An asymptotic relationship between coupling methods for stochastically modeled population processes

David F. Anderson and Masanori Koyama

March 14, 2014

Abstract

This paper is concerned with elucidating a relationship between two common coupling methods for the continuous time Markov chain models utilized in the cell biology literature. The couplings considered here are primarily used in a computational framework by providing reductions in variance for different Monte Carlo estimators, thereby allowing for significantly more accurate results for a fixed amount of computational time. Common applications of the couplings include the estimation of parametric sensitivities via finite difference methods and the estimation of expectations via multi-level Monte Carlo algorithms. While a number of coupling strategies have been proposed for the models considered here, and a number of articles have experimentally compared the different strategies, to date there has been no mathematical analysis describing the connections between them. Such analyses are critical in order to determine the best use for each. In the current paper, we show a connection between the common reaction path (CRP) method and the split coupling (SC) method, which is termed coupled finite differences (CFD) in the parametric sensitivities literature. In particular, we show that the two couplings are both limits of a third coupling strategy we call the “local-CRP” coupling, with the split coupling method arising as a key parameter goes to infinity, and the common reaction path method arising as the same parameter goes to zero. The analysis helps explain why the split coupling method often provides a lower variance than does the common reaction path method, a fact previously shown experimentally.

1 Introduction

Models of biochemical reaction networks with stochastic dynamics have become increasingly popular in the science literature over the previous fifteen years where they are often studied via computational methods and, in particular, Monte Carlo methods. These computational methods tend to be extremely expensive and time-consuming without the use of variance reduction techniques. One of the most common ways to achieve a large reduction of variance is to couple two relevant processes in order to increase their covariance. There are three main couplings found in the relevant literature: (i) the use of common random numbers

1Department of Mathematics, University of Wisconsin, Madison, Wi. 53706, anderson@math.wisc.edu, grant support from NSF-DMS-1009275 and NSF-DMS-1318832.

2Department of Systems Science, Graduate School of Informatics, Kyoto University, Integrated Systems Biology Lab, koyama.masanori@gmail.com, grant support from NSF-DMS-1009275, NSF-DMS-0805793, and NSF-DMS-1318832. Corresponding Author.

0AMS 2000 subject classifications: Primary 60H35, 65C99; Secondary 92C40
(CRN), (ii) the common reaction path (CRP) method [16], and (iii) a “split coupling” (SC) method termed coupled finite differences in the setting of parametric sensitivities [1, 3]. It has been observed in the literature that both the CRP and SC methods are far superior to the CRN method [1, 16, 17]. It has also been observed through examples that the SC method tends to perform much better than the CRP method, though some exceptions exist [1, 17]. To the best of the authors’ knowledge there has to date been no analytical work on understanding the connections between these two couplings. In the present paper we prove that both the CRP and SC couplings arise naturally as different limits of a third family of couplings we term the “local-CRP” coupling. In particular, the CRP coupling arises as a limit in which the local-CRP coupling is as loosely coupled as possible, whereas the SC coupling arises from a limit of the local-CRP “recoupling” as often as possible. Such an analysis sheds light on why the split coupling often provides a lower variance than does the CRP coupling.

The outline for the remainder of the paper is as follows. In Section 2, we formally present the mathematical models considered in this paper, together with a brief description of the computational methods that serve as motivation for the present work. In Section 3, we present the different coupling strategies for the models presented in Section 2. In Section 4, we state and prove our main results. In Section 5, we provide numerical examples demonstrating our main results and in Section 6 we conclude with some brief remarks.

2 Mathematical model and motivating computational methods

Motivated by models in biochemistry, we consider continuous time Markov chain models in $\mathbb{Z}^d$, in which the $i$th component of the process typically represents the number of molecules of “species” $i$ present in the system. The transitions of the chain are specified by vectors, $\zeta_k \in \mathbb{Z}^d$, for $k \in \{1, \ldots, R\}$ with $R < \infty$, determining the net change in the chain due to the occurrence of a single “reaction,” and by the intensity functions $\lambda_k : \mathbb{Z}^d \to \mathbb{R}_{\geq 0}$, which determine the rate at which the different reactions are occurring.\footnote{Intensity functions are termed “propensity” functions in the biochemistry literature.} Specifically, letting $N_k(t)$ be the number of times transition $k \in \{1, \ldots, R\}$ has occurred by time $t \geq 0$, we will consider the continuous time Markov chain $X$ satisfying the equation

$$X(t) = X(0) + \sum_{k=1}^{R} N_k(t)\zeta_k,$$

where $N_k$ is a counting process with local intensity function $\lambda_k$. That is, $\{N_k\}$ are the counting processes for which the processes

$$N_k(t) - \int_0^t \lambda_k(X(s))ds$$

are local martingales. One useful representation for the counting processes $N_k(t)$ is via time-changed unit-rate Poisson processes [5, 9, 14],

$$N_k(t) = Y_k \left( \int_0^t \lambda_k(X(s))ds \right),$$

where $Y_k$ are unit-rate Poisson processes.
yielding the stochastic equation

$$X(t) = X(0) + \sum_{k=1}^{R} Y_k \left( \int_0^t \lambda_k(X(s))ds \right) \zeta_k$$  \hspace{1cm} (2.1)$$

where \( \{Y_k\}_{k=1}^{R} \) is a collection of independent unit-rate Poisson processes. Note that \( X \) can also be specified by its infinitesimal generator,

$$\(Af\)(x) = \sum_{k=1}^{R} \lambda_k(x)(f(x + \zeta_k) - f(x)),$$  \hspace{1cm} (2.2)$$

where \( f \) is any bounded function with compact support.

We denote \( Z \) as the process on \( \mathbb{Z}^d \) with the same transition directions \( \{\zeta_k\} \) as \( X \), but with intensities \( \lambda_k: \mathbb{Z}^d \to \mathbb{R}_{\geq 0} \). That is, \( Z \) is the Markov process with infinitesimal generator

$$\(Bf\)(x) = \sum_{k=1}^{R} \lambda_k(x)(f(x + \zeta_k) - f(x)),$$  \hspace{1cm} (2.3)$$

and which satisfies the stochastic equation

$$Z(t) = Z(0) + \sum_{k=1}^{R} Y_k \left( \int_0^t \lambda_k(Z(s))ds \right) \zeta_k,$$  \hspace{1cm} (2.4)$$

where \( \{Y_k\}_{k=1}^{R} \) is a collection of independent unit-rate Poisson processes. In the remainder of the paper, we consider different ways to couple \( X \) and \( Z \) and provide an asymptotic relationship between two of the couplings.

2.1 Motivating computational methods

We briefly present two computational methods that serve as the motivation for much of the analysis of different coupling strategies: finite difference methods for parametric sensitivity analysis and multi-level Monte Carlo for the estimation of expectations.

2.1.1 Parametric sensitivity analysis

Suppose that \( \{X^\theta\} \) is a parametric family of processes about \( \theta \) on a state space \( E \), and \( f: E \to \mathbb{R} \) is some statistic of interest. For example, \( f(X(t)) = X_i(t) \) may provide the abundance of species \( i \) at time \( t \geq 0 \). It is common to wish to evaluate

$$\frac{d}{d\theta} \mathbb{E}[f(X^\theta(t))] \approx \frac{\mathbb{E}[f(X^{\theta+h}(t))] - \mathbb{E}[f(X^\theta(t))]}{h}$$  \hspace{1cm} (2.5)$$

as a measurement of the sensitivity of \( \mathbb{E}f(X^\theta(t)) \) with respect to \( \theta \). Such a strategy is usually called a finite difference method. We would like to empirically evaluate the right-hand side of (2.5) in as efficient a manner as possible. By coupling the processes \( (X^{\theta+h}, X^\theta) \), we may evaluate

$$h^{-1}\mathbb{E}[f(X^{\theta+h}(t)) - f(X^\theta(t))],$$

with the magnitude of \( \text{Var}(f(X^{\theta+h}(t)) - f(X^\theta(t))) \) determining the quality of the coupling. In particular, we wish to minimize \( \text{Var}(f(X^{\theta+h}(t)) - f(X^\theta(t))) \) without greatly increasing
the computational cost of producing realizations of the coupled processes \((X^{\theta+h}, X^{\theta})\). We explicitly note that in the setting of the previous section, we have

\[
\lambda_k(\cdot) = \eta_k(\theta, \cdot), \quad \tilde{\lambda}_k(\cdot) = \eta_k(\theta + h, \cdot),
\]

where for each \(k\), \(\{\eta_k(\theta, \cdot) : \mathbb{R}^d \to \mathbb{R}_{\geq 0}\}\) is a parametric family of functions about \(\theta\). In this case, we have

\[
X = X^{\theta}, \quad \text{and} \quad Z = X^{\theta+h}.
\]

As mentioned in Section 1, there has been a large amount of work in the literature on developing good coupling strategies for the estimation of parametric sensitivities via finite differences (2.5); see, for example, [1, 6, 16, 17]. To the best of the authors’ knowledge there has been no mathematical analysis detailing the connection between the different couplings used, though see the discussion in Section 6 for details pertaining to a recent work by Arampatzis and Katsoulakis [7].

### 2.2 Multi-level Monte Carlo

In [10], Mike Giles introduced the multi-level Monte Carlo (MLMC) method for the approximation of expectations of diffusion processes. Specifically, if \(X\) is the diffusion process of interest and \(\{Z_\ell\}\) are a family of approximations to \(X\), with higher values of \(\ell\) corresponding to better approximations, then we observe that for any function \(f\) of interest,

\[
E[f(X(t))] \approx E[f(Z_L)] = \sum_{\ell=1}^{L} E[f(Z_\ell(t)) - f(Z_{\ell-1}(t))] + E[f(Z_0)],
\]

where \(L\) is chosen large enough so that \(|E[f(X(t))] - E[f(Z_L(t))]|\) is below some target accuracy. It is typical to choose \(Z_\ell\) to be the process produced by Euler-Maruyama with a step size of \(M^{-\ell}\) for some \(M \in \{2, 3, \ldots, 7\}\). If each term \(f(Z_\ell(t)) - f(Z_{\ell-1}(t))\) is tightly coupled, then the variance of each of the intermediate estimators will be low, thereby moving the computational cost to the lowest level, \(E[f(Z_0)]\), which can be estimated quickly via Euler-Maruyama with large time-steps.

In [3], Anderson and Higham extended the multi-level Monte Carlo method to the setting of this paper by utilizing the split coupling detailed in Section 3. They further noted that an unbiased estimator can be produced for jump models by coupling the exact process \(X\) with the approximate process with the finest time-step

\[
E[f(X(t))] = E[f(X(t)) - f(Z_L(t))] + \sum_{\ell=1}^{L} E[f(Z_\ell(t)) - f(Z_{\ell-1}(t))] + E[f(Z_0)],
\]

where, again, it is the quality of the coupling at each level that determines the overall quality of the method.

We point out that in the diffusive case the most natural coupling is to re-use the Brownian path for each of the coupled processes. This is relatively easy to do via the Brownian bridge. However, as will be noted in the next section, there are multiple natural couplings to choose from in the context of jump processes with state dependent intensity functions, and different choices lead to computational methods with vastly different computational complexities and, hence, runtimes.
3 Different Couplings

We return to the notation introduced at the beginning of Section 2 and focus our discussion on ways to couple $X$ and $Z$ with intensities $\lambda_k$ and $\tilde{\lambda}_k$, respectively.

3.1 Split coupling

We will begin by introducing the split coupling (SC), which first appeared as an analytic tool in [15] and later appeared in the context of computational methods in [1, 2, 3, 4, 7]. Let $a \wedge b \equiv \min\{a, b\}$, and let $R$ and $L$ be any càdlàg processes on $\mathbb{R}^d$. Then for each $k \in \{1, \ldots, R\}$ we define the operators $r_{1k}, r_{2k},$ and $r_{3k}$ via

$$r_{1k}(\lambda_k, \tilde{\lambda}_k, R, L)(s) \overset{\text{def}}{=} \lambda_k(R(s)) \wedge \tilde{\lambda}_k(L(s))$$
$$r_{2k}(\lambda_k, \tilde{\lambda}_k, R, L)(s) \overset{\text{def}}{=} \lambda_k(R(s)) - r_{1k}(\lambda_k, \tilde{\lambda}_k, R, L)(s)$$
$$r_{3k}(\lambda_k, \tilde{\lambda}_k, R, L)(s) \overset{\text{def}}{=} \tilde{\lambda}_k(L(s)) - r_{1k}(\lambda_k, \tilde{\lambda}_k, R, L)(s).$$

(3.1)

The split coupling of the processes $X$ and $Z$ is then given by

$$X_{\text{sc}}(t) = X(0) + \sum_{k=1}^{R} \left\{ Y_{1k} \left( \int_0^t r_{1k}(\lambda_k, \tilde{\lambda}_k, X_{\text{sc}}, Z_{\text{sc}})(s) ds \right) \right\} + Y_{2k} \left( \int_0^t r_{2k}(\lambda_k, \tilde{\lambda}_k, X_{\text{sc}}, Z_{\text{sc}})(s) ds \right) \xi_k$$
$$Z_{\text{sc}}(t) = Z(0) + \sum_{k=1}^{R} \left\{ Y_{1k} \left( \int_0^t r_{1k}(\lambda_k, \tilde{\lambda}_k, X_{\text{sc}}, Z_{\text{sc}})(s) ds \right) \right\} + Y_{3k} \left( \int_0^t r_{3k}(\lambda_k, \tilde{\lambda}_k, X_{\text{sc}}, Z_{\text{sc}})(s) ds \right) \xi_k,$$

(3.2)

where $\{Y_{1k}\}_{k=1}^{R} \cup \{Y_{2k}\}_{k=1}^{R} \cup \{Y_{3k}\}_{k=1}^{R}$ are mutually independent unit-rate Poisson processes. Note that $X_{\text{sc}}$ and $Z_{\text{sc}}$ share the family of counting processes determined by the Poisson processes $Y_{1k}$. Further note that $(X, Z)$ satisfying the stochastic equation (3.2) is simply a continuous time Markov chain on $\mathbb{Z}^d \times \mathbb{Z}^d$ with infinitesimal generator

$$\mathcal{L}_{\text{sc}} g(x, z) = \sum_{k=1}^{R} \min\{\lambda_k(x), \tilde{\lambda}_k(z)\} (g(x + \xi_k, z + \xi_k) - g(x, z))$$
$$+ \sum_{k=1}^{R} (\lambda_k(x) - \min\{\lambda_k(x), \tilde{\lambda}_k(z)\}) (g(x + \xi_k, z) - g(x, z))$$
$$+ \sum_{k=1}^{R} (\tilde{\lambda}_k(z) - \min\{\lambda_k(x), \tilde{\lambda}_k(z)\}) (g(x, z + \xi_k) - g(x, z)),$$

where $g : \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{R}$ is any bounded function with compact support.

3.2 Common random numbers

In the common random numbers (CRN) coupling, we simply simulate the embedded discrete time Markov chain for each process concurrently with the exponential holding time for each
transition. The processes \( X \) and \( Z \) are then coupled by using (i) the same stream of random variables for the generation of the embedded discrete time chain, and (ii) the same stream of random variables for the exponential holding times.

More explicitly, let \( \{U_i\}_{i=0}^{\infty} \) be a sequence of uniform random variables over the interval \([0, 1]\), and let \( \eta: \mathbb{R}_0^\infty \times [0, 1] \rightarrow \{\zeta_1, \ldots, \zeta_R\} \) be defined via

\[
\eta(c_1, \ldots, c_R, u) = \zeta_k \quad \text{if} \quad \sum_{i=1}^{k-1} c_i \leq u < \sum_{i=1}^{k} c_i / \sum_{i=1}^{R} c_i,
\]

which is a categorical random variable parametrized by \( c_1, \ldots, c_R \). Also, let us denote

\[
\lambda_0(x) = \sum_{k=1}^{R} \lambda_k(x) \quad \text{and} \quad \tilde{\lambda}_0(x) = \sum_{k=1}^{R} \tilde{\lambda}_k(x).
\]

Then for a common unit-rate poisson process \( Y \), which will determine the exponential holding times, we consider the following system:

\[
R_X(t) = Y \left( \int_0^t \lambda_0(X_{\text{crn}}(s))ds \right) \\
R_Z(t) = Y \left( \int_0^t \tilde{\lambda}_0(Z_{\text{crn}}(s))ds \right)
\]

\[
X_{\text{crn}}(t) = X_{\text{crn}}(0) + \int_0^t \eta(\lambda_1(X_{\text{crn}}(s-)), \ldots, \lambda_R(X_{\text{crn}}(s-)), U_{R_X(s-)}), dR_X(s) \\
Z_{\text{crn}}(t) = Z_{\text{crn}}(0) + \int_0^t \eta(\tilde{\lambda}_1(Z_{\text{crn}}(s-)), \ldots, \tilde{\lambda}_R(Z_{\text{crn}}(s-)), U_{R_Z(s-)}), dR_Z(s),
\]

where we note that the processes shared not just the Poisson process \( Y \), but also the sequence of uniform \([0, 1]\) random variables \( \{U_i\}_{i=0}^{\infty} \). The solution to this system exists and is unique by construction [5, 11, 12]. We note that while the representations are different, the marginal processes \( X_{\text{crn}} \) and \( X_{\text{sc}} \) have the same distribution, while the coupled processes \((X_{\text{crn}}, Z_{\text{crn}})\) and \((X_{\text{sc}}, Z_{\text{sc}})\) obviously do not.

### 3.3 Common reaction path coupling and the local common reaction path coupling

The common reaction path (CRP) coupling arises by simply noting that we may couple the processes \( (2.1) \) and \( (2.4) \) via the Poisson processes \( \{Y_k\} \). That is, in the CRP coupling \((X_{\text{crp}}, Z_{\text{crp}})\) satisfies

\[
X_{\text{crp}}(t) = X_{\text{crp}}(0) + \sum_{k=1}^{R} Y_k \left( \int_0^t \lambda_k(X_{\text{crp}}(s))ds \right) \zeta_k \\
Z_{\text{crp}}(t) = Z_{\text{crp}}(0) + \sum_{k=1}^{R} Y_k \left( \int_0^t \tilde{\lambda}_k(Z_{\text{crp}}(s))ds \right) \zeta_k,
\]

where the \( Y_k \) are independent unit-rate Poisson processes.

Numerical experiments have shown that this coupling is significantly tighter than the CRN coupling, in that it produces a lower variance between the coupled processes, for many
we also have

\[ \int_0^{t_0} \tilde{\lambda}_k(Z_{\text{crp}}(s)) ds \ll \int_0^{t_0} \lambda_k(X_{\text{crp}}(s)) ds \quad (3.6) \]

for all \( k \). Then the time and type of the next jump of \( X_{\text{crp}} \) is nearly uncorrelated from the time and type of the next jump of \( Z_{\text{crp}} \). This is true even if \( X_{\text{crp}}(t_0) \) and \( Z_{\text{crp}}(t_0) \) are very close or equal. This problem does not occur with the split coupling since the next jump time of \( X \) and \( Z \) are always correlated via the counting processes with intensity

\[ \lambda_k(X_{\text{sc}}(s)) \wedge \tilde{\lambda}_k(Z_{\text{sc}}(s)). \]

The above discussion motivates us to reconsider the CRP coupling and “reset” the Poisson processes at small time intervals in order to overcome the problem of \( X_{\text{crp}} \) and \( Z_{\text{crp}} \) can fully decouple from each other. In particular, we wish to ensure that whenever the processes \( X_{\text{crp}}(s) \) and \( X_{\text{crp}}(s) \) are equal, the processes recouple. We will elaborate on this strategy. Let \( \pi = \{ 0 = s_0 < s_1 \cdots < s_N = T \} \) be a partition of \([0,T]\). Also let \( \{ Y_{km} : k = 1, \ldots, R, m = 0, 1, 2, \ldots \} \) be a set of independent, unit-rate Poisson processes. Then we define the local-CRP coupling over \([0,T]\) with respect to \( \pi \) as the solution of

\[
X_{\text{crp}}^\pi(t) = X(0) + \sum_{k=1}^{R} \sum_{m=0}^{\infty} Y_{km} \left( \int_{t \wedge s_m}^{t \wedge s_{m+1}} \lambda_k(X_{\text{crp}}^\pi(s)) ds \right) \zeta_k \\
Z_{\text{crp}}^\pi(t) = Z(0) + \sum_{k=1}^{R} \sum_{m=0}^{\infty} Y_{km} \left( \int_{t \wedge s_m}^{t \wedge s_{m+1}} \lambda_k(Z_{\text{crp}}^\pi(s)) ds \right) \zeta_k. \quad (3.7)
\]

We remark that, irrespective of \( \pi \), the marginal distribution of \( X_{\text{crp}}^\pi \) is the same as that of \( X \), our process of interest, and the same goes for \( Z_{\text{crp}}^\pi \) and \( Z \). Also, when \( \pi \) is a trivial partition with \( N = 1 \), the coupling (3.7) is precisely the CRP coupling of (3.5). In the next section, we will consider the limit of the set family of local-CRP couplings as \( N \to \infty \) and prove that under reasonable conditions the coupled processes converge weakly to the processes coupled via the split coupling (3.2).

4 Limit of the local-CRP

We begin this section by specifying some notation. First, when \( X \) and \( Z \) are stochastic processes built on the probability space \((\Omega, \mathcal{F}, P)\), we denote by \( X(s, \omega) \) the process \( X \) evaluated at time \( s \) for a given choice \( \omega \in \Omega \). Further, by \( (X, Z)(s, \omega) \) we mean \((X(s, \omega), Z(s, \omega))\), a vector of random variables evaluated at time \( s \). As is usual, we will often omit \( \omega \) from the notation when no confusion is expected. Finally, when \( t = (t_1, \ldots, t_K) \) is a \( K \) dimensional vector of times points, we denote

\[ X(t) = [X(t_1), \ldots, X(t_K)]. \]

Also, throughout the section, we assume that \( X(0) = Z(0) \).
4.1 Weak convergence at finite coordinates

We first have to articulate what we mean by taking \(N \to \infty\) in the context of the last section.

**Definition.** Let \(\pi_N = \{0 = s_0 \leq s_1 \leq \cdots \leq s_N = T\}\) be a partition of \([0,T]\). For \(m \in \{0, \ldots, N-1\}\) let
\[
\Delta_m \pi_N = s_{m+1} - s_m.
\]
The mesh of \(\pi_N\) is defined as
\[
\text{mesh}(\pi_N) \overset{\text{def}}{=} \max\{\Delta_m \pi_N : m \in \{0, \ldots, N-1\}\}.
\]

Supposing that \(\text{mesh}(\pi_N) \to 0\) as \(N \to \infty\), the limit of interest to us is the weak limit of \((X^{\pi_N}, Z^{\pi_N})\) as \(N \to \infty\). We begin with Proposition 4.1 showing the weak convergence of \(X_{\pi N}^{\pi N}\) to \(X_{sc}\) over finite coordinates as \(N \to \infty\). In Subsection 4.2 we prove weak convergence at the process level.

**Proposition 4.1.** Suppose that neither of the nominal processes \(X, Z\) are explosive and let \((X_{sc}(t), Z_{sc}(t))\) be coupled in the way of (3.2). Let
\[
\pi_n = \{0 = s_0 \leq s_1 \leq \cdots \leq s_{N(n)} = T\}
\]
be a sequence of partitions such that \(\text{mesh}(\pi_n) \to 0\), as \(n \to \infty\), and for each \(n\) let \((X^{\pi_n}(t), Z^{\pi_n}(t))\) be coupled in the way of (3.7). Then for any \(K \in \mathbb{Z}_{\geq 0}\) and \(t \in [0,T]^K\), and any bounded Lipchitz \(f : (\mathbb{R}^d \times \mathbb{R}^d)^K \to \mathbb{R}\),
\[
\mathbb{E}[f((X^{\pi_n}_{\pi_{crp}}, Z^{\pi_n}_{\pi_{crp}})(t))] \to \mathbb{E}[f((X_{sc}, Z_{sc})(t))], \quad \text{as } n \to \infty.
\]

We will briefly outline the proof of 4.1. For a fixed \(n\), let
\[
\{Y^n_{ikm} : i = 1, 2, 3, \; k = 1, \ldots, R, \; m = 0, 1, 2, \ldots\} \quad (4.1)
\]
and
\[
\{Y^n_{km} : k = 1, \ldots, R, \; m = 0, 1, 2, \ldots\} \quad (4.2)
\]
be two sets of independent unit-rate Poisson processes. At this point, we do not make any assumption on the correlation between the processes in the set (4.1) and the processes in the set (4.2), except to note that they will not be independent. In fact, we will construct the Poisson processes of (4.1) as functions of the Poisson processes of (4.2). For now, simply consider the processes built using the Poisson processes of (4.1)
\[
X^{\pi_n}_{sc}(t) = X_{sc}(0) + \sum_{m=0}^{\infty} \sum_{k=1}^{R} Y^n_{1km} \left( \int_{s_m \wedge t}^{s_{m+1} \wedge t} r_{1k}(\lambda_k, \tilde{\lambda}_k, X^{\pi_n}_{sc}, Z^{\pi_n}_{sc})(s)ds \right)
\]
\[
+ Y^n_{2km} \left( \int_{s_m \wedge t}^{s_{m+1} \wedge t} r_{2k}(\lambda_k, \tilde{\lambda}_k, X^{\pi_n}_{sc}, Z^{\pi_n}_{sc})(s)ds \right) \zeta_k,
\]
\[
Z^{\pi_n}_{sc}(t) = X_{sc}(0) + \sum_{m=0}^{\infty} \sum_{k=1}^{R} Y^n_{1km} \left( \int_{s_m \wedge t}^{s_{m+1} \wedge t} r_{1k}(\lambda_k, \tilde{\lambda}_k, X^{\pi_n}_{sc}, Z^{\pi_n}_{sc})(s)ds \right)
\]
\[
+ Y^n_{3km} \left( \int_{s_m \wedge t}^{s_{m+1} \wedge t} r_{3k}(\lambda_k, \tilde{\lambda}_k, X^{\pi_n}_{sc}, Z^{\pi_n}_{sc})(s)ds \right) \zeta_k,
\]
\[
\text{as } n \to \infty.
\]
which are built with the Poisson processes (4.2). Note that $(X_{\text{crp}}(t), Y_{\text{crp}}(t))$ and $(\tilde{X}_{\text{crp}}(t), \tilde{Y}_{\text{crp}}(t))$ are independent unit-rate Poisson processes. We generate $(\tilde{X}_{\text{crp}}(t), \tilde{Y}_{\text{crp}}(t))$ up to time $s$ using them according to (4.3).

We further suppose that we have constructed the relevant Poisson processes $Y_{\text{crp}}^n$ and $\tilde{Y}_{\text{crp}}^n$ up to time $s$. We now turn our attention to constructing the required Poisson processes $X_{\text{crp}}^n$ and $\tilde{X}_{\text{crp}}^n$ built using them according to (4.3).

Inductively arguing on $m$, suppose we have already generated $(X_{\text{crp}}^n, Z_{\text{crp}}^n)$ given by (4.3) up to time $s_m \geq 0$. We further suppose that we have constructed the relevant Poisson processes $Y_{\text{crp}}^{\tilde{m}}$ for all $\tilde{m} < m$. We must now describe how to construct $Y_{\text{crp}}^{m}$ for each valid pair $(i, k)$. We define the following random times for each $i \in \{1, 2, 3\}$ and $k \in \{1, \ldots, R\}$:

$$T_{ikm} \overset{\text{def}}{=} \tilde{r}_{ik}(\lambda_k, \tilde{\lambda}_k, X_{\text{crp}}^n, Z_{\text{crp}}^n)(s_m) \cdot \Delta_m(\pi_n)$$

and

$$T_{\text{crp}}^{km} \overset{\text{def}}{=} \left( \int_{s_m}^{s_{m+1}} \lambda(X_{\text{crp}}^n(s)) \, ds \right) \vee \left( \int_{s_m}^{s_{m+1}} \tilde{\lambda}(Z_{\text{crp}}^n(s)) \, ds \right).$$

We must now describe how to construct $Y_{\text{crp}}^{km}$ for each valid pair $(i, k)$. We define the following random times for each $i \in \{1, 2, 3\}$ and $k \in \{1, \ldots, R\}$:

$$T_{ikm} \overset{\text{def}}{=} \tilde{r}_{ik}(\lambda_k, \tilde{\lambda}_k, X_{\text{crp}}^n, Z_{\text{crp}}^n)(s_m) \cdot \Delta_m(\pi_n)$$

and

$$T_{\text{crp}}^{km} \overset{\text{def}}{=} \left( \int_{s_m}^{s_{m+1}} \lambda(X_{\text{crp}}^n(s)) \, ds \right) \vee \left( \int_{s_m}^{s_{m+1}} \tilde{\lambda}(Z_{\text{crp}}^n(s)) \, ds \right).$$

We must still describe the specific construction alluded to above that will allow us to conclude (4.4). For each $n$, let

$$\{Y_{km}^n, Y_{ikm}^{n,aug}, i = 1, 2, 3, k = 1, \ldots, R, m = 0, 1, 2, \ldots\}$$

be independent unit-rate Poisson processes. We generate $(X_{\text{crp}}^n, Z_{\text{crp}}^n)$ up to time $T$ using the processes $Y_{km}^n$ according to (4.4). We now turn our attention to constructing the required independent unit-rate Poisson processes $Y_{ikm}^{n,aug}$, and the coupled processes $(X_{\text{crp}}^n, Z_{\text{crp}}^n)$ built using them according to (4.3).

Inductively arguing on $m$, suppose we have already generated $(X_{\text{crp}}^n, Z_{\text{crp}}^n)$ given by (4.3) up to time $s_m \geq 0$. We further suppose that we have constructed the relevant Poisson processes $Y_{ikm}^{n,aug}$ for all $\tilde{m} < m$. We must now describe how to construct $Y_{ikm}^{n,aug}$ for each valid pair $(i, k)$. We define the following random times for each $i \in \{1, 2, 3\}$ and $k \in \{1, \ldots, R\}$:

$$T_{ikm} \overset{\text{def}}{=} \tilde{r}_{ik}(\lambda_k, \tilde{\lambda}_k, X_{\text{crp}}^n, Z_{\text{crp}}^n)(s_m) \cdot \Delta_m(\pi_n)$$

and

$$T_{\text{crp}}^{km} \overset{\text{def}}{=} \left( \int_{s_m}^{s_{m+1}} \lambda(X_{\text{crp}}^n(s)) \, ds \right) \vee \left( \int_{s_m}^{s_{m+1}} \tilde{\lambda}(Z_{\text{crp}}^n(s)) \, ds \right).$$

for any $\gamma > 0$. We can then appeal to a standard Portmanteau type argument to finish the proof of Proposition 4.1: let $\epsilon > 0$, and consider any bounded continuous map $f : (E \times E)^K \to \mathbb{R}$ with Lipshitz constant $L$. Then

$$|E f((X_{\text{sc}}, Z_{\text{sc}})(t)) - E f((X_{\text{crp}}^n, Z_{\text{crp}}^n)(t))| = |E f((X_{\text{sc}}^n, Z_{\text{sc}}^n)(t)) - E f((X_{\text{crp}}^n, Z_{\text{crp}}^n)(t))| \leq L \Delta_m(\pi_n).$$

We can first choose $\gamma < \epsilon/(2LK)$. With this $\gamma$ fixed, we may choose $n$ large enough so that the second piece can be bounded by $\epsilon/2$, and the claim is achieved.
where, as usual, \( a \lor b \overset{\text{def}}{=} \max\{a, b\} \), and we recall that \((X_{\text{crp}}^n, Z_{\text{crp}}^n)\) has already been generated up to time \( T \). For notational clarity we refrain from using \( n \) in the notation above for the random times. We now define \( Y_{1km}^n \) in the following manner

\[
Y_{1km}^n(s) = \begin{cases} 
    Y_{km}^n(s) & \text{for } s \leq T_{1km} \\
    Y_{1km}(s) + Y_{1km}^{n,aug}(s - T_{1km}) & \text{for } s > T_{1km},
\end{cases}
\]

Having defined \( Y_{1km}^n \), we turn to the construction of \( Y_{2km}^n \) and \( Y_{3km}^n \). The construction is based on which of two of the following cases hold.

1. If \( \lambda(Z_{\text{sc}}^n(s_m)) \leq \lambda(X_{\text{sc}}^n(s_m)) \), then let \( Y_{2km}^n \) satisfy

\[
Y_{2km}^n(s) = Y_{km}^n(s + T_{1km}) - Y_{km}^n(T_{1km}) \quad \text{for } s \leq T_{2km}
\]

\[
Y_{2km}^n(s) = Y_{2km}^n(T_{2km}) + Y_{2km}^{n,aug}(s - T_{2km}) \quad \text{for } s > T_{2km},
\]

and let \( Y_{3km}^n(s) = Y_{3km}^{n,aug} \) for all \( s \geq 0 \).

2. If \( \lambda(Z_{\text{sc}}^n(s_m)) > \lambda(X_{\text{sc}}^n(s_m)) \), then let \( Y_{3km}^n \) satisfy

\[
Y_{3km}^n(s) = Y_{km}^n(s + T_{1km}) - Y_{km}^n(T_{1km}) \quad \text{for } s \leq T_{3km}
\]

\[
Y_{3km}^n(s) = Y_{3km}^n(T_{3km}) + Y_{3km}^{n,aug}(s - T_{3km}) \quad \text{for } s > T_{3km},
\]

and let \( Y_{2km}^n(s) = Y_{2km}^{n,aug} \) for all \( s \geq 0 \).

Note that the strong Markov property guarantees that the processes \( \{Y_{ikm}^n\} \) so constructed are independent, unit-rate Poisson processes. We then generate \((X_{\text{sc}}^n, Z_{\text{sc}}^n)\) between times \( s_m \) and \( s_{m+1} \) according to (4.3) with the processes \( \{Y_{ikm}^n\} \). Note that in so doing, we have also created a between \((X_{\text{crp}}^n, Z_{\text{crp}}^n)\) and \((X_{\text{crp}}^n, Z_{\text{crp}}^n)\).

Note that for each \( i, k, \) and \( m \), the value \( T_{ikm} \) as defined in (4.7) is an approximation to

\[
T_{ikm}^{sc} \overset{\text{def}}{=} \int_{s_m}^{s_{m+1}} r_{ik}(\lambda_k, X_{sc}^n, Z_{sc}^n)(s) \, ds.
\]

We would like to make a few observations about this approximation before proceeding further.

**Lemma 4.2.** Fix \( n \), and let \( m \in \{0, 1, \ldots\} \). If

\[
\sum_{k=1}^{R} \sum_{i=1}^{3} Y_{ikm}^n(T_{ikm} \lor T_{ikm}^{sc}) = 1 \quad (4.8)
\]

then there is a unique \( j \in \{1, 2, 3\} \) and \( \ell \in \{1, \ldots, R\} \) for which

\[
Y_{j\ell m}^n(T_{j\ell m} \land T_{j\ell m}^{sc}) = 1.
\]

Note the difference between \( \land \) and \( \lor \) in the above statement.
Proof. For each \((i, k)\), define

\[
Q_{ik}(t) \overset{\text{def}}{=} Y_{nk} \left( \int_{s_m}^{t+s_m} r_{ik}(\lambda_k, \bar{\lambda}_k, X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s) ds \right),
\]

for \(t \geq 0\). Note that (4.8) implies that \(Y_{j\ell m}(T_{j\ell m} \lor T_{sc}^{j\ell m}) = 1\) for some \(j\) and \(\ell\) and \(Y_{ikm}(T_{ikm} \lor T_{sc}^{ikm}) = 0\) for all \((i, k) \neq (j, \ell)\). In particular, this implies \(Q_{j\ell}\) is the first one among the set of counting processes \(\{Q_{ik}\}\) to jump. (This follows since for all \((i, k)\), \(r_{ik}(\lambda_k, \bar{\lambda}_k, X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s) \) will not change from \(r_{ik}(\lambda_k, \bar{\lambda}_k, X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s_m)\) until the first jump of \((X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})\).) By the definitions of \(T_{j\ell m}\) and \(T_{sc}^{j\ell m}\), it easily follows that \(Y_{j\ell m}(T_{j\ell m} \land T_{sc}^{j\ell m}) = 1\). It is trivial that no other \((i, k)\) pair can satisfy this relation. □

The following is an analogue to Lemma (4.2).

Lemma 4.3. If \((X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(s_m) = (X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s_m)\) and

\[
\sum_k Y_{nk} \left( \left( \sum_{i=1}^3 T_{ikm} \right) \lor T_{crp}^{km} \right) = 1
\]

then there is a unique \(j\) for which

\[
Y_{njm} \left( \left( \sum_{i=1}^3 T_{ijm} \right) \land T_{crp}^{jm} \right) = 1.
\]

Further, the first jump time of the Poisson process \(Y_{njm}\) occurs at some \(t_0\) satisfying

\[
t_0 < \left( \lambda_j(X_{\text{crp}}^{\pi_n}(s_m)) \lor \bar{\lambda}_j(Z_{\text{crp}}^{\pi_n}(s_m)) \right) \Delta_m.
\]

Proof. Because the two processes are equal at time \(s_m\), we have that

\[
\sum_{i=1}^3 T_{ikm} = \left( \lambda_{k_0}(X_{\text{crp}}^{\pi_n}(s_m)) \lor \bar{\lambda}_{k_0}(Z_{\text{crp}}^{\pi_n}(s_m)) \right) \Delta_m.
\]

As neither \(Z_{\text{crp}}^{\pi_n}\) nor \(X_{\text{crp}}^{\pi_n}\) changes until the first firing of \(Y_{njm}\), the claim follows. □

Based on the last two observations, we have the following lemma which will be useful in proving Proposition 4.1.

Lemma 4.4. Fix \(n\) and suppose that, for a given path of \((X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(\omega)\), \((X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(\omega)\) coupled in the way we described above,

\[
H_{m,n}(\omega) \overset{\text{def}}{=} \sum_{k=1}^R \max \left\{ \sum_{i=1}^3 Y_{ikm}(T_{ikm} \lor T_{sc}^{ikm}), Y_{km} \left( \left( \sum_{i=1}^3 T_{ikm} \right) \lor T_{crp}^{km} \right) \right\} \leq 1,
\]

for all \(m\). Then for all \(m = 0, ..., N(n)\),

\[
(X_{\text{sc}}^{\pi_n}, Z_{\text{sc}}^{\pi_n})(s_m, \omega) = (X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(s_m, \omega)
\]
Proof. We will omit \( \omega \) in the expressions. We have
\[
(X_{sc}^{\pi_n}, Z_{sc}^{\pi_n})(s_0) = (X_{crp}^{\pi_n}, Z_{crp}^{\pi_n})(s_0)
\]
by assumption. Arguing inductively, assume that
\[
(X_{sc}^{\pi_n}, Z_{sc}^{\pi_n})(s_m) = (X_{crp}^{\pi_n}, Z_{crp}^{\pi_n})(s_m).
\]
We will show that
\[
(X_{sc}^{\pi_n}, Z_{sc}^{\pi_n})(s_{m+1}) = (X_{crp}^{\pi_n}, Z_{crp}^{\pi_n})(s_{m+1})
\]
when (4.9) holds. If \( H_{m,n} = 0 \) for this \( m \), then
\[
(X_{sc}^{\pi_n}, Z_{sc}^{\pi_n})(s_{m+1}) = (X_{sc}^{\pi_n}, Z_{sc}^{\pi_n})(s_m) = (X_{crp}^{\pi_n}, Z_{crp}^{\pi_n})(s_m) = (X_{crp}^{\pi_n}, Z_{crp}^{\pi_n})(s_{m+1}),
\]
and there is nothing to do. Therefore we consider the case in which \( H_{m,n} = 1 \). More specifically, suppose that for some \( k_0 \),
\[
\max \left\{ \sum_{i=1}^{3} Y_{i k_0 m} (T_{i k_0 m} \lor T_{sc_{i k_0 m}}) Y_{k_0 m} \left( \left( \sum_{i=1}^{3} T_{i k_0 m} \lor T_{crp_{i k_0 m}} \right) \right) \right\} = 1.
\]
This means that
\[
\max \left\{ \sum_{i=1}^{3} Y_{i k m} (T_{i m} \lor T_{sc_{i k m}}) Y_{k m} \left( \left( \sum_{i=1}^{3} T_{i k m} \lor T_{crp_{i k m}} \right) \right) \right\} = 0
\]
for all \( k \neq k_0 \) by the condition (4.9). Combined with Lemmas 4.2 and 4.3, these conditions guarantee that each of the processes \( X_{sc}^{\pi_n}, Z_{sc}^{\pi_n}, X_{crp}^{\pi_n}, Z_{crp}^{\pi_n} \) jump precisely one time in the time interval \([s_m, s_{m+1}]\), and the jump happens according to reaction channel \( k_0 \) (see [13] for more details). That is, we have
\[
X_{sc}^{\pi_n}(s_{m+1}) = Z_{sc}^{\pi_n}(s_{m+1}) = X_{crp}^{\pi_n}(s_{m+1}) = Z_{crp}^{\pi_n}(s_{m+1}) = X_{sc}^{\pi_n}(s_m) + \zeta_{k_0},
\]
and we are done. \( \square \)

It is not too difficult to see that if \( \lambda_k \) and \( \bar{\lambda}_k \) are uniformly bounded for all \( k \), then we can make the condition in Lemma 4.4 hold with a probability greater than \( 1 - \epsilon \) for any \( \epsilon > 0 \) by setting \( mesh(\pi_n) \) small enough. Of course, we do not have such a uniform bound on the intensity functions. Also, note that Lemma 4.4 does not imply that
\[
(X_{crp}^{\pi_n}, Z_{crp}^{\pi_n})(t) = (X_{sc}^{\pi_n}, Z_{sc}^{\pi_n})(t) \text{ for } t \in [s_m, s_{m+1}]
\]
even if the conditions of the lemma are met, as the processes may (and most likely will) jump at slightly different times. However, we trivially note that under the conditions of Lemma 4.4,
\[
(X_{crp}^{\pi_n}, Z_{crp}^{\pi_n})(t) = (X_{sc}^{\pi_n}, Z_{sc}^{\pi_n})(t) \text{ for all } t \in [s_m, s_{m+1}]
\]
if neither \( (X_{crp}^{\pi_n}, Z_{crp}^{\pi_n}) \) nor \( (X_{sc}^{\pi_n}, Z_{sc}^{\pi_n}) \) jump at all in \([s_m, s_{m+1}]\).

We are now in a position to prove Proposition 4.1.
Proof of Proposition 4.1. We first recall that \( t = (t_1, \ldots, t_K) \) for some \( K \in \{1, 2, \ldots\} \). Next, we define
\[
K_0^n \overset{\text{def}}{=} \{ m \in \{0, \ldots, N(n) - 1\} : \{ t_j \}_{j=1}^K \cap [s_m, s_{m+1}) \neq \emptyset \}.
\]

Fix \( \epsilon > 0 \). As we remarked around (4.5), it suffices to show that, for large enough \( n \),
\[
P \left( \max_{i=0,\ldots,K} |(X^{\pi_n}_{sc}(t_i), Z^{\pi_n}_{sc}(t_i)) - (X^{\pi_n}_{crp}(t_i), Z^{\pi_n}_{crp}(t_i))| > 0 \right) < \epsilon,
\]
where we converted the \( \gamma \) in (4.5) to a zero as our processes take values in \( \mathbb{Z}^d \).

We will resort to a localization argument and take advantage of the fact that \( X \) and \( Z \) are both nonexplosive. Let \( M > 0 \), and let \( H_{m,n} \) be defined as in Lemma 4.4. Define
\[
A_n(t) \overset{\text{def}}{=} \{ \omega : H_{m,n}(\omega) \leq 1 \text{ if } m \notin K_0^n \text{ and } H_{m,n}(\omega) = 0 \text{ if } m \in K_0^n \} \tag{4.11}
\]
and
\[
B_{M,n} \overset{\text{def}}{=} \{ \omega : \max_{s \leq T} \sup_{s \leq T} \lambda_k(X^{\pi_n}_{sc}(s)), \sup_{s \leq T} \lambda_k(Z^{\pi_n}_{sc}(s)), \sup_{s \leq T} \lambda_k(Z^{\pi_n}_{crp}(s)) \leq M \}. \tag{4.12}
\]

Note that by the non-explosivity of the processes, the supremums are achieved everywhere they appear above. By Lemma 4.4 and the arguments in and around (4.10), we have that
\[
A_n(t) \subset \{(X^{\pi_n}_{sc}, Z^{\pi_n}_{sc})(t) = (X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(t)\}.
\]

Therefore
\[
P((X^{\pi_n}_{sc}, Z^{\pi_n}_{sc})(t) \neq (X^{\pi_n}_{crp}, Z^{\pi_n}_{crp})(t)) \leq P(A_n^C(t)) = P(A_n^C(t) \cap B_{M,n}) + P(A_n^C(t) \cap B_{M,n}^C). \tag{4.13}
\]

We handle the two pieces on the right hand side of (4.13) separately.

For the second term in (4.13), we first note that
\[
B_{M,n}^C \subset \{ \sup_{s \leq T} \lambda_k(X^{\pi_n}_{sc}(s)) > M \} \cup \{ \sup_{s \leq T} \lambda_k(Z^{\pi_n}_{sc}(s)) > M \} \cup
\{ \sup_{s \leq T} \lambda_k(X^{\pi_n}_{crp}(s)) > M \} \cup \{ \sup_{s \leq T} \lambda_k(Z^{\pi_n}_{crp}(s)) > M \}.
\]

Now, recall that the marginal distributions of \( X^{\pi_n}_{crp} \) and \( X^{\pi_n}_{sc} \) are the same as the marginal distribution of \( X \), and that the same goes for \( Z^{\pi_n}_{crp} \) and \( Z^{\pi_n}_{sc} \) compared with \( Z \). Therefore, for all \( n \) we have
\[
P(B_{M,n}^C) \leq 2 \times \left[ P(\sup_{s \leq T} \lambda_k(X_s)) > M \right] + P(\sup_{s \leq T} \lambda_k(Z_s)) > M \}. \tag{4.14}
\]

By the monotone convergence theorem and the fact that the processes are all non explosive, the right hand side of (4.14) will tend to 0 as \( M \to \infty \). Therefore, we can take \( M \) large enough so that the second piece of (4.13) is smaller than \( \epsilon/2 \). We fix this \( M \), and turn attention to the first term on the right hand side of (4.13).
We consider the localized version of $H$. In particular, for our fixed $M > 0$ let

$$H^M_{m,n}(\omega) \equiv \sum_{k=1}^R \max \left\{ 3 \sum_{i=1}^{\frac{3}{4}} Y^n_{ikm}(M \Delta_m(\pi_n)), Y^n_{km}(3M \Delta_m(\pi_n)) \right\}.$$ 

Then it is clear that, for any $q > 0$,

$$\{H_{m,n} > q\} \cap B_{M,n} \subset \{H^M_{m,n} > q\} \cap B_{M,n} \subset \{H^M_{m,n} > q\},$$

and therefore

$$P(A^C_n(t) \cap B_{M,n}) \leq P(H^M_{m,n} > 1 \text{ for some } m \notin K^n_0 \text{ OR } H^M_{m,n} > 0 \text{ for some } m \in K^n_0) \leq \sum_{m \notin K^n_0} P(H^M_{m,n} > 1) + \sum_{m \in K^n_0} P(H^M_{m,n} > 0). \quad (4.15)$$

To handle these two pieces, we recall two basic facts pertaining to Poisson random variables. First, if we denote by $W(\Lambda) \sim \text{Poisson}(\Lambda)$ then

$$P(W(\Lambda) > 1) = 1 - \exp(-\Lambda)(1 + \Lambda) \leq 1 - (1 - \Lambda)(1 + \Lambda) = \Lambda^2,$$

where we used the inequality $\exp(-x) \geq 1 - x$. Second, and using the same inequality,

$$P(W(\Lambda) > 0) = 1 - \exp(-\Lambda) \leq \Lambda.$$

Now note that

$$P(\{H^M_{m,n} > q\}) \leq P(W(6RM\Delta_m(\pi_n)) > q).$$

Hence, if $\text{mesh}(\pi_n) = \delta_n$ then by the two facts above and (4.15), we have

$$P(A^C_n(t) \cap B_{M}) \leq \sum_{m \notin K^n_0} (6RM\Delta_m(\pi_n))^2 + \sum_{m \in K^n_0} (6RM\Delta_m(\pi_n)) \leq (6RM)^2 \delta_n \sum_{m \notin K^n_0} \frac{\Delta_m(\pi_n)}{\delta_n} \Delta_m(\pi_n) + 6RM|K^n_0|\delta_n \leq (6RM)^2 \delta_n T + 6RM|K^n_0|\delta_n, \quad (4.16)$$

where in the third inequality we used that $\frac{\Delta_m(\pi_n)}{\delta_n} < 1$, which follows by the definition of mesh. We can now take $n$ large enough so that (4.16) is less than $\epsilon/2$. We may now conclude that for such $n$,

$$P(|(X^{\pi_n}_{\text{sc}}, Z^{\pi_n}_{\text{sc}})(t) - (X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})(t)| > 0) < \epsilon,$$

as required. \Box

Of course, it is to be hoped that Proposition 4.1 implies the weak convergence of $(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})$ to $(X_{\text{sc}}, Z_{\text{sc}})$ in Skorohod’s sense. However, in order to achieve such convergence we require that $\{(X^{\pi_n}_{\text{crp}}, Z^{\pi_n}_{\text{crp}})\}$ be precompact in the Skorohod topology. The following result is available on page 128 of [9].
Theorem 4.5. Suppose that \( \{ X_n \} \) is a series of processes with sample paths in \( D_E[0, \infty) \). Then the law of \( \{ X_n \} \) is precompact if and only if the following two conditions hold:

1. For each \( \eta, t \) there is a compact set \( \Gamma_{\eta,t} \) on \( E \) such that
   \[
   \inf_n P(X_n(t) \in \Gamma_{\eta,t}) \geq 1 - \eta
   \]
2. For each \( \eta > 0 \) and \( T > 0 \) there exists a \( \delta > 0 \) such that
   \[
   \sup_n P(w'(X_n, \delta, T) \geq \eta) < \eta
   \]
   where
   \[
   w'(X, \delta, T) \overset{def}{=} \inf_{\pi} \max_i \sup_{a,b \in [t_i, t_{i+1})} |X(a) - X(b)|
   \]
   with \( \pi \) ranging over all partitions of \([0, T]\) with all \( t_{i+1} - t_i > \delta \).

Unfortunately the conditions of Theorem 4.5 do not hold in general for our set of processes \( \{ (X^{\pi_n}_{crp}, Z^{\pi_n}_{crp}) \} \). To see this, first note two things: (i) that for jump processes whose jump sizes are bounded below, for example by integer values in our present setting, for small enough \( \eta > 0 \) we have
   \[
   \{ w'((X, Z), \delta, T) < \eta \} = \{ w'((X, Z), \delta, T) = 0 \},
   \]
and (ii) the event \( w'((X, Z), \delta, T) = 0 \) can be achieved only if the minimum time between jumps of \((X, Z)\) is greater than \( \delta \). To understand this last statement, simply note that if the minimum time between jumps is less than \( \delta \), then for any partition \( \pi \) satisfying \( t_{i+1} - t_i > \delta \) for all \( i \), the process must change by at least the smallest jump size \( \min_k |\zeta_k| \) in our case in some interval of the partition. Conversely, if the minimum holding time of the process is greater than \( \delta \), then we achieve a value of 0 for \( w' \) by choosing \( \pi \) so that the jump times correspond with a subset of the partition times \( t_i \). Next, in order to definitively show how the above facts cause Theorem 4.5 to not hold in our situation, we consider the following simple linear growth system

\[
A \rightarrow 2A,
\]
which models increases in \( A \) as a counting process with a linear intensity. We consider the corresponding processes \( (X_{sc}, Z_{sc}) \) and \( (X^{\pi_n}_{crp}, Z^{\pi_n}_{crp}) \) with
   \[
   \lambda_1(x) = \theta x, \quad \tilde{\lambda}_1(x) = (\theta + h)x,
   \]
and initial condition
   \[
   X_{sc}(0) = Z_{sc}(0) = X_{crp}^{\pi_n}(0) = Z_{crp}^{\pi_n}(0) > 0.
   \]
For any \( \delta > 0 \), the probability that the processes \( X_{sc} \) and \( Z_{sc} \) jump simultaneously in the time period \([0, \delta]\) and that their simultaneous jump is the first jump for either process is

\[
\alpha_\delta \overset{def}{=} \frac{\theta}{\theta + h} \left( 1 - e^{-(\theta + h)X(0)\delta} \right) > 0.
\]
By the arguments we made in the proof above, for any \( \epsilon > 0 \) there exists some \( M \) such that if \( n > M \), then with probability greater than \( \alpha_\delta - \epsilon \), both \( X_{\text{crp}}^{\pi_n} \) and \( Z_{\text{crp}}^{\pi_n} \) will also make a first jump in \([0, \delta]\). However, with a probability of one, \( X_{\text{crp}}^{\pi_n} \) and \( Z_{\text{crp}}^{\pi_n} \) never jump at the exact same time. Hence, when they jump in the time interval \([0, \delta]\), we have

\[
\sup_{a, b \in [0, \delta]} |(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(a) - (X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})(b)| \geq 1.
\]

This in particular means that for any \( 0 < \eta < \frac{\alpha_\delta}{\delta} \),

\[
\sup_n P(\{w'((X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n}), \delta, T) \geq \eta\}) \geq \alpha_\delta,
\]

and the law of \( \{(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})\} \) fails to be precompact.

### 4.2 Weak Convergence in the product Skorohod topology

We can achieve the weak convergence of \((X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})\) to \((X_{\text{sc}}, Z_{\text{sc}})\) in a weaker topology than that considered in the previous subsection. In this subsection, we prove the following.

**Proposition 4.6.** Suppose \( X \) and \( Z \) are both non-explosive, càdlàg process as given above. Let \( D_{\mathbb{R}^2}[0, \infty) \) be the Skorohod Space as defined in [9]. Consider the product topology on \( D := D_{\mathbb{R}^d}[0, \infty) \times D_{\mathbb{R}^d}[0, \infty) \).

Also, let \( \pi_n = \{s^n_j\} \) be a sequence of partitions of \([0, \infty)\) such that

\[
\text{mesh}(\pi_n) = \max_{j < \infty} (s^n_j - s^n_{j-1}) \to 0, \quad \text{as } n \to \infty.
\]

Then for all \( f : D \to \mathbb{R} \) that are bounded and continuous,

\[
\mathbb{E}[f(X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n})] \to \mathbb{E}[f(X_{\text{sc}}, Z_{\text{sc}})].
\]

That is, \((X_{\text{crp}}^{\pi_n}, Z_{\text{crp}}^{\pi_n}) \to (X_{\text{sc}}, Z_{\text{sc}})\) weakly in the product Skorohod topology.

We would like to emphasize that the test function \( f \) considered above maps a path in \( D \) to \( \mathbb{R} \). The test functions for Proposition 4.1, on the other hand, are evaluated at discrete time points. The following theorem in [9] (Section 3.7, Proposition 7.1) shows that more sets are open in the topology of \( D \) than in the Skorohod topology on \( D_{\mathbb{R}^2}[0, \infty) \) (That is, the topology of \( D \) is a stronger topology, and permits fewer continuous test functions).

**Theorem 4.7.** For each \( t \geq 0 \) and a polish space \( E \), define \( \pi_t : D_{E}[0, \infty) \to E \) by \( \pi_t(x) = x(t) \). Then the Borel \( \sigma \)-algebra \( \mathcal{W} \) of \( D_{E}[0, \infty) \) is given by

\[
\mathcal{W} = \sigma(\pi_t, t \in D).
\]

where \( D \) is any dense subset of \([0, \infty)\).

**Corollary 4.8.** The Borel \( \sigma \)-algebra of \( D \) is given by

\[
\sigma(\pi_t \times \pi_s, s, t \in D).
\]

where \( (\pi_t \times \pi_s)(x, z) = (x(t), z(s)) \).
Let\( f \) be a uniform bound on the (finite) set of bounded functions \( \{h \in B: \|h\| < M\} \). Then

\[
\sup_{(s_i, t_j)} |f_{\pi}(s_i, t_j)| < \frac{1}{M}
\]

where the limit follows from the proof of Proposition 4.1.

**Proof.** We first note that since \( (X_{\text{crp}}, Z_{\text{crp}}) \) and \( (X_{\text{sc}}, Z_{\text{sc}}) \) have the same distribution, we can conclude

\[
\mathbb{E} \left[ \prod_{i=0}^{m_1} f_i((X_{\text{crp}}(s_i)) \prod_{j=0}^{m_2} g_j(Z_{\text{crp}}(t_j))) \right] = \mathbb{E} \left[ \prod_{i=0}^{m_1} f_i((X_{\text{sc}}(s_i)) \prod_{j=0}^{m_2} g_j(Z_{\text{sc}}(t_j))) \right].
\]

Let \( M \) be a uniform bound on the (finite) set of bounded functions \( \{f_i, g_j\} \). Then,

\[
\left| \mathbb{E} \left[ \prod_{i=0}^{m_1} f_i((X_{\text{crp}}(s_i)) \prod_{j=0}^{m_2} g_j(Z_{\text{crp}}(t_j))) - \prod_{i=0}^{m_1} f_i((X_{\text{sc}}(s_i)) \prod_{j=0}^{m_2} g_j(Z_{\text{sc}}(t_j))) \right] \right| 
\leq 2M^{m_1 + m_2} P \left( \max_{\tau \in (s_i) \cup (t_j)} \left| (X_{\text{crp}}(\tau), Z_{\text{crp}}(\tau)) - (X_{\text{sc}}(\tau), Z_{\text{sc}}(\tau)) \right| \neq 0 \right)
\rightarrow 0,
\]

where the limit follows from the proof of Proposition 4.1. \( \square \)

Now put everything together to prove Proposition 4.6.

**Proof of Proposition 4.6.** We would like to show that \( (X_{\text{crp}}, Z_{\text{crp}}) \) satisfies the conditions in Theorem 4.5 with respect to the space \( D \). To show that \( (X_{\text{crp}}, Z_{\text{crp}}) \) is precompact, we need to show that, for any \( \epsilon > 0 \), there exists a compact set \( C^* \) such that

\[
\inf_n P((X_{\text{crp}}, Z_{\text{crp}}) \in C^*) > 1 - \epsilon.
\]

Recall that the marginals satisfy \( X \sim X_{\text{crp}} \) and \( Z \sim Z_{\text{crp}} \). If \( P(X \in A^*) > 1 - \epsilon/2 \) and \( P(Z \in B^*) > 1 - \epsilon/2 \) for compact \( A^*, B^* \subset D[0, \infty) \), then \( P((X, Z) \in A^* \times B^*) > 1 - \epsilon \) with \( A^* \times B^* \) compact in the product topology. Therefore it suffices to show that \( X \) and \( Z \) both separately satisfy the conditions in Theorem 4.5. Since \( X \) is a nonexplosive pure jump process, it clearly passes the first condition. Also, recall that \( X \) is constructed with \( R \) channels of Poisson processes. Hence for any \( T > 0 \) and \( M > 0 \),

\[
P(w'(X, \delta, T) > 0) \leq P \left( w'(X, \delta, T) > 0, \sup_{k=1,..,R,s<T} \lambda_k(X(s)) \leq M \right) + P \left( \sup_{k=1,..,R,s<T} \lambda_k(X(s)) > M \right)
\]

\[
= P \left( w'(Y(MR^*), \delta, T) > 0 \right) + P \left( \sup_{k=1,..,R,s<T} \lambda_k(X(s)) > M \right)
\]

(4.17)
where \( Y(MR) \) is a rate \( MR \) poisson process. Since \( X \) is non explosive, we may take \( M \) large enough to control the second piece, and for this \( M \) we can choose \( \delta \) small enough to control the first piece. That is, \( \lim_{\delta \to 0} P(w'(X,\delta,T) > 0) = 0 \). This tells us that \( X \) also passes the second condition. The same procedure works for \( Z \). Thus, \( \{(X_{\text{crp}},Z_{\text{crp}})\} \) is precompact in the product topology. Together with Corollary 4.9, we are done since
\[
\mathbb{E} \left[ \prod_{i=0}^{m_1} f_i(X_{sc}(s_i)) \prod_{j=0}^{m_2} g_j(Z_{sc}(t_j)) \right], \quad \forall \ \{s_i\}, \{t_j\} \subset [0, \infty), f_i,g_i \in \overline{C}(\mathbb{R}^d)
\]
characterizes the law of \( (X_{sc},Z_{sc}) \). To see this, suppose that for some other process \( (X^*,Z^*) \),
\[
\mathbb{E} \left[ \prod_{i=0}^{m_1} f_i(X_{sc}(s_i)) \prod_{j=0}^{m_2} g_j(Z_{sc}(t_j)) \right] = \mathbb{E} \left[ \prod_{i=0}^{m_1} f_i(X^*(s_i)) \prod_{j=0}^{m_2} g_j(Z^*(t_j)) \right],
\]
for all \( \{s_i\}, \{t_j\} \subset [0, \infty), f_i,g_i \in \overline{C}(\mathbb{R}^d) \). Note that we can approximate the indicator functions of open sets \( \{A_i\}, \{B_j\} \) of \( \mathbb{R}^d \) with increasing sequences of functions \( f_i^n, g_i^n \), respectively. By the monotone convergence theorem, we have
\[
\mathbb{E} \left[ \prod_{i=0}^{m_1} 1_{A_i}(X_{sc}(s_i)) \prod_{j=0}^{m_2} 1_{B_j}(Z_{sc}(t_j)) \right] = \mathbb{E} \left[ \prod_{i=0}^{m_1} 1_{A_i}(X^*(s_i)) \prod_{j=0}^{m_2} 1_{B_j}(Z^*(t_j)) \right],
\]
for any open sets \( \{A_i\}, \{B_j\} \) in \( \mathbb{R}^d \). Now we apply a version of monotone class theorem:

**Theorem 4.10.** Let \( H \) be a linear space of bounded functions on \( \Omega \) that contains constants, and let \( \mathcal{L} \) be a collection of subsets of \( \Omega \) such that if \( A,B \in \mathcal{L} \) then \( A \cap B \in \mathcal{L} \). Suppose \( \chi_A \in H \) for any \( A \in \mathcal{L} \). Also suppose that \( \{f_n\} \subset H, f_1 \leq f_2 \leq \cdots \) and \( \sup_n f_n \leq c \) for some constant \( c \) imply that \( f \equiv \text{bp}-\lim_{n \to \infty} f_n \in H \). Then \( H \) contains all bounded \( \sigma(\mathcal{L}) \) measurable functions.

For us, we put \( H \) to be the set of all Borel measurable functions \( f : D_E[0,\infty) \times D_E[0,\infty) \to \mathbb{R} \) for which
\[
\mathbb{E}[f(X,Z)] = \mathbb{E}[f(X^*,Z^*)].
\]
(4.18)

Next,
\[
\mathcal{L} = \{ \text{Any finite intersections of } (\pi_t \times \pi_s)^{-1}(A) : t \in [0,\infty), A \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d) \}.
\]

Corollary 4.9 shows that \( \chi_U \subset H \) for any \( U \in \mathcal{L} \). Moreover, the last condition on \( H \) is guaranteed by the monotone convergence theorem. This shows that any \( \sigma(\mathcal{L}) \) measurable function is in \( H \). By Corollary 4.8 this implies that (4.18) holds for any function \( f \) that is borel measurable on \( D_E[0,\infty) \times D_E[0,\infty) \).

While the results presented so far pertain to the specific couplings found in the numerical analysis literature, a slightly more general theorem can be achieved by following an identical line of reasoning.

---

\footnote{For the definition of bounded point wise convergence, we refer the reader to the Appendices in [9]. For our specific case, all we need is \( \|f_n\|_{\infty} < \infty \) and \( f_n(x) \to f(x) \) for every \( x \in E \).}
Theorem 4.11. For \( i \in \{1, 2, 3\} \) and \( k \in \{1, \ldots, R\} \), let \( r_{ik} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0} \) be a nonnegative measurable function. Suppose that \( \{\pi_n\} \) is a sequence of partitions of \([0, \infty)\) for which \( \text{mesh}(\pi_n) \to 0 \), as \( n \to \infty \). Define \((X_{sc}, Z_{sc})\) and \((X_{crp}^n, Z_{crp}^n)\) via

\[
X_{sc}(t) = X(0) + \sum_{k=1}^{R} \left\{ Y_{1k} \left( \int_0^t r_{1k}(X_{sc}, Z_{sc})(s) ds \right) + Y_{2k} \left( \int_0^t r_{2k}(X_{sc}, Z_{sc})(s) ds \right) \right\} \zeta_k
\]

\[
Z_{sc}(t) = Z(0) + \sum_{k=1}^{R} \left\{ Y_{1k} \left( \int_0^t r_{1k}(X_{sc}, Z_{sc})(s) ds \right) + Y_{3k} \left( \int_0^t r_{3k}(X_{sc}, Z_{sc})(s) ds \right) \right\} \zeta_k,
\]

and

\[
X_{crp}^n(t) = X(0) + \sum_{m=0}^{\infty} \sum_{k=1}^{R} Y_{km} \left( \int_{t \wedge s_m}^{t \wedge s_{m+1}} \{ r_{1k}(X_{crp}^n, Z_{crp}^n)(s) + r_{2k}(X_{crp}^n, Z_{crp}^n)(s) \} ds \right) \zeta_k
\]

\[
Z_{crp}^n(t) = Z(0) + \sum_{m=0}^{\infty} \sum_{k=1}^{R} Y_{km} \left( \int_{t \wedge s_m}^{t \wedge s_{m+1}} \{ r_{1k}(X_{crp}^n, Z_{crp}^n)(s) + r_{3k}(X_{crp}^n, Z_{crp}^n)(s) \} ds \right) \zeta_k,
\]

where all notations are as before. Finally, we suppose that all processes are non-explosive. Then, \((X_{crp}^n, Z_{crp}^n) \to (X_{sc}, Z_{sc})\) weakly in the product Skorohod topology.

Proposition 4.6 is therefore a special case of Theorem 4.11 in which each \( r_{ik} \) depends on \( \lambda_k \) and \( \tilde{\lambda}_k \) in a specific way.

5 Numerical examples

In this section, we provide two numerical examples demonstrating the convergence of the local-CRP coupling to that of the split coupling. Based upon our motivation in terms of variance reduction, we focus upon the convergence of the variance between the coupled processes.

Example 1. We begin by considering a basic model of gene transcription and translation, where the model tracks the counts for the numbers of genes \((G)\), mRNA molecules \((M)\), and proteins \((P)\) in the system. We suppose that the system can undergo the following possible reactions,

\[
\begin{align*}
G & \to G + M \\
M & \to M + P \\
M & \to 0 \\
P & \to 0,
\end{align*}
\]

where, for example, reaction \((R1)\) implies a net change to the system of one extra mRNA molecule. Since no reaction changes the number of genes present in the system, we may take that to be a fixed quantity. Hence, there are two dynamic components, and the stochastic model for this system is

\[
X(t) = X(0) + Y_1 \left( \int_0^t \lambda_1(X(s)) ds \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + Y_2 \left( \int_0^t \lambda_2(X(s)) ds \right) \begin{bmatrix} 0 \\ 1 \end{bmatrix} + Y_3 \left( \int_0^t \lambda_3(X(s)) ds \right) \begin{bmatrix} -1 \\ 0 \end{bmatrix} + Y_4 \left( \int_0^t \lambda_4(X(s)) ds \right) \begin{bmatrix} 0 \\ -1 \end{bmatrix},
\]

19
where $X_1$ counts the numbers of mRNA molecules, and $X_2$ counts the numbers of proteins. We now let $X$ be the process with intensity functions

$$\lambda_1(x) = 2, \quad \lambda_2(x) = 10x_1, \quad \lambda_3(x) = (1/4 + 1/80)x_1, \quad \lambda_4(x) = x_2,$$

and let $Z$ be the process with intensity functions

$$\lambda_1(x) = 2, \quad \lambda_2(x) = 10x_1, \quad \lambda_3(x) = (1/4 - 1/80)x_1, \quad \lambda_4(x) = x_2.$$

These are reasonable choices, for example, if we were attempting to estimate the sensitivity of some statistic with respect to the rate parameter for the third intensity function evaluated at $1/4$.

Let $\pi_n$ be a partition of $[0, 30]$ into $n$ equally sized intervals. In Figure 1, we plot numerical estimates of $\text{Var}(X_{sc}(t) - Z_{sc}(t))$, $\text{Var}(X_{crp}(t) - Z_{crp}(t))$, and $\text{Var}(X_{\pi n_{crp}}(t) - Z_{\pi n_{crp}}(t))$, for $n \in \{2, 6, 30, 300\}$, over the time period $[0, 30]$. The estimates were achieved via Monte Carlo methods with 10,000 sample paths. We observe the uniform convergence of $\text{Var}(X_{\pi n_{crp}}(\cdot) - Z_{\pi n_{crp}}(\cdot))$ to $\text{Var}(X_{sc}(\cdot) - Z_{sc}(\cdot))$ as $\text{mesh}(\pi_n) \to 0$. We also observe a sharp drop in the variance of $X_{\pi n_{crp}}(\cdot) - Z_{\pi n_{crp}}$ at the “resetting” of the Poisson processes, which occur at the end of each interval of the discretization $\pi_n$.

**Example 2.** Consider a simple quadratic birth and death model

$$\emptyset \to 2A \quad \text{(r1)}$$
$$2A \to \emptyset \quad \text{(r2)}$$
with initial count $X(0)$ given by a Poisson random variable with parameter 15. We can model the dynamics of this system with the stochastic equations

$$X(t) = X(0) + 2Y_1 \left( \int_0^t \lambda_1(X(s))ds \right) - 2Y_2 \left( \int_0^t \lambda_2(X(s))ds \right),$$

where

$$\lambda_1(x) = 400, \quad \lambda_2(x) = kx(x-1),$$

and where $k$ is a parameter of the model. We consider the model $X$ with $k = 0.1 + 1/25$ and the model $Z$ with $k = 0.1 - 1/25$. Further, we let the initial conditions of $X$ and $Z$ be independent Poisson random variables with a parameter of 15 (that is, the initial conditions of $X$ and $Z$ are independent from each other). Let $\pi_n$ be a partition of $[0,1]$ into $n$ equally sized intervals. In Figure 2, we plot numerical estimates of $\text{Var}(X_{sc}(t) - Z_{sc}(t))$, $\text{Var}(X_{crp}(t) - Z_{crp}(t))$, and $\text{Var}(X_{crp}(t) - Z_{crp}(t))$, for $n \in \{2, 4, 8, 100\}$, over the time period $[0,1]$. The estimates were achieved via Monte Carlo methods with 5,000 sample paths. We again observe the sharp drop in variance at the “resetting” times of the processes.

6 Discussion

The stochastic models finding widespread use in the cell biology literature are typically immensely complicated, and computational methods often provide the only effective way
to probe the dynamics. As Persi Diaconis recently noted [8], this presents mathematicians with an opportunity to make contributions by explicitly studying the different simulation and computational algorithms themselves. Such analyses will not only shed light on which methods to use in different contexts, but will inevitably lead to a deeper understanding of the underlying processes, and hence to better computational methods.

In this work we have clarified the connection between two couplings commonly found in the computational cell biology literature and, in particular, showed that the split coupling can be regarded as a natural limit of a localized version of the CRP coupling. There are other interesting ways to understand the split coupling. For example, Arampatzis and Katsoulakis [7] recently studied a group of couplings that is included in the family of general split couplings considered in Theorem 4.11. They note that for each test function \( f \) there is an optimal choice for the function \( r_{1k}(\lambda_k, \tilde{\lambda}_k, R, L)(s) \) in (3.1) that minimizes the variance of the finite difference \( \mathbb{E}[f(X_t) - f(Z_t)] \) in the setting of (3.2). When the test function is \( f(x) = x \), the correct choice of \( r_{1k} \) is the one given in (3.1), which yields the split coupling.

Acknowledgments. We thank Thomas Kurtz for several illuminating discussions and for reading an early version of this work.

References

[1] David F. Anderson, *An efficient finite difference method for parameter sensitivities of continuous time Markov chains*, SIAM Journal on Numerical Analysis 50 (2012), no. 5, 2237 – 2258.

[2] David F. Anderson, Arnab Ganguly, and Thomas G. Kurtz, *Error analysis of tau-leap simulation methods*, Annals of Applied Probability 21 (2011), no. 6, 2226 – 2262.

[3] David F. Anderson and Desmond J. Higham, *Multi-level Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics*, SIAM: Multiscale Modeling and Simulation 10 (2012), no. 1, 146 – 179.

[4] David F. Anderson, Desmond J. Higham, and Yu Sun, *Complexity analysis of multilevel monte carlo tau-leaping*, submitted, 2013.

[5] David F. Anderson and Thomas G. Kurtz, *Continuous time markov chain models for chemical reaction networks*, Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology (H. Koeppel et al., ed.), Springer, 2011, pp. 3–42.

[6] David F. Anderson and Elizabeth Skubak Wolf, *A finite difference method for estimating second order parameter sensitivities of discrete stochastic chemical reaction networks*, J. Chem. Phys. 137 (2012), no. 22, 224112.

[7] Georgios Arampatzis and Markos A. Katsoulakis, *Goal-oriented sensitivity analysis for lattice kinetic Monte Carlo simulations*, Available on arXiv:1311.5710v2.

[8] Persi Diaconis, *Some things we’ve learned (about markov chain monte carlo)*, Bernoulli 19 (2013), no. 4, 1294 – 1305.

[9] Stewart N. Ethier and Thomas G. Kurtz, *Markov processes: Characterization and convergence*, John Wiley & Sons, New York, 1986.
[10] Mike B. Giles, *Multilevel Monte Carlo path simulation*, Operations Research 56 (2008), 607–617.

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

[12] ———, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem. 81 (1977), no. 25, 2340–2361.

[13] Masanori Koyama, *Analysis of stochastically modeled biochemical processes with applications to numerical methods*, Ph.D. thesis, University of Wisconsin at Madison, December, 2013.

[14] Thomas G. Kurtz, *Approximation of population processes*, CBMS-NSF Reg. Conf. Series in Appl. Math.: 36, SIAM, 1981.

[15] ———, *Representation and approximation of counting processes*, Lecture Notes in Control and Information Science 42 (1982), 177–191.

[16] Muruhan Rathinam, Patrick W. Sheppard, and Mustafa Khammash, *Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks*, Journal of Chemical Physics 132 (2010), 034103.

[17] Rishi Srivastava, David F. Anderson, and James B. Rawlings, *Comparison of finite difference based methods to obtain sensitivities of stochastic chemical kinetic models*, Journal of Chemical Physics 138 (2013), no. 7, 074110.