COMPLEXITY OF MULTILEVEL MONTE CARLO TAU-LEAPING[*]

DAVID F. ANDERSON[†], DESMOND J. HIGHAM[‡], AND YU SUN[†]

Abstract. Tau-leaping is a popular discretization method for generating approximate paths of continuous time, discrete space Markov chains, notably for biochemical reaction systems. To compute expected values in this context, an appropriate multilevel Monte Carlo form of tau-leaping has been shown to improve efficiency dramatically. In this work we derive new analytic results concerning the computational complexity of multilevel Monte Carlo tau-leaping that are significantly sharper than previous ones. We avoid taking asymptotic limits and focus on a practical setting where the system size is large enough for many events to take place along a path, so that exact simulation of paths is expensive, making tau-leaping an attractive option. We use a general scaling of the system components that allows for the reaction rate constants and the abundances of species to vary over several orders of magnitude, and we exploit the random time change representation developed by Kurtz. The key feature of the analysis that allows for the sharper bounds is that when comparing relevant pairs of processes we analyze the variance of their difference directly rather than bounding via the second moment. Use of the second moment is natural in the setting of a diffusion equation, where multilevel Monte Carlo was first developed and where strong convergence results for numerical methods are readily available, but is not optimal for the Poisson-driven jump systems that we consider here. We also present computational results that illustrate the new analysis.

Key words. computational complexity, multilevel Monte Carlo, variance, tau-leaping, continuous time Markov chain, coupling

AMS subject classifications. 60H35, 65C05, 92C40, 65C20, 65C40

DOI. 10.1137/130940761

1. Introduction. Many modeling scenarios give rise to continuous time, discrete space Markov chains. Notable application areas include chemistry, systems biology, epidemiology, population dynamics, queuing theory, and several branches of physics [9, 17, 26, 27, 28]. It is straightforward to simulate sample paths for this class of processes, but in many realistic contexts the computational cost of performing Monte Carlo is prohibitive. This work focuses on the commonly arising task of computing an expected value of some feature of the solution; for example, the mean level of a chemical species at some specified time or the correlation between two population levels. We study the method proposed in [4], which combined ideas from tau-leaping and multilevel Monte Carlo in a manner that can dramatically improve computational complexity. Our main aim is to provide further analytical support for the method in the form of substantially sharper bounds for the variances of the coupled processes, and hence for the overall computational complexity of the resulting multilevel Monte Carlo estimator. The sharper analytic bounds will naturally inform implementation.

Tau-leaping, proposed by Gillespie [16], is an Euler-style discretization technique for generating paths that approximate those of the underlying process. Given a step-

---

size, h, the method proceeds by freezing the intensity functions over a subinterval of length h and then updating with an approximate summary of the events that would have taken place. Intuitively, this approach is attractive when (a) many events occur over each subinterval, so that exact simulation would be expensive, and (b) the relative change in the system state is small over each subinterval, so that the discretization is accurate [2, 16, 25]. However, in our case, where the aim is to combine sample paths in order to approximate an expected value, it should be kept in mind that

- we are concerned with the overall accuracy of the expected value approximation, not the accuracy of the individual paths, and
- forming a Monte Carlo style sample average automatically introduces a statistical sampling error, so we should focus on carefully balancing sampling and discretization effects.

In the case of simulating diffusion processes, these two points were highlighted in the multilevel Monte Carlo approach of Giles [12], with related earlier work by Heinrich [19], which delivers remarkable improvements in computational complexity in comparison with standard Monte Carlo. A key ingredient of multilevel Monte Carlo is to combine paths at different discretization levels, using relatively fewer paths at the more expensive resolution scales. The accuracy of the final estimate relies on a recursive control variate construction, exploiting the fact that pairs of paths at neighboring resolution levels can be made to have a low variance.

The multilevel philosophy was adapted in [4] for the case of tau-leaping. This required a novel simulation approach to generate tightly coupled pairs of Poisson-driven paths at neighboring resolutions. The resulting multilevel Monte Carlo tau-leaping algorithm is straightforward to implement and, unlike the standard multilevel methods utilized in the diffusion case, can be made unbiased by using an exact simulator at the most refined level. (We note, however, that the recent work of Rhee and Glynn produces an unbiased estimator in the diffusive setting utilizing a randomization idea [18].) Computational complexity was analyzed in [4] for a generic system scaling, and experimental results confirmed the potential of the method.

Our aim here is to refine the complexity analysis of [4]. We do this by directly estimating the variance between relevant pairs of processes, rather than bounding via the second moment. Second moment bounds arise naturally in the setting of a diffusion equation, where multilevel was first developed and where strong convergence results for numerical methods are readily available, but they are not optimal for the Poisson-driven jump systems that we consider here.

The next section sets up the notation, defines the simulation methods, and discusses some issues that arise in developing realistic results. Section 3 presents the new analytical results. The consequent bounds on computational complexity are derived in section 4. Section 5 then provides some numerical confirmation of the findings and section 6 gives conclusions. Some technical results required for the analysis are collected the appendix.

2. Background and notation. For concreteness, we follow [4] by using the language of chemical kinetics. We consider a system with d constituent species, $\{S_1, \ldots, S_d\}$, taking part in $K < \infty$ reactions, with the kth such reaction written as

$$\sum_{i=1}^{d} \nu_{ki} S_i \rightarrow \sum_{i=1}^{d} \nu'_{ki} S_i.$$
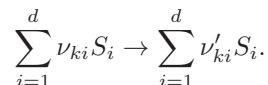
Here, $\nu_{ki} \in \mathbb{Z}_{\geq 0}$ specifies the number of molecules of type S_i required as input for the kth reaction, and $\nu'_{ki} \in \mathbb{Z}_{\geq 0}$ specifies the number of molecules of type S_i produced during the kth reaction. If the kth reaction occurs at time t, the system is updated via

$$X(t) = X(t-) + \nu'_k - \nu_k,$$

where ν_k and ν'_k are the vectors whose ith components are ν_{ki} and ν'_{ki}, respectively, and $X_i(t)$ denotes the number of S_i molecules at time t. For notational compactness, we let

$$\zeta_k \stackrel{\text{def}}{=} \nu'_k - \nu_k,$$

which is commonly termed the *reaction vector* for the kth reaction. Then $X(t)$ satisfies

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

where we have enumerated over the reactions, and $R_k(t)$ is the number of times the kth reaction has occurred by time t. We typically drop the K from the summation when this will not lead to confusion.

By far the most widely used stochastic model for this system assumes the existence of an *intensity*, or *propensity*, function $\lambda_k : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ for the kth reaction so that

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

where the collection $\{Y_k\}$ are independent unit-rate Poisson processes; see, for example, [1, 6, 22, 23]. Thus, $X(t)$ is the solution to the stochastic equation

$$(1) \qquad X(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X(s))ds \right) \zeta_k.$$

The standard choice for the intensity functions λ_k is that of mass-action kinetics, which assumes that for $x \in \mathbb{Z}_{\geq 0}^d$

$$(2) \qquad \lambda_k(x) = \kappa_k \prod_{i=1}^{d} \frac{x_i!}{(x_i - \nu_{ki})!} 1_{\{x_i \geq \nu_{ki}\}}.$$

We note that the continuous time Markov chain (1) is equivalent to the model derived by Gillespie [14, 15]. The corresponding forward Kolmogorov equation is often referred to as the Chemical Master Equation [28], and solving this very large ODE system forms the basis of an alternative computational approach that is appropriate when detailed information is required about the distribution of the solution process [21, 24].

In analyzing computational complexity, it is important to account for the fact that simulation becomes expensive when many reactions take place along a path, so some sort of "large system parameter" is required. Following [4, 7], we denote this parameter by N and scale the model by setting $X_i^N = N^{-\alpha_i}X_i$, where $\alpha_i \geq 0$ is chosen so that $X_i^N = O(1)$. The scaled variable then satisfies

$$X_i^N(t) = X_i^N(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X(s))ds \right) \zeta_{ki}^N,$$

where $\zeta_{ki}^N = N^{-\alpha_i}\zeta_{ki}$. Also following [4, 7], we set

$$\rho_k \stackrel{\text{def}}{=} \min\{\alpha_i \; : \; \zeta_{ki}^N \neq 0\},$$

so that $\zeta_{ki}^N = O(N^{-\rho_k})$, with the order achieved for at least one component of ζ_k^N (i.e., there can be j for which $\zeta_{kj}^N = o(N^{-\rho_k})$). We also set $\rho \stackrel{\text{def}}{=} \min\{\rho_k\} \geq 0$.

Accounting for the fact that the rate parameters of (2), κ_k, may also vary over several orders of magnitude, for each k there is an r_k for which $\lambda_k(X) = N^{r_k}\lambda_k^N(X^N)$ so that $\lambda_k^N(X^N)$ is $O(1)$. This yields the stochastic equation

$$X^N(t) = X^N(0) + \sum_k Y_k \left(\int_0^t N^{r_k}\lambda_k^N(X^N(s))ds \right) \zeta_k^N.$$

Define now

$$\gamma \stackrel{\text{def}}{=} \max_k\{r_k - \rho_k\},$$

which is the natural time-scale for the model. For $\gamma > 0$ the shortest timescale in the problem is much smaller than 1, and for $\gamma < 0$ it is much larger. Our analysis will be relevant to the case where $\gamma \leq 0$.

Setting $r_k = \gamma + c_k$, we arrive at the stochastic equation for our general scaled model,

$$(3) \qquad X^N(t) = X^N(0) + \sum_k Y_k \left(N^\gamma \int_0^t N^{c_k}\lambda_k^N(X^N(s))ds \right) \zeta_k^N.$$

Note that, by construction, $c_k \leq \rho_k$. We may now regard

$$(4) \qquad \overline{N} = N^\gamma \sum_k N^{c_k}$$

as an order of magnitude for the number of computations required to generate a single path using an exact algorithm. We refer to [4, 7] for further details and illustrative examples pertaining to the scaling, and point out that the *classical scaling* is covered in this framework by taking $c_k \equiv \rho_k \equiv 1$ and $\gamma = 0$; see [3, 22].

Given a stepsize, h, the tau-leaping approximation for the scaled system (3) takes the form

$$(5) \qquad Z_h^N(t) = Z_h^N(0) + \sum_k Y_k \left(N^\gamma \int_0^t N^{c_k}\lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N,$$

where $\eta(s) \stackrel{\text{def}}{=} \lfloor \frac{s}{h} \rfloor h$. Note that we have defined the process over continuous time.

The multilevel tau-leaping method for approximating $\mathbb{E}[f(X^N(T))]$ uses levels $\ell = \ell_0, \ell_0 + 1, \ldots, L$, where both ℓ_0 and L are to be determined. The characteristic stepsize at level ℓ is given by $h_\ell = T \cdot M^{-\ell}$ for a fixed positive integer M, typically between 2 and 7. Using Z_ℓ^N to denote the tau-leaping process (5) generated with a step-size of h_ℓ, we consider the telescoping sum identity

$$(6) \qquad \mathbb{E}[f(Z_L^N(T))] = \mathbb{E}[f(Z_{\ell_0}^N(T))] + \sum_{\ell=\ell_0+1}^L \mathbb{E}[f(Z_\ell^N(T)) - f(Z_{\ell-1}^N(T))].$$

Exploiting the right-hand side of this identity, we introduce the estimators

$$(7) \quad \widehat{Q}_{\ell_0} \overset{\text{def}}{=} \frac{1}{n_0} \sum_{i=1}^{n_0} f(Z_{\ell_0,[i]}^N(T)) \quad \text{and} \quad \widehat{Q}_\ell \overset{\text{def}}{=} \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} (f(Z_{\ell,[i]}^N(T)) - f(Z_{\ell-1,[i]}^N(T))),$$

where the pairs $Z_{\ell,[i]}^N$ and $Z_{\ell-1,[i]}^N$ are to be generated in such a way that $\mathsf{Var}(\widehat{Q}_\ell)$ is small. We will then let

$$(8) \qquad \qquad \widehat{Q}_B \overset{\text{def}}{=} \sum_{\ell=\ell_0}^{L} \widehat{Q}_\ell$$

be the unbiased estimator for $\mathbb{E}[f(Z_L^N(T))]$. Note that $\widehat{Q}_B$ is, therefore, a biased estimator for the quantity $\mathbb{E}[f(X^N(T))]$.

We emphasize at this stage that the number of levels, $L - \ell_0 + 1$, and the number of paths per level, n_0 and n_ℓ, have not yet been specified, and will be determined by the required accuracy.

In a similar manner, we may exploit the fact that exact simulation is possible and use the identity

$$\mathbb{E}[f(X^N(T))] = \mathbb{E}[f(X^N(T)) - f(Z_L^N(T))] + \sum_{\ell=\ell_0+1}^{L} \mathbb{E}[f(Z_\ell^N) - f(Z_{\ell-1}^N)] + \mathbb{E}[f(Z_{\ell_0}^N(T))].$$

We then define estimators for the three types of terms on the right-hand side via (7) and

$$(9) \qquad \qquad \widehat{Q}_E \overset{\text{def}}{=} \frac{1}{n_E} \sum_{i=1}^{n_E} (f(X_{[i]}^N(T)) - f(Z_{L,[i]}^N(T))),$$

leading to

$$(10) \qquad \qquad \widehat{Q}_{UB} \overset{\text{def}}{=} \widehat{Q}_E + \sum_{\ell=\ell_0}^{L} \widehat{Q}_\ell,$$

which is an *unbiased* estimator for $\mathbb{E}[f(X^N(T))]$.

In the multilevel framework, rather than single paths, we generally compute pairs of paths, and tight coupling is the key to success [5]. Defining $a \wedge b \overset{\text{def}}{=} \min\{a, b\}$, the pair $(Z_\ell^N, Z_{\ell-1}^N)$ appearing in (7) is defined as the solution to the stochastic equation

(11)
$$Z_\ell^N(t) = Z_\ell^N(0) + \sum_k \zeta_k^N \left[Y_{k,1} \left(N^\gamma N^{c_k} \int_0^t \lambda_k^N(Z_\ell^N(\eta_\ell(s))) \wedge \lambda_k^N(Z_{\ell-1}^N(\eta_{\ell-1}(s))) ds \right) \right.$$
$$\left. + Y_{k,2} \left(N^\gamma N^{c_k} \int_0^t [\lambda_k^N(Z_\ell^N(\eta_\ell(s))) - \lambda_k(Z_\ell^N(\eta_\ell(s))) \wedge \lambda_k(Z_{\ell-1}^N(\eta_{\ell-1}(s)))] ds \right) \right],$$

(12)
$$Z_{\ell-1}^N(t) = Z_{\ell-1}^N(0) + \sum_k \zeta_k^N \left[Y_{k,1} \left(N^\gamma N^{c_k} \int_0^t \lambda_k(Z_\ell^N(\eta_\ell(s))) \wedge \lambda_k(Z_{\ell-1}^N(\eta_{\ell-1}(s))) ds \right) \right.$$
$$\left. + Y_{k,3} \left(N^\gamma N^{c_k} \int_0^t [\lambda_k(Z_{\ell-1}^N(\eta_{\ell-1}(s))) - \lambda_k(Z_\ell^N(\eta_\ell(s))) \wedge \lambda_k(Z_{\ell-1}^N(\eta_{\ell-1}(s)))] ds \right) \right],$$

where the $Y_{k,i}$, $i \in \{1, 2, 3\}$, are independent, unit-rate Poisson processes, and for each ℓ, we define $\eta_\ell(s) \overset{\text{def}}{=} \lfloor s/h_\ell \rfloor h_\ell$.

Similarly, for (9), the pair (X^N, Z_L^N) is defined as the solution to the stochastic equation

(13)

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

(14)

$$
\begin{aligned}
Z_L^N(t) = Z_L^N(0) &+ \sum_k Y_{k,1}\left(N^\gamma N^{c_k} \int_0^t \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_L^N(\eta_L(s)))ds \right) \zeta_k^N \\
&+ \sum_k Y_{k,3}\left(N^\gamma N^{c_k} \int_0^t [\lambda_k^N(Z_L^N(\eta_L(s))) - \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_L^N(\eta_L(s)))]ds \right) \zeta_k^N.
\end{aligned}
$$

In practice, both the biased and the unbiased multilevel Monte Carlo methods described above can be implemented straightforwardly; full pseudocode descriptions are given in [4]. In this setting, the exact paths X^N are essentially being simulated by the *next reaction method* [10], which has a natural relation to the random change of time representation (1); see [1].

To analyze the computational complexity of these methods, we assume that $\mathbb{E}[f(X^N(T))]$ is required to an accuracy of ϵ, in the sense of a confidence interval. Hence, we have in mind that ϵ is small. As discussed earlier, we also wish to regard the system parameter, N, as large. It is important, however, to keep in mind that we have a "moving target." As N increases the properties of the underlying system may vary. In particular, for the classical scaling, in the *thermodynamic limit* $N \to \infty$ the stochastic fluctuations become negligible [6, 22] and the system approaches a deterministic ODE. We are implicitly assuming that the problem is specified in a regime where fluctuations are of interest and the task is computationally challenging, so we regard ϵ as small and N as large without committing ourselves to asymptotic limits. A key aim in our analysis is to capture the effect that the prescribed task can become less costly as N increases, in the sense that fluctuations decay.

The levelwise estimators $\widehat{Q}_\ell$ and $\widehat{Q}_E$ appearing in (8) and (10) are independent, and hence their variances add. To make the overall variance achieve the desired value of $O(\epsilon^2)$, our aim is to choose the number of paths, n_ℓ, so that each level contributes equally to the overall variance. The goal, in terms of obtaining a good upper bound on the computational complexity of the method, is therefore to develop tight bounds on the variances of $f(Z_{\ell,[i]}^N(T)) - f(Z_{\ell-1,[i]}^N(T))$ and $f(X_{[i]}^N(T)) - f(Z_{L,[i]}^N(T))$.

We note that multilevel was first developed and analyzed for diffusion processes [12]. Here the classical and well-studied concept of L^2 strong error of a numerical method can be used. Two processes that are both close to the underlying exact process in the L^2 norm must also be close to each other, and the second moment trivially bounds the variance. This approach has proved useful in connection with Brownian motion [11, 13, 20] and is optimal in the sense that complexity results can be derived that match the residual $O(\epsilon^{-2})$ cost that would remain if a simulable expression for the exact solution were known. The L^2 approach was also used in [4] for the Possion-

driven case that we consider here, but our aim is to show that improved bounds can be obtained from a more refined analysis that studies the variances directly.

3. Analyzing the variance of the coupled processes. We begin by explicitly detailing the *running assumption* we make on the model. We will suppose that for each k, the intensity function for the normalized process $\lambda_k^N(x)$, along with the components of its gradient and Hessian, are bounded.

RUNNING ASSUMPTION. *We suppose there is a constant $C > 0$ such that*

$$\sum_k |\lambda_k^N(x)| \leq C, \quad \sum_k \left| \frac{\partial \lambda_k^N(x)}{\partial x_j} \right| \leq C,$$

$$and \ \sum_k \left| \frac{\partial^2 \lambda_k^N(x)}{\partial x_i \partial x_l} \right| \leq C \quad for \ any \ i, j, l = 1, \dots, d.$$

Hence, letting

$$F^N(x) \stackrel{def}{=} \sum_k N^\gamma N^{c_k} \lambda_k^N(x) \zeta_k^N,$$

we have

$$\left| \frac{\partial F_i^N}{\partial x_j} \right| \leq CN^\gamma \quad and \quad \left| \frac{\partial^2 F_i^N}{\partial x_k \partial x_l} \right| \leq CN^\gamma \quad for \ any \ i, j, k, l = 1, 2, \dots, d.$$

We will also assume throughout that our initial condition is fixed. That is, $X^N(0) \equiv x(0) \in \mathbb{R}_{\geq 0}$ for all choices of N.

We note that not all reaction networks satisfy the above running assumption. However, any model for which there is a $w \in \mathbb{R}_{>0}^d$ with $w \cdot \zeta_k \leq 0$ for all k will satisfy the running assumption so long as λ_k^N is defined to satisfy it outside the positive orthant. (Note that the tau-leap process may leave the positive orthant.) In particular, any process which satisfies a conservation relation will satisfy the running assumption. We further wish to emphasize that the boundedness assumption is being made on the *scaled* intensity function, and not the intensity function for the unscaled process. Since the scaled intensity function is explicitly constructed to satisfy $\lambda_k^N(x) = O(1)$ *in our region of interest*, it is not an unrealistic assumption.

Example 1. Consider the family of models, indexed by N, with the following network, rate parameters, which are placed alongside the reaction arrows, and initial condition,

$$2A \underset{1}{\overset{1/N}{\rightleftarrows}} B, \qquad X_1(0) = X_2(0) = M \cdot N,$$
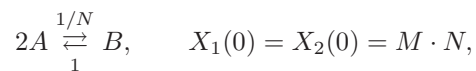
where $X_1(t)$ and $X_2(t)$ denote the abundances of the A and B molecules, respectively, at time t, $M > 0$ is some constant, and only those N for which $M \cdot N \in \mathbb{Z}$ are considered. Note that $w = [1, 2]^T$ satisfies $w \cdot \zeta_k = 0$ for each $k \in \{1, 2\}$. Choosing $\alpha_i \equiv 1$, so that $\rho_k \equiv r_k \equiv c_k \equiv \rho = 1$, and $\gamma = 0$, the normalized process satisfies

(15)
$$X^N(t) = X^N(0) + \frac{1}{N} Y_1 \left(N \int_0^t \lambda_1^N(X^N(s)) ds \right) \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$
$$+ \frac{1}{N} Y_2 \left(N \int_0^t \lambda_2^N(X^N(s)) ds \right) \begin{bmatrix} 2 \\ -1 \end{bmatrix},$$

with

$$\lambda_1^N(x) = \begin{cases} x_1(x_1 - N^{-1}) & \text{if } x_1, x_2 \geq 0, \\ h_1^N(x) & \text{if } x_1 < 0 \text{ or } x_2 < 0, \end{cases}$$

where $h_1^N(x)$ is any function which ensures λ_1^N is differentiable at the boundary of the positive orthant and satisfies our running assumptions and

$$\lambda_2^N(x) = \begin{cases} x_2 & \text{if } x_1, x_2 \geq 0, \\ h_2^N(x) & \text{if } x_1 < 0 \text{ or } x_2 < 0, \end{cases}$$

where $h_2^N(x)$ is also chosen to ensure the satisfaction of our running assumptions. For example, we require only that $h_2^N(x) = 0$ when $x_2 = 0$, $h_2^N(x) = (3/2)\lfloor M \cdot N \rfloor$ when $x_1 = 0$, and $h_2^N(x)$ itself satisfies the running assumptions outside the positive orthant. Note that it is necessary to define each λ_k^N outside $\mathbb{Z}_{\geq 0}^2$ for the tau-leaping process, which does not satisfy the same constraints as the exact process. Also note that a bound of $9M^2$ can be used to satisfy our running assumption for this (nonlinear) model.

3.1. Statements and proofs of our main results. We couple X^N and Z_h^N via (13) and (14) and are interested in the problem of estimating

$$\text{Var}\left(f(X^N) - f(Z_h^N)\right)$$

for Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$. Our main result is the following.

THEOREM 1. *Suppose the model satisfies our running assumption and that* X^N *and* Z_h^N *satisfy the coupling* (13) *and* (14). *Assume that* $f : \mathbb{R}^d \to \mathbb{R}$ *has continuous second derivative and there exists a constant* M *such that*

$$\left|\frac{\partial f}{\partial x_i}\right| \leq M \quad and \quad \left|\frac{\partial^2 f}{\partial x_i \partial x_j}\right| \leq M \quad for \; any \; i, j = 1, 2, \ldots, d.$$

Then, for $0 \leq t \leq T$,

$$\begin{aligned} \text{Var}(f(X^N(t)) - f(Z_h^N(t))) \leq & \bar{C}_1(N^\gamma T, d, M) N^{-\rho}(N^\gamma h)^2 \\ & + \bar{C}_2(N^\gamma T, d, M) N^{-\rho}(N^\gamma h), \end{aligned}$$

where $\bar{C}_1$ *is defined as* (30) *and* $\bar{C}_2$ *is defined as* (31).

Note that the functional dependence of $\bar{C}_1$ and $\bar{C}_2$ on N and T is via the product $N^\gamma T$, so in the case of T fixed and $\gamma \leq 0$, the values of $\bar{C}_1$ and $\bar{C}_2$ may be bounded above uniformly in N.

Remark 1. If it is the case that X remains in a bounded region with a probability of one, for example if there is a $w \in \mathbb{R}_{>0}^d$ with $w \cdot \zeta_k \leq 0$ for all k, then the assumption pertaining to the derivative bounds of f is automatically satisfied for any smooth f.

The main results in [4], which used the L^2 norm of the difference of the coupled processes, present an upper bound for the variance of $f(X^N(t)) - f(Z_h^N(t))$ of the form $\bar{C}_{11}(N^\gamma h)^2 + \bar{C}_{22}N^{-\rho}(N^\gamma h)$ for constants $\bar{C}_{11}$ and $\bar{C}_{22}$ which are similar to $\bar{C}_1$ and $\bar{C}_2$ of Theorem 1. Hence, a key improvement in the present work lies in the multiplication by $N^{-\rho}$ of the term $(N^\gamma h)^2$. Importantly, if $\gamma \leq 0$ and if $\bar{C}_1$ is not significantly larger than $\bar{C}_2$, we may now conclude that the variance has an upper bound that is dominated by a term of the form $N^{-\rho}(N^\gamma h)$ multiplied by a constant.

This is a conclusion that the analysis in [4] does not allow us to reach under similarly reasonable assumptions.

An immediate corollary to Theorem 1 is that the coupled tau-leaping processes satisfy a similar bound.

COROLLARY 1. *Under the assumptions of Theorem 1, Z_ℓ^N and $Z_{\ell-1}^N$ in (11) and (12) satisfy, for $0 \le t \le T$,*

$$
\begin{aligned}
\mathsf{Var}(f(Z_\ell^N(t)) - f(Z_{\ell-1}^N(t))) \le\ & \bar{D}_1(N^\gamma T, d, M)N^{-\rho}(N^\gamma h_\ell)^2 \\
& + \bar{D}_2(N^\gamma T, d, M)N^{-\rho}(N^\gamma h_\ell).
\end{aligned}
$$

Proof. This is an immediate consequence of the following simple inequality:

$$
\mathsf{Var}(f(Z_\ell^N(t)) - f(Z_{\ell-1}^N(t))) = \mathsf{Var}(f(Z_\ell^N(t)) - f(X^N(t)) + f(X^N(t)) - f(Z_{\ell-1}^N(t)))
$$

$$
\le 2\mathsf{Var}(f(Z_\ell^N(t)) - f(X^N(t))) + 2\mathsf{Var}(f(X^N(t)) - f(Z_{\ell-1}^N(t)))
$$

and Theorem 1.   □

In order to prove Theorem 1, we first present some preliminary calculations on the variance of the processes X^N and Z_h^N, and some useful lemmas.

By our running assumption, we know that

$$
|F_i^N(x) - F_i^N(y)| \le \sqrt{d}CN^\gamma |x - y|
$$

for all x, y and $i = 1, 2, \ldots, d$, where $|\cdot|$ is the 2-norm. We let x^N be the solution to the *deterministic* equation

$$
(16) \qquad\qquad x^N(t) = x(0) + \int_0^t F^N(x^N(s))ds,
$$

where $x(0)$ is the initial condition for each member of our family of processes.

LEMMA 1. *Under our running assumption, for any $T > 0$ we have*

$$
\mathbb{E}\left[\sup_{s \le T} |X^N(s) - x^N(s)|^2\right] \le \left(C_1 e^{C_2 \cdot (TN^\gamma)^2}\right)\left(TN^\gamma N^{-\rho}\right),
$$

where C_1 and C_2 are two positive constants that do not depend on N, γ, T.

Proof. Defining $\tilde{Y}_k(u) := Y(u) - u$ to be the centered Poisson process we have

$$
\begin{aligned}
X^N(t) - x^N(t) =\ & \sum_k \tilde{Y}_k\left(N^\gamma N^{c_k}\int_0^t \lambda_k^N(X^N(s))ds\right)\zeta_k^N \\
& + \int_0^t F^N(X^N(s)) - F^N(x^N(s))ds,
\end{aligned}
$$

and so using the trivial bound $(a + b)^2 \le 2a^2 + 2b^2$ plus our running assumption,

$$
\begin{aligned}
|X^N(t) - x^N(t)|^2 & \\
\le\ & 2\left|\sum_k \tilde{Y}_k\left(N^\gamma N^{c_k}\int_0^t \lambda_k^N(X^N(s))ds\right)\zeta_k^N\right|^2 \\
& + 2C^2 t d N^{2\gamma}\int_0^t |X^N(s) - x^N(s)|^2 ds.
\end{aligned}
$$

Hence, by the Burkholder–Davis–Gundy inequality [8] and our running assumption

$$
\mathbb{E}\left[\sup_{s\leq t}|X^N(s)-x^N(s)|^2\right]
$$

$$
\leq 2N^\gamma \sum_k |\zeta_k^N|^2 \int_0^t N^{c_k}\mathbb{E}\left[\lambda_k^N(X^N(r))\right]dr
$$

$$
+ 2C^2 t d N^{2\gamma}\int_0^t \mathbb{E}\left[\sup_{s\leq r}|X^N(s)-x^N(s)|^2\right]dr
$$

$$
\leq C_1 N^\gamma N^{-\rho}t + 2C^2 t d N^{2\gamma}\int_0^t \mathbb{E}\left[\sup_{s\leq r}|X^N(s)-x^N(s)|^2\right]dr,
$$

and the result follows from Gronwall's inequality with $C_2 = 2C^2 d$.   □

Now we let z be the deterministic solution to

(17)
$$
z_h^N(t) = x(0) + \int_0^t F^N(z_h^N(\eta(s)))ds,
$$

which is the Euler approximate solution to the ODE (16).

LEMMA 2. *Under our running assumptions, for any $T > 0$*

$$
\mathbb{E}\left[\sup_{0\leq s\leq T}|Z_h^N(s)-z_h^N(s)|^2\right] \leq \left(C_1 e^{C_2\cdot(TN^\gamma)^2}\right)(TN^\gamma N^{-\rho}),
$$

where C_1 and C_2 are the same constants which appear in Lemma 1.

Proof. Following the proof of Lemma 1, we have

$$
Z_h^N(t) - z_h^N(t) = \sum_k \tilde{Y}_k\left(N^\gamma N^{c_k}\int_0^t \lambda_k^N(Z_h^N(\eta(s)))ds\right)\zeta_k^N
$$

$$
+ \int_0^t F^N(Z_h^N(\eta(s))) - F^N(z_h^N(\eta(s)))ds,
$$

and, again as a result of the Burkholder–Davis–Gundy inequality [8] and our running assumptions,

$$
\mathbb{E}\left[\sup_{s\leq t}\left|Z_h^N(s)-z_h^N(s)\right|^2\right]
$$

(18)
$$
\leq 2N^\gamma \sum_k |\zeta_k^N|^2 \int_0^t N^{c_k}\mathbb{E}\left[\lambda_k^N(Z_h^N(\eta(s)))\right]ds
$$

$$
+ 2C^2 t d N^{2\gamma}\int_0^t \mathbb{E}\left[\sup_{s\leq r}|Z_h^N(\eta(s))-z_h^N(\eta(s))|^2\right]dr
$$

$$
\leq C_1 N^\gamma N^{-\rho}t + C_2 t N^{2\gamma}\int_0^t \mathbb{E}\left[\sup_{s\leq r}|Z_h^N(s)-z_h^N(s)|^2\right]dr,
$$

where C_1 and C_2 are the same constants which appear in Lemma 1. An application of Gronwall's inequality now completes the proof.   □

Our proof of Theorem 1 will begin by considering the difference between X^N and Z_h^N. It is therefore useful to get bounds on the second moment of the quadratic

variation of a certain martingale which will arise naturally. We therefore let

$$
M^N(t) \overset{\text{def}}{=} \sum_k \tilde{Y}_{k,2} \left(N^\gamma N^{c_k} \int_0^t [\lambda_k^N(X^N(s)) - \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_h^N(\eta(s)))]ds \right) \zeta_k^N
$$

(19)

$$
- \sum_k \tilde{Y}_{k,3} \left(N^\gamma N^{c_k} \int_0^t [\lambda_k^N(Z_h^N(\eta(s))) - \lambda_k^N(X^N(s)) \wedge \lambda_k^N(Z_h^N(\eta(s)))]ds \right) \zeta_k^N,
$$

where, again, $\tilde{Y}_k$ is the centered Poisson process. The proof of the following lemma can be found in [4].

LEMMA 3. *Under our running assumption,*

$$
\mathbb{E}[|M^N(t)|^2] \le c_1 T e^{c_2 N^\gamma T} N^{2\gamma}(N^{-\rho}h),
$$

where c_1, c_2 are independent of N, γ, and T.

We are now ready to prove our main result.

Proof of Theorem 1. We first prove the result in the case that $f_i(x) = x_i$. We have

$$
X_i^N(t) - Z_{h,i}^N(t) = M_i^N(t) + \int_0^t F_i^N(X^N(s)) - F_i^N(Z_h^N(\eta(s)))ds
$$

(20)

$$
= M_i^N(t) + \int_0^t F_i^N(X^N(s)) - F_i^N(Z_h^N(s))ds
$$

$$
+ \int_0^t F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s)))ds,
$$

where $M^N(t)$ is defined in (19).

We will prove the result by first considering the variance of the first and third terms of (20). Consideration of the second term will then lead naturally to an application of Gronwall's inequality.

To begin, we note that Lemma 3 implies that

$$
\mathsf{Var}(M_i^N(t)) \le c_1 T e^{c_2 N^\gamma T} N^{2\gamma}(N^{-\rho}h)
$$

for some c_1 and c_2 which are positive and do not depend on N^γ and T.

Turning to the third term of (20), we have the following lemma.

LEMMA 4.

$$
\mathsf{Var}\left(\int_0^t F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s)))ds \right)
$$

$$
\le d\hat{C}_1 \cdot (TN^\gamma)^2 N^{-\rho}(N^\gamma h)^2 + d\hat{C}_2 \cdot (TN^\gamma)^2 N^{-\rho}N^\gamma h,
$$

where $\hat{C}_1$ is a constant defined via (26), which depends upon the product TN^γ, and $\hat{C}_2$ is a constant independent of N, γ, and T.

Proof. From Lemma 8 in the appendix, we have

(21)
$$
F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s)))
$$

$$
= \int_0^1 \nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s) - Z_h^N(\eta(s))))dr \cdot (Z_h^N(s) - Z_h^N(\eta(s))).
$$

In order to bound the right-hand side of (21), we will apply Lemma 6 in the appendix with $A^{N,h}$ as the jth component of $\int_0^1 \nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s) - Z_h^N(\eta(s))))dr$ and $B^{N,h}$ as the jth component of $(Z_h^N(s) - Z_h^N(\eta(s)))$. Hence, we must find appropriate bounds on these components.

We begin with $Z_h^N(s) - Z_h^N(\eta(s))$. As

$$(22) \qquad Z_h^N(s) = Z^N(0) + \sum_k Y_k \left(N^\gamma N^{c_k} \int_0^t \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N,$$

we see

$$Z_h^N(s) - Z_h^N(\eta(s)) = \sum_k Y_k \left(N^\gamma N^{c_k} \int_0^s \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N$$

$$- \sum_k Y_k \left(N^\gamma N^{c_k} \int_0^{\eta(s)} \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N$$

$$\stackrel{d}{=} \sum_k Y_k{}' \left(N^\gamma N^{c_k} \int_{\eta(s)}^s \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_k^N,$$

where the collection $\{Y_k{}'\}$ are independent unit-rate Poisson processes and the last equality is in the sense of distributions. This implies

$$\left| \mathbb{E}[Z_h^N(s) - Z_h^N(\eta(s))] \right| = \left| \mathbb{E}\left[\mathbb{E}[Z_h^N(s) - Z_h^N(\eta(s))|Z_h^N(\eta(s))] \right] \right| \leq CN^\gamma h.$$

Similarly, using Lemma 2 and the law of total variance, we may conclude

$$\mathsf{Var}(Z_{h,j}^N(s) - Z_{h,j}^N(\eta(s))) = \mathbb{E}\left[\mathsf{Var}(Z_{h,j}^N(s) - Z_{h,j}^N(\eta(s)) \mid Z_h^N(\eta(s))) \right]$$

$$+ \mathsf{Var}\left(\mathbb{E}[Z_{h,j}^N(s) - Z_{h,j}^N(\eta(s))] \mid Z_h^N(\eta(s)) \right)$$

$$= \mathbb{E}\left[\mathsf{Var}\left(\sum_k \tilde{Y}_k \left(N^\gamma N^{c_k} \int_{\eta(s)}^s \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_{kj}^N \,\Big|\, Z_h^N(\eta(s)) \right) \right]$$

$$+ \mathsf{Var}\left(\mathbb{E}\left[\sum_k \tilde{Y}_k \left(N^\gamma N^{c_k} \int_{\eta(s)}^s \lambda_k^N(Z_h^N(\eta(s)))ds \right) \zeta_{kj}^N \,\Big|\, Z_h^N(\eta(s)) \right] \right)$$

$$(23) \qquad = \sum_k (s - \eta(s))N^\gamma N^{c_k}\mathbb{E}\left[\lambda_k^N(Z_h^N(\eta(s))) \right] (\zeta_{kj}^N)^2$$

$$+ \mathsf{Var}\left(\sum_k (s - \eta(s))N^\gamma N^{c_k}\lambda_k^N(Z_h^N(\eta(s)))\zeta_{kj}^N \right)$$

$$\leq CN^{-\rho}N^\gamma h + h^2 N^{2\gamma}K \sum_k \mathsf{Var}\left(\lambda_k^N(Z_h^N(\eta(s))) \right)$$

$$\leq CN^{-\rho}N^\gamma h + h^2 N^{2\gamma}K \sum_k \mathbb{E}\left[\left| \lambda_k^N(Z_h^N(\eta(s))) - \lambda_k^N(z_h^N(\eta(s))) \right|^2 \right]$$

$$\leq CN^{-\rho}N^\gamma h + dN^{2\gamma}K^2C^2 \left(C_1 e^{C_2 \cdot (TN^\gamma)^2} \right) \left(TN^\gamma N^{-\rho} \right) h^2,$$

where we recall that K is the number of reaction channels, and Lemma 2 was utilized in the final inequality.

Turning to $\int_0^1 \nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s) - Z_h^N(\eta(s)))) dr$, we apply Lemma 5 in the appendix with $X_1(s) = Z_h^N(s)$, $X_2(s) = Z_h^N(\eta(s))$, $x_1(s) = z_h^N(s)$, $x_2(s) = z_h^N(\eta(s))$, and $u(x) = [\nabla F_i^N(x)]_j$ to obtain

$$\mathsf{Var}\left(\int_0^1 [\nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s)) - Z_h^N(\eta(s)))]_j \, dr\right) \le N^{2\gamma} \bar{C} N^{-\rho},$$

where

$$(24) \qquad \bar{C} = dC^2 C_1 e^{2C_2 \cdot (TN^\gamma)^2}(TN^\gamma N^{-\rho}).$$

In order to apply Lemma 6 with

$$A^{N,h} = \int_0^1 [\nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s) - Z_h^N(\eta(s))))]_j \, dr$$

and $B^N = Z_{h,j}^N(s) - Z_{h,j}^N(\eta(s))$, we note $\left|[\nabla F_i^N]_j\right| \le CN^\gamma$ from our running assumptions to see

$$\mathsf{Var}\left(\int_0^1 [\nabla F_i^N(Z_h^N(\eta(s)) + r(Z_h^N(s) - Z_h^N(\eta(s))))]_j \, dr \cdot (Z_{h,j}^N(s) - Z_{h,j}^N(\eta(s)))\right)$$

$$\le 3C^2 \bar{C} N^{2\gamma} N^{-\rho}(N^\gamma h)^2 + 15C^2 N^{2\gamma} \mathsf{Var}(Z_{h,j}^N(s) - Z_{h,j}^N(\eta(s)))$$

$$(25)$$

$$\le \hat{C}_1 N^{2\gamma} N^{-\rho}(N^\gamma h)^2 + \hat{C}_2 N^{2\gamma} N^{-\rho} N^\gamma h,$$

where the final inequality follows from (23) with

$$(26) \qquad \begin{aligned} \hat{C}_1 &= 3C^2 \bar{C} + 15dK^2 C^4 C_1 e^{C_2 \cdot (TN^\gamma)^2} TN^\gamma, \\ \hat{C}_2 &= 15C^3. \end{aligned}$$

Returning to (21), the above allows us to conclude

$$\mathsf{Var}(F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s))))$$
$$\le d^2 \hat{C}_1 N^{2\gamma} N^{-\rho}(N^\gamma h)^2 + d^2 \hat{C}_2 N^{2\gamma} N^{-\rho} N^\gamma h.$$

Finally, by Lemma 7 in the appendix,

$$\mathsf{Var}\left(\int_0^t F_i^N(Z_h^N(s)) - F_i^N(Z_h^N(\eta(s))) ds\right)$$
$$\le d^2 T^2 \hat{C}_1 N^{2\gamma} N^{-\rho}(N^\gamma h)^2 + d^2 T^2 \hat{C}_2 N^{2\gamma} N^{-\rho} N^\gamma h,$$

as desired. □

We now turn to the middle term of (20). We first write

$$F_i^N(X^N(s)) - F_i^N(Z_h^N(s)) = DF_i^N(s) \cdot (X^N(s) - Z_h^N(s)),$$

where

$$DF_i^N(s) = \int_0^1 \nabla F_i^N(Z_h^N(s) + r(X^N(s) - Z_h^N(s))) dr.$$

We will again apply Lemma 6 to get the necessary bounds. Therefore, we let $A^{N,h} = [DF_i^N(s)]_j$ and $B^{N,h} = X_j^N(s) - Z_{h,j}^N(s)$.

Letting $X_1(s) = X^N(s)$, $X_2(s) = Z^N(s)$, $x_1(s) = x^N(s)$, $x_2(s) = z_h^N(s)$, and $u(x) = [\nabla F_i^N(x)]_j$ for an application of Lemma 5, we have

$$\mathsf{Var}(A^{N,h}) \leq N^{2\gamma} \bar{C} N^{-\rho},$$

where $\bar{C}$ is defined in (24) and where we use our running assumption that $[\nabla F_i]_j$ is uniformly bounded by CN^γ. From [4, Lemma 3] we know

$$\mathbb{E}\left[|B^{N,h}|\right] \leq \tilde{c}_1(e^{\tilde{c}_2 N^\gamma T} - 1)N^\gamma h,$$

where $\tilde{c}_1$ and $\tilde{c}_2$ are constants independent of N, γ, and T. Hence, applying Lemma 6 we see

$$\mathsf{Var}([DF_i^N(s)]_j(X_j^N(s) - Z_{h,j}^N(s)))$$
$$\leq 3N^{2\gamma}\bar{C}\tilde{c}_1^2(e^{\tilde{c}_2 N^\gamma T} - 1)^2 N^{-\rho}(N^\gamma h)^2 + 15C^2 N^{2\gamma}\mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s))$$

and

$$\mathsf{Var}(F_i^N(X^N(s))F_i^N(Z_h^N(s))) \leq 15C^2 dN^{2\gamma}\sum_{j=1}^d \mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s))$$
$$+ d^2\bar{c}_1 N^{2\gamma}N^{-\rho}(N^\gamma h)^2,$$

where

(27) $$\bar{c}_1 = 3\bar{C}\tilde{c}_1^2(e^{\tilde{c}_2 TN^\gamma} - 1)^2.$$

Finally returning to (20), we may combine all of the above to see

$$\mathsf{Var}(X_i^N(t) - Z_{h,i}^N(t)) \leq 3c_1 Te^{c_2 N^\gamma T} N^{2\gamma}(N^{-\rho}h)$$

$$+ 3\left[15C^2 dN^{2\gamma}t\sum_{j=1}^d \int_0^t \mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s))\,ds + d^2\bar{c}_1 N^{2\gamma}N^{-\rho}(N^\gamma h)^2 T^2\right]$$

$$+ 3\left[d^2\hat{C}_1 \cdot (TN^\gamma)^2 N^{-\rho}(N^\gamma h)^2 + d^2\hat{C}_2 \cdot (TN^\gamma)^2 N^{-\rho}N^\gamma h\right]$$

$$\leq 3d^2 T^2(\hat{C}_1 + \bar{c}_1)N^{2\gamma}N^{-\rho}(N^\gamma h)^2 + 3(d^2 T^2\hat{C}_2 N^{2\gamma} + c_1 N^\gamma Te^{c_2 N^\gamma T})N^{-\rho}N^\gamma h$$

$$+ 45C^2 TdN^{2\gamma}\sum_{j=1}^d \int_0^t \mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s))ds.$$

Thus, setting

$$g(t) = \max_{i \in \{1,\ldots,d\}}\{\mathsf{Var}([X^N(t)]_i - [Z_h^N(t)]_i)\},$$

we have

$$g(t) \leq 3d^2 T^2(\hat{C}_1 + \bar{c}_1)N^{2\gamma}N^{-\rho}(N^\gamma h)^2 + 3(d^2 T^2\hat{C}_2 N^{2\gamma} + c_1 N^\gamma Te^{c_2 N^\gamma T})N^{-\rho}N^\gamma h$$

$$+ 45C^2 Td^2 N^{2\gamma}\int_0^t g(s)ds,$$

and the result under the assumption that $f_i(x) = x_i$ is shown by an application of Gronwall's inequality,

$$\mathsf{Var}(X_i^N(s) - Z_{h,i}^N(s)) \leq 3e^{45C^2T^2d^2N^{2\gamma}}d^2T^2(\hat{C}_1 + \bar{c}_1)N^{2\gamma}N^{-\rho}(N^\gamma h)^2$$

$$+ 3e^{45C^2T^2d^2N^{2\gamma}}(d^2T^2\hat{C}_2N^{2\gamma} + c_1N^\gamma Te^{c_2N^\gamma T})N^{-\rho}N^\gamma h$$

$$(28) \qquad = \tilde{C}_1 \cdot N^{-\rho}(N^\gamma h)^2 + \tilde{C}_2 \cdot N^{-\rho}N^\gamma h,$$

where $\tilde{C}_1$ and $\tilde{C}_2$ are defined by the above equality.

To show the general case, note that from Lemma 8 in the appendix we have

$$f(X^N(t)) - f(Z_h^N(t)) = \int_0^1 \nabla f(Z_h^N(t) + r(X^N(t) - Z_h^N(t)))dr \cdot (X^N(t) - Z_h^N(t)).$$

We let $X_1(t) = X^N(t)$, $X_2(t) = Z_h^N(t)$, $x_1(t) = x^N(t)$, $x_2(t) = z_h^N(t)$, and $u(x) = [\nabla f(x)]_j$ for an application of Lemma 5, which yields

$$(29) \qquad \mathsf{Var}\left(\int_0^1 [\nabla f(Z_h^N(t) + r(X^N(t) - Z^N(t)))]_j dr\right) \leq dM^2\bar{C}N^{-\rho},$$

where M is the uniform bound of $\nabla_{ij}f(x)$ and the factor of d appears in the application of Lemma 5 since $|[\nabla f(x)]_j - [\nabla f(y)]_j| \leq M\sqrt{d}|x - y|$ for each j. Hence, by an application of Lemma 6 and the work above we see

$$\mathsf{Var}\left(\int_0^1 \nabla_j f(Z_h^N(t) + r(X^N(t) - Z_h^N(t)))dr \cdot (X_j^N(s) - Z_{h,j}^N(s))\right)$$

$$\leq dM^2\bar{c}_1N^{-\rho}(N^\gamma h)^2 + 15M^2\mathsf{Var}(X_j^N(s) - Z_{h,j}^N(s)).$$

Thus, utilizing (28), we have

$$\mathsf{Var}(f(X^N(t)) - f(Z_h^N(t))) \leq \bar{C}_1 \cdot N^{-\rho}(N^\gamma h)^2 + \bar{C}_2 \cdot N^{-\rho}N^\gamma h,$$

where

$$(30) \qquad \bar{C}_1 = 15d^2M^2\tilde{C}_1 + d^3m^2\bar{c}_1$$

and

$$(31) \qquad \bar{C}_2 = 15d^2M^2\tilde{C}_2.$$

Note that the functional dependence of $\bar{C}_1$ and $\bar{C}_2$ on N and T is via the product $N^\gamma T$. □

4. Consequences for complexity. In the present section we assume that $\gamma \leq 0$ (which holds, for example, in the classical scaling). We further assume that $\bar{C}_1$ is not significantly larger than $\bar{C}_2$, where $\bar{C}_1$ and $\bar{C}_2$ are the constants presented in Theorem 1. As detailed in the discussion following Theorem 1, under these assumptions our upper bounds on the variances given in Theorem 1 and Corollary 1 are well approximated by constants multiplied by $N^{-\rho}(N^\gamma h)$. This scaling is verified numerically on an example in section 5. Theoretically demonstrating the validity of the assumption pertaining to the relative sizes of the constants will require a finer analysis than is carried out here.

Our theoretical results, combined with the assumptions outlined above, imply that the variances in the levelwise estimators $\widehat{Q}_\ell$ and $\widehat{Q}_E$ appearing in (7) and (9) satisfy

$$(32) \qquad \mathsf{Var}(\widehat{Q}_\ell) \leq C \cdot \frac{N^{-\rho}N^{\gamma}h_\ell}{n_\ell} \qquad \text{and} \qquad \mathsf{Var}(\widehat{Q}_E) \leq C \cdot \frac{N^{-\rho}N^{\gamma}h_L}{n_E},$$

where C is a generic constant. Compared with the original analysis in [4, Theorems 1 and 2], the important refinement is the deletion of the h_ℓ^2 terms found in the similar bounds of [4]. These h_ℓ^2 terms dominate in most cases, including when $h_\ell > N^{-1}$ and the system satisfies the classical scaling (recall that $\rho = 1$ in the classical scaling).

Further, in the basic inequality

$$\mathsf{Var}(f(Z_{\ell_0}^N(t))) \leq \left(\sqrt{\mathsf{Var}(f(Z_{\ell_0}^N(t)) - f(X^N(t)))} + \sqrt{\mathsf{Var}(f(X^N(t)))} \right)^2,$$

we may use Theorem 1 with $h = h_{\ell_0}$ to control $\mathsf{Var}(f(Z_{\ell_0}) - f(X^N))$ and Lemma 1 to control $\mathsf{Var}(f(X^N))$ to find that

$$(33) \qquad \mathsf{Var}(\widehat{Q}_0) \leq C \cdot \frac{N^{-\rho}N^{\gamma}}{n_0}.$$

Taking a variance in (8) it then follows from (32) and (33) that to leading order

$$\mathsf{Var}(\widehat{Q}_B) \leq CN^{-\rho}N^{\gamma} \left(\frac{1}{n_0} + \sum_{\ell=\ell_0+1}^{L} \frac{h_\ell}{n_\ell} \right).$$

To achieve the required overall variance of ϵ^2, and to spread the variance budget fairly evenly across the levels, we may then use, to order of magnitude,

$$(34) \qquad n_0 = N^{-\rho}N^{\gamma}(L - \ell_0 + 1)\epsilon^{-2} \qquad \text{and} \qquad n_\ell = N^{-\rho}N^{\gamma}(L - \ell_0 + 1)h_\ell\epsilon^{-2}.$$

For the biased estimator (8) we take $h_L = O(\epsilon)$ in order for the discretization error to be within our target accuracy. This gives $L = O(|\ln(\epsilon)|)$ levels. The computational cost of each pair of tau-leap paths is proportional to h_ℓ^{-1}, and hence the overall complexity is of magnitude

$$(35) \quad n_0 h_{\ell_0}^{-1} + \sum_{\ell=\ell_0+1}^{L} n_\ell h_\ell^{-1} = N^{-\rho}N^{\gamma}\epsilon^{-2} \left((L - \ell_0 + 1)h_{\ell_0}^{-1} + (L - \ell_0 + 1)(L - \ell_0) \right).$$

Based on this analysis, which we recall is performed under the assumption that $\gamma \leq 0$, we take $\ell_0 = 0$. Doing so yields a computational complexity of leading order

$$N^{-\rho}N^{\gamma}\epsilon^{-2}\ln(\epsilon)^2.$$

Still assuming that $\ell_0 = 0$, for the unbiased estimator (10) we may take, to leading order of magnitude,

$$n_0 = N^{-\rho}N^{\gamma}(L + 2)\epsilon^{-2}, \quad n_\ell = N^{-\rho}N^{\gamma}(L + 2)h_\ell\epsilon^{-2},$$
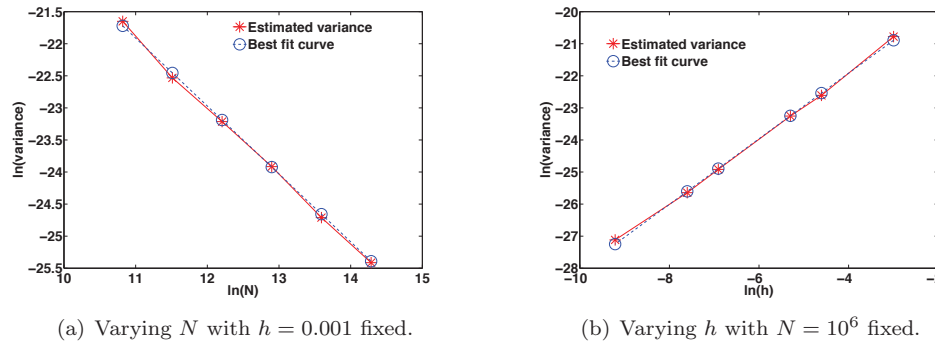$$\text{and} \quad n_E = N^{-\rho}N^{\gamma}(L + 2)h_L\epsilon^{-2}.$$

(a) Varying N with $h = 0.001$ fixed.　　(b) Varying h with $N = 10^6$ fixed.

FIG. 1. *Log-log plots of* $\mathsf{Var}(X_1^N(t) - Z_{h,1}^N(t))$ *for the model in Example 2. In (a), h is held constant while N is changed. In (b), N is fixed while h is varied. The best fit curve for all the data is overlain in the dashed line.*

If we again use $h_L = O(\epsilon)$, then the computational complexity for the unbiased method is bounded in magnitude by

$$(36) \qquad N^{-\rho}N^{\gamma}\epsilon^{-2}\ln(\epsilon)^2 + \max\{N^{-\rho}N^{\gamma}|\log(\epsilon)| \cdot h_L\epsilon^{-2}, 1\} \cdot \overline{N},$$

where we recall that $\overline{N}$, defined in (4), is the cost of computing a sample path with the exact method, and we note that the maximum appears because some exact paths are necessarily computed for the unbiased method. We also note that we are implicitly assuming that h_L^{-1} is not appreciably larger than $\overline{N}$, which we believe is reasonable. Finally, while we computed the above under the assumption that $h_L = O(\epsilon)$, we kept $h_L\epsilon^{-2}$ in the second term of (36) instead of simply writing ϵ^{-1} in order to explicitly point out the dependence on each term.

We note that the complexity bounds derived in [4], which considered only the L^2 norm of the difference of the coupled processes, have another term of order $\max\{N^{2\gamma}h_L^2\epsilon^{-2}, 1\} \cdot \overline{N}$ added to (36). This term was often the dominating one and has been removed by the direct analysis on the variance presented here.

To finish this section we point out that the analysis produces upper bounds on the computational complexity—in particular, the choices for the number of samples paths per level are sufficient to achieve the required accuracy, based on bounds on the individual variances and with the strategy of spreading the cost evenly between levels, but we have not shown that they are optimal. In practice, and as described more fully in [4], for a given problem, and with a small amount of further computation, it is possible to perform an initial optimization in order to choose these key parameters adaptively. Hence, practical performance may outstrip these complexity bounds.

5. Computational test. The efficiency of multilevel Monte Carlo tau-leaping was demonstrated computationally in [4], so we restrict ourselves here to testing the sharpness of our new analytical results on a simple nonlinear model. We consider a particular case from the family of models presented in Example 1.

Example 2. Consider the case where $M = 0.2$ in the model of Example 1, defined through (15). In Figure 1, we provide log-log plots of $\mathsf{Var}(X_1^N(T) - Z_{h,1}^N(T))$ for the coupled processes with $T = 0.3$, and varying values for N and h. The plots are consistent with the functional form

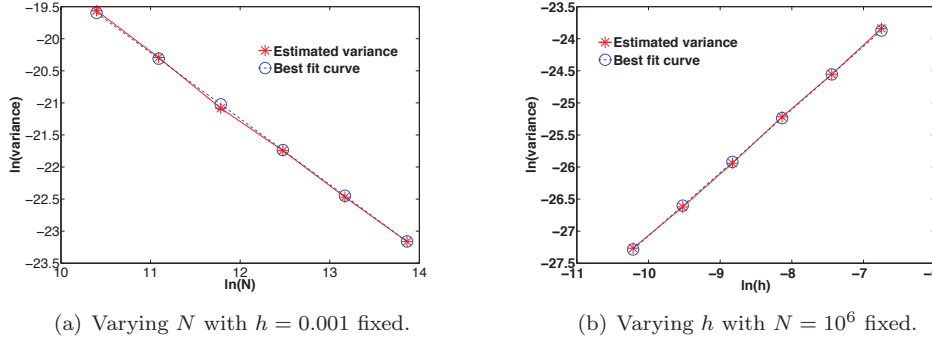$$(37) \qquad \mathsf{Var}(X_1^N(t) - Z_{h,1}^N(t)) \approx CN^{-1}h,$$

(a) Varying N with $h = 0.001$ fixed.　　　　(b) Varying h with $N = 10^6$ fixed.

FIG. 2. *Log-log plots of* $\mathsf{Var}(Z_{\ell,1}^N(t) - Z_{\ell-1,1}^N(t))$ *for the model in Example 2. In* (a), h *is held constant while* N *is changed. In* (b), N *is fixed while* h *is varied. The best fit curve for all the data is overlain in the dashed line.*

matching the bound arising from Theorem 1. The best fit curve for the data, obtained by a least squares approximation and which is shown in each image, is $\mathsf{Var}(X_1^N(t) - Z_{h,1}^N(t)) \approx 0.0408 \cdot N^{-1.0588} h^{1.0228}$.

In Figure 2, we provide log-log plots of $\mathsf{Var}(Z_{\ell,1}^N(T) - Z_{\ell-1,1}^N(T))$ for the coupled processes with $T = 0.3$, and varying values for N and h_ℓ. These plots also follow the functional form of (37), matching the bound arising from Corollary 1, with best fit curve of $\mathsf{Var}(Z_{\ell,1}^N(T) - Z_{\ell-1,1}^N(T)) \approx 0.1038 \cdot N^{-1.0279} h_\ell^{0.9845}$.

6. Conclusions. The main contribution of this work is to add further theoretical support for multilevel Monte Carlo tau-leaping by developing new complexity bounds that behave well for large values of the system size parameter. To do this, we took the novel step of directly estimating the variance between pairs of paths, rather than proceeding via a mean-square convergence property. We also provided numerical support showing our estimates for the variances are sharp.

Stochastic simulation of continuous time, discrete space, Markov chains is a bottleneck across a range of application areas, and there are many promising directions for further study of multilevel Monte Carlo in this context. In particular, specific instances of the very general scaling considered here could be used in order to develop more customized strategies, and complexity bounds, in suitable model classes—for example, where there is a known separation of scale.

The analysis presented is valid for $\gamma \leq 0$. For the case $\gamma > 0$, which is the regime of "stiff" systems, it is often possible to generate, via averaging techniques, a reduced model satisfying $\gamma \leq 0$. Taking this reduced model as the "finest level" in a multilevel Monte Carlo framework is then a natural way to proceed in the construction of an efficient Monte Carlo method. This procedure was carried out successfully in section 9 of [4] for an example of viral growth.

Appendix. Technical details from the analysis. We provide here some technical lemmas which were used in section 3.

LEMMA 5. *Suppose* $X_1(s)$ *and* $X_2(s)$ *are stochastic processes on* $\mathbb{R}^d$ *and that* $x_1(s)$ *and* $x_2(s)$ *are deterministic processes on* $\mathbb{R}^d$. *Further, suppose that*

(38)
$$\sup_{s \leq T} \mathbb{E}\left[|X_1(s) - x_1(s)|^2\right] \leq \widehat{C}_1(N^\gamma T)N^{-\rho}, \quad \sup_{s \leq T} \mathbb{E}\left[|X_2(s) - x_2(s)|^2\right] \leq \widehat{C}_2(N^\gamma T)N^{-\rho}$$

for some $\widehat{C}_1, \widehat{C}_2$ depending upon $N^\gamma T$. Assume that $u : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz with Lipschitz constant L. Then,

$$\sup_{s \leq T} \mathsf{Var}\left(\int_0^1 u(X_2(s) + r(X_1(s) - X_2(s)))dr\right) \leq L^2 \max(\widehat{C}_1, \widehat{C}_2)N^{-\rho}.$$

Proof. First, we know

$$\mathsf{Var}\left(\int_0^1 u(X_2(s) + r(X_1(s) - X_2(s)))dr\right)$$

$$= \mathsf{Var}\left(\int_0^1 u(X_2(s) + r(X_1(s) - X_2(s))) - u(x_2(s) + r(x_1(s) - x_2(s)))dr\right)$$

$$\leq \mathbb{E}\left(\int_0^1 u(X_2(s) + r(X_1(s) - X_2(s))) - u(x_2(s) + r(x_1(s) - x_2(s)))dr\right)^2$$

$$\leq \int_0^1 \mathbb{E}[u(X_2(s) + r(X_1(s) - X_2(s))) - u(x_2(s) + r(x_1(s) - x_2(s)))]^2 dr.$$

Using that u is Lipschitz, we may continue

$$\mathsf{Var}\left(\int_0^1 u(X_2(s) + r(X_1(s) - X_2(s)))dr\right)$$

$$\leq L^2 \int_0^1 \mathbb{E}\left[|X_2(s) + r(X_1(s) - X_2(s)) - (x_2(s) + r(x_1(s) - x_2(s)))|^2\right] dr$$

$$= L^2 \int_0^1 \mathbb{E}\left[|r(X_1(s) - x_1(s)) + (1 - r)(X_2(s) - x_2(s))|^2\right] dr$$

$$\leq L^2 \int_0^1 r\mathbb{E}\left[|X_1(s) - x_1(s)|^2\right] + (1 - r)\mathbb{E}\left[|X_2(s) - x_2(s)|^2\right] dr$$

$$\leq L^2 \max(\widehat{C}_1, \widehat{C}_2)N^{-\rho},$$

where the second to last inequality follows from convexity of the quadratic function, and the final inequality holds from applying (38). □

LEMMA 6. *Suppose that $A^{N,h}$ and $B^{N,h}$ are families of random variables determined by scaling parameters N and h. Further, suppose that there are $C_1 > 0, C_2 > 0$, and $C_3 > 0$ such that for all $N > 0$ the following three conditions hold:*

1. $\mathsf{Var}(A^{N,h}) \leq C_1 N^{-\rho}$ *uniformly in h.*
2. $|A^{N,h}| \leq C_2 N^\gamma$ *uniformly in h.*
3. $|\mathbb{E}[B^{N,h}]| \leq C_3 N^\gamma h$.

Then

$$\mathsf{Var}(A^{N,h}B^{N,h}) \leq 3C_3^2 C_1 N^{-\rho}(N^\gamma h)^2 + 15C_2^2 N^{2\gamma} \mathsf{Var}(B^{N,h}).$$

Proof. Via a direct expansion, the variance of the product can be represented in the following manner:

$$\mathsf{Var}(A^{N,h}B^{N,h}) = \mathbb{E}[(\mathbb{E}[B^{N,h}])(A^{N,h} - \mathbb{E}[A^{N,h}]) + (\mathbb{E}[A^{N,h}])(B^N - \mathbb{E}[B^{N,h}])$$
$$+ (A^{N,h} - \mathbb{E}[A^{N,h}])(B^{N,h} - \mathbb{E}[B^{N,h}])$$
$$- \mathbb{E}[(A^{N,h} - \mathbb{E}[A^{N,h}])(B^{N,h} - \mathbb{E}[B^{N,h}])]]^2.$$

Using the basic bound $(a + b + c)^2 \leq 3a^2 + 3b^2 + 3c^2$, we have

$$\mathsf{Var}(A^{N,h}B^{N,h}) \leq 3(\mathbb{E}[B^{N,h}])^2\mathsf{Var}(A^{N,h}) + 3(\mathbb{E}[A^{N,h}])^2\mathsf{Var}(B^{N,h})$$
$$+ 3\mathsf{Var}((A^{N,h} - \mathbb{E}[A^{N,h}])(B^{N,h} - \mathbb{E}[B^{N,h}])).$$

Using our assumptions in the statement of the lemma, the following two inequalities are immediate:

$$(39) \qquad\qquad 3(\mathbb{E}[B^{N,h}])^2\mathsf{Var}(A^{N,h}) \leq 3C_3^2 C_1 N^{-\rho}(N^\gamma h)^2$$

and

$$(40) \qquad\qquad 3(\mathbb{E}[A^{N,h}])^2\mathsf{Var}(B^{N,h}) \leq 3C_2^2 N^{2\gamma}\mathsf{Var}(B^{N,h}).$$

For the final term we bound the variance by the second moment to achieve

$$(41)$$
$$3\mathsf{Var}((A^{N,h} - \mathbb{E}[A^{N,h}])(B^{N,h} - \mathbb{E}[B^{N,h}])) \leq 3\mathbb{E}((A^{N,h} - \mathbb{E}[A^{N,h}])(B^{N,h} - \mathbb{E}[B^{N,h}]))^2$$
$$\leq 12C_2^2 N^{2\gamma}\mathsf{Var}(B^{N,h}).$$

Combining (39), (40), and (41) gives the desired result. $\square$

LEMMA 7. *Let $Q(s)$ be a stochastic process for which $\sup_{s\in[a,b]} \mathsf{Var}(Q(s)) < \infty$. Then*

$$\mathsf{Var}\left(\int_a^b Q(s)ds\right) \leq (b-a)\int_a^b \mathsf{Var}(Q(s))ds.$$

Proof. The proof is straightforward.

$$\mathsf{Var}\left(\int_a^b Q(s)ds\right) = \mathbb{E}\left(\int_a^b Q(s)ds - \mathbb{E}\left[\int_a^b Q(s)ds\right]\right)^2$$
$$= \mathbb{E}\left(\int_a^b (Q(s) - \mathbb{E}[Q(s)])ds\right)^2$$
$$\leq (b-a)\int_a^b \mathbb{E}\left[(Q(s) - \mathbb{E}Q(s))^2\right]ds = (b-a)\int_a^b \mathsf{Var}(Q(s))ds. \quad \square$$

LEMMA 8. *Let $f : \mathbb{R}^d \to \mathbb{R}$ have continuous first derivative. Then, for any $x, y \in \mathbb{R}^d$,*

$$f(x) = f(y) + \int_0^1 \nabla f(sx + (1-s)y)ds \cdot (x - y).$$

Proof. Let $H(t) = f(tx + (1-t)y)$. Then $H'(t) = \nabla f(tx + (1-t)y) \cdot (x - y)$, and by the fundamental theorem of calculus, $H(1) = H(0) + \int_0^1 H'(s)ds$, which is equivalent to the statement of the lemma. $\square$

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), 214107.

[2] D. F. Anderson, *Incorporating postleap checks in tau-leaping*, J. Chem. Phys., 128 (2008), 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*, Multiscale Model. Simul., 10 (2012), pp. 146–179.

[5] D. F. Anderson and M. Koyama, *An asymptotic relationship between coupling methods for stochastically modeled population processes*, preprint, arXiv:1403.3127, 2014; available online at http://imajna.oxfordjournals.org/content/early/2014/09/18/imanum.dru044.full.pdf+html, IMA J. Numer. Anal., accepted.

[6] 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. Koeppl, D. Densmore, G. Setti, and M. di Bernardo, eds., Springer, New York, 2011, pp. 3–42.

[7] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala, *Asymptotic analysis of multiscale approximations to reaction networks*, Ann. Appl. Probab., 16 (2006), pp. 1925–1961.

[8] D. L. Burkholder, B. J. Davis, and R. F. Gundy, *Integral inequalities for convex functions of operators on martingales*, in Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2, University of California Press, Berkeley, CA, 1972, pp. 223–240.

[9] C. W. Gardiner, *Handbook of Stochastic Methods: For Physics, Chemistry and the Natural Sciences*, Springer, Berlin, 2002.

[10] M. A. 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.

[11] M. B. Giles, *Improved multilevel Monte Carlo convergence using the Milstein scheme*, in Monte Carlo and Quasi-Monte Carlo Methods 2006, A. Keller, S. Heinrich, and H. Niederreiter, eds., Springer, Berlin, 2008, pp. 343–358.

[12] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

[13] M. Giles, D. J. Higham, and X. Mao, *Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff*, Finance Stoch., 13 (2009), pp. 403–413.

[14] 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.

[15] D. T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem., 81 (1977), pp. 2340–2361.

[16] D. T. Gillespie, *Approximate accelerated simulation of chemically reacting systems*, J. Chem. Phys., 115 (2001), pp. 1716–1733.

[17] D. T. Gillespie, A. Hellander, and L. Petzold, *Perspective: Stochastic algorithms for chemical kinetics*, J. Chem. Phys., 138 (2013), 170901.

[18] C. han Rhee and P. W. Glynn, *A new approach to unbiased estimation for SDE's*, in Proceedings of the 2012 Winter Simulation Conference, C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A.M Uhrmacher, eds., IEEE Press, Piscataway, NJ, 2012, pp. 201–207.

[19] S. Heinrich, *Multilevel Monte Carlo methods*, in Large-Scale Scientific Computing, Lecture Notes in Comput. Sci. 2179, Springer, New York, 2001, pp. 58–67.

[20] D. J. Higham, X. Mao, M. Roj, Q. Song, and G. Yin, *Mean exit times and the multilevel Monte Carlo method*, SIAM/ASA J. Uncertainty Quantification, 1 (2013), pp. 2–18.

[21] T. Jahnke, *On reduced models for the chemical master equation*, Multiscale Model. Simul., 9 (2011), pp. 1646–1676.

[22] T. G. Kurtz, *The relationship between stochastic and deterministic models for chemical reactions*, J. Chem. Phys., 57 (1972), pp. 2976–2978.

[23] T. G. Kurtz, *Representations of Markov processes as multiparameter time changes*, Ann. Probab., 8 (1980), pp. 682–715.

[24] S. MacNamara, K. Burrage, and R. B. Sidje, *Multiscale modeling of chemical kinetics via the master equation*, Multiscale Model. Simul., 6 (2007), pp. 1146–1168.

[25] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, *Consistency and stability of tau-leaping schemes for chemical reaction systems*, Multiscale Model. Simul., 4 (2005), pp. 867–895.

[26] E. Renshaw, *Stochastic Population Processes. Analysis, Approximations, Simulations*, Oxford University Press, Oxford, UK, 2011.

[27] S. M. Ross, *Simulation*, 4th ed., Academic Press, Burlington, MA, 2006.

[28] D. J. Wilkinson, *Stochastic Modelling for Systems Biology*, 2nd ed., Chapman and Hall/CRC Press, Boca Raton, FL, 2011.