Modeling delay in genetic networks: From delay birth-death processes to delay stochastic differential equations

Chinmaya Gupta,[1] José Manuel López,[1] Robert Azencott,[1] Matthew R. Bennett,[2,3] Krešimir Josić,[1,4] and William Ott[1]

[1)] *Department of Mathematics, University of Houston, Houston, TX*
[2)] *Department of Biochemistry & Cell Biology, Rice University, Houston, TX*
[3)] *Institute of Biosciences & Bioengineering, Rice University, Houston, TX*
[4)] *Department of Biology & Biochemistry, University of Houston, Houston, TX*

Delay is an important and ubiquitous aspect of many biochemical processes. For example, delay plays a central role in the dynamics of genetic regulatory networks as it stems from the sequential assembly of first mRNA and then protein. Genetic regulatory networks are therefore frequently modeled as stochastic birth-death processes with delay. Here we examine the relationship between delay birth-death processes and their appropriate approximating delay chemical Langevin equations. We prove a quantitive bound on the error between the pathwise realizations of these two processes. Our results hold for both fixed delay and distributed delay. Simulations demonstrate that the delay chemical Langevin approximation is accurate even at moderate system sizes. It captures dynamical features such as the oscillatory behavior in negative feedback circuits, cross-correlations between nodes in a network, and spatial and temporal information in two commonly studied motifs of metastability in biomechemical systems. Overall, these results provide a foundation for using delay stochastic differential equations to approximate the dynamics of birth-death processes with delay.

I. INTRODUCTION

Gene regulatory networks play a central role in cellular function by translating genotype into phenotype. By dynamically controlling gene expression, gene regulatory networks provide cells with a mechanism for responding to environmental challenges. Therefore, creating accurate mathematical models of gene regulation is a central goal of mathematical biology.

Delay in protein production can significantly affect the dynamics of gene regulatory networks. For example, delay can induce oscillations in systems with negative feedback[1–7], and has been implicated in the production of robust, tunable oscillations in synthetic gene circuits containing linked positive and negative feedback[8,9]. Indeed, delayed negative feedback is thought to govern the dynamics of circadian oscillators[10,11], a hypothesis experimentally verified in mammalian cells[12].

In genetic regulatory networks, noise and delay interact in subtle and complex ways. Delay can affect the stochastic properties of gene expression and hence the phenotype of the cell[2,13–16]. It is well known that noise can induce switching in bistable genetic circuits[17–26]; the infusion of delay dramatically enhances the stability of such circuits[27] and can induce an analog of stochastic resonance[28,29]. Variability in the delay time (distributed delay) can accelerate signaling in transcriptional signaling cascades[30].

Given the importance of delay in gene regulatory networks, it is necessary to develop methods to simulate and analyze such systems across spatial scales. In the absence of delay, it is well known that chemical reaction networks are accurately modeled by ordinary differential equations (ODEs) in the thermodynamic limit, *i.e.* when molecule numbers are sufficiently large. When molecule numbers are small, however, stochastic effects can dominate. In this case, the chemical master equation (CME) describes the evolution of the probability density function over all states of the system. Gillespie's stochastic simulation algorithm (SSA)[31] samples trajectories from the probability distribution described by the CME.

While exact, the CME is difficult to analyze and the SSA can be computationally expensive. To address these issues, a hierarchy of coarse-grained approximations of the SSA has been developed[32] (see Figure 1): Spatially discrete approximations, such as τ-leaping[33–36] and K-leaping[37] trade exactness for efficiency. At the next level are chemical Langevin equations (CLEs), which are stochastic differential equations of dimension equal to the number of species in the biochemical system. CLEs offer two advantages. First, unlike the SSA, the well-developed ideas from random dynamical systems and stochastic differential equations apply to CLEs. Second, it is straightforward to simulate large systems using CLEs. Finally, in the thermodynamic limit, one arrives at the end of the Markovian hierarchy: the reaction rate equation (RRE).

The Markovian hierarchy above (no delay) is well-understood[32,40], but a complete analogue of the Markovian theory does not yet exist for systems with delay. The SSA has been generalized to a delay version - the dSSA - to allow for both fixed[2,41] and variable[30,38] delay. Notably, the modification of the SSA to incorporate non-Markov processes predates the dSSA[42]; here many exponential steps from a linear chain are replaced by a gamma-distributed delay. Some analogues of τ-leaping exist for systems with delay; see *e.g.* D-leaping[43].

Several methods have been used to formally derive a delay chemical Langevin equation (dCLE) from the delay chemical master equation (dCME); see Section IV for details. Brett and Galla[39] use the path integral formalism of Martin, Siggia, Rose, Janssen, and de Dominicis to derive a dCLE approximation without relying on a mas-

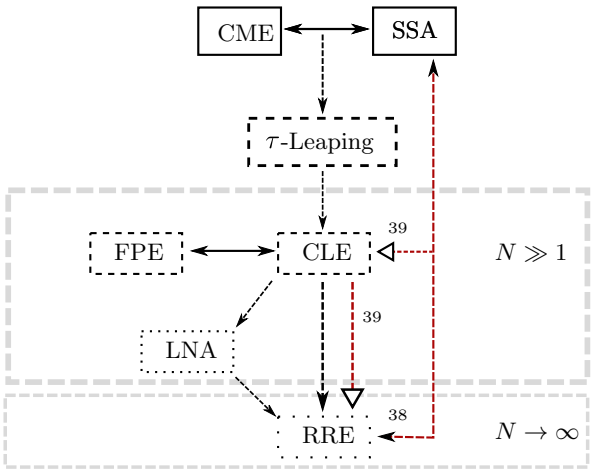

FIG. 1. Schematic of the modeling hierarchy for biochemical systems. The black arrows link the various components of the theory for Markov systems (no delay); red arrows link the corresponding delay components. Numbers attached to arrows refer to papers that establish the corresponding links. Empty arrowheads denote heuristic derivations; in this paper, we rigorously establish the dSSA to dCLE and dCLE to dRRE links.

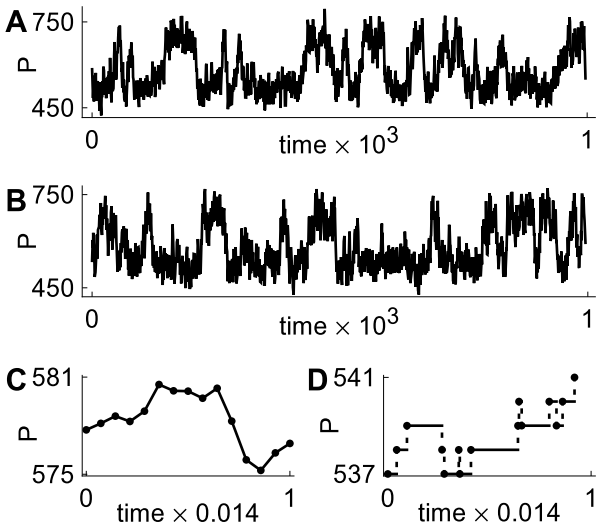

FIG. 2. (A) Typical trace obtained by using the dCLE. (B) Corresponding trace obtained using dSSA. By zooming into a small time segment, we see that the trace obtained from the dCLE (C) consists of equi-spaced time points (corresponding to an Euler discretization) and is continuous. The corresponding segment of the trace from the dSSA (D) consists of events that occur at random times and has jump discontinuities. The displayed traces were gathered after a long transient. The model simulated is the single-gene positive feedback model with $N = 500$; see Section II C 1.

ter equation. The Brett and Galla derivation produces the 'correct' dCLE approximation of the underlying delay birth-death (dBD) process in the sense that the first and second moments of the dCLE match those of the dBD process. However, their derivation has some limitations (see Section IV). In particular, it gives no rigorous quantitative information about the distance between the dBD process and the dCLE; it does not clarify for what types of deterministic dynamics (in the thermodynamic limit) the dCLE accurately approximates the delay birth-death process, and on what timescales; and it does not clarify in what sense (distributionally or pathwise) the dCLE approximates the delay birth-death process. Our analysis answers these questions.

In this paper, we establish a rigorous link between dBD processes and dCLEs by proving that the distance between the dBD process and the correct approximating dCLE process converges to zero as system size tends to infinity (as measured by expectations of functionals of the processes). In particular, this result applies to all moments. It is natural to express distance in terms of expectations of functionals because the dBD process is spatially discrete while the correct dCLE produces continuous trajectories (see Figure 2). This first result is just a corollary of our main mathematical theorem. We rigorously estimate the probability that a trajectory of the correct dCLE remains inside a narrow tube around the corresponding trajectory of the dBD process. This quantitative tube estimate will allow practitioners to determine the accuracy of simulations of concrete biochemical reaction networks. The mathematical results hold for both fixed delay and distributed delay (see Figure 3A).

The correct dCLE approximation is distinguished within the class of Gaussian approximations of the dBD process by the fact that it matches both the first and second moments of the dBD process. As we will see, it performs remarkably well at moderate system sizes in a number of dynamical settings: steady state dynamics, oscillatory dynamics, and certain metastable switches. We will demonstrate via simulation and argue mathematically using characteristic functions that no other Gaussian process with appropriately scaled noise performs as well. In the following, the term 'dCLE' shall refer specifically to the dCLE derived by Brett and Galla and expressed by (15), unless specifically stated otherwise. We provide a proof of our mathematical results in the supplement[44].

II. SIMULATIONS/OUTLINE AND INTERPRETATION OF RESULTS

Genetic regulatory networks may be simulated using an exact dSSA to account for transcriptional delay[2,30,38,41]. Here we provide a heuristic derivation of a related dCLE, and show that in a number of concrete examples it provides an excellent approximation of the system (see Figure 3). These simulations raise the fol-

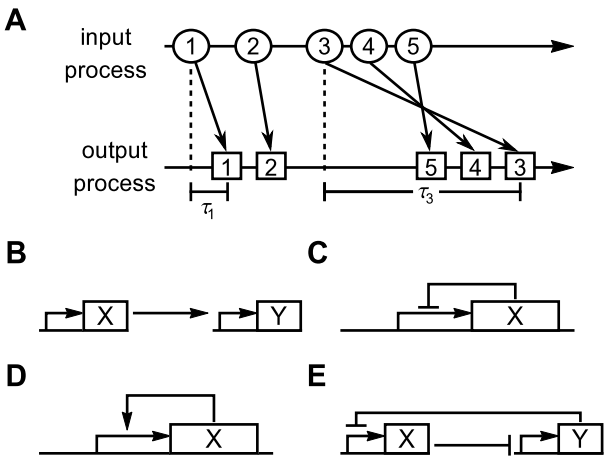

FIG. 3. (A) The effect of distributed delay on the protein production process. The "input process" is the first step in the transcription process, while the "output process" is the final mature product that enters the population. The time delay τ accounts for the lag between the initialization of transcription and the production of mature product. In a system with distributed delay, different production events can have different delay times; the order of the output process may therefore not match that of the input process. (B–E) Simulated gene regulatory network motifs: a transcriptional cascade (B), oscillators (C) and metastable systems (D–E).

lowing questions: Is the dCLE approximation valid in general? Can the expected quality of the approximation be quantified in general? We answer these questions mathematically in Section III.

We will adopt the following notation for reactions with delay,

$$X + Y \xrightarrow[\mu]{\alpha(X,Y)} Z$$

Here $\alpha(X,Y)$ denotes the rate of the reaction, the dashed arrow indicates a reaction with delay, and μ is a probability measure that describes the delay distribution. Solid arrows indicate reactions without delay.

A. A transcriptional cascade

First we consider a transcriptional cascade with two genes that code for proteins X and Y. Protein X is produced at a basal rate; production of Y is induced by the presence of X. The state of the system is represented by an ordered pair (X,Y). Note that we use X and Y to denote both protein names and protein numbers. The reactions in the network, and the associated state change vectors v_i, are given by

$$\emptyset \xrightarrow[\mu]{a} X \qquad v_1 = (1,0) \tag{1}$$

$$X \xrightarrow{b_1 X} \emptyset \qquad v_2 = (-1,0) \tag{2}$$

$$\emptyset \xrightarrow[\mu]{\psi(X)} Y \qquad v_3 = (0,1) \tag{3}$$

$$Y \xrightarrow{b_2 Y} \emptyset \qquad v_4 = (0,-1) \tag{4}$$

This system can be simulated exactly using the dSSA: Suppose the state of the system and the reactions in the queue are known at time t_0 (the queued reactions can be thought of as the "input process"; see Figure 3A), and that the delay kernel μ is supported on a finite interval $[0,\tau_0]$[45].

(1) Sample a waiting time t_w from an exponential distribution with parameter $A_0 := a + b_1 X + \psi(X) + b_2 Y$.

(2) If there is a reaction in the queue that is scheduled to exit at time $t_q < t_0 + t_w$, advance to time t_q and set $t_0 \to t_q$ and $(X,Y) \mapsto (X,Y) + (v_i^{(1)}, v_i^{(2)})$, where $v_i = (v_i^{(1)}, v_i^{(2)})$ is the change in the system due to the scheduled reaction. Finally, sample a new waiting time for the next reaction.

(3) If no reaction exits before $t_0 + t_w$, set $t_0 \mapsto t_0 + t_w$ and sample a reaction type from the set $\{1,2,3,4\}$ with probabilities proportional to $\{a, b_1 X, \psi(X), b_2 Y\}$, respectively. If the reaction chosen is a non-delayed reaction, perform the update $(X,Y) \mapsto (X-1,Y)$ (or $(X,Y) \mapsto (X,Y-1)$). However, if the reaction chosen is a delayed reaction, the state change vector $(1,0)$ (or $(0,1)$) is put into the queue along with an exit time t_q. The difference $\tau = t_q - t_0$ between the current and the exit time is sampled from the delay distribution μ.

We now heuristically derive the dCLE for the feedforward system from this spatially discrete process.

Suppose the delay kernel μ is given by a probability density function κ supported on $[0,\tau_0]$ ($\mathrm{d}\mu(s) = \kappa(s)\mathrm{d}s$). We first approximate the number of reactions that produce Y (Eq. (3)) that will be completed within the interval $[t_0, t_0+\Delta]$, where t_0 denotes the current time and Δ is a small increment. Since the production of Y involves delay, a reaction of this type that is completed within $[t_0, t_0+\Delta]$ must have been initiated at some time within $[t_0-\tau_0, t_0]$. Let $t_0 > t_0 - \Delta > t_0 - 2\Delta > \cdots$ be the partition of $[t_0-\tau_0, t_0]$ into intervals of length Δ. The (random) number of reactions completed within $[t_0, t_0+\Delta]$ *and* initiated within $[t_0-(i+1)\Delta, t_0-i\Delta]$ may be approximated by a Poisson random variable with mean

$$\psi(X_{t_0-(i+1)\Delta})\Delta \cdot \kappa((i+1)\Delta)\Delta.$$

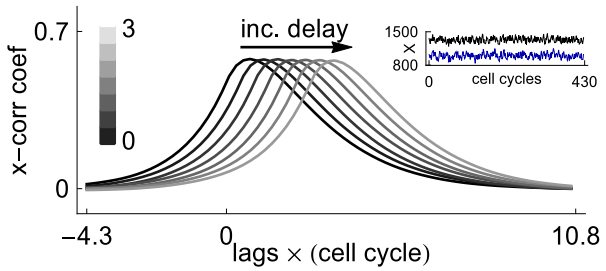

FIG. 4. Cross correlation functions for the two-species feedforward architecture at system size $N = 1000$. The inset shows sample trajectories for X (black; top) and Y (blue; bottom). The mean field model has a fixed point, and for this reason, stochastic dynamics stay within a neighborhood of this fixed point (see Theorem 1). However, the effect of increasing delay is clearly seen in the cross correlation function. The color bar displays the delay size corresponding to the cross correlation curves. Parameter values are given by $\tilde{\psi}(X) = \tilde{a}_1 x^3/(m^3+x^3)$, $\tilde{a} = 0.92$, $\tilde{a}_1 = 1.39$, $b_1 = b_2 = \ln(2)$, $m = 1.33$, $\mu = \delta_\tau$. The cross correlations have been normalized by dividing by the standard deviations σ_X and σ_Y of X and Y. Time has been normalized to cell cycle length.

Summing over i, the (random) number of reactions completed within $[t_0, t_0+\Delta]$ may be approximated by a Poisson random variable with mean

$$\Delta \sum_i \left[\psi(X_{t_0-(i+1)\Delta})(\kappa((i+1)\Delta)\Delta)\right];$$

this is a Riemann sum that approximates the integral

$$\Delta \int_0^{\tau_0} \psi(X_{t_0-s})\kappa(s)\,\mathrm{d}s.$$

Known as τ-leaping, this line of reasoning produces a Poissonian approximation of the dBD process:

$$\delta X_t = \mathrm{Poisn}(a\Delta) - \mathrm{Poisn}(b_1 X_t \Delta)$$
$$\delta Y_t = \mathrm{Poisn}\left(\Delta \int_0^{\tau_0} \psi(X_{t-s})\,\mathrm{d}\mu(s)\right) - \mathrm{Poisn}(b_2 Y_t \Delta).$$

Here $\mathrm{Poisn}(\eta)$ denotes a Poisson random variable with mean η.

If these Poisson random variables have large mean, they can be approximated by normal random variables. For example, the Poisson variable representing the number of reactions that produce Y can be approximated by a normal random variable with mean and variance equal to $\Delta \int_0^{\tau_0} \psi(X_{t-s})\,\mathrm{d}\mu(s)$. Since each reaction changes the state of either X or Y (but never both), it follows that the evolution of the system can be approximated by the stochastic difference equation

$$\delta X_t = \Delta(a - b_1 X_t) + \sqrt{\Delta(a + b_1 X_t)}\mathcal{N}(0,1) \tag{5a}$$
$$\delta Y_t = \Delta\left(\int_0^{\tau_0} \psi(X_{t-s})\,\mathrm{d}\mu(s) - b_2 Y_t\right) + \sqrt{\Delta\left(\int_0^{\tau_0} \psi(X_{t-s})\,\mathrm{d}\mu(s) + b_2 Y_t\right)}\mathcal{N}(0,1), \tag{5b}$$

where $\mathcal{N}(0,1)$ is the standard normal random variable.

System (5) may be written in terms of concentrations. Let N be a system size parameter. We think of N as a characteristic system size; X_t/N and Y_t/N therefore represent fractions of this characteristic value. Writing $x_t = X_t/N$, $y_t = Y_t/N$, $\tilde{\psi}(x) = \psi(Nx)/N$, and assuming that the basal production rate a scales with N as $a = \tilde{a}N$, we obtain

$$\delta x_t = \Delta(\tilde{a} - b_1 x_t) + \frac{1}{\sqrt{N}}\sqrt{\Delta(\tilde{a} + b_1 x_t)}\mathcal{N}(0,1) \tag{6a}$$
$$\delta y_t = \Delta\left(\int_0^{\tau_0} \tilde{\psi}(x_{t-s})\,\mathrm{d}\mu(s) - b_2 y_t\right) + \frac{1}{\sqrt{N}}\sqrt{\Delta\left(\int_0^{\tau_0} \tilde{\psi}(x_{t-s})\,\mathrm{d}\mu(s) + b_2 y_t\right)}\mathcal{N}(0,1). \tag{6b}$$

Eq. (6) is the Euler–Maruyama type discretization of a delay stochastic differential equation. Replacing Δ with $\mathrm{d}t$ and $\sqrt{\Delta}\mathcal{N}(0,1)$ with $\mathrm{d}W_t$ in (6), we obtain

$$\mathrm{d}x_t = (\tilde{a} - b_1 x_t)\,\mathrm{d}t + \frac{1}{\sqrt{N}}\sqrt{(\tilde{a} + b_1 x_t)}\,\mathrm{d}W_t^1 \tag{7a}$$
$$\mathrm{d}y_t = \left(\int_0^{\tau_0} \tilde{\psi}(x_{t-s})\,\mathrm{d}\mu(s) - b_2 y_t\right)\mathrm{d}t + \frac{1}{\sqrt{N}}\sqrt{\left(\int_0^{\tau_0} \tilde{\psi}(x_{t-s})\,\mathrm{d}\mu(s) + b_2 y_t\right)}\,\mathrm{d}W_t^2. \tag{7b}$$

This is the dCLE for the transcriptional cascade in this section.

Taking the formal thermodynamic limit, $N \to \infty$, in Eq. (7) yields the reaction rate equations derived in[38]:

$$\mathrm{d}x_t = (\tilde{a} - b_1 x_t)\,\mathrm{d}t \tag{8a}$$
$$\mathrm{d}y_t = \left(\int_0^{\tau_0} \tilde{\psi}(x_{t-s})\,\mathrm{d}\mu(s) - b_2 y_t\right)\mathrm{d}t. \tag{8b}$$

The dynamics described by Eq. (8) are quite simple; if $b_1 > 0$ and $b_2 > 0$, then (8) has a globally attracting stable stationary point.

To test the validity of the dCLE approximation given in Eq. (7), we examine if it captures the interaction between the two proteins in our transcriptional cascade network. Figure 4 shows the cross correlation functions obtained by simulating the system with $N = 1000$ using dSSA. From left to right, the curves correspond to fixed delay τ increasing from 0 to 3. The corresponding cross correlation curves for the dCLE approximation (7) are indistinguishable from those obtained using dSSA.

In the heuristic derivation above, we first fix N and let $\Delta \to 0$ to obtain the dCLE; we then separately let $N \to \infty$ to obtain the thermodynamic limit. Brett and Galla[39] also derive the dCLE by first fixing N and then sending $\Delta \to 0$. However, the two limits, $\Delta \to 0$ and $N \to \infty$, cannot be taken independently; this is a common problem with heuristic derivations of stochastic differential equations, even in the absence of delay[46]. The time discretization, Δ, can be thought of as a sampling frequency, while the system size, N, determines the rate at which reactions fire. If N becomes too large for a given Δ, then the number of reactions that fire within $[t, t+\Delta]$ no longer follows a Poisson distribution with mean dependent only on the state of the system at time t. On the other hand, if N is too small for a given Δ, then the Poisson distribution cannot be approximated by a normal distribution. In order to rigorously derive the Langevin approximation and estimate the distance between the dBD and dCLE processes, we will have to take a careful limit by relating Δ to N (with $\Delta \to 0$ as $N \to \infty$). We describe the proper scaling in Section III.

Applied to the transcriptional cascade, our main mathematical result (Theorem 1 in Section III) provides an explicit bound on the probability that the dCLE trajectory deviates from a narrow tube around the dBD trajectory (provided Δ is correctly scaled with respect to N).

In the previous example, the limiting deterministic system has a fixed point. Time series for the stochastic system therefore stay within a small neighborhood of this fixed point (see inset, Fig. 4). In the next example, we show that the dCLE approximation remains excellent even when the limiting deterministic dynamics are non-trivial. We consider a degrade-and-fire oscillator for which the limiting deterministic system has a limit cycle. The dCLE correctly captures the peak height and the inter-peak times for the dSSA realization of the degrade and fire oscillator, in addition to statistics such as the mean and variance. The approximation does not break down at small instantaneous protein numbers. Indeed, the mathematical theory developed in this work makes an important point: protein numbers at any particular time do not limit the quality of the dCLE approximation (in the presence of delay, or otherwise). Instead, the quality of the dCLE approximation depends on the latent parameter N. Theorem 1 makes this more precise: if one fixes the allowable error ε in the approximation of the dBD process by the dCLE process, then the time T during which the approximation error stays smaller than ε increases with N.

In the analysis of the previous example, and in the rest of this work, we make an assumption: There exists an integer N which describes the characteristic number of proteins in the system and all protein numbers and propensity functions can be scaled by this characteristic number. Because of this assumption, our setting is different from that of Gillespie[47], who used the term *chemical Langevin equation* to differentiate his approximating Langevin equation from previous work such as that of van Kampen[48]. While we assume the existence of a system size parameter, we obtain the approximating Langevin equation without reference to a chemical master equation.

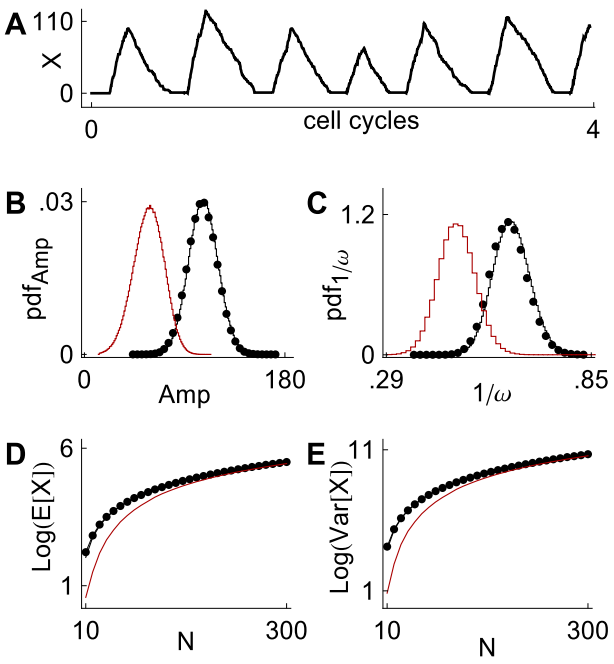

FIG. 5. Comparison of dSSA results to dSDE approximations for the degrade and fire oscillator. (A) depicts a stochastic realization of the oscillator generated by dSSA. We compare dSSA statistics (black dots in (B)–(E)) to those generated by the dCLE approximation given by Eq. (10) (black curves). We also show results for the dSDE approximation given by Eq. (11), obtained by removing delay from the diffusion term in Eq. (10) (red curves). At system size $N = 50$, the dCLE approximation given by Eq. (10) closely matches dSSA with respect to spike height distribution (B) and interspike interval distribution (C). In contrast, removing delay from the diffusion term results in a poor approximation of these distributions, as shown by the sizable shifts affecting the red curves. (D) and (E) illustrate mean repressor protein level and repressor protein variance, respectively, as functions of system size N. Eq. (10) provides a good approximation for all simulated values of N while the performance of Eq. (11) improves as N increases. The quantity P represents protein number, not protein concentration. Parameter values are $\alpha = 20.8$, $C_1 = 0.04$, $\beta = \ln(2)$, $V_{max} = 5.55$, $\gamma_0 = 0.01$, $\mu = \delta_{0.14}$. A soft boundary was added at 0 to ensure positivity.

B. Degrade and fire oscillator

The degrade and fire oscillator depicted schematically in Figure 3C consists of a single autorepressive gene and

corresponds to the reaction network

$$\emptyset \xrightarrow[\mu]{\psi(X)} X \xrightarrow{\gamma X} \emptyset$$

$$X + E \xrightarrow{\eta(X)} E$$

The production rate $\psi(X)$ is given by $\psi(X) = Nf(X/N)$, where f is the propensity function

$$f(x) = \frac{\alpha}{1 + (x/C_1)^4};$$

the enzymatic degradation rate $\eta(X)$ is given by $\eta(X) = Ng(X/N)$. Here $g(x) = V_{max}x/(K + x)$; K is the Michaelis-Menten constant, V_{max} the maximal enzymatic degradation rate, and γ the dilution rate coefficient. In the thermodynamic limit, the system is modeled by the delay differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \int_0^{\tau_0} \frac{\alpha}{1 + \left(\frac{x(t-s)}{C_1}\right)^4} \,\mathrm{d}\mu(s) - \gamma x(t) - \frac{V_{max}x(t)}{K + x(t)}. \quad (9)$$

As before, $x(t)$ denotes the concentration of protein X. We model the formation of functional repressor protein using distributed delay (described by the probability measure μ); this delayed negative feedback can induce oscillations[5]. Figure 5A depicts a sample realization of the stochastic version of the degrade and fire oscillator (the finite system size regime) generated by dSSA.

The dCLE approximation is in this case given by

$$\begin{aligned}\mathrm{d}x_t = &\int_0^{\tau_0} f(x_{t-s}) \,\mathrm{d}\mu(s) - \gamma x_t - g(x_t) \,\mathrm{d}t \\ &+ \frac{1}{\sqrt{N}} \left[\int_0^{\tau_0} f(x_{t-s}) \,\mathrm{d}\mu(s) + \gamma x_t + g(x_t)\right]^{\frac{1}{2}} \mathrm{d}W_t.\end{aligned} \quad (10)$$

Figure 5 illustrates that Eq. (10) provides a good approximation of the dBD dynamics, even when system size is relatively small. At system size $N = 50$, the spike height distribution and interspike interval distribution obtained using the dSSA (black dots in Figure 5B–5C) are nearly indistinguishable from those obtained using Eq. (10) (black curves in Figure 5B–5C). Further, we see a close match with respect to mean repressor protein level and repressor protein variance across a range of system sizes (Figure 5D–5E).

Interestingly, the dCLE approximation is very good even though the protein number approaches zero during part of the oscillation. This illustrates a central feature of the theory: the quality of the dCLE approximation is a function of a latent parameter N, not of the number of molecules present at any given time.

The exact form of the diffusion term is crucial to the accuracy of the dCLE approximation. If we remove delay from the diffusion term in Eq. (10), we obtain

$$\begin{aligned}\mathrm{d}x_t = &\int_0^{\tau_0} f(x_{t-s}) \,\mathrm{d}\mu(s) - \gamma x_t - g(x_t) \,\mathrm{d}t \\ &+ \frac{1}{\sqrt{N}} \left[f(x_t) + \gamma x_t + g(x_t)\right]^{1/2} \mathrm{d}W_t.\end{aligned} \quad (11)$$

At system size $N = 50$ (red curves in Figure 5B–5C), dSDE (11) produces dramatically different results from those generated by the correct dCLE approximation. The performance of Eq. (11) improves as N increases (Figure 5D–5E). This is expected, as both Eq. (10) and Eq. (11) converge weakly to Eq. (9) as $N \to \infty$.

C. Metastable systems

Understanding metastability in stochastic systems is of fundamental importance in the study of biological switches[17–26]. While metastability is well understood mathematically in the absence of delay, understanding the impact of delay on metastability remains a major theoretical and computational challenge[27,49,50]. We examine two canonical examples to show that for certain systems, the dCLE can be used to study the impact of delay on metastability: a positive feedback circuit and a co-repressive genetic toggle switch.

We offer a cautionary note before we proceed with the examples. Even in the Markovian context, SDE and Fokker-Planck modeling does not always accurately capture large deviations statistics associated with an underlying spatially-discrete stochastic process. Existing work for systems without delay (see *e.g.*[51–53]) illustrates the delicate relationship between continuum and discrete treatments of hitting times, extinction times, and the like. Determining when the dCLE accurately approximates the large deviations statistics of the underlying dBD process is an open problem.

1. Single species positive feedback circuit

The simplest metastable system consists of a single protein that drives its own production (Figure 3D). The chemical reaction network is given by

$$\emptyset \xrightarrow[\mu]{\psi(X)} X \xrightarrow{\gamma X} \emptyset$$

with $\psi(X) = Nf(X/N)$ for the propensity

$$f(x) = \alpha + \frac{\beta x^b}{c^b + x^b}.$$

In the thermodynamic limit, the dynamics of this model are described by the DDE

$$\mathrm{d}x_t = \int_0^{\tau_0} \alpha + \beta \frac{x(t-s)^b}{c^b + x(t-s)^b} \,\mathrm{d}\mu(s) - \gamma x(t) \,\mathrm{d}t. \quad (12)$$

Here x represents protein concentration and b is the Hill coefficient. In the thermodynamic limit, there are two stable stationary states, x_l and x_h, as well as an unstable stationary state x_s. These states satisfy $x_l < x_s < x_h$.

In the stochastic (finite N) regime, the stationary states x_l and x_h become metastable. We simulate the

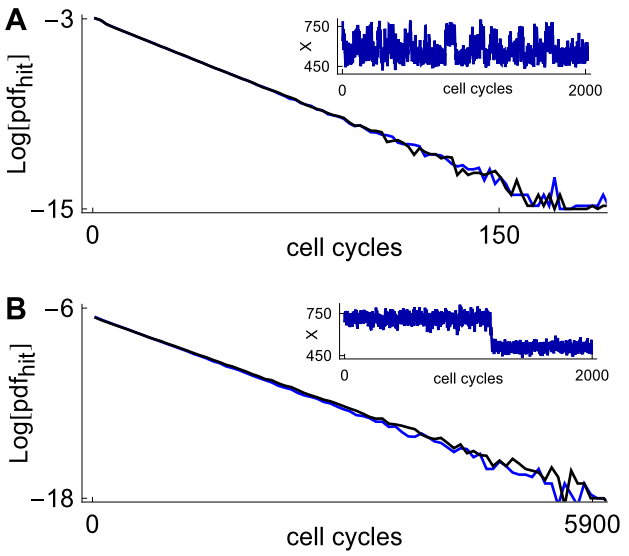

FIG. 6. Hitting time distributions for the positive feedback circuit. Black curves represent dSSA data; blue curves represent data from the dCLE approximation. The top panel corresponds to the Markov case $\tau = 0$ and a Hill coefficient of $b = 15$. The bottom row corresponds to a delay $\tau = 0.75$ and $b = 25$. The tail of the hitting times distribution becomes longer with both increasing delay and increasing Hill coefficient b. The dCLE captures the lengthening of the tail due to both effects. Parameter values are $\alpha = 0.35$, $\beta = 0.15$, $c = 0.615$, $\gamma = \ln(2)$, $\mu = \delta_\tau$, $N = 1000$.

metastable dynamics using dSSA and the dCLE approximation (15) given in this case by

$$\begin{aligned} \mathrm{d}x_t = &\int_0^{\tau_0} f(x_{t-s})\,\mathrm{d}\mu(s) - \gamma x_t\,\mathrm{d}t \\ &+ \frac{1}{\sqrt{N}}\left[\int_0^{\tau_0} f(x_{t-s})\,\mathrm{d}\mu(s) + \gamma x_t\right]^{\frac{1}{2}} \mathrm{d}W_t. \end{aligned} \tag{13}$$

Figure 6 displays hitting time distributions for the dSSA simulations (black curves) and Eq. (13) (blue curves). A hitting time is defined as follows: We choose neighborhoods $(x_l - \delta_l, x_l + \delta_l)$ and $(x_s - \delta_s, x_s + \delta_s)$ of x_l and x_s, respectively. We start the clock when a trajectory enters $(x_l - \delta_l, x_l + \delta_l)$ from the right. The clock is stopped when that trajectory first enters $(x_s - \delta_s, x_s + \delta_s)$. A hitting time is the amount of time that elapses from clock start to clock stop. We see that for no delay (Figure 6, top) and fixed delay $\tau = 1$ (bottom), the dCLE approximation accurately captures the hitting time distributions for Hill coefficients increasing from 15 to 25. Hence, the dCLE approximation accurately captures the rare events associated with a spatially-discrete delay stochastic process. This is significant because dSDEs are more amenable to large deviations theoretical analysis than their spatially-discrete counterparts.

Hitting times increase dramatically as the delay increases from 0 to 1, in accord with earlier analysis[27]. A dramatic increase is also seen as the Hill coefficient increases. This is due to the fact that the potential wells around x_l and x_h deepen as b increases.

2. Co-repressive toggle switch

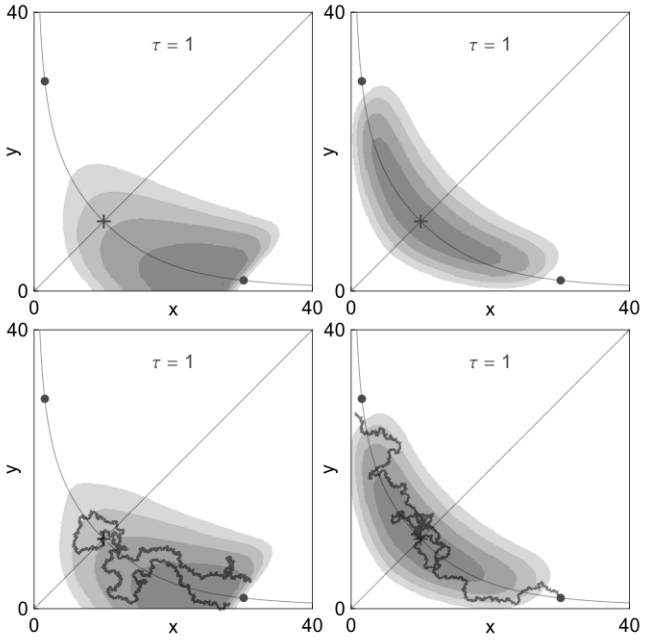

FIG. 7. Conditional density plots for the co-repressive toggle switch. The two left panels illustrate trajectories that leave a neighborhood of the stable point (x_h, y_l) and fall back into the same neighborhood before transitioning into a neighborhood of the stable point (x_l, y_h); displayed densities are conditioned on this set of trajectories. The two right panels illustrate trajectories that leave a neighborhood of (x_h, y_l) and transition to a neighborhood of (x_l, y_h) before falling back into the first neighborhood; displayed densities are conditioned on the set of such trajectories. Top panels correspond to dSSA; bottom panels correspond to dCLE. Cartoons of typical trajectories corresponding to failed transitions (left) and successful transitions (right) are shown for the dCLE process. Plots are shown for $N = 30$. The values of the other parameters are $\beta = 0.73$, $k = 0.05$, $\gamma = \ln(2)$.

The co-repressive toggle switch (Figure 3E) is a two-dimensional metastable system described in the thermodynamic limit by the DDEs

$$\mathrm{d}x_t = \int_0^{\tau_0} \frac{\beta}{1 + y(t-s)^2/k}\,\mathrm{d}\mu_1(s) - \gamma x\,\mathrm{d}t \tag{14a}$$

$$\mathrm{d}y_t = \int_0^{\tau_0} \frac{\beta}{1 + x(t-s)^2/k}\,\mathrm{d}\mu_1(s) - \gamma y\,\mathrm{d}t. \tag{14b}$$

The measure μ_1 describes the delay associated with production in this symmetric circuit. Eq. (14) has two stable stationary points (x_l, y_h) and (x_h, y_l) separated by the unstable manifold associated with a saddle equilibrium point (x_s, y_s). In the stochastic (finite system size) regime, the stable stationary points become metastable. In this regime a typical trajectory spends most of its time near the metastable points, occasionally moving between them.

Figure 7 displays density plots corresponding to trajectories that either successfully transition between metastable states (four panels on the right) or make failed transition attempts (four panels on the left). Even for the moderate system size $N = 30$, the density plots generated by dSSA (top four panels) closely match those generated by the dCLE approximation in Eq. (15) (bottom four panels).

Given the importance of rare events throughout stochastic dynamics, it is encouraging that the dCLE approximation captures well the dynamics of two popular biochemical motifs of metastability.

III. MAIN MATHEMATICAL RESULTS

The simulations thus far described suggest that the dCLE closely approximates the dBD process provided that the time increment Δ scales properly with N. Here we mathematically quantify the quality of the approximation and we determine the optimal scaling of Δ.

The general biochemical setup is as follows. Consider a reaction network of D biochemical species and M possible reactions. We are interested in describing the dynamics as a function of a latent system parameter N, the system size. The state of the system at time t is described by the vector $B_N(t) \in \mathbb{Z}^D$ that lists the number of molecules of each species present at time t.

Each reaction R_j is described by the following:

(a) A propensity function $f_j : \mathbb{R}^D \to \mathbb{R}^+$. The firing rate of reaction R_j is given by $Nf_j(B_N(t)/N)$. It is reasonable to use propensity functions to model firing rates in well-mixed chemical systems[47]. We assume the propensities satisfy mild technical conditions[44]; these conditions are essentially always satisfied by chemical reaction networks. Examples include reactions of first and second order, Michaelis-Menten kinetics, and propensities of Hill type.

(b) A state-change vector $\boldsymbol{v}_j \in \mathbb{Z}^D$. The vector $\boldsymbol{v}_j$ describes the change in the number of molecules of each species that results from the completion of a reaction of type j.

(c) A probability measure μ_j that models the delay between the initiation and completion of a reaction of type j. We allow the delay to be distributed (μ_j has a density) or fixed (μ_j is a Dirac mass δ_τ at a fixed delay τ). For instantaneous reactions, $\mu_j = \delta_0$.

In order for the dCLE to closely approximate the dBD process, the time step Δ must be small enough so that the propensity functions remain essentially constant on time intervals of length Δ and large enough so that many reactions fire on such intervals. Gillespie[47] calls such a Δ a *macroscopic infinitesimal time* for the reaction network. The primary mathematical result of this paper provides a scaling of Δ as a function of N for which Δ is rigorously shown to be a macroscopic infinitesimal time for the reaction network. Further, we quantify the quality of the dCLE approximation: with high probability, the solution trajectory of the dCLE will remain inside a narrow tube around the solution trajectory of the dBD process. We state the main result precisely and then discuss implications.

Since $B_N(t)$ is a vector of molecule numbers, we rescale by N to allow for a direct comparison of the dBD process and the dCLE. The correct dCLE approximation of $(b_N(t)) = (B_N(t)/N)$ is given for $1 \leqslant k \leqslant D$ by

$$\begin{aligned} \mathrm{d}x_k = &\left(\sum_{j=1}^{M} \int_0^{\tau_0} v_{jk} f_j(\boldsymbol{x}(t-s)) \, \mathrm{d}\mu_j(s) \right) \mathrm{d}t \\ &+ \frac{1}{\sqrt{N}} (\Sigma \, \mathsf{dW})_k, \end{aligned} \tag{15}$$

where W is a D-dimensional vector of independent standard Brownian motions and Σ^2 is given by

$$\Sigma^2_{lm} = \sum_{j=1}^{M} v_{jl} v_{jm} \int_0^{\tau_0} f_j(\boldsymbol{x}(t-s)) \, \mathrm{d}\mu_j(s).$$

Let ℓ_N denote the stochastic process defined by (15).

Theorem 1. *For any fixed time $T > 0$, $\Delta(N) = N^{-1/4}$ is a macroscopic infinitesimal time on $[0, T]$: There exist constants K_1 and K_2, depending only on T and the chemical reaction network, such that*

$$\mathbb{P}\left(\max_{t \in [0,T]} \| b_N(t) - \ell_N(t) \| > \frac{K_1}{N^{1/8}} \right) \leqslant K_2 e^{-\frac{1}{3} N^{1/4}}. \tag{16}$$

Theorem 1 has several interesting corollaries. First, for every continuous observable Ψ (real-valued function of trajectories of the processes), the difference between the expected value of Ψ with respect to the dBD and dCLE processes converges to zero as $N \to \infty$. This implies in particular that the distance between every moment of the processes converges to zero as $N \to \infty$. Second, since the right side of Eq. (16) is summable with respect to N, it follows from the Borel-Cantelli lemma that a trajectory of the dCLE process will, with probability one, lie within the tube of radius $K_1 T \zeta / N^{1/8}$ around the corresponding trajectory of the dBD process for all but finitely many values of N.

One would use Theorem 1 in a concrete situation by first fixing a desirable probability bound, say 1/1000. Using (16), one would then solve for the radius of the tube

around the trajectory of the dBD process for which the dCLE trajectory will remain in this tube up to time T with probability at least 999/1000.

We prove Theorem 1 in three steps[44]. First, we prove that the family of measures on trajectories associated with b_N is a tight family, roughly meaning that they are all approximately supported on a common compact set of trajectories. Second, we prove a version of (16) for 'discrete-time tubes' wherein the maximum on the left side of (16) is taken not over all $t \in [0, T]$ but rather over only $0, \Delta, 2\Delta, \ldots, (T/\Delta)\Delta$. Third, we study the size of fluctuations of the dBD and dCLE processes on time intervals of length Δ in order to upgrade the discrete-time tube estimate to the continuous-time tube estimate (16). We determine the optimal scaling of Δ during this final step. At the beginning of the proof, we assume that Δ scales with N as $\Delta(N) = N^{-\alpha}$ and then look for the optimal value of α. In order to both prove the discrete-time tube estimate and control fluctuations of the dCLE process on intervals of length Δ, we scale the radius $r(N)$ of the continuous-time tube as

$$r(N) = \frac{C_1}{N^{\alpha/2}} + \frac{C_2}{N^{(1-3\alpha)/2}}.$$

The first term in the sum measures local fluctuations of the dCLE while the second comes from the proof of the discrete-time tube estimate. In order to minimize $r(N)$, it is asymptotically optimal to have $\alpha = 1 - 3\alpha$, giving $\alpha = 1/4$.

We conclude this section with a cautionary note about what the main theorem does *not* say. Theorem 1 is formulated for fixed time intervals: one must first fix T and only then allow N to vary. As a consequence, this theorem does not apply when one scales T with N such that $T \to \infty$ as $N \to \infty$, as one does in the large deviations context.

IV. DISCUSSION

Stochastic differential equations (SDEs) are one of our main tools for modeling noisy processes in nature. Interactions between the components of a system or network are frequently not instantaneous. It is therefore natural to include such delay into corresponding stochastic models. However, the relationship between delay SDEs (dSDEs) and the processes they model has not been fully established.

Delay stochastic differential equations have previously been formally derived from the delay chemical master equation (dCME). Unlike the chemical master equation, however, the dCME is not closed; this complicates the derivation of dSDE approximations. Closure in this context means the following: Let $P(n, t)$ denote the probability that the stochastic system is in state n at time t. The dCME expresses the time derivative of $P(n, t)$ in terms of joint probabilities of the form $P(j, t; k, t - \tau)$ - the probability that the system is in state j at time t and was in state k at time $t - \tau$, where τ is the delay. The one-point probability distribution $P(\cdot, t)$ is therefore expressed in terms of two-point joint distributions, resulting in a system that is not closed. Timescale separation assumptions have been used to close the dCME. If the delay time is large compared to the other timescales in the system, one may assume that events that occur at time $t - \tau$ are decoupled from those that occur at time t and close the dCME[2,54] by assuming the joint probabilities may be written as products:

$$P(j, t; k, t - \tau) = P(j, t)P(k, t - \tau).$$

Having closed the dCME, one may then derive dSDE approximations[54] as well as useful expressions for autocorrelations and power spectra[2].

Approximations of dSDE type have also been derived using system size expansions such as van Kampen expansions[48,55] and Kramers-Moyal expansions for both fixed delay[56] and distributed delay[57].

Brett and Galla[39] use the path integral formalism of Martin, Siggia, Rose, Janssen, and de Dominicis to derive the delay chemical Langevin equation (dCLE) without relying on the dCME. Using this formalism, a moment generating functional may be expressed in terms of the system size parameter N and the sampling rate Δ. In the continuous-time limit, $\Delta \to 0$, the dCLE may be inferred from the moment generating functional. However, the Brett and Galla derivation has some limitations. First, the $\Delta \to 0$ limit cannot be taken without simultaneously letting $N \to \infty$. Intuitively, this is because as $\Delta \to 0$, the Gaussian approximation to the Poisson distribution with mean $N\lambda\Delta$ breaks down unless the parameter $N\lambda$ simultaneously diverges to infinity. Second, the derivation gives no quantitative information about the distance between the dCLE and the original delay birth-death (dBD) process.

Here, we addressed these shortcomings. We proved rigorously that the dBD process can be approximated by a class of Gaussian processes that includes the dCLE. In particular, we established that for most biophysically relevant propensity functions, the dCLE process will approximate all moments of the dBD process. Our proof includes bounds on the quality of the approximation in terms of the time T for which the approximation is desired to hold and the characteristic system size N (see Theorem 1). The error bounds also indicate that the quality of the dCLE approximation worsens with increasing upper bounds on the reaction propensity functions and state-change vectors. Physically, this means that high reaction rates and reactions that cause large changes in the protein populations are detrimental to the quality of the dCLE approximation.

The dCLE is one of many Gaussian processes that approximate the dBD process. Among all Gaussian approximations with noise components that scale as $1/\sqrt{N}$, the dCLE is optimal because it is the only such approximation that exactly matches the first and second moments of the dBD process. We formally justify this assertion in the

supplement[44] using characteristic functions. As our simulations of the degrade and fire oscillator demonstrate, the dCLE can significantly outperform other Gaussian approximations at moderate system sizes.

Nevertheless, the quantitative tube estimates in Theorem 1 apply to *any* Gaussian approximation of the dBD process provided the noise scales as $1/\sqrt{N}$. This is significant because it is often advantageous to use linear noise approximations of the dCLE. Delay appears in the drift component of a linear noise approximation but not in the diffusion component. Linear noise approximations are therefore easier to analyze than their dCLE counterparts. In particular, elements of the theory of large deviations for Markovian systems can be extended to SDEs with delay in the drift[58].

For metastable systems, our simulations indicate that the dCLE may capture both temporal information (such as hitting times for the positive feedback model; see Fig. 6) and spatial information (such as densities for trajectories corresponding to failed and successful transitions; see Fig. 7) for some examples. While SDE approximations (without delay) do not in general capture the dynamics in the presence of metastability even for Markov systems[51–53], our work suggests that dCLE approximations may be used, on a case by case basis, to study rare events for metastable biochemical systems.

We have made a number of simplifying assumptions in order to be able to prove the existence of a macroscopic infinitesimal time postulated by Gillespie[47]. In particular we assumed the existence of a latent system size parameter N, in terms of which this macroscopic infinitesimal time Δ can be expressed. However, for many biochemical systems, it is often unclear what the system size is, and in this case, a more careful analysis will be required to rigorously justify the use of delay stochastic differential equations to approximate the dynamics of the underlying discrete systems. Further complications can arise when the system size itself changes as a function of time or if there is a scale separation between the fast and slow components of the system. In the non-delayed case, hybrid systems[59–61] are often used to address these challenges. Developing a mathematical theory for hybrid systems in the presence of delays is an important future direction.

We have shown that the dCLE provides an accurate approximation of a number of stochastic processes. Although we chose gene regulatory networks in our examples, the theory is applicable to general birth-death processes with delayed events. SDEs, and the chemical Langevin equation in particular, are fundamental in modeling and understanding the behavior of natural and engineered systems. We therefore expect that the dCLE will be widely applicable when delays impact system dynamics.

Acknowledgements. This work was funded by the National Institutes of Health, through the joint National Science Foundation/National Institute of General Medical Sciences Mathematical Biology Program Grant R01GM104974 (to M.R.B., K.J., and W.O.), and the Robert A. Welch Foundation Grant C-1729 (to M.R.B.).

[1] A. Amir, S. Meshner, T. Beatus, and J. Stavans, Molecular Microbiology **76**, 428 (2010).
[2] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, Proceedings of the National Academy of Sciences of the United States of America **102**, 14593 (2005), http://www.pnas.org/content/102/41/14593.full.pdf+html.
[3] B. C. Goodwin, Advances in Enzyme Regulation **3**, 425 (1965).
[4] J. Lewis, Current Biology **13**, 1398 (2003).
[5] W. Mather, M. R. Bennett, J. Hasty, and L. S. Tsimring, Phys. Rev. Lett. **102**, 068105 (2009).
[6] N. A. Monk, Current Biology **13**, 1409 (2003).
[7] P. Smolen, D. Baxter, and J. Byrne, American Journal of Physiology-Cell Physiology **277**, C777 (1999).
[8] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, Nature **456**, 516 (2008).
[9] M. Tigges, T. T. Marquez-Lago, J. Stelling, and M. Fussenegger, Nature **457**, 309 (2009).
[10] P. Smolen, D. A. Baxter, and J. H. Byrne, Biophysical Journal **83**, 2349 (2002).
[11] K. Sriram and M. Gopinathan, Journal of Theoretical Biology **231**, 23 (2004).
[12] M. Ukai-Tadenuma, R. G. Yamada, H. Xu, J. A. Ripperger, A. C. Liu, and H. R. Ueda, Cell **144**, 268 (2011).
[13] A. Grönlund, P. Lötstedt, and J. Elf, Proceedings of the National Academy of Sciences **107**, 8171 (2010).
[14] A. Grönlund, P. Lötstedt, and J. Elf, Nature Communications **2** (2011), 10.1038/ncomms1422.
[15] R. Maithreye, R. R. Sarkar, V. K. Parnaik, and S. Sinha, PLoS ONE **3**, e2972 (2008).
[16] M. Scott, Phys. Rev. E **80**, 031129 (2009).
[17] M. Wu, R.-Q. Su, X. Li, T. Ellis, Y.-C. Lai, and X. Wang, Proceedings of the National Academy of Sciences (2013).
[18] T. Hong, J. Xing, L. Li, and J. J. Tyson, BMC Syst Biol **6**, 66 (2012).
[19] E. He, O. Kapuy, R. A. Oliveira, F. Uhlmann, J. J. Tyson, and B. Novák, Proc Natl Acad Sci USA **108**, 10016 (2011).
[20] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. Van Oudenaarden, Nature **427**, 737 (2004).
[21] T. B. Kepler and T. C. Elston, Biophysical Journal **81**, 3116 (2001).
[22] E. Aurell and K. Sneppen, Phys Rev Lett **88**, 048101 (2002).
[23] P. B. Warren and P. R. ten Wolde, J Phys Chem B **109**, 6812 (2005).
[24] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, Science **305**, 1622 (2004).
[25] T. S. Gardner, C. R. Cantor, and J. J. Collins, Nature **403**, 339 (2000).
[26] D. Nevozhay, R. M. Adams, E. Van Itallie, M. R. Bennett, and G. Balázsi, PLoS Comp Biol **8**, e1002480 (2012).
[27] C. Gupta, J. M. López, W. Ott, K. c. v. Josić, and M. R. Bennett, Phys. Rev. Lett. **111**, 058104 (2013).
[28] M. Fischer and P. Imkeller, Stoch. Dyn. **5**, 247 (2005).
[29] M. Fischer and P. Imkeller, Stoch. Anal. Appl. **24**, 135 (2006).
[30] K. Josić, J. M. López, W. Ott, L. Shiau, and M. R. Bennett, PLoS Comput Biol **7**, e1002264 (2011).
[31] D. Gillespie, Journal of Physical Chemistry **81**, 2340 (1977).
[32] D. J. Higham, SIAM Rev. **50**, 347 (2008).
[33] Y. Cao, D. Gillespie, and L. Petzold, Journal of Chemical Physics **124** (2006).
[34] Y. Cao, D. Gillespie, and L. Petzold, Journal of Chemical Physics **126** (2007).
[35] D. Gillespie, Journal of Chemical Physics **115**, 1716 (2001).
[36] Z. Xu and X. Cai, Journal of Chemical Physics **128** (2008).
[37] X. Cai and Z. Xu, Journal of Chemical Physics **126** (2007).
[38] R. Schlicht and G. Winkler, J. Math. Biol. **57**, 613 (2008).
[39] T. Brett and T. Galla, Phys. Rev. Lett. **110**, 250601 (2013).
[40] D. Gillespie, A. Hellander, and L. Petzold, Journal of Chemical Physics **138** (2013).

[41] M. Barrio, K. Burrage, A. Leier, and T. Tian, PLoS Comput Biol **2**, e117 (2006).
[42] M. A. Gibson and J. Bruck, The Journal of Physical Chemistry A **104**, 1876 (2000).
[43] B. Bayati, P. Chatelain, and P. Koumoutsakos, Journal of Computational Physics **228**, 5908 (2009).
[44] C. Gupta, J. M. López, R. Azencott, M. R. Bennett, K. Josić, and W. Ott, "Supplementary material," (2013).
[45] "While such an assumption is not needed for the dSSA, it is biochemically realistic and used in our mathematical analysis.".
[46] S. N. Ethier and T. G. Kurtz, Markov processes, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics (John Wiley & Sons Inc., New York, 1986) pp. x+534, characterization and convergence.
[47] D. T. Gillespie, The Journal of Chemical Physics **113**, 297 (2000).
[48] N. G. van Kampen, Canadian Journal of Physics **39**, 551 (1961).
[49] C. Masoller, Phys. Rev. Lett. **90**, 020601 (2003).
[50] L. S. Tsimring and A. Pikovsky, Phys. Rev. Lett. **87**, 250602 (2001).
[51] C. R. Doering, K. V. Sargsyan, and L. M. Sander, Multiscale Model. Simul. **3**, 283 (2005).
[52] P. Hanggi, H. Grabert, P. Talkner, and H. Thomas, Phys. Rev. A (3) **29**, 371 (1984).
[53] H. J. Kushner, SIAM J. Appl. Math. **44**, 160 (1984).
[54] T. Tian, K. Burrage, P. M. Burrage, and M. Carletti, Journal of Computational and Applied Mathematics **205**, 696 (2007), special issue on evolutionary problems.
[55] N. G. van Kampen, Stochastic processes in physics and chemistry, Vol. 1 (Elsevier, 1992).
[56] T. Galla, Phys. Rev. E **80**, 021909 (2009).
[57] L. F. Lafuerza and R. Toral, Phys. Rev. E **84**, 021128 (2011).
[58] I. B. Schwartz, T. W. Carr, L. Billings, and M. Dykman, arXiv preprint arXiv:1207.7278 (2012).
[59] E. L. Haseltine and J. B. Rawlings, The Journal of chemical physics **117**, 6959 (2002).
[60] H. Salis and Y. Kaznessis, The Journal of chemical physics **122**, 054103 (2005).
[61] H. Salis, V. Sotiropoulos, and Y. N. Kaznessis, BMC bioinformatics **7**, 93 (2006).