

## COMPUTATIONAL COMPLEXITY ANALYSIS FOR MONTE CARLO APPROXIMATIONS OF CLASSICALLY SCALED POPULATION PROCESSES\*

DAVID F. ANDERSON<sup>†</sup>, DESMOND J. HIGHAM<sup>‡</sup>, AND YU SUN<sup>§</sup>

**Abstract.** We analyze and compare the computational complexity of different simulation strategies for Monte Carlo in the setting of classically scaled population processes. This allows a range of widely used competing strategies to be judged systematically. Our setting includes stochastically modeled biochemical systems. We consider the task of approximating the expected value of some path functional of the state of the system at a fixed time point. We study the use of standard Monte Carlo when samples are produced by exact simulation and by approximation with tau-leaping or an Euler–Maruyama discretization of a diffusion approximation. Appropriate modifications of recently proposed multilevel Monte Carlo algorithms are also studied for the tau-leaping and Euler–Maruyama approaches. In order to quantify computational complexity in a tractable yet meaningful manner, we consider a parameterization that, in the mass action chemical kinetics setting, corresponds to the classical system size scaling. We base the analysis on a novel asymptotic regime where the required accuracy is a function of the model scaling parameter. Our new analysis shows that, under the specific assumptions made in the manuscript, if the bias inherent in the diffusion approximation is smaller than the required accuracy, then multilevel Monte Carlo for the diffusion approximation is most efficient, besting multilevel Monte Carlo with tau-leaping by a factor of a logarithm of the scaling parameter. However, if the bias of the diffusion model is greater than the error tolerance or if the bias cannot be bounded analytically, multilevel versions of tau-leaping are often the optimal choice.

**Key words.** computational complexity, variance, reaction networks, classical scaling, coupling, multilevel Monte Carlo

**AMS subject classifications.** 60H35, 65C05, 92C40

**DOI.** 10.1137/17M1138169

**1. Introduction.** For some large  $N_0 > 0$ , we consider a continuous time Markov chain satisfying the stochastic equation

$$(1) \quad X^{N_0}(t) = X^{N_0}(0) + \sum_{k=1}^K \frac{1}{N_0} Y_k \left( N_0 \int_0^t \lambda_k(X^{N_0}(s)) ds \right) \zeta_k,$$

where  $X^{N_0}(t) \in \mathbb{R}^d$ ,  $K < \infty$ , the  $Y_k$  are independent unit Poisson processes, and, for each  $k$ ,  $\zeta_k \in \mathbb{R}^d$  and  $\lambda_k : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$  satisfies mild regularity conditions. For a given path functional  $f$ , we consider the task of numerically approximating  $\mathbb{E}[f(X^{N_0}(\cdot))]$ , in the sense of confidence intervals, to some fixed tolerance  $\varepsilon_0 < 1$ . Specifically, we consider the computational complexity, as quantified by the number of random

---

\*Received by the editors July 12, 2017; accepted for publication (in revised form) May 30, 2018; published electronically August 7, 2018.

<http://www.siam.org/journals/mms/16-3/M113816.html>

**Funding:** The first author was supported by grant support from NSF-DMS-1318832 and Army Research Office W911NF-14-1-0401. The second author was supported by grant support from a Wolfson/Royal Society Research Merit Award and an EPSRC Fellowship with reference EP/M00158X/1. EPSRC data statement: Not relevant.

<sup>†</sup>Department of Mathematics, University of Wisconsin, Madison, WI 53706 (anderson@math.wisc.edu).

<sup>‡</sup>Department of Mathematics and Statistics, University of Strathclyde, UK (d.j.higham@maths.strath.ac.uk).

<sup>§</sup>Facebook, Inc., 1 Hacker Way, Menlo Park, CA 94025 (ysunfb@fb.com).

variables utilized, required by different Monte Carlo schemes to achieve a root mean squared error of  $\varepsilon_0$ . For concreteness, we will assume throughout that the path functional  $f$  depends upon  $X^{N_0}(\cdot)$  only on the compact time interval  $[0, T]$ .

The class of models of the form (1) has a long history in terms of modeling [12, 13, 14, 31], analysis [9, 27, 28], and computation [18, 19]. The framework covers many application areas, including population dynamics [32], queueing theory [33], and several branches of physics [15]. In recent years, chemical and biochemical kinetics models in systems biology [34] have been the driving force behind a resurgence of activity in algorithmic developments, including tau-leaping [20] and its multilevel extension [4, 5]. In this setting, the parameter  $N_0$  in (1) can represent Avogadro's number multiplied by the volume, and in this classical scaling, species are measured in moles per liter. More generally, however,  $N_0$  can just be considered a large number, often of the order 100s or 1000s.

In section 2, we discuss some of the issues involved in quantifying computational complexity in the present setting, and introduce a novel scaling regime in which clear-cut comparisons can be made. Further, the specific assumptions utilized throughout the manuscript are presented, and a high-level summary of our main conclusions is presented. In section 3, we summarize two widely used approximation methods for the model (1): the tau-leap discretization method and the Langevin or diffusion approximation. In section 4, we quantify the computational complexity of using exact simulation, tau-leaping, and simulation of the diffusion equation with standard Monte Carlo for approximating  $\mathbb{E}[f(X^{N_0}(\cdot))]$  to a desired tolerance under our assumptions. Further, in subsection 4.2, we review the more recent multilevel methods and quantify the benefits of their use in both the tau-leaping and the diffusion scenarios. In section 5, we provide numerical examples demonstrating our main conclusions. In section 6, we close with some brief conclusions.

This paper makes use of results from two recent papers:

- In [5], an analysis was carried out to determine the variance of the difference between coupled paths in the jump process setting under a more general scaling than is considered here.
- In [6], an analysis was carried out to determine the variance of the difference between coupled paths in the setting of stochastic differential equations with small noise.

Our goals here are distinct from those of these two papers. First, the analysis in [5] allowed such a general scaling that no modified versions of Euler-based tau-leaping, such as midpoint or trapezoidal tau-leaping, could be considered. Here, we consider a particular scaling (which is the most common in the literature) and present a unified computational complexity analysis for a range of Monte Carlo-based methods. This allows us to make what we believe are the first concrete conclusions pertaining to the relative merits of current methods in a practically relevant asymptotic regime. Moreover, an open question in the literature involves the selection of the finest time step in the unbiased version of multilevel Monte Carlo (since it is not constrained by the accuracy requirement). By carrying out our analysis in this particular scaling regime, we are able to determine the asymptotics for the optimal selection of this parameter. Selecting the finest time step according to this procedure is shown to lower the computational complexity of the method by a nontrivial factor. See the end of section 4.2.2 for this derivation and the end of section 5 for a numerical example. Second, it has become part of the “folk wisdom” surrounding these models that in the particular scaling considered here, properly implemented numerical methods applied to the diffusion approximation are the best choice. This idea was somewhat exacerbated

by the analysis in [6], which applied to a key aspect of the algorithm. There it was shown that the variance between the coupled paths of a diffusion approximation is asymptotically smaller than the variance between the properly scaled jump processes. However, here we show that the actual difference in overall complexity between properly implemented versions of multilevel Monte Carlo for the diffusion approximation and for the jump process never differ by more than a logarithm term. If one combines this conclusion with the fact that the bias of the diffusion approximation itself is often unknown, whereas multilevel Monte Carlo applied to the jump process is naturally unbiased, then the folk wisdom is overturned and unbiased multilevel Monte Carlo is seen as a competitive choice.

**2. Scaling, assumptions, and a summary of results.** In order to motivate our analysis and computations, we begin with a brief high-level overview. In particular, we discuss the entries in Table 1, which summarizes the key conclusions of this work. Full details are given later in the manuscript; however, we point out here that the terms in Table 1 include assumptions on the variances of the constituent processes that will be detailed below.

A natural approach to approximate the desired expectation is to simulate paths exactly, for example, with the stochastic simulation algorithm [18, 19] or the next reaction method [1, 16], in order to obtain independent sample paths  $\{X_{[i]}^{N_0}\}_{i=1}^n$  that can be combined into a sample average:

$$(2) \quad \hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n f(X_{[i]}^{N_0}(\cdot)).$$

This becomes problematic if the cost of each sample path is high—to follow a path exactly, we must take account of each individual transition in the process. This is a serious issue when many jumps take place, which is the case when  $N_0$  is large.

The essence of the Euler tau-leaping approach is to fix the system intensities over time intervals of length  $h$  and thereby only require the generation of  $K$  Poisson random variables per time interval [20]. In order to analyze the benefit of tau-leaping and related methods, Anderson, Ganguly, and Kurtz [3] considered a family of models, parameterized by  $N \geq N_0$  (see (3) below), and considered the limit  $N \rightarrow \infty$  and  $h \rightarrow 0$  with  $h = N^{-\beta}$  for some  $\beta > 0$ . To see why such a limit is useful, we note two facts:

- If, instead, we allow  $N \rightarrow \infty$  with  $h$  fixed, then the stochastic fluctuations become negligible [8, 27]. In this *thermodynamic limit*, the model reduces to a deterministic ODE, so a simple deterministic numerical method could be used.
- If, instead, we allow  $h \rightarrow 0$  with  $N_0$  fixed, then tau-leaping becomes arbitrarily inefficient. The “empty” waiting times between reactions, which have nonzero expected values, are being needlessly refined by the discretization method.

The relation  $h = N^{-\beta}$  brings together the large system size effect (where exact simulation is expensive and tau-leaping offers a computational advantage) with the small  $h$  effect (where the accuracy of tau-leaping can be analyzed). This gives a realistic setting where the benefits of tau-leaping can be quantified. It may then be shown [3, Theorem 4.1] that the bias arising from Euler tau-leaping is  $O(h) = O(N^{-\beta})$  in a wide variety of cases. Higher-order alternatives to the original tau-leaping method [20] are available. For example, a midpoint discretization [3, Theorem 4.2] or a trapezoidal method [7] both achieve  $O(h^2) = O(N^{-2\beta})$  bias for a wide variety of cases.

As an alternative to tau-leap discretizations, we could replace the continuous-time Markov chain by a diffusion approximation and use a numerical stochastic differential

equation (SDE) simulation method to generate approximate paths [9]. This approximation is detailed in section 3.2 below. While higher-order methods are available for the simulation of diffusion processes, we restrict ourselves to Euler–Maruyama as the perturbation in the underlying model has already created a difficult to quantify bias. Thus, higher-order numerical schemes are hard to justify in this setting.

For our purposes, rather than the step size  $h$  of a particular approximate method, it is more natural to work in terms of the system size,  $N_0$ , and accuracy parameter  $\varepsilon_0$ . Let  $\varepsilon_0 = N_0^{-\alpha}$  for some fixed  $\alpha > 0$ . A larger value of  $\alpha$  corresponds to a more stringent accuracy requirement. Next, consider the following family of models parameterized by  $N \geq N_0$ ,

$$(3) \quad X^N(t) = X^N(0) + \sum_{k=1}^K \frac{1}{N} Y_k \left( N \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k,$$

with initial conditions satisfying  $\lim_{N \rightarrow \infty} X^N(0) = x_0 \in \mathbb{R}_{>0}^d$ . We will study the asymptotic behavior, as  $N \rightarrow \infty$ , of the computational complexity required of various schemes to approximate  $\mathbb{E}[f(X^N(\cdot))]$  to a tolerance of

$$(4) \quad \varepsilon_N = N^{-\alpha},$$

where  $f$  is a desired path functional. Specifically, we require that both the bias and the standard deviation of the resulting estimator are less than  $\varepsilon_N$ .

We emphasize at this stage that we are no longer studying a fixed model. Instead we look at the family of models (3) parameterized through the system size  $N$  and consider the limit, as  $N \rightarrow \infty$ , of the computational complexity of the different methods under the accuracy requirement (4). The computed results then tell us, to leading order, the costs associated with solving our fixed problem (1) with accuracy requirement  $N_0^{-\alpha}$ .

**2.1. Specific assumptions and a brief summary of results.** Instead of giving specific assumptions on the intensity functions  $\lambda_k$  and the functional  $f$ , we give assumptions pertaining to the cost of different path simulation strategies, the bias of those strategies, and the variance of different relevant terms. We then provide citations for when the assumptions are valid. We expect the assumptions to be valid for a wider class of models and functionals than has been proven in the literature, and discovering such classes is an active area of research.

To quantify computational complexity, we define the “expected cost-per-path” to be the expected value of the number of random variables generated in the simulation of a single path. Standard  $\Theta$  notation is used (providing an asymptotic upper and lower bound in  $N$  or  $h$ ). We emphasize that computations take place over a fixed time interval  $[0, T]$ .

*Assumption 2.1.* We assume the following expected cost-per-path for different methods.

<i>Method</i>	<i>Expected cost-per-path</i>
Exact simulation	$\Theta(N)$
Euler tau-leaping	$\Theta(h^{-1})$
Midpoint tau-leaping	$\Theta(h^{-1/2})$
Euler–Maruyama for diffusion	$\Theta(h^{-1})$

We make the following assumptions on the bias,  $|\mathbb{E}[f(X^N(\cdot))] - \mathbb{E}[f(Z^N(\cdot))]|$ , of the different approximation methods, where  $Z^N$  is a generic placeholder for the different methods.

*Assumption 2.2.* We assume the following biases.

<i>Method</i>	<i>Bias</i>	<i>Reference</i>
Exact simulation	0	N.A.
Euler tau-leaping	$\Theta(h)$	[3]
Midpoint tau-leaping	$\Theta(h^2)$	[3]
Euler–Maruyama for diffusion	$\Theta(h)$	[6, 26]

A bias of  $\Theta(h)$  for Euler–Maruyama applied to a diffusion approximation is extremely generous, as it assumes that the bias of the underlying diffusion approximation is negligible. However, analytical results pertaining to the bias of the diffusion approximation for general functionals  $f$  are sparse. A startling result of the present analysis is that even with such generosity, the complexity of the *unbiased* version of multilevel tau-leaping is still often within a factor of a logarithm of the complexity of the multilevel version of Euler–Maruyama applied to the diffusion approximation.

We provide our final assumption, pertaining to the variances of relevant terms. Below,  $Z_h^N$  is a tau-leap process with step size  $h$ ,  $Z_h^N$  is a midpoint tau-leap process with step size  $h$ , and  $D_h^N$  is an Euler–Maruyama approximation of the diffusion approximation with step size  $h$ . The coupling methods utilized are described later in the paper. Finally,  $h_\ell = M^{-\ell}$  for some integer  $M > 1$ .

*Assumption 2.3.* We assume the following relevant variances per realization/path.

<i>Method</i>	<i>Variance</i>	<i>Reference</i>
Exact simulation	$\text{Var}(f(X^N(\cdot))) = \Theta(N^{-1})$	[5]
Euler tau-leaping	$\text{Var}(f(Z_h^N(\cdot))) = \Theta(N^{-1})$	[5]
Coupled exact/tau-leap	$\text{Var}(f(X^N(\cdot)) - f(Z_h^N(\cdot))) = \Theta(h \cdot N^{-1})$	[5]
Coupled tau-leap	$\text{Var}(f(Z_{h_\ell}^N(\cdot)) - f(Z_{h_{\ell-1}}^N(\cdot))) = \Theta(h_\ell \cdot N^{-1})$	[5]
Midpt. or trap. tau-leaping	$\text{Var}(f(Z_h^N(\cdot))) = \Theta(N^{-1})$	[5]
Euler–Maruyama for diffusion	$\text{Var}(f(D_h^N(\cdot))) = \Theta(N^{-1})$	[6]
Coupled diffusion approx.	$\text{Var}(f(D_{h_\ell}^N(\cdot)) - f(D_{h_{\ell-1}}^N(\cdot))) = \Theta(N^{-1}h_\ell^2 + N^{-2}h_\ell)$	[6]

The results presented in Table 1 can now start coming into focus. For example, we immediately see that in order to get both the bias and the standard deviation under control, i.e., below  $\varepsilon_N$ , we have the following.

**Monte Carlo plus exact simulation:** We require  $\Theta(N^{-1}\varepsilon_N^{-2} + 1)$  paths for the standard deviation to be order  $\varepsilon_N^2$  at a cost of  $\Theta(N)$  per path. This totals a computational complexity of  $\Theta(\varepsilon_N^{-2} + N)$  or  $\Theta(N^{2\alpha} + N)$ .

**Monte Carlo plus tau-leaping:**  $\Theta(N^{-1}\varepsilon_N^{-2} + 1)$  paths at a cost of  $\Theta(\varepsilon_N^{-1})$  per path (required to achieve a bias of  $O(\varepsilon)$ ), totaling a computational complexity of  $\Theta(N^{-1}\varepsilon_N^{-3} + \varepsilon_N^{-1})$  or  $\Theta(N^{3\alpha-1} + N^\alpha)$ .

This is summarized in the first two rows of Table 1. Note that the “+1” terms above account for the requirement that we cannot generate less than one path. In this regime, we see that tau-leaping is beneficial for  $\alpha < 1$ . This makes sense intuitively. If we ask for too much accuracy relative to the system size ( $\alpha > 1$  in (4)), then tau-leaping’s built-in bias outweighs its cheapness, or, equivalently, the required step size is so small that tau-leaping works harder than exact simulation. The remainder of the table will be considered in section 4.

We also mention that a crude and inexpensive approximation to the required expected value can be computed by simply simulating the deterministic mass action ODE approximation to (1), which is often referred to as the reaction rate equation [8, 9]. Depending upon the choice of functional  $f$  and the underlying model (3), the bias from the ODE approximation can range from zero (in the case of a linear  $\lambda_k$  and

TABLE 1

Computational cost for different Monte Carlo methods as  $N \rightarrow \infty$ . The final column indicates when each method is most efficient, in terms of the parameter  $\alpha$ , up to factors involving logarithms.

Monte Carlo method	Computational complexity	Unbiased?	Most efficient
MC + exact simulation	$\Theta(N^{2\alpha} + N)$	Yes	Never
MC + tau-leaping	$\Theta(N^{3\alpha-1} + N^\alpha)$	No	Never
MC + midpt. or trap. tau-leap	$\Theta(N^{2.5\alpha-1} + N^{\alpha/2})$	No	$\frac{1}{2} < \alpha \leq \frac{2}{3}$
MC + Euler for diff. approx.	$\Theta(N^{3\alpha-1} + N^\alpha)$	No	Never
MLMC + E-M for diff. approx.	$\Theta(N^{2\alpha-1} + N^\alpha)$	No	$\alpha \geq \frac{2}{3}$
biased MLMC tau-leaping	$\Theta(N^{2\alpha-1}(\log N)^2 + N^\alpha)$	No	$\alpha \geq \frac{2}{3}$
unbiased MLMC tau-leaping	$\Theta(N^{2\alpha-1}(\log N)^2 + N)$	Yes	$\alpha \geq 1$

linear function  $f$ ), to order  $N^{-1/2}$  (for example, when  $f(X^N(\cdot)) = \sup_{t \leq T} |X^N(t) - c(t)|$ , where  $c$  is the ODE approximation itself). As we are interested in the fluctuations inherent to the stochastic model, we view  $\alpha = \frac{1}{2}$  as a natural cutoff in the relationship (4).

In addition to the asymptotic complexity counts in Table 1, another important feature of a method is the availability of computable a posteriori confidence interval information. As indicated in the table, two of the methods considered here, exact simulation with Monte Carlo and an appropriately constructed multilevel tau-leaping, are unbiased. The sample mean, accompanied by an estimate of the overall variance, can then be delivered with a computable confidence interval. By contrast, the remaining methods in the table are biased: Tau-leaping and Euler–Maruyama introduce discretization errors, and the diffusion approximation perturbs the underlying model. Although the asymptotic leading order of these biases can be estimated, useful a posteriori upper bounds cannot be computed straightforwardly in general, making these approaches much less attractive for reliably achieving a target accuracy.

Based on the range of methods analysed here in an asymptotic regime that couples system size and target accuracy, three key messages are the following:

- Simulating exact samples alone is never advantageous.
- Even assuming there is no bias to the underlying model, simulating at the level of the the diffusion approximation is only marginally advantageous.
- Tau-leaping can offer advantages over exact simulation, and an appropriately designed version of multilevel tau-leaping (which combines exact and tau-leaped samples) offers an unbiased method that is efficient over a wide range of accuracy requirements.

**3. Approximation methods.** In this section, we briefly review the two alternatives to exact simulation of (3) we study in this paper: tau-leaping and an Euler–Maruyama discretization of a diffusion approximation.

**3.1. Tau-leaping.** Tau-leaping [20] is a computational method that generates Euler-style approximate paths for the continuous-time Markov chain (3). The basic idea is to hold the intensity functions fixed over a time interval  $[t_n, t_n + h]$  at the values  $\lambda_k(X^N(t_n))$ , where  $X^N(t_n)$  is the state of the system at time  $t_n$ , and, under this simplification, compute the number of times each reaction takes place over this period. As the waiting times for the reactions are exponentially distributed, this leads to the following algorithm, which simulates up to a time of  $T > 0$ . For  $x \geq 0$ , we will write  $\text{Poisson}(x)$  to denote a sample from the Poisson distribution with parameter  $x$ , with all such samples being independent of each other and of all other sources of randomness used.

ALGORITHM 3.1 (Euler tau-leaping). Fix  $h > 0$ . Set  $Z_h^N(0) = x_0, t_0 = 0, n = 0$ , and repeat the following until  $t_n = T$ :

- (i) Set  $t_{n+1} = t_n + h$ . If  $t_{n+1} \geq T$ , set  $t_{n+1} = T$  and  $h = T - t_n$ .
- (ii) For each  $k$ , let  $\Lambda_k = \text{Poisson}(\lambda_k(Z_h^N(t_n))h)$ .
- (iii) Set  $Z_h^N(t_{n+1}) = Z_h^N(t_n) + \sum_k \Lambda_k \zeta_k$ .
- (iv) Set  $n \leftarrow n + 1$ .

Analogously to (3), a pathwise representation of Euler tau-leaping defined for all  $t \geq 0$  can be given through a random time change of Poisson processes:

$$(5) \quad Z_h^N(t) = Z_h^N(0) + \sum_k \frac{1}{N} Y_k \left( N \int_0^t \lambda_k(Z_h^N(\eta_h(s))) ds \right) \zeta_k,$$

where the  $Y_k$  are as before and  $\eta_h(s) \stackrel{\text{def}}{=} \lfloor \frac{s}{h} \rfloor h$ . Thus,  $Z_h^N(\eta_h(s)) = Z_h^N(t_n)$  if  $t_n \leq s < t_{n+1}$ . As the values of  $Z_h^N$  can go negative, the functions  $\lambda_k$  must be defined outside of  $\mathbb{Z}_{\geq 0}^d$ . One option is to simply define  $\lambda_k(x) = 0$  for  $x \notin \mathbb{Z}_{\geq 0}^d$ , though other options exist [2].

**3.2. Diffusion approximation.** The tau-leaping algorithm utilizes a time-stepping method to directly approximate the underlying model (3). Alternatively, a diffusion approximation arises by perturbing the underlying model into one which can be discretized more efficiently.

Define the function  $F$  via

$$F(x) = \sum_k \lambda_k(x) \zeta_k.$$

By the functional central limit theorem,

$$(6) \quad \frac{1}{\sqrt{N}} [Y_k(Nu) - Nu] \approx W_k(u),$$

where  $W_k$  is a standard Brownian motion. Applying (6) to (3) yields

$$X^N(t) \approx X^N(0) + \int_0^t F(X^N(s)) ds + \sum_k \frac{1}{\sqrt{N}} W_k \left( \int_0^t \lambda_k(X^N(s)) ds \right) \zeta_k,$$

where the  $W_k$  are independent standard Brownian motions. This implies that  $X^N$  can be approximated by the process  $D^N$  satisfying

$$(7) \quad D^N(t) = D^N(0) + \int_0^t F(D^N(s)) ds + \sum_k \frac{1}{\sqrt{N}} W_k \left( \int_0^t \lambda_k(D^N(s)) ds \right) \zeta_k,$$

where  $D^N(0) = X^N(0)$ . An equivalent and more prevalent way to represent  $D^N$  is via the Itô representation

$$(8) \quad D^N(t) = D^N(0) + \int_0^t F(D^N(s)) ds + \sum_k \frac{1}{\sqrt{N}} \zeta_k \int_0^t \sqrt{\lambda_k(D^N(s))} dW_k(s),$$

which is often written in the differential form

$$(9) \quad dD^N(t) = F(D^N(t)) dt + \sum_k \frac{1}{\sqrt{N}} \zeta_k \sqrt{\lambda_k(D^N(t))} dW_k(t),$$

where the  $W_k$  of (9) are not necessarily the same as those in (7).

The SDE system (9) is known as a *Langevin* approximation in the biology and chemistry literature and as a *diffusion* approximation in probability [9, 34]. We note the following points:

- The diffusion coefficient, often termed the “noise” in the system, is  $\Theta(\frac{1}{\sqrt{N}})$  and hence, in our setting, is small relative to the drift.
- The diffusion coefficient involves square roots. Hence, it is critical that the intensity functions  $\lambda_k$  only take values in  $\mathbb{R}_{\geq 0}$  on the domain of the solution. This is of particular importance in the population process setting where the solutions of the underlying model (3) naturally satisfy a nonnegativity constraint, whereas the SDE solution paths cannot be guaranteed to remain nonnegative in general. In this case, one reasonable representation, of many, would be

$$(10) \quad dD^N(t) = F(D^N(t))dt + \sum_k \frac{1}{\sqrt{N}} \zeta_k \sqrt{[\lambda_k(D^N(s))]^+} dW_k(s),$$

where  $[x]^+ = \max\{x, 0\}$ . Another reasonable option would be to use a process with reflection [30].

- The coefficients of the SDE are not globally Lipschitz in general, and hence standard convergence theory for numerical methods, such as that in [26], is not applicable. Examples of nonlinear SDEs for which standard Monte Carlo and multilevel Monte Carlo, when combined with an Euler–Maruyama discretization with a uniform time step, fail to produce a convergent algorithm have been pointed out in the literature [22, 23]. The question of which classes of reaction systems lead to well-defined SDEs and which discretizations converge at the traditional rate therefore remains open.

In this work, to get a feel for the best possible computational complexity that can arise from the Langevin approximation, we will study the case where the bias that arises from switching models from  $X^N$  to  $D^N$  is zero. We will also assume that, even though the diffusion coefficients involve square roots and are therefore not generally globally Lipschitz, the Euler–Maruyama method has a bias of order  $\Theta(h)$ . We will find that even in this idealized light, the asymptotic computational complexity of Euler–Maruyama on a diffusion approximation combined with either a standard or a multilevel implementation is only marginally better than the corresponding computational complexity bounds for multilevel tau-leaping. In particular, they differ only in a factor of a logarithm of the scaling parameter.

Finally, due to the fact that the diffusion approximation itself already has a difficult to quantify bias, we will not consider higher-order methods [10] or even unbiased methods [21] for this process.

**4. Complexity analysis.** In this section, we establish the results given in Table 1. In subsection 4.1, we derive the first four rows, whereas in subsection 4.2, we discuss the multilevel framework and establish rows 5, 6, and 7.

**4.1. Complexity analysis of standard Monte Carlo approaches.**

**4.1.1. Exact sampling and Monte Carlo.** By Assumption 2.1, the expected number of system updates required to generate a single exact sample path is  $\Theta(N)$ . Letting

$$\delta_N = \text{Var}(f(X^N(\cdot))),$$

in order to get a standard deviation below  $\varepsilon_N$ , we require

$$n^{-1}\delta_N \leq \varepsilon_N^2 \implies n \geq \delta_N \varepsilon_N^{-2} + 1.$$

Thus, the total computational complexity of making the desired approximation is

$$\Theta(nN) = \Theta(\delta_N \varepsilon_N^{-2} N + N) = \Theta(\delta_N N^{2\alpha+1} + N).$$

By Assumption 2.3,  $\delta_N = \Theta(N^{-1})$ , yielding an overall complexity of  $\Theta(N^{2\alpha} + N)$ , as given in the first row of Table 1.

**4.1.2. Tau-leaping and Monte Carlo.** Suppose now that we use  $n$  paths of the tau-leaping process (5) to construct the Monte Carlo estimator  $\hat{\mu}_n$  for  $\mathbb{E}[f(X^N(\cdot))]$ . By Assumption 2.2, the bias is  $\Theta(h)$ , so we constrain ourselves to  $h = \varepsilon_N$ . Letting

$$\delta_{N,h} = \text{Var}(f(Z_h^N(\cdot))),$$

we again require  $n \geq \delta_{N,h} \varepsilon_N^{-2} + 1$  to control the statistical error. Since by Assumption 2.1 there are  $\Theta(h^{-1})$  expected operations per path generation, the total computational complexity for making the desired approximation is

$$\Theta(nh^{-1}) = \Theta(\delta_{N,h} \varepsilon_N^{-3} + \varepsilon_N^{-1}).$$

By Assumption 2.3,  $\text{Var}(f(Z_{h,i}^N(\cdot))) = \Theta(N^{-1})$ , giving an overall complexity of  $\Theta(N^{3\alpha-1} + N^\alpha)$ , as reported in the second row of Table 1.

Weakly second-order extensions to the tau-leaping method can lower the computational complexity dramatically. For example, if we use the midpoint tau-leaping process  $Z_h^N$  from [3], by Assumption 2.2, we can set  $h = \sqrt{\varepsilon_N}$  and still achieve a bias of  $\Theta(\varepsilon_N)$ . Since by Assumption 2.3 we need  $n \geq N^{-1} \varepsilon_N^{-2} + 1$  paths to control the standard deviation, the complexity is

$$\Theta(n \cdot h^{-1}) = \Theta(N^{-1} \varepsilon_N^{-2.5} + \varepsilon_N^{-1/2}) = \Theta(N^{2.5\alpha-1} + N^{\alpha/2}),$$

as stated in the third row of Table 1. The same conclusion can also be drawn for the trapezoidal method in [7].

If methods are developed that are higher order in a weak sense, then further improvements can be gained. In general, if a method is developed that is weakly of order  $\rho$ , then we may set  $h = \varepsilon_N^{1/\rho}$  to achieve a bias of  $\Theta(\varepsilon_N)$ . Still supposing a per-path variance of  $\Theta(N^{-1})$ , we again choose  $n \geq N^{-1} \varepsilon_N^{-2} + 1$  paths and find a complexity of

$$\Theta(n \cdot h^{-1}) = \Theta(N^{-1} \varepsilon_N^{-(2+\frac{1}{\rho})} + \varepsilon_N^{-1/\rho}) = \Theta(N^{(2+\frac{1}{\rho})\alpha-1} + N^{\alpha/\rho}).$$

For example, if a third-order method is developed, i.e.,  $\rho = 3$ , then this method becomes optimal for  $\frac{1}{2} \leq \alpha \leq \frac{3}{4}$ . To the best of the authors' knowledge, no such methods have yet been designed.

**4.1.3. Diffusion approximation and Monte Carlo.** Given Assumptions 2.1, 2.2, and 2.3, the complexity analysis for the diffusion approximation with Euler-Maruyama is exactly the same as for Euler tau-leaping. Hence, we can again give an overall complexity of  $\Theta(N^{3\alpha-1} + N^\alpha)$ , as reported in the fourth row of Table 1.

**4.2. Multilevel Monte Carlo and complexity analysis.** In this section, we study multilevel Monte Carlo approaches and derive the results summarized in rows 5, 6, and 7 of Table 1.

**4.2.1. Multilevel Monte Carlo and diffusion approximation.** Here we specify and analyze an Euler-based multilevel method for the diffusion approximation, following the original framework of Giles [17].

For some fixed  $M > 1$ , we let  $h_\ell = T \cdot M^{-\ell}$  for  $\ell \in \{0, \dots, L\}$ , where  $T > 0$  is a fixed terminal time. Reasonable choices for  $M$  include  $M \in \{2, 3, 4, \dots, 7\}$ , and  $L$  is determined below. Let  $D_{h_\ell}^N$  denote the approximate process generated by Euler–Maruyama applied to (9) with a step size of  $h_\ell$ . By Assumption 2.2, we may set  $h_L = \varepsilon_N$ , giving  $L = \Theta(|\log \varepsilon_N|)$ , so that the finest level achieves the required order of magnitude for the bias.

Noting that

$$(11) \quad \mathbb{E}[f(D_{h_L}^N(\cdot))] = \mathbb{E}[f(D_{h_0}^N(\cdot))] + \sum_{\ell=1}^L \mathbb{E}\left[f(D_{h_\ell}^N(\cdot)) - f(D_{h_{\ell-1}}^N(\cdot))\right],$$

we use  $i$  as an index over sample paths and let

$$\widehat{Q}_0^N \stackrel{\text{def}}{=} \frac{1}{n_0} \sum_{i=1}^{n_0} f\left(D_{h_0, [i]}^N(\cdot)\right) \quad \text{and} \quad \widehat{Q}_\ell^N \stackrel{\text{def}}{=} \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} \left(f\left(D_{h_\ell, [i]}^N(\cdot)\right) - f\left(D_{h_{\ell-1}, [i]}^N(\cdot)\right)\right)$$

for  $\ell = 1, \dots, L$ , where  $n_0$  and the different  $n_\ell$  have yet to be determined. Note that the form of the estimator  $\widehat{Q}_\ell^N$  above implies that the processes  $D_{h_\ell}^N$  and  $D_{h_{\ell-1}}^N$  will be *coupled*, or constructed on the same probability space. We consider here the case when  $(D_{h_\ell}^N, D_{h_{\ell-1}}^N)$  are coupled in the usual way by using the same Brownian path in the generation of each of the marginal processes. Our (biased) estimator is then

$$\widehat{Q}^N \stackrel{\text{def}}{=} \widehat{Q}_0^N + \sum_{\ell=1}^L \widehat{Q}_\ell^N.$$

Set

$$\delta_{N,\ell} = \text{Var}(f(D_{h_\ell}^N(\cdot)) - f(D_{h_{\ell-1}}^N(\cdot))).$$

By Assumption 2.3,  $\delta_{N,\ell} = \Theta(N^{-1}h_\ell^2 + N^{-2}h_\ell)$  and  $\text{Var}(f(D_{h_0}^N(\cdot))) = \Theta(N^{-1})$ . In [6], it is shown that under these circumstances, the computational complexity required is  $\Theta(\varepsilon_N^{-2}N^{-1} + \varepsilon_N^{-1})$ . In the regime (4), this translates to  $\Theta(N^{2\alpha-1} + N^\alpha)$ , as reported in the fifth row of Table 1.

**4.2.2. Multilevel Monte Carlo and tau-leaping.** The use of multilevel Monte Carlo with tau-leaping for continuous-time Markov chains of the form considered here was proposed in [4], where effective algorithms were devised. Complexity results were given in a nonasymptotic multiscale setting, with follow-up results in [5]. Our aim here is to customize the approach in the scaling regime (4) and thereby develop easily interpretable complexity bounds that allow straightforward comparison with other methods. In this section,  $Z_{h_\ell}^N$  denotes a tau-leaping process generated with a step size of  $h_\ell = T \cdot M^{-\ell}$  for  $\ell \in \{0, \dots, L\}$ .

A major step in [4] was to show that a coupling technique used for analytical purposes in [3, 29] can also form the basis of a practical simulation algorithm. Letting  $Y_{k,i}$ ,  $i \in \{1, 2, 3\}$ , denote independent, unit rate Poisson processes, we couple the exact and approximate tau-leaping processes in the following way:

$$(12) \quad \begin{aligned} X^N(t) = X^N(0) &+ \sum_k \frac{1}{N} Y_{k,1} \left( N \int_0^t \lambda_k(X^N(s)) \wedge \lambda_k(Z_{h_L}^N(\eta_L(s))) ds \right) \zeta_k \\ &+ \sum_k \frac{1}{N} Y_{k,2} \left( N \int_0^t [\lambda_k(X^N(s)) - \lambda_k(X^N(s)) \wedge \lambda_k(Z_{h_L}^N(\eta_L(s)))] ds \right) \zeta_k \end{aligned}$$

$$\begin{aligned}
 Z_{h_L}^N(t) &= Z_{h_L}^N(0) + \sum_k \frac{1}{N} Y_{k,1} \left( N \int_0^t \lambda_k(X^N(s)) \wedge \lambda_k(Z_{h_L}^N(\eta_L(s))) ds \right) \zeta_k \\
 &+ \sum_k \frac{1}{N} Y_{k,3} \left( N \int_0^t [\lambda_k(Z_{h_L}^N(\eta_L(s))) - \lambda_k(X^N(s)) \wedge \lambda_k(Z_{h_L}^N(\eta_L(s)))] ds \right) \zeta_k,
 \end{aligned}
 \tag{13}$$

where  $a \wedge b$  denotes  $\min\{a, b\}$  and  $\eta_L(s) = \lfloor s/h_L \rfloor h_L$ . Sample paths of (12)–(13) can be generated with a natural extension of the next reaction method or Gillespie’s algorithm (see [4]), and for  $h_L \geq N^{-1}$ , the complexity required for the generation of a realization  $(X^N, Z_{h_L}^N)$  remains at the  $\Theta(N)$  level. The coupling of two approximate processes,  $Z_{h_\ell}^N$  and  $Z_{h_{\ell-1}}^N$ , takes the similar form

$$\begin{aligned}
 Z_{h_\ell}^N(t) &= Z_{h_\ell}^N(0) + \sum_k \frac{1}{N} Y_{k,1} \left( N \int_0^t \lambda_k(Z_{h_\ell}^N(\eta_\ell(s))) \wedge \lambda_k(Z_{h_{\ell-1}}^N(\eta_{\ell-1}(s))) ds \right) \zeta_k \\
 &+ \sum_k \frac{1}{N} Y_{k,2} \left( N \int_0^t [\lambda_k(Z_{h_\ell}^N(\eta_\ell(s))) - \lambda_k(Z_{h_\ell}^N(\eta_\ell(s))) \wedge \lambda_k(Z_{h_{\ell-1}}^N(\eta_{\ell-1}(s)))] ds \right) \zeta_k
 \end{aligned}
 \tag{14}$$

$$\begin{aligned}
 Z_{h_{\ell-1}}^N(t) &= Z_{h_{\ell-1}}^N(0) + \sum_k \frac{1}{N} Y_{k,1} \left( N \int_0^t \lambda_k(Z_{h_\ell}^N(\eta_\ell(s))) \wedge \lambda_k(Z_{h_{\ell-1}}^N(\eta_{\ell-1}(s))) ds \right) \zeta_k \\
 &+ \sum_k \frac{1}{N} Y_{k,3} \left( N \int_0^t [\lambda_k(Z_{h_{\ell-1}}^N(\eta_{\ell-1}(s))) - \lambda_k(Z_{h_\ell}^N(\eta_\ell(s))) \wedge \lambda_k(Z_{h_{\ell-1}}^N(\eta_{\ell-1}(s)))] ds \right) \zeta_k,
 \end{aligned}
 \tag{15}$$

where  $\eta_\ell(s) \stackrel{\text{def}}{=} \lfloor s/h_\ell \rfloor h_\ell$ . The pair (14)–(15) can be sampled at the same  $\Theta(h_\ell^{-1})$  cost as a single tau-leaping path (see [4]).

For  $L$  as yet to be determined and noting the identity

$$\begin{aligned}
 \mathbb{E}[f(X^N(\cdot))] &= \mathbb{E}[f(X^N(\cdot)) - f(Z_L^N(\cdot))] \\
 &+ \sum_{\ell=1}^L \mathbb{E}[f(Z_{h_\ell}^N(\cdot)) - f(Z_{h_{\ell-1}}^N(\cdot))] + \mathbb{E}[f(Z_{h_0}^N(\cdot))],
 \end{aligned}
 \tag{16}$$

we define estimators for the three terms above via

$$\begin{aligned}
 \widehat{Q}_E^N &\stackrel{\text{def}}{=} \frac{1}{n_E} \sum_{i=1}^{n_E} (f(X_{[i]}^N(\cdot)) - f(Z_{h_L, [i]}^N(\cdot))), \\
 \widehat{Q}_\ell^N &\stackrel{\text{def}}{=} \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} (f(Z_{h_\ell, [i]}^N(\cdot)) - f(Z_{h_{\ell-1}, [i]}^N(\cdot))) \quad \text{for } \ell \in \{1, \dots, L\}, \\
 \widehat{Q}_0^N &\stackrel{\text{def}}{=} \frac{1}{n_0} \sum_{i=1}^{n_0} f(Z_{h_0, [i]}^N(\cdot)),
 \end{aligned}
 \tag{17}$$

so that

$$\widehat{Q}^N \stackrel{\text{def}}{=} \widehat{Q}_E^N + \sum_{\ell=1}^L \widehat{Q}_\ell^N + \widehat{Q}_0^N
 \tag{18}$$

is an unbiased estimator for  $\mathbb{E}[f(X^N(\cdot))]$ . Here,  $\widehat{Q}_E^N$  uses the coupling (12)–(13) between exact paths and tau-leaped paths of step size  $h_L$ ,  $\widehat{Q}_\ell^N$  uses the coupling (14)–(15) between tau-leaped paths of step sizes  $h_\ell$  and  $h_{\ell-1}$ , and  $\widehat{Q}_0^N$  involves single tau-leaped paths of step size  $h_0$ . Note that the algorithm implicit in (18) produces an unbiased estimator, whereas the estimator is biased if  $\widehat{Q}_E^N$  is left off, as will sometimes be desirable. Hence, we will refer to estimator  $\widehat{Q}^N$  in (18) as the unbiased estimator and will refer to

$$(19) \quad \widehat{Q}_B^N \stackrel{\text{def}}{=} \sum_{\ell=1}^L \widehat{Q}_\ell^N + \widehat{Q}_0^N$$

as the biased estimator. For both the biased and the unbiased estimators, the number of paths at each level,  $n_0$ ,  $n_\ell$ , and  $n_E$ , will be chosen to ensure an overall estimator standard deviation of  $\varepsilon_N$ .

We consider the biased and unbiased versions of tau-leaping multilevel Monte Carlo separately.

**Biased multilevel Monte Carlo tau-leaping**

Here we consider the estimator  $\widehat{Q}_B^N$  defined in (19). By Assumption 2.2, we have  $|\mathbb{E}[f(X^N(\cdot))] - \mathbb{E}[f(Z_{h_L}^N(\cdot))]| = \Theta(h_L)$ . Hence, in order to control the bias, we begin by choosing  $h_L = \varepsilon_N$  and so  $L = \Theta(\log(1/\varepsilon_N)) = \Theta(\log N)$ .

For  $\ell \in \{1, \dots, L\}$ , let  $C_\ell$  be the expected number of random variables required to generate a single pair of coupled trajectories at level  $\ell$ , and let  $\delta_{N,\ell}$  be the variance of the relevant processes on level  $\ell$ . Let  $C_0$  be the expected number of random variables required to generate a single trajectory at the coarsest level. To find  $n_\ell, \ell \in \{0, \dots, L\}$ , we solve the following optimization problem, which ensures that the variance of  $\widehat{Q}_B^N$  is no greater than  $\varepsilon_N^2$ :

$$(20) \quad \underset{n_\ell}{\text{minimize}} \quad \sum_{\ell=0}^L n_\ell C_\ell$$

$$(21) \quad \text{subject to} \quad \sum_{\ell=0}^L \frac{\delta_{N,\ell}}{n_\ell} = \varepsilon_N^2.$$

We use Lagrange multipliers. Since we have  $C_\ell = K \cdot h_\ell^{-1}$ , for some fixed constant  $K$ , the optimization problem above is solved at solutions to

$$\nabla_{n_0, \dots, n_L, \lambda} \left( \sum_{\ell=0}^L n_\ell K \cdot h_\ell^{-1} + \lambda \left( \sum_{\ell=0}^L \frac{\delta_{N,\ell}}{n_\ell} - \varepsilon_N^2 \right) \right) = 0.$$

Taking derivatives with respect to  $n_\ell$  and setting each derivative to zero yields

$$(22) \quad n_\ell = \sqrt{\frac{\lambda}{K} \delta_{N,\ell} h_\ell} \quad \text{for } \ell \in \{0, 1, 2, \dots, L\}$$

for some  $\lambda \geq 0$ . Plugging (22) into (21) gives us

$$(23) \quad \sum_{\ell=0}^L \sqrt{\frac{\delta_{N,\ell}}{h_\ell}} = \sqrt{\frac{\lambda}{K}} \cdot \varepsilon_N^2$$

and hence, by Assumption 2.3,

$$(24) \quad \sqrt{\frac{\lambda}{K}} = \sum_{\ell=0}^L \sqrt{\frac{\delta_{N,\ell}}{h_\ell}} \leq CL\varepsilon_N^{-2}N^{-1/2},$$

where  $C$  is a constant. Noting that  $L = \Theta(\log(\varepsilon_N^{-1}))$ , we have

$$\frac{\lambda}{K} = \Theta\left(\varepsilon_N^{-4}(\log \varepsilon_N)^2 N^{-1}\right).$$

Plugging this back into (22) and recognizing that at least one path must be generated to achieve the desired accuracy, we find

$$n_\ell = \Theta(\varepsilon_N^{-2}N^{-1}h_\ell L + 1).$$

Hence, the overall computational complexity is

$$\begin{aligned} \sum_{\ell=0}^L n_\ell K h_\ell^{-1} &= \Theta\left(\sum_{\ell=0}^L \varepsilon_N^{-2}N^{-1}h_\ell L h_\ell^{-1} + \sum_{\ell=0}^L h_\ell^{-1}\right) = \Theta\left(\varepsilon_N^{-2}N^{-1}(\log \varepsilon_N)^2 + \varepsilon_N^{-1}\right) \\ &= \Theta\left(N^{2\alpha-1}(\log N)^2 + N^\alpha\right), \end{aligned}$$

recovering row 6 of Table 1.

Note that the computational complexity reported for this biased version of multilevel Monte Carlo tau-leaping is, up to logarithms, the same as that for multilevel Monte Carlo on the diffusion approximation. However, none of the generous assumptions we made for the diffusion approximation were required.

### Unbiased multilevel Monte Carlo tau-leaping

The first observation to make is that the telescoping sum (16) implies that the method which utilizes  $\mathbb{E}[f(X^N(\cdot)) - f(Z_{h_L}^N(\cdot))]$  at the finest level is unbiased for *any* choice of  $h_L$ . That is, we are no longer constrained to choose  $L = \Theta(|\log \varepsilon_N|)$ .

Assume that  $h_L \geq N^{-1}$ . Let  $C_E$  be the expected number of random variables required to generate a single pair of the coupled exact and tau-leaped processes when the tau-leap discretization is  $h_L$ . To determine  $n_\ell$  and  $n_E$ , we still solve an optimization problem,

$$(25) \quad \underset{n_\ell}{\text{minimize}} \quad \sum_{\ell=0}^L n_\ell C_\ell + n_L C_E$$

$$(26) \quad \text{subject to} \quad \sum_{\ell=0}^L \frac{\delta_{N,\ell}}{n_\ell} + \frac{\delta_{N,E}}{n_E} = \varepsilon_N^2,$$

where  $C_\ell$  and  $\delta_{N,\ell}$  are as before and  $\delta_{N,E} = \text{Var}(f(X^N(\cdot)) - f(Z_{h_L}^N(\cdot)))$ .

Using Lagrange multipliers again, we obtain

$$(27) \quad n_\ell = \sqrt{\frac{\lambda \delta_{N,\ell}}{C_\ell}} \quad \text{for } \ell \in \{0, 1, 2, \dots, L\}$$

and

$$(28) \quad n_E = \sqrt{\frac{\lambda \delta_{N,E}}{C_E}}.$$

Plugging back into (26) and noting that by Assumption 2.1,  $C_\ell = \Theta(h_\ell^{-1})$  and  $C_E = \Theta(N)$  yields

$$(29) \quad \sqrt{\lambda} = \varepsilon_N^{-2} \left( \sum_{\ell=0}^L \sqrt{\delta_{N,\ell} C_\ell} + \sqrt{\delta_{N,E} C_E} \right) \leq C(L\varepsilon_N^{-2} N^{-1/2} + \varepsilon_N^{-2} \sqrt{h_L}).$$

Therefore, plugging (29) back into (27) and (28) and noting  $n_\ell \geq 1$  and  $n_E \geq 1$ , we get

$$n_\ell = \sqrt{\frac{\lambda \delta_{N,\ell}}{C_\ell}} + 1 = O \left( \left( L\varepsilon_N^{-2} N^{-1} + \varepsilon_N^{-2} \sqrt{\frac{h_L}{N}} \right) h_\ell + 1 \right) \quad \text{for } \ell \in \{0, 1, 2, \dots, L\}$$

and

$$(30) \quad n_E = \sqrt{\frac{\lambda \delta_{N,E}}{C_E}} + 1 = O(L\varepsilon_N^{-2} N^{-3/2} h_L^{1/2} + \varepsilon_N^{-2} N^{-1} h_L + 1).$$

As a result, the total complexity is

$$\begin{aligned} g(h_L) &= O \left( \varepsilon_N^{-2} N^{-1} L^2 + \varepsilon_N^{-2} \sqrt{\frac{h_L}{N}} L + h_L^{-1} + \varepsilon_N^{-2} \sqrt{\frac{h_L}{N}} L + \varepsilon_N^{-2} h_L + N \right) \\ &\leq O \left( \varepsilon_N^{-2} N^{-1} L^2 + 2\varepsilon_N^{-2} \sqrt{\frac{h_L}{N}} L + \varepsilon_N^{-2} h_L + 2N \right) \\ &= O(2\varepsilon_N^{-2} N^{-1} L^2 + 2\varepsilon_N^{-2} h_L + 2N), \end{aligned}$$

where the inequality follows since  $h_L^{-1} \leq N$  and we used that  $2ab \leq a^2 + b^2$  in the final equality. It is relatively easy to show that the last line above is minimized at

$$(31) \quad h_L = \frac{2}{(\log 2)^2 N} \text{LambertW} \left( \frac{N}{2/(\log 2)^2} \right) \approx \frac{2}{(\log 2)^2 N} \log \left( \frac{N}{2/(\log 2)^2} \right).$$

Hence, taking  $h_L = \Theta(N^{-1} \log N)$ , we have  $(\log h_L)^2 = \Theta((\log N)^2)$ , and this method achieves a total computational complexity of leading order

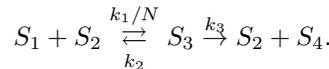
$$\begin{aligned} \Theta(\varepsilon_N^{-2} N^{-1} (\log N)^2 + \varepsilon_N^{-2} N^{-1} \log N + N) &= \Theta(\varepsilon_N^{-2} N^{-1} (\log N)^2 + N) \\ &= \Theta(N^{2\alpha-1} (\log N)^2 + N), \end{aligned}$$

as reported in the last row of Table 1.

Note here that if we choose  $h_L = \frac{1}{N}$ , we get the same order of magnitude for the computational complexity. However, the  $h_L$  in (31) is the optimized solution, meaning that the leading order constant should be better, and we will see this in Figures 3 and 4 in the next section.

**5. Computational results.** In this section, we provide numerical evidence for the sharpness of the computational complexity analyses provided in Table 1. We will measure complexity by total number of random variables utilized. We emphasize that these experiments use extreme parameter choices solely for the purpose of testing the sharpness of the delicate asymptotic bounds.

*Example 5.1.* We consider the classically scaled stochastic model for the following reaction network (see [9]):



Letting  $X_i(t)$  give the number of molecules of species  $S_i$  at time  $t$  and letting  $X^N(t) = X(t)/N$ , the stochastic equations are

$$\begin{aligned}
X^N(t) = X^N(0) &+ \frac{1}{N} Y_1 \left( Nk_1 \int_0^t X_1^N(s) X_2^N(s) ds \right) \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \\
&+ \frac{1}{N} Y_2 \left( Nk_2 \int_0^t X_3^N(s) ds \right) \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} \\
&+ \frac{1}{N} Y_3 \left( Nk_3 \int_0^t X_3^N(s) ds \right) \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix},
\end{aligned}$$

where we assume  $X^N(0) \rightarrow (0.2, 0.2, 0.2, 0.2)^T$  as  $N \rightarrow \infty$ . Note that the intensity function  $\lambda_1(x) = \kappa_1 x_1 x_2$  is globally Lipschitz on the domain of interest, as that domain is bounded (mass is conserved in this model).

We implemented different Monte Carlo simulation methods for the estimation of  $\mathbb{E}[X_1^N(T)]$  to an accuracy of  $\varepsilon_N = N^{-\alpha}$  for both  $\alpha = 1$  and  $\alpha = 5/4$ . Specifically, for each of the order one methods we chose a step size of  $h = \varepsilon_N$  and required the variance of the estimator to be  $\varepsilon_N^2$ . For midpoint tau-leaping, which has a weak order of two, we chose  $h = \sqrt{\varepsilon_N}$ . For the unbiased multilevel Monte Carlo method, we chose the finest time step according to (31). We do not provide estimates for Monte Carlo combined with exact simulation, as those computations were too intensive to complete to the target accuracy.

For our numerical example, we chose  $T = 1$  and  $X(0) = \lceil N \cdot [0.2, 0.2, 0.2, 0.2]^T \rceil$  with  $X^N(0) = X(0)/N$ . Finally, we chose  $k_1 = k_2 = k_3 = 1$  as our rate constants. In Figure 1, we provide log-log plots of the computational complexity required to solve this problem for the different Monte Carlo methods to an accuracy of  $\varepsilon_N = N^{-1}$  for each of

$$N \in \{2^{13}, 2^{14}, 2^{15}, 2^{16}, 2^{17}\}.$$

In Figure 2, we provide log-log plots for the computational complexity required to solve this problem for the different methods to an accuracy of  $\varepsilon_N = N^{-5/4}$  for each of

$$N \in \{2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}\}.$$

Tables 2 and 3 provide the estimator standard deviations for the different Monte Carlo methods with  $\varepsilon_N = N^{-1}$  and  $\varepsilon_N = N^{-5/4}$ , respectively. The top line provides the target standard deviations.

The specifics of the implementations and results for the different Monte Carlo methods are detailed below.

**Diffusion approximation plus Monte Carlo.** We took a time step of size  $h = \varepsilon_N$  to generate our independent samples. See Figure 1, where the best fit line is  $y = 1.94x - 0.88$ , and Figure 2, where the best fit line is  $y = 2.73x - 1.37$ , which are consistent with the exponent  $\alpha$  in Table 1.

**Monte Carlo tau-leaping.** We took a time step of size  $h = \varepsilon_N$  to generate our independent samples. See Figure 1, where the best fit line is  $y = 1.96x - 1.02$ , and Figure 2, where the best fit line is  $y = 2.76x - 1.63$ , which are consistent with the exponent  $\alpha$  in Table 1.

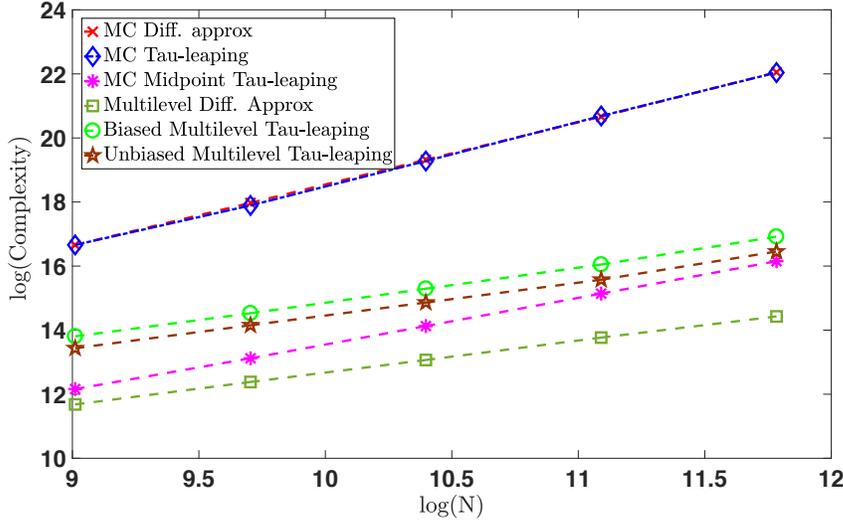


FIG. 1. Log-log plots of the computational complexity for the different Monte Carlo methods with varying  $N \in \{2^{13}, 2^{14}, 2^{15}, 2^{16}, 2^{17}\}$  and  $\varepsilon_N = N^{-1}$ .

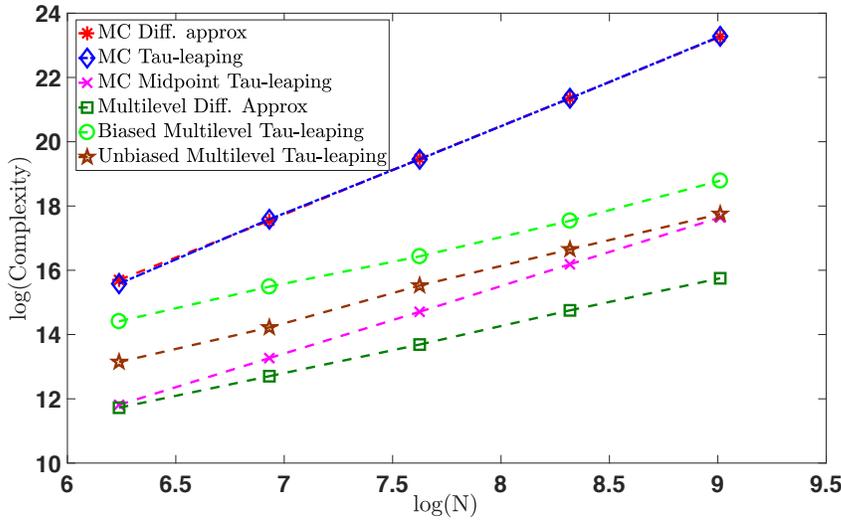


FIG. 2. Log-log plots of the computational complexity for the different Monte Carlo methods with varying  $N \in \{2^9, 2^{10}, 2^{11}, 2^{12}, 2^{13}\}$  and  $\varepsilon_N = N^{-\frac{5}{4}}$ .

TABLE 2  
Actual estimator standard deviations when  $\varepsilon_N = N^{-1}$ .

Method	Estimator standard deviations
	$2^{-13}, 2^{-14}, 2^{-15}, 2^{-16}, 2^{-17}$
MC and diff. approx	$2^{-13.10}, 2^{-14.02}, 2^{-15.02}, 2^{-16.01}, 2^{-17.00}$
MC and tau-leaping	$2^{-13.09}, 2^{-14.01}, 2^{-15.01}, 2^{-16.01}, 2^{-17.00}$
MC and midpoint tau-leaping	$2^{-13.09}, 2^{-14.04}, 2^{-15.03}, 2^{-16.00}, 2^{-17.01}$
Multilevel diff. approx	$2^{-13.20}, 2^{-14.15}, 2^{-15.11}, 2^{-16.09}, 2^{-17.07}$
Biased multilevel tau-leaping	$2^{-13.44}, 2^{-14.39}, 2^{-15.39}, 2^{-16.38}, 2^{-17.32}$
Unbiased multilevel tau-leaping	$2^{-13.29}, 2^{-14.28}, 2^{-15.26}, 2^{-16.21}, 2^{-17.18}$

TABLE 3  
*Actual estimator standard deviations when  $\varepsilon_N = N^{-5/4}$ .*

Method	Estimator standard deviations
$\varepsilon_N = N^{-5/4}$	$2^{-11.25}, 2^{-12.50}, 2^{-13.75}, 2^{-15.00}, 2^{-16.25}$
MC and diff. approx	$2^{-11.27}, 2^{-12.51}, 2^{-13.75}, 2^{-15.00}, 2^{-16.25}$
MC and tau-leaping	$2^{-11.26}, 2^{-12.52}, 2^{-13.76}, 2^{-15.00}, 2^{-16.25}$
MC and midpoint tau-leaping	$2^{-11.26}, 2^{-12.52}, 2^{-13.76}, 2^{-15.00}, 2^{-16.25}$
Multilevel diff. approx	$2^{-11.46}, 2^{-12.63}, 2^{-13.85}, 2^{-15.06}, 2^{-16.29}$
Biased multilevel tau-leaping	$2^{-11.62}, 2^{-12.81}, 2^{-13.99}, 2^{-15.19}, 2^{-16.41}$
Unbiased multilevel tau-leaping	$2^{-11.34}, 2^{-12.57}, 2^{-13.79}, 2^{-15.03}, 2^{-16.26}$

**Monte Carlo midpoint tau-leaping.** We took a time step of size  $h = \sqrt{\varepsilon_N}$ . See Figure 1, where the best fit line is  $y = 1.44 - 0.86$ , and Figure 2, where the best fit line is  $y = 2.10x - 3.53$ , which are consistent with the exponent  $\alpha$  in Table 1.

Our implementation of the multilevel methods proceeded as follows. We chose  $h_\ell = 2^{-\ell}$ , and for  $\varepsilon_N > 0$ , we fixed  $h_L = \varepsilon_N$  and  $L = \lceil \log(h_L)/\log(2) \rceil$  for the biased methods. For each level, we generated  $N_0$  independent sample trajectories in order to estimate  $\delta_{N,\ell}$ , as defined in section 3. Then we selected

$$n_\ell = \left\lceil \varepsilon_N^{-2} \sqrt{\delta_{N,\ell} h_\ell} \sum_{j=0}^L \sqrt{\frac{\delta_{N,j}}{h_j}} \right\rceil + 1 \quad \text{for } \ell \in \{0, 1, 2, \dots, L\}$$

to ensure that the overall variance is below the target  $\varepsilon_N^2$ .

**Multilevel Monte Carlo diffusion approximation** We used  $N_0 = 400$  for our precalculation of the variances. See Figure 1, where the best fit line is  $y = 0.99x + 2.75$ , and Figure 2, where the best fit line is  $y = 1.45x + 2.61$ , which are consistent with the exponent  $\alpha$  in Table 1.

**Multilevel Monte Carlo tau-leaping.** We used  $N_0 = 100$  for our precalculation of the variances. See Figure 1, where the best fit line is  $y = 1.12x + 3.70$ , and Figure 2, where the best fit line is  $y = 1.56x + 4.64$ , which are, up to a log factor, consistent with the exponent  $\alpha$  in Table 1.

**Unbiased tau-leaping multilevel Monte Carlo.** For our implementation of unbiased multilevel tau-leaping, we set  $h_L = \frac{2}{N} \text{LambertW}\left(\frac{N}{2}\right)$  and  $L = \lceil \log(h_L)/\log(2) \rceil$ . For each level, we utilized  $N_0 = 100$  independent sample trajectories in order to estimate  $\delta_{N,\ell}, C_\ell, \delta_{N,E}$ , and  $C_E$ , as defined in section 3. We then selected

$$n_\ell = \left\lceil \varepsilon_N^{-2} \sqrt{\frac{\delta_{N,\ell}}{C_\ell}} \left( \sum_{\ell=0}^L \sqrt{\delta_{N,\ell} C_\ell} + \sqrt{\delta_{N,E} C_E} \right) \right\rceil + 1 \quad \text{for } \ell \in \{0, 1, 2, \dots, L\}$$

and

$$n_E = \left\lceil \varepsilon_N^{-2} \sqrt{\frac{\delta_{N,E}}{C_E}} \left( \sum_{\ell=0}^L \sqrt{\delta_{N,\ell} C_\ell} + \sqrt{\delta_{N,E} C_E} \right) \right\rceil + 1$$

to ensure that the overall estimator variance is below our target  $\varepsilon_N^2$ . See Figure 1, where the best fit line is  $y = 1.08x + 3.71$ , and Figure 2, where the best fit line is  $y = 1.68x + 2.65$ , which are, up to a log factor, consistent with the exponent  $\alpha$  in Table 1.

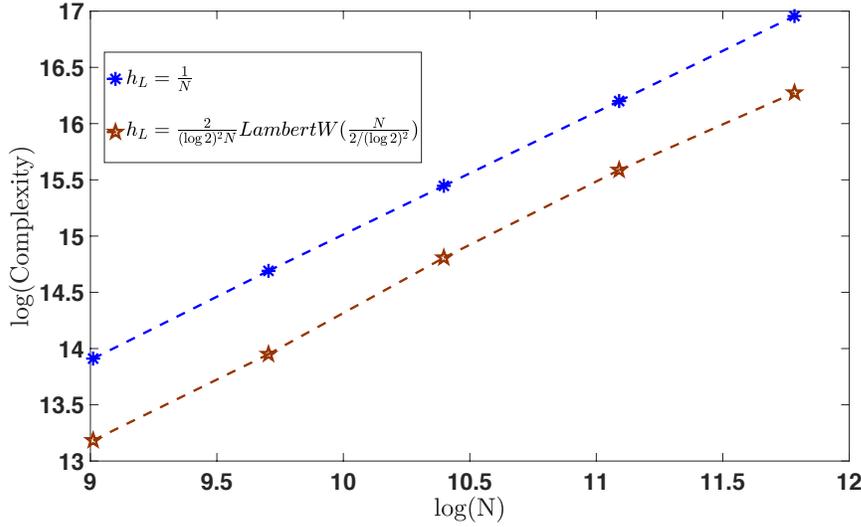


FIG. 3. Complexity comparison of unbiased multilevel Monte Carlo tau-leaping when  $h_L = \frac{1}{N}$  and  $h_L = \frac{2}{(\log 2)^2 N} \text{LambertW}(\frac{N}{2/(\log 2)^2})$  with  $\varepsilon_N = N^{-1}$ .

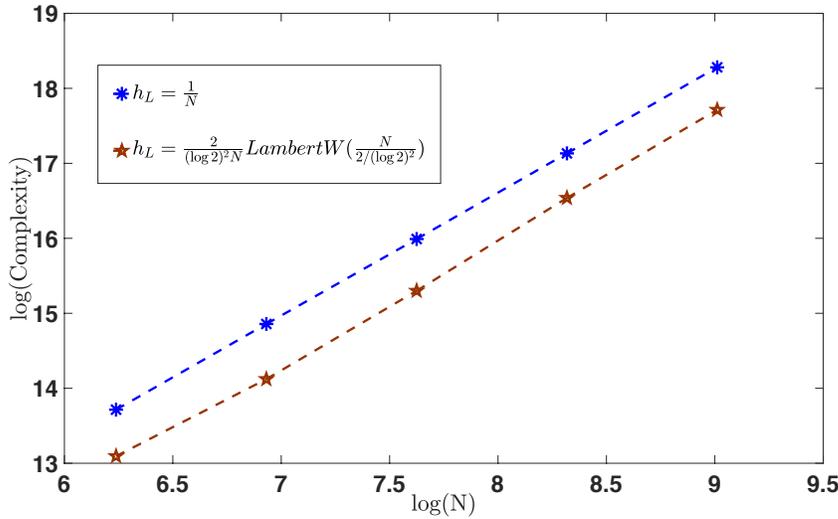


FIG. 4. Complexity comparison of unbiased multilevel Monte Carlo tau-leaping when  $h_L = \frac{1}{N}$  and  $h_L = \frac{2}{(\log 2)^2 N} \text{LambertW}(\frac{N}{2/(\log 2)^2})$  with  $\varepsilon_N = N^{-\frac{5}{4}}$ .

We also used the unbiased tau-leaping multilevel Monte Carlo method with  $h_L = N^{-1}$  to estimate  $\mathbb{E}[X_1(1)]$  to accuracy  $\varepsilon_N = N^{-\alpha}$  for both  $\alpha = 1$  and  $\alpha = 5/4$ . See Figures 3 and 4 for log-log plots of the required complexity when  $h_L = N^{-1}$  and  $h_L = \frac{2}{(\log 2)^2 N} \text{LambertW}(\frac{N}{2/(\log 2)^2})$ . As predicted in section 4.2.2, the complexity required when  $h_L = \frac{2}{(\log 2)^2 N} \text{LambertW}(\frac{N}{2/(\log 2)^2})$  is lower by some constant factor.

**6. Conclusions.** Many researchers have observed in practice that approximation methods can lead to computational efficiency relative to exact path simulation. However, *meaningful, rigorous justification for whether and under what circumstances*

approximation methods offer computational benefit has proved elusive. Focusing on the classical scaling, we note that a useful analysis must resolve two issues:

- (1) Computational complexity is most relevant for “large” problems, where many events take place. However, as the system size grows, the problem converges to a simpler, deterministic limit that is cheap to solve.
- (2) On a fixed problem, in the traditional numerical analysis setting where mesh size tends to zero, discretization methods become arbitrarily more expensive than exact simulation because the exact solution is piecewise constant.

In this work, we offer what we believe to be the first rigorous complexity analysis that allows for systematic comparison of simulation methods. The results, summarized in Table 1, apply under the classical scaling for a family of problems parametrized by the system size,  $N$ , with accuracy requirement  $N^{-\alpha}$ . In this regime, we can study performance on “large” problems when fluctuations are still relevant.

A simple conclusion from our analysis is that standard tau-leaping does offer a concrete advantage over exact simulation when the accuracy requirement is not too high ( $\alpha < 1$ ); see the first two rows of Table 1. Also, “second-order” midpoint or trapezoidal tau-leaping improves on exact simulation for  $\alpha < 2$ ; see row 3 of Table 1. Furthermore, in this framework, we were able to analyze the use of a diffusion, or Langevin, approximation and the multilevel Monte Carlo versions of tau-leaping and diffusion simulation. Our overall conclusion is that in this scaling regime, using exact samples alone is never worthwhile. For low accuracy ( $\alpha < 2/3$ ), second-order tau-leaping with standard Monte Carlo is the most efficient of the methods considered. At higher accuracy requirements ( $\alpha > 2/3$ ), multilevel Monte Carlo with a diffusion approximation is best so long as the bias inherent in perturbing the model is provably lower than the desired error tolerance. When no such analytic bounds can be achieved, multilevel versions of tau-leaping are the methods of choice. Moreover, for high accuracy ( $\alpha > 1$ ), the *unbiased* version is the most efficient, as it does not need to take a time step smaller than  $\varepsilon_N$  as the biased version must.

Possibilities for further research along the lines opened up by this work include the following:

- Analyzing other methods within this framework, for example, (a) multilevel Monte Carlo for the diffusion approximation using discretization methods customized for small noise systems or (b) methods that tackle the Chemical Master Equation directly using large-scale deterministic ODE technology [24, 25].
- Development of tau-leaping methods with weak order greater than two.
- Coupling the required accuracy to the system size in other scaling regimes, for example, to study specific problem classes with multiscale structure [11].
- Determining conditions on the system for when the diffusion approximation and Euler–Maruyama scheme achieve the  $\Theta(h)$  bias given in Assumption 2.2.
- Determining wider classes of models and functionals  $f$  for which Assumptions 2.1, 2.2, and 2.3 hold. In particular, most of the results in the literature require  $\lambda_k$  to be Lipschitz and for  $f$  to be a scalar valued function with domain  $\mathbb{Z}^d$  and bounded second derivatives.

#### REFERENCES

- [1] D. F. ANDERSON, *A modified next reaction method for simulating chemical systems with time dependent propensities and delays*, J. Chem. Phys., 127 (2007), p. 214107.
- [2] D. F. ANDERSON, *Incorporating postleap checks in tau-leaping*, J. Chem. Phys., 128 (2008), p. 054103.

- [3] D. F. ANDERSON, A. GANGULY, AND T. G. KURTZ, *Error analysis of tau-leap simulation methods*, Ann. Appl. Probab., 21 (2011), pp. 2226–2262.
- [4] D. F. ANDERSON AND D. J. HIGHAM, *Multilevel Monte-Carlo for continuous time Markov chains, with applications in biochemical kinetics*, SIAM Multiscale Model. Simul., 10 (2012), pp. 146–179.
- [5] D. F. ANDERSON, D. J. HIGHAM, AND Y. SUN, *Complexity of multilevel Monte Carlo tau-leaping*, SIAM J. Numer. Anal., 52 (2014), pp. 3106–3127.
- [6] D. F. ANDERSON, D. J. HIGHAM, AND Y. SUN, *Multilevel Monte Carlo for stochastic differential equations with small noise*, SIAM J. Numer. Anal., 54 (2016), pp. 505–529.
- [7] D. F. ANDERSON AND M. KOYAMA, *Weak error analysis of numerical methods for stochastic models of population processes*, SIAM Multiscale Model. Simul., 10 (2012), pp. 1493–1524.
- [8] D. F. ANDERSON AND T. G. KURTZ, *Continuous time Markov chain models for chemical reaction networks*, in Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology, H. Koepl, D. Densmore, G. Setti, and M. di Bernardo, eds., Springer, Berlin, 2011, pp. 3–42.
- [9] D. F. ANDERSON AND T. G. KURTZ, *Stochastic analysis of biochemical systems*, Stochastics in Biological Systems, 1.2, Springer International Publishing, Cham, 2015.
- [10] D. F. ANDERSON AND J. C. MATTINGLY, *A weak trapezoidal method for a class of stochastic differential equations*, Communi. Math. Sci., 9 (2011), pp. 301–318.
- [11] 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.
- [12] A. F. BARTHOLOMAY, *Stochastic models for chemical reactions. I. Theory of the unimolecular reaction process*, Bull. Math. Biophys., 20 (1958), pp. 175–190.
- [13] A. F. BARTHOLOMAY, *Stochastic models for chemical reactions. II. The unimolecular rate constant*, Bull. Math. Biophys., 21 (1959), pp. 363–373.
- [14] M. DELBRÜCK, *Statistical fluctuations in autocatalytic reactions*, J. Chem. Phys., 8 (1940), pp. 120–124, <https://doi.org/10.1063/1.1750549>, <http://link.aip.org/link/?JCP/8/120/1>.
- [15] C. W. GARDINER, *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences*, Springer, Berlin, 2002.
- [16] M. GIBSON AND J. BRUCK, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem. A, 105 (2000), pp. 1876–1889.
- [17] M. GILES, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.
- [18] 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.
- [19] D. T. GILLESPIE, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem., 81 (1977), pp. 2340–2361.
- [20] D. T. GILLESPIE, *Approximate accelerated simulation of chemically reacting systems*, J. Chem. Phys., 115 (2001), pp. 1716–1733.
- [21] P. GLYNN AND C. HAN RHEE, *Unbiased estimation with square root convergence for SDE models*, Oper. Res., 63 (2015), pp. 1026–1043.
- [22] M. HUTZENTHALER, A. JENTZEN, AND P. E. KLOEDEN, *Strong and weak divergence in finite time of Euler’s method for stochastic differential equations with non-globally Lipschitz continuous coefficients*, Proc. Roy. Soc. A, 467 (2011), pp. 1563–1576.
- [23] M. HUTZENTHALER, A. JENTZEN, AND P. E. KLOEDEN, *Divergence of the multilevel Monte Carlo Euler method for nonlinear stochastic differential equations*, Ann. Appl. Probab., 23 (2013), pp. 1913–1966.
- [24] T. JAHNKE, *On reduced models for the chemical master equation*, SIAM Multiscale Model. Simul., 9 (2011), pp. 1646–1676.
- [25] V. KAZEEV, M. KHAMMASH, M. NIP, AND C. SCHWAB, *Direct solution of the chemical master equation using quantized tensor trains*, PLOS Comput. Biol., (2014), <http://dx.doi.org/10.1371/journal.pcbi.1003359>.
- [26] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics 23, Springer-Verlag, Berlin, 1992.
- [27] T. G. KURTZ, *The relationship between stochastic and deterministic models for chemical reactions*, J. Chem. Phys., 57 (1972), pp. 2976–2978.
- [28] T. G. KURTZ, *Strong approximation theorems for density dependent Markov chains*, Stoch. Proc. Appl., 6 (1978), pp. 223–240.
- [29] T. G. KURTZ, *Representation and Approximation of Counting Processes*, Advances in Filtering and Optimal Stochastic Control 42, Springer, Berlin, 1982.
- [30] S. C. LEITE AND R. J. WILLIAMS, *A constrained Langevin approximation for chemical reaction networks*, <http://www.math.ucsd.edu/~williams/biochem/biochem.pdf>, 2016.

- [31] D. A. MCQUARRIE, *Stochastic approach to chemical kinetics*, J. Appl. Probab., 4 (1967), pp. 413–478.
- [32] E. RENSHAW, *Stochastic Population Processes*, Oxford University Press, Oxford, 2011.
- [33] S. M. ROSS, *Simulation*, 4th ed., Academic Press, Burlington, MA, 2006.
- [34] D. J. WILKINSON, *Stochastic Modelling for Systems Biology*, 2nd ed., Chapman and Hall/CRC Press, Boca Raton, FL, 2011.