1^{(r)}$ may be taken as zero provided that \hat{t} is a multiple of h .

Summing this over r we get

$$\begin{aligned} \mathbf{E} \|(Y - X)(\hat{t})\|^2 &\leq C_1 h + C_2 h \hat{t} \exp(C\hat{t}) \\ &\quad + \underbrace{\|\mathbb{N}\|^2 \sum_{r=1}^R (4C_0^{(r)} \hat{t} + 1) L_r}_{=:L(\hat{t})=L(\hat{t},P)} \mathbf{E} \left[\int_0^{\hat{t}} \|(Y - X)(s)\|^2 ds \right], \end{aligned}$$

where the “integer inequality”, $n \leq n^2$ for $n \in \mathbf{Z}$, was used. By Grönwall’s inequality this implies the bound

$$\mathbf{E} \|(Y - X)(\hat{t})\|^2 \leq h [C_1 + C_2 \hat{t} \exp(C\hat{t})] \exp(\hat{t}L(\hat{t})).$$

Using brackets to denote characteristic functions we write

$$\mathbf{E} \|(Y - X)(t)\|^2 = \mathbf{E} [\|(Y - X)(t)\|^2 [t < \tau_P]] + \underbrace{\mathbf{E} [\|(Y - X)(t)\|^2 [t \geq \tau_P]]}_{=:M}.$$

To bound M , note first that by Cauchy–Schwarz’s inequality,

$$M \leq (\mathbf{E} \|(Y - X)(t)\|^4)^{1/2} (\mathbf{P}[t \geq \tau_P])^{1/2}.$$

Since $\mathbf{P}[t \geq \tau_P] = \mathbf{P}[\sup_{s \in [0,t]} \|X_s\|_{\mathbf{l}} \vee \|Y_s\|_{\mathbf{l}} > P]$ we get from Markov’s inequality

$$M \leq (\mathbf{E} \|(Y - X)(t)\|^4)^{1/2} P^{-1/2} \left(\mathbf{E} \sup_{s \in [0,t]} \|X_s\|_{\mathbf{l}} \vee \|Y_s\|_{\mathbf{l}} \right)^{1/2}.$$

By Theorems 2.2 and 3.5 we find that M converges to zero as $P \rightarrow \infty$. In particular, for any given $\varepsilon \in (0, 1)$, we can find P large enough (and hence also C_1 , C_2 , and $L(t)$) such that

$$M \leq \varepsilon \mathbf{E} \|(Y - X)(t)\|^2.$$

Combining we thus get

$$(3.28) \quad \mathbf{E} \|(Y - X)(t)\|^2 \leq \frac{h}{1 - \varepsilon} [C_1 + C_2 t \exp(Ct)] \exp(tL(t)),$$

where C_1 , C_2 , and $L(t)$ depend on ε via the choice of P . \square

We now offer some brief comments on this result. The estimated error *growth* in any given interval of time $[0, t]$ is clearly quite pessimistic, but is on the other hand also quite general, relying as it does mainly on the existence of local Lipschitz constants. Also, the fact that the constant C_1 may be disregarded when t is a multiple of the split step h , is perhaps mostly of theoretical interest as the other terms seem to dominate by far. One may simply think of this as a smaller error when the split-step solution is viewed at the discrete mesh $[0, h, 2h, \dots]$.

Finally we comment also that Theorem 3.8 may be strengthened in the direction of convergence in

$$(3.29) \quad \mathbf{E} \sup_{s \in [0,t]} \|(Y - X)(t)\|^2,$$

by bounding the martingale part and relying on Burkholder’s inequality. Since this is somewhat tedious and does not add much insight we have chosen to omit it altogether.

4. Numerical examples. To illustrate the theory developed in the previous sections, but also as a means to investigate the sharpness of some of the bounds, a few selected numerical examples are presented here. In section 4.1 we briefly discuss

the actual implementation of the method implied by (3.6), and in sections 4.2 and 4.3 the convergence of the split-step method applied to two one-dimensional examples is investigated. Concretely, our computational analysis investigates the convergence dependency in the presence of nonlinearities, the weak convergence behavior, and the qualities of higher order split-step methods. In section 4.4 convergence over longer time intervals is discussed and in section 4.5, finally, we look at a split which violates Assumption 3.1.

4.1. Implementation. The idea to use the operational time representation (2.4) to devise computational algorithms has been employed previously for time-parallel algorithms [17] and for parameter sensitivity computations [7, 35]. The *common reaction path method* [35] simulates (2.4) using R separate streams of random numbers and can be regarded as a careful implementation of the *next reaction method* [20] such that a consistent operational time is always defined. An additional improvement reported for spatial problems in [7] is to handle rates that become zero in a somewhat careful way. By explicitly storing the operational time τ^{old} and associated rate w^{old} before the rate vanished we can “activate” the reaction channel by using the rescaling technique [20],

$$(4.1) \quad \tau^{\text{new}} = t_{\text{current}} + (\tau^{\text{old}} - t_{\text{current}}) w^{\text{old}} / w^{\text{new}},$$

where w^{new} is the first nonzero rate encountered. As opposed to drawing a new random number whenever the channel is reactivated, this technique preserves the consistency of the current operational time.

All these implementation techniques transfer nicely to the split-step formulation (3.6). The kernel function σ_h may be thought of as generating deterministic events at points in time which are multiples of $h/2$, where the active and passive pathways simply change roles. Besides being able to compare trajectories pathwise for different values of h , a great feature with this implementation strategy is of course that the limit $h \rightarrow 0$ is directly computable.

For the computations presented below we used the estimator

$$(4.2) \quad E[(Y - X)(t)]^2 \approx M \equiv \frac{1}{N} \sum_{i=1}^N (Y - X)(t; \omega_i)^2,$$

where ω_i indicate independent trajectories coupled and computed according to the previous description. To estimate the uncertainty in (4.2) we compute

$$(4.3) \quad S^2 \equiv \frac{1}{N-1} \sum_{i=1}^N [(Y - X)(t; \omega_i)^2 - M]^2,$$

and use a suitable multiple of S/\sqrt{N} as a measure of the uncertainty.

4.2. Linear birth-death process. We first consider the model



with $[k, \mu] = [5, 0.05]$ for time $t \in [0, 100]$ and with $X(0) = 50$. This model approaches a steady-state Poissonian distribution around the mean value 100 and executes about 10^3 events in the given time interval. Sample illustrations of this process are displayed in Figure 2.

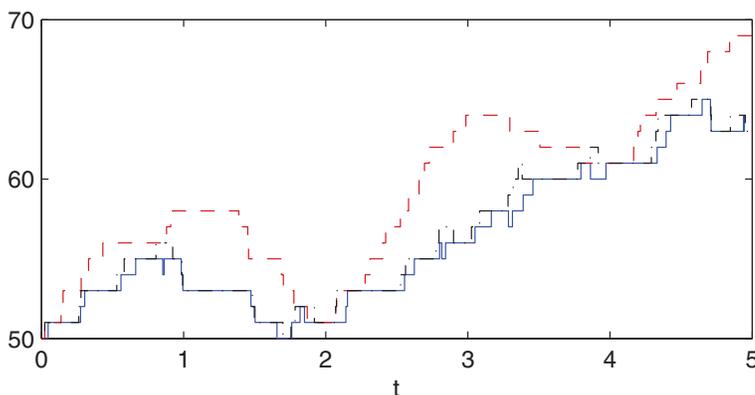


FIG. 2. Sample trajectories of the linear birth-death model. Solid: direct simulation; dashed: split-step solution ($h = 2$); dash-dot: split-step solution ($h = 1/4$). For $h = 2$ the periodic effect of the kernel function σ_h is clearly visible.

In the notation of Assumption 2.1 the birth-death model is characterized by $\alpha = -\mu$, $\beta_1 = \mu/2$, and $\beta_2 = 0$. By the negative sign of α , the dynamics is dissipative for states $x > k/\mu$ and we note also that the propensities are globally Lipschitz with constant $L = \mu$. Computational results are reported in comparison with the dimerization model, next to be introduced.

4.3. Dimerization. As a simple nonlinear extension of (4.4) we consider

$$(4.5) \quad \emptyset \xrightarrow{k} X, \quad X + X \xrightarrow{\nu X(X-1)} \emptyset.$$

We normalize the model such that the equilibrium mean $\sqrt{k/(2\nu)}$ coincides with that of (4.4) and we also use the same birth constant k . This means that for both models, about the same number of events is generated in the considered interval of time.

It is instructive to try to reason about the effect of the nonlinearity in (4.5) in comparison with the fully linear model (4.4). Although by necessity $\alpha = 0$ in Assumption 2.1(i), it still holds true that the dynamics is dissipative for states larger than the equilibrium mean value. Also, due to the quadratic reaction, there is no longer a global Lipschitz constant in Assumption 2.1(iii), but one may note that, close to the equilibrium mean value, $L \sim \nu \times (x - 1) \sim \mu^2/(2k) \times k/\mu = \mu/2$. Since the quadratic reaction involves two molecules one may argue that the effective Lipschitz constant for (4.5) approximately matches that of (4.4).

A more striking difference between the two models can be found in Assumption 2.1(ii). Namely, for (4.5) the left side reads

$$(4.6) \quad k/2 [1 + 4\nu/k \times x(x - 1)] \sim 3k/2,$$

whereas for (4.4) we have

$$(4.7) \quad k/2 [1 + \mu/k \times x] \sim k,$$

again, assuming that x is close to the equilibrium mean value.

4.3.1. Strong convergence. We first consider the strong convergence of the split-step method. The mean square error as a function of the split step h is shown in Figure 3 and the results show that, although both the birth-death model (4.4) and

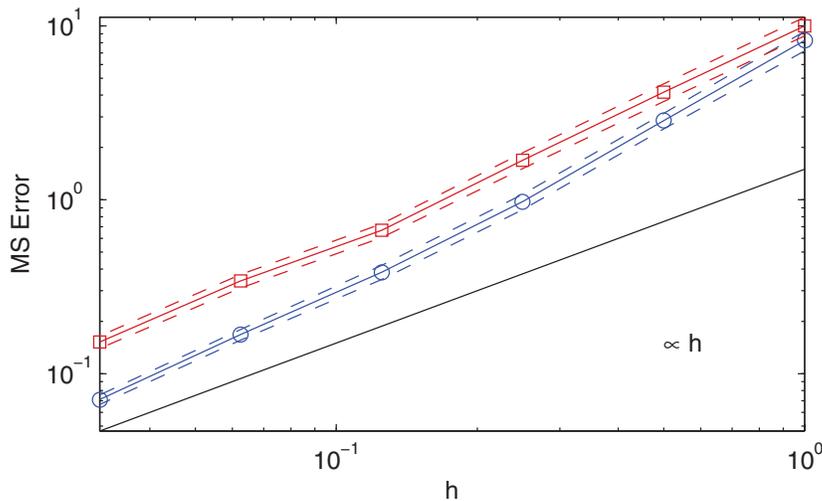


FIG. 3. Convergence in mean square of the split-step method with decreasing h for the linear birth-death model (4.4) (circles) and for the dimerization (4.5) (squares). Dashed lines indicate the estimator's uncertainty $\pm S/\sqrt{N}$ (cf. (4.3)). For both cases the method converges faster initially but approaches the strong order $1/2$ as h becomes smaller.

the dimerization model (4.5) are unbounded problems, the strong order is still $1/2$ as predicted by Theorem 3.8.

The interesting observation to be made is that, with the exception of the case $h = 1$, the mean square error for the dimerization problem is consistently almost a factor of two larger than for the birth-death problem (the measured factor falls in the range $[1.5, 2.1]$ for the cases studied). It is not easy to strictly analyze the reasons behind this phenomenon but we may argue heuristically as follows.

Let us take (3.28) in Theorem 3.8 for some fixed value of ε as an estimate of the error. By the setup of the measurements, $C_1 = 0$ in (3.28), and we have also argued previously that the effective Lipschitz constants for the two models are about the same. The difference between the two models has therefore been isolated to the expression $C_2 h t \exp(Ct)$ in the right side of (3.28), which can be traced back to the bound of B_r in (3.27). In turn, this estimate comes from the bound on the quadratic variation in Lemma 3.4 and relies indirectly also on the moment estimate from Theorem 3.2. In the present case the first few moments of (4.4) and (4.5) are of comparable magnitude so we thus focus our attention on the constant C in (3.15) of Lemma 3.4. In the proof, this constant emerges in (3.19) and is a direct consequence of Assumption 2.1(ii).

This assumption was investigated in (4.6)–(4.7) above where we found a difference between the two models in the form of an overall factor of $3/2$ and a factor of 2 when considering the nonconstant part, in striking agreement with the measured factor $\in [1.5, 2.1]$.

It is an interesting and challenging question if the above heuristic way of reasoning can somehow be put on firm grounds. In particular, it would be very useful if the use of “effective” Lipschitz constants could render a consistent analysis.

4.3.2. Weak convergence. We briefly look also at the weak convergence of the split-step scheme. Hence we estimate for varying split step h ,

$$(4.8) \quad |\mathbb{E}[f(Y_t)] - \mathbb{E}[f(X_t)]|$$

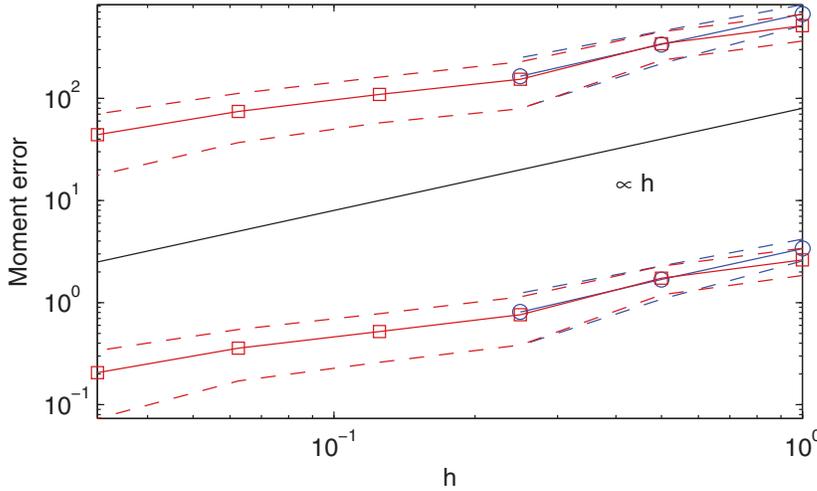


FIG. 4. Weak convergence of the split-step method with decreasing h . Circles: linear birth-death model (4.4) and squares: dimerization (4.5). Top: second order factorial moment; bottom: first order factorial moment. For the birth-death model, only data for the larger values of h are available as the variance was too large to determine the sign of the error within the target uncertainty.

with f some smooth function and otherwise following the computational procedure in section 4.1. To be concrete we took

$$(4.9) \quad f_1(x) = x,$$

$$(4.10) \quad f_2(x) = x(x - 1),$$

that is, the first two factorial moments. In analogy with Figure 3, in Figure 4 we report the error (4.8) for these two cases. As expected we find a first order weak error when measured in the split step h . Perhaps more interesting is the observation that the errors for the two models are very similar and that there is no visible impact from the nonlinearity.

4.3.3. The second order Strang split. The second order Strang split is traditionally written in operator form in the style of (3.3)–(3.4):

$$(4.11) \quad \tilde{p}_h(x, t + h/2) = p_h(x, t) + \int_t^{t+h/2} \sum_{r \in \mathcal{R}_1} w_r(x + \mathbb{N}_r) \tilde{p}_h(x + \mathbb{N}_r, s) - w_r(x) \tilde{p}_h(x, s) ds,$$

$$(4.12) \quad \tilde{\tilde{p}}_h(x, t + h) = \tilde{p}_h(x, t + h/2) + \int_t^{t+h} \sum_{r \in \mathcal{R}_2} w_r(x + \mathbb{N}_r) \tilde{\tilde{p}}_h(x + \mathbb{N}_r, s) - w_r(x) \tilde{\tilde{p}}_h(x, s) ds,$$

$$(4.13) \quad p_h(x, t + h) = \tilde{\tilde{p}}_h(x, t + h) + \int_{t+h/2}^{t+h} \sum_{r \in \mathcal{R}_1} w_r(x + \mathbb{N}_r) p_h(x + \mathbb{N}_r, s) - w_r(x) p_h(x, s) ds,$$

which via two intermediate steps takes us from time t to $t + h$. A moments consideration gives that this is just the same thing as substituting the kernel function $\sigma_h(s)$

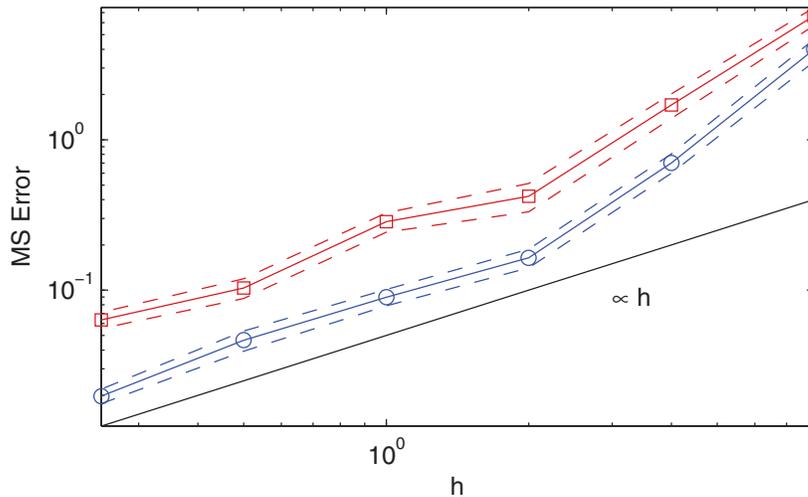


FIG. 5. Convergence in mean square of the Strang split method (4.11)–(4.13), following closely the notation in Figure 3. Although the asymptotic strong order of convergence is still $1/2$, the error is considerably smaller for the same values of h .

in (3.6) with the time shifted version $\sigma_h(s + h/4)$. Importantly, the resulting compact notation is open to the same kind of analysis performed in section 3 and we draw the conclusion that also this method can be expected to converge at strong order $1/2$. We should mention though, that to strictly prove that the convergence order is not actually higher might require some more work.

A feature with the split method (4.11)–(4.13) is that it can be expected to be second order weakly convergent. This follows heuristically from the fact that it is a second order operator splitting method in the finite-dimensional case, and hence can be expected to perform as such also in cases that are of effectively bounded character.

In Figure 5 we display the mean square strong error as a function of the split step h for the two models considered in this section. The asymptotic behavior is quite similar to Figure 3 in that we still approach the strong order $1/2$ as $h \rightarrow 0$ and in that the error reduction factor between the two models is about the same, or perhaps slightly larger $\in [1.7, 3.2]$.

By far the most striking difference is that the mean square error is now between 1.5 to 2 orders smaller than before. It is unfortunately quite difficult to explain this in the setting of section 3.4 since the analysis there would be quite similar upon substituting $\sigma_h(s) \mapsto \sigma_h(s + h/4)$. In particular, assuming the Strang split to be second order weakly convergent, it is difficult to see where this feature could enter the analysis effectively.

4.4. Elementary bimolecular reaction, case of no equilibrium. As a slightly more involved model we consider the following bimolecular birth-death system,



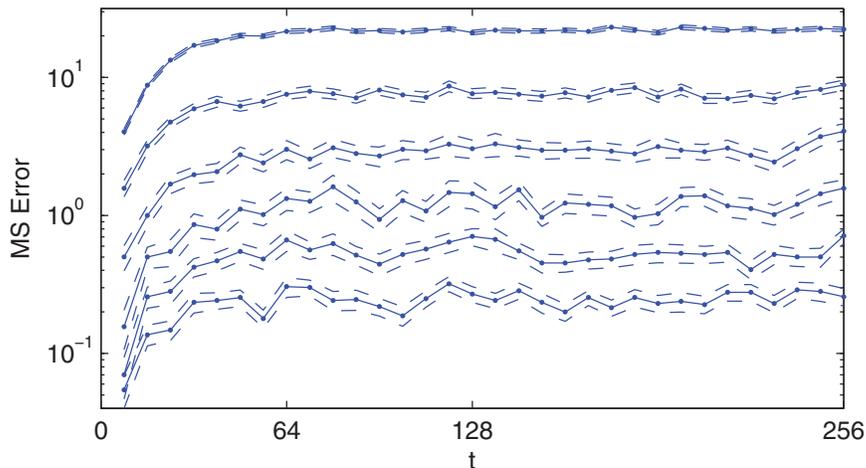


FIG. 6. Mean square error of the split-step method for the bimolecular problem (4.14) as a function of time t . From top to bottom: $h = [4, 2, 1, 1/2, 1/4, 1/8]$. The asymptotic strong order of the method is $q \approx [0.76, 0.67, 0.67, 0.61, 0.55]$ for the cases shown in the figure and averaging over the data for $t \in [128, 256]$.

with the stoichiometric matrix

$$(4.15) \quad \mathbb{N} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix}$$

and reaction propensities $w(x) = [k_1, k_1, k_2 x_1 x_2]^T$ for $x = [x_1, x_2]^T$.

To get some feeling for the behavior of (4.14), we define the difference process $U(t) = (X - Y)(t) \in \mathbf{Z}$. From Itô’s formula (2.9) with $f(x) = x_1 - x_2$ we get

$$(4.16) \quad dU_t = df(X_t) = -[-1, 1, 0] \boldsymbol{\mu}(dt),$$

which is equivalent to the model



which is just a constant intensity random walk on all of \mathbf{Z} . One can solve this explicitly in terms of two independent Poisson distributions,

$$(4.18) \quad U_t = U_0 + \Pi_1(k_1 t) - \Pi_2(k_1 t) \sim U_0 + \mathcal{N}(U_0, 2k_1 t), \quad t \rightarrow \infty,$$

for \mathcal{N} a Gaussian random variable of the indicated mean and variance. Hence the system takes longer and longer excursions away from the origin and we have an example of an equilibrium density which clearly does not exist.

We split the problem (4.14) with $\mathcal{R}_1 = \{1, 2\}$ and $\mathcal{R}_2 = \{3\}$, that is, with the two birth processes evolved simultaneously. In Figure 6 we display the mean square error as a function of time for several values of the split-step parameter h . Experimentally we find that the strong order of convergence is still 1/2 even for quite long simulation times and despite the fact that the behavior of the underlying process is open, visiting states further and further away.

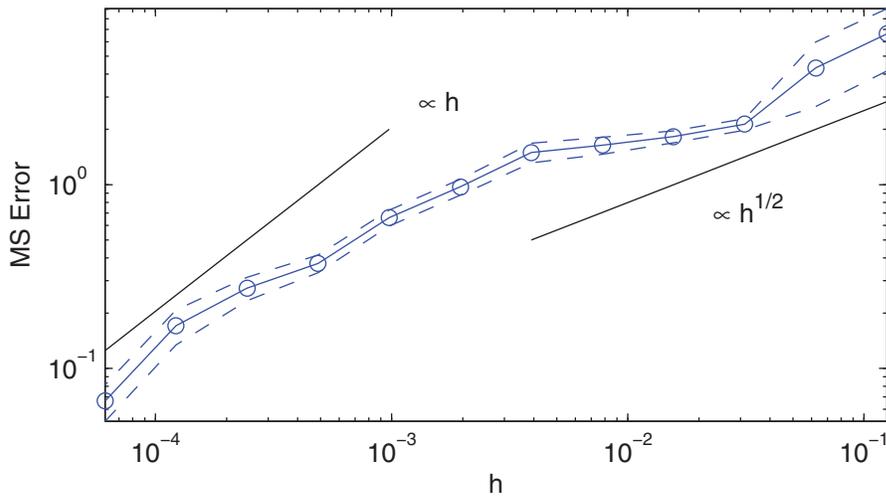
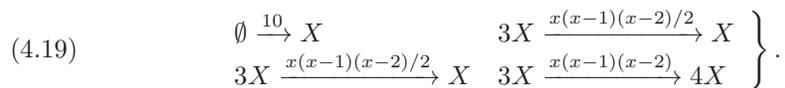


FIG. 7. Mean square convergence for the split-step method applied to the problem (4.19), for which one of the subsystems is ill-posed in the sense of unbounded second moments in finite time. The order of convergence is roughly 1/4 for the range of h studied here but appears to improve somewhat towards the smaller values.

4.5. An ill-posed split. In this section we apply the split-step method to a model which does not comply with Assumption 3.1. This is quite challenging computationally and thus the model settled for was selected after trying several quite different cases. Although somewhat artificial looking, the rather simple system chosen was



The stoichiometric vector is given by $\mathbb{N} = [-1, -2, -2, 1]$ and hence the problem is very strongly dissipative for states large enough in view of the fact that the ingoing cubic reactions dominates. However, by splitting the system according to $\mathcal{R}_1 = \{1, 2\}$ and $\mathcal{R}_2 = \{3, 4\}$, we have that the second pair violates Assumption 3.1. In fact, this subsystem can be shown to explode in the second moment for $t \lesssim X(0)^{-3}/3$ whenever $X(0) \geq 3$ despite the fact that the net drift is zero [15, Proposition 4.1].

In the simulation we worked with the stopped process $\hat{X}(t) = X(t) \wedge P$, where $P = 10^3$ was used and where $t \in [0, 1]$ with $X(0) = 10$ was considered. In Figure 7 we report the results of a convergence study for a selection of comparably small split-step sizes h . The convergence behavior is rather intriguing, with an initial much slower strong order convergence rate of about 1/4, but which picks up speed for h smaller than about $10^{-3}/3$, which is also approximately the maximum time interval over which the second subsystem does not explode in variance. Although the method still seems to converge, the effects of choosing a split containing an ill-posed subsystem are clearly visible.

5. Conclusions. We have proposed a framework for analyzing certain popular split-step methods for jump stochastic differential equations. The framework consists of a formulation of the methods in operational time via the split-step kernel function, enabling a meaningful pathwise coupling even in the limit $h \rightarrow 0$. We have also

presented concrete assumptions and a priori results which together with the theoretical convergence results form a basis for the sound use and further development of these types of methods.

In performing the computational experiments we also developed an actual implementation of our formulation which is useful in assessing the split-step solution quality as a function of various parameters. For example, this is important when experimentally approaching setups for which the proposed sufficient conditions for convergence are violated.

The numerical experiments illustrate the theory in different ways and also open up some intriguing questions. For instance, how can the surprisingly good relative efficiency of the second order Strang method versus the simpler first order method best be explained in the setting of strong convergence? We also observed a better convergence over longer time intervals than perhaps intuitively expected, and we noted that the method converges even in a case for which one of the split subsystems was ill-posed in the sense of finite-time moment explosions.

More practically, an important and natural extension is when the different subsystems are evolved by specially designed approximative methods, a quite common approach in multiscale and multiphysics problems. The proposed framework should remain a viable approach as long as these methods may also be analyzed in a pathwise sense, a point to which we intend to return to in future work.

Acknowledgments. The writing of this paper was initiated at the inspiring BIRS workshop “*Particle-Based Stochastic Reaction-Diffusion Models in Biology*,” in Banff, Alberta, Canada, in November 2014. The author also likes to acknowledge several fruitful discussions with Augustin Chevallier as well as constructive referee comments.

REFERENCES

- [1] D. F. ANDERSON, *Incorporating postleap checks in tau-leaping*, J. Chem. Phys., 128 (2008), 054103.
- [2] D. F. ANDERSON, A. GANGULY, AND T. G. KURTZ, *Error analysis of tau-leap simulation methods*, Ann. Appl. Probab., 21 (2011), pp. 2226–2262.
- [3] D. APPLEBAUM, *Lévy Processes and Stochastic Calculus*, Cambridge Stud. Adv. Math. 93, Cambridge University Press, Cambridge, 2004.
- [4] G. ARAMPATZIS, M. A. KATSOLAKIS, AND P. PLECHÁČ, *Parallelization, processor communication and error analysis in lattice kinetic Monte Carlo*, SIAM J. Numer. Anal., 52 (2014), pp. 1156–1182.
- [5] 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.
- [6] N. BARKAI AND S. LEIBLER, *Circadian clocks limited by noise*, Nature, 403 (2000), pp. 267–268.
- [7] P. BAUER AND S. ENGBLOM, *Sensitivity estimation and inverse problems in spatial stochastic models of chemical kinetics*, in Numerical Mathematics and Advanced Applications: ENUMATH 2013, A. Abdulle, S. Deparis, D. Kressner, F. Nobile, and M. Picasso, eds., Lecture Notes Comput. Sci. Eng. 103, Springer, Berlin, 2015, pp. 519–527.
- [8] B. S. BAYATI, *Fractional diffusion-reaction stochastic simulations*, J. Chem. Phys., 138 (2013), 104117.
- [9] P. BRÉMAUD, *Point Processes and Queues: Martingale Dynamics*, Springer Ser. Statist., Springer, New York, 1981.
- [10] P. BRÉMAUD, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Texts Appl. Math. 31, Springer, New York, 1999.
- [11] C. BRIAT, A. GUPTA, AND M. KHAMMASH, *A scalable computational framework for establishing long-term behavior of stochastic reaction networks*, PLoS Comput. Biol., 10 (2014), 1003669.

- [12] Y. CAO, D. GILLESPIE, AND L. PETZOLD, *Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems*, J. Comput. Phys., 206 (2005), pp. 395–411.
- [13] A. CHEVALLIER AND S. ENGBLOM, *Pathwise Error Bounds in Multiscale Variable Splitting Methods for Spatial Stochastic Kinetics*, manuscript, 2015.
- [14] W. E, D. LIU, AND E. VANDEN-EIJNDEN, *Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates*, J. Chem. Phys., 123 (2005), 194107.
- [15] S. ENGBLOM, *On the stability of stochastic jump kinetics*, Appl. Math., 5 (2014), pp. 3217–3239.
- [16] 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.
- [17] S. ENGBLOM, *Parallel in time simulation of multiscale stochastic chemical kinetics*, Multiscale Model. Simul., 8 (2009), pp. 46–68.
- [18] S. N. ETHIER AND T. G. KURTZ, *Markov Processes: Characterization and Convergence*, Wiley Ser. Probab. Math. Stat., John Wiley & Sons, New York, 1986.
- [19] A. GANGULY, D. ALTINTAN, AND H. KOEPL, *Jump-diffusion approximation of stochastic reaction dynamics: Error bounds and algorithms*, Multiscale Model. Simul., 13 (2015), pp. 1390–1419.
- [20] 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.
- [21] 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.
- [22] D. T. GILLESPIE, *A rigorous derivation of the chemical master equation*, Phys. A, 188 (1992), pp. 404–425.
- [23] D. T. GILLESPIE, *Approximate accelerated stochastic simulation of chemically reacting systems*, J. Chem. Phys., 115 (2001), pp. 1716–1733.
- [24] J. GOUTSIAS, *Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems*, J. Chem. Phys., 122 (2005), 184102.
- [25] 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.
- [26] A. HELLANDER, M. J. LAWSON, B. DRAWERT, AND L. PETZOLD, *Local error estimates for adaptive simulation of the reaction-diffusion master equation via operator splitting*, J. Comput. Phys., 266 (2014), pp. 89–100.
- [27] T. JAHNKE AND D. ALTINTAN, *Efficient simulation of discrete stochastic reaction systems with a splitting method*, BIT, 50 (2010), pp. 797–822.
- [28] N. G. VAN KAMPEN, *Stochastic Processes in Physics and Chemistry*, 2nd ed., Elsevier, Amsterdam, 2004.
- [29] T. G. KURTZ, *Strong approximation theorems for density dependent Markov chains*, Stochastic Process. Appl., 6 (1978), pp. 223–240.
- [30] T. LI, *Analysis of explicit tau-leaping schemes for simulating chemically reacting systems*, Multiscale Model. Simul., 6 (2007), pp. 417–436.
- [31] 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.
- [32] P. E. PROTTER, *Stochastic Integration and Differential Equations*, in Stoch. Model. Appl. Probab. 21, 2nd ed., Springer, Berlin, 2005.
- [33] M. RATHINAM, *Moment growth bounds on continuous time Markov processes on non-negative integer lattices*, Quart. Appl. Math., 73 (2015), pp. 347–364.
- [34] 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.
- [35] M. RATHINAM, P. W. SHEPPARD, AND M. KHAMMASH, *Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks*, J. Chem. Phys., 132 (2010), 034103.
- [36] R. SITU, *Theory of Stochastic Differential Equations with Jumps and Applications*, Math. Anal. Tech. Appl. Eng., Springer, New York, 2005.
- [37] G. STRANG, *On the construction and comparison of difference schemes*, SIAM J. Numer. Anal., 5 (1968), pp. 506–517.
- [38] 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), 020901.