Elimination of fast variables in chemical Langevin equations

Yueheng Lan,[1] Timothy C. Elston,[2] and Garegin A. Papoian[1,a)]
[1]*Department of Chemistry, University of North Carolina at Chapel Hill, North Carolina 27599-3290, USA*
[2]*Department of Pharmacology, University of North Carolina at Chapel Hill,
North Carolina 27599, USA*

Internal and external fluctuations are ubiquitous in cellular signaling processes. Because biochemical reactions often evolve on disparate time scales, mathematical perturbation techniques can be invoked to reduce the complexity of stochastic models. Previous work in this area has focused on direct treatment of the master equation. However, eliminating fast variables in the chemical Langevin equation is also an important problem. We show how to solve this problem by utilizing a partial equilibrium assumption. Our technique is applied to a simple birth-death-dimerization process and a more involved gene regulation network, demonstrating great computational efficiency. Excellent agreement is found with results computed from exact stochastic simulations. We compare our approach with existing reduction schemes and discuss avenues for future improvement. © *2008 American Institute of Physics.* [DOI: 10.1063/1.3027499]

I. INTRODUCTION

With the decoding of the genomes and proteomes of various organisms, a key focus of cell biology has shifted to uncovering the organization and functioning principles of regulatory and signaling networks. Recent high throughput approaches have accelerated this transition not only by identifying biomolecular species involved in various cellular processes but also by revealing and quantifying the complex interactions between genes, proteins, and metabolites that compose these systems.[1,2] The analysis and interpretation of these data pose a great challenge to theoreticians working in the area of systems biology. Novel analysis and simulation techniques based on physical and chemical principles are needed to uncover the design principles of signaling and regulatory networks. At the mesoscopic level, many different techniques are used to characterize the activity of these systems, including inference networks, Boolean maps, differential equations, and stochastic models, the latter playing a special role.[3–10]

Cells live in noisy environments and are subject to constant internal and external fluctuations. To maintain normal function in the face of these changing conditions, regulatory networks must possess a certain level of robustness while retaining enough sensitivity to pick up useful signals from the noisy background.[11–15] When molecular species are present in low abundances, fluctuations in molecule number can generate large variability in the cell's response to an external or internal signal.[16–18] Such intrinsic fluctuations have been suggested to play important roles in evolution and embryonic development.[19,20] Thus, describing and evaluating the effects of molecular level noise are important problems in cell modeling.

The chemical master equation describes how the probabilities for the number of molecules of each chemical species evolve in time.[21,22] In the limit of large particle numbers, the law of mass action for chemical kinetics is recovered from the master equation as a set of ordinary differential equations (ODEs).[23,24] At intermediate particle numbers, the chemical Langevin equations or the equivalent Fokker–Planck equation can be derived under quite general assumptions.[21,23] The analysis of these equations provides an important approach for characterizing noisy cellular behavior.[6,25,26]

Only in a few special cases[21,27] can stochastic models be solved analytically. Therefore, numerical simulations have been the approach used to analyze and quantitatively explain many biochemical processes.[3,21] The large size of the interaction networks determined by new experimental approaches and the vastly different time scales associated with the biochemical processes that make up the network often necessitate large amounts of computer time and memory. Thus, in many modeling studies the computational load turns out to be the bottleneck for fruitful analysis. Consequently, model reduction and simplification techniques are a necessity.

In physics and chemistry, many techniques have been developed to treat small amplitude fluctuations or noise with short correlation times.[21,27,28] More recently, considerable effort has been devoted to accelerating stochastic simulations.[5,7–29,38] In this paper, we report on a new reduction technique for the chemical Langevin equation, which is widely used in modeling the stochastic behavior of biochemical processes and has been studied extensively in previous work. The chemical Langevin equations consist of a deterministic part plus associated noise terms. Its derivation is based on the assumption that the fluctuations occur on a much faster time scale than the deterministic motion. Hence, in the chemical Langevin equation, the noise is taken to be white because its correlation time is assumed to be negligible relative to the deterministic time scale.[21,23] Nevertheless,

a)Electronic mail: gpapoian@unc.edu.

when time scale separations exist in the fluctuations, it may be necessary to distinguish noise from different sources to achieve model reduction.

When using ODEs derived from the law of mass action,[2,21,22,39,40] singular perturbation techniques that take advantage of separation of time scales can be used to eliminate fast variables that are entrained to the slow variables. This type of reduction is not directly applicable to Langevin equations because the noise terms always represent fast motions and cannot be formally separated into groups with different time scales. Here we propose a practical approach to achieve the desired simplification that is based on physical considerations of chemical reactions. For intrinsic fluctuations in chemical reactions, not only the deterministic motion but also the fluctuations evolve on separate time scales: fast motions are associated with fast intrinsic noise and slow motions with slow intrinsic noise. The chemical Langevin equation masks this distinction. Partial equilibrium assumptions in the context of the master equation have been widely used to reduce multiscale systems.[30,31,35,36,38,41] Similarly, in order to take advantage of separations in time scale within the Langevin formalism, we first determine the slow degrees of freedom and assume that the fast variables reach a quasi-steady state, where an adiabatic steady probability distribution exists and parametrically evolves with the slow variables. Chemical Langevin equations are written for the slow variables in which only averages of fast variables appear and the noise term comes purely from the slow motion. This reduction ensures that we only need to simulate the slow reactions, achieving high computational efficiency. Total fluctuations, which include the fast noise, may be computed at any desired time instant from the adiabatic steady state distribution.

In Sec. II, we use a simple birth-death-dimerization process to illustrate our technique. The slow variables are constructed and the associated chemical Langevin equation is derived. Through solution of the steady state dimerization reaction, the adiabatic probability distribution for the fast variable is derived containing the slow variable as a parameter. In Sec. III, we compare the results from our reduction with exact Gillespie simulations. Next, a related gene regulation model is introduced in which the dimers act as transcriptional activators for the protein and thus change the production rate of the monomers. After a small modification in the reduction scheme to account for the discrete states of the promoter, our model shows excellent agreement with the benchmark computations. In Sec. IV, separation of time scale and model reduction problems are discussed, and our scheme is compared with existing methods. In Sec. V, we summarize our work and point to possible future directions.

II. FORMULATING THE PROBLEM THROUGH A SIMPLE DIMERIZATION PROCESS

Consider the following simple dimerization process to be used later as part of a gene regulatory network,

$$\varnothing \underset{k}{\overset{g}{\rightleftharpoons}} \mathcal{M},$$

$$2\mathcal{M} \underset{\lambda}{\overset{\mu}{\rightleftharpoons}} \mathcal{D}. \tag{1}$$

The monomers $\mathcal{M}$ are produced at a rate g by a Poisson process. The monomers dimerize into $\mathcal{D}$ with a rate μ. At the same time, $\mathcal{M}$ and $\mathcal{D}$ decay back to their original reactants with rates k and λ, respectively. In the limit of large molecular abundances, the deterministic evolution of the molecule numbers evolves according to the following chemical kinetic equations:

$$\frac{dn}{dt} = \frac{\mu}{2}m(m-1) - \lambda n,$$

$$\frac{dm}{dt} = 2\lambda n - \mu m(m-1) + g - km, \tag{2}$$

where n denotes the number of $\mathcal{D}$ and m the number of $\mathcal{M}$. It turns out that the deterministic dynamics given by Eq. (2) can be simplified by projecting the dynamics onto a slow manifold[42–46] which is a smooth hypersurface of lower dimension than the original phase space. However, for stochastic dynamics dictated by the corresponding chemical Langevin equations the reduction is not obvious due to the mixing of deterministic and stochastic time scales, and therefore must be done very carefully. Below, we elaborate on such a reduction technique.

To eliminate the fast variable, it is necessary to define the variables that correspond to the slow manifold. The total particle number $N = m + 2n$ is a good slow variable[31] and satisfies a simpler equation,

$$\frac{dN}{dt} = g - km. \tag{3}$$

However, Eq. (3) is not closed because it contains m, the monomeric protein number. To close this equation, we need to find an algebraic relation between m and N. If $\mu, \lambda \gg g, k$, an adiabatic assumption may be made that the dimerization reactions always reach a quasisteady state (i.e., $dn/dt = 0$ in the lowest order approximation). Under this assumption, we have the identity

$$N = m + \frac{\mu}{\lambda}m(m-1), \tag{4}$$

which defines the slow manifold in the phase space of Eq. (2) and also gives a relation between the time evolution of m and N,

$$\frac{dm}{dt} = \left(1 + \frac{2\mu m}{\lambda} - \frac{\mu}{\lambda}\right)^{-1}\frac{dN}{dt}. \tag{5}$$

Substituting Eq. (3) into Eq. (5) results in a closed equation for m. Thus, we have successfully eliminated the fast variable by utilizing the slow manifold represented by Eq. (4).

Can the same elimination techniques be applied in the presence of noise? Caution must be exercised because the

approximation given by Eq. (4) is only valid near the slow manifold which is determined by the deterministic part of the equation. Noise coming from the fast dimerization reactions can potentially drive the solution away from the manifold, which makes Eq. (4) and thus Eq. (5) inapplicable. With noise, we have to use a probabilistic description. When all species are present in large quantities, the chemical Langevin equations[23] accurately generate individual realizations of the reaction trajectories. In the limit of fast dimerization, the distribution for the monomers $\mathcal{M}$ can be assumed to be in steady state for each value of the slowly evolving N. Although the monomer number m itself fluctuates rapidly, $\bar{m}$, the average conditioned on N, remains close to the slow manifold, i.e., Eqs. (4) and (5) are still satisfied with high accuracy if m is replaced by $\bar{m}$. On the other hand, the slow birth-death reaction effectively only sees the average $\bar{m}$, which leads to the chemical Langevin equation

$$\frac{dN}{dt} = g - k\bar{m} + \sqrt{g}\Gamma_1(t) - \sqrt{k\bar{m}}\Gamma_2(t), \tag{6}$$

where $\Gamma_i(t)$, $i=1,2$, are Gaussian white noise terms with $\langle\Gamma_i\rangle=0$, $\langle\Gamma_i(t)\Gamma_j(t')\rangle=\delta_{ij}(t-t')$. Although the noise in Eq. (6) seems to have zero time correlation, we have to keep in mind that it is a mathematical abstraction of the stochasticity intrinsic to the chemical reactions and it is "slow" compared with fluctuations from the dimerization process. Substituting Eq. (5) into Eq. (6) results in

$$\frac{d\bar{m}}{dt} = \left(1 + \frac{2\mu\bar{m}}{\lambda} - \frac{\mu}{\lambda}\right)^{-1}(g - k\bar{m} + \sqrt{g}\Gamma_1(t) - \sqrt{k\bar{m}}\Gamma_2(t)), \tag{7}$$

which gives the time evolution of the conditional average $\bar{m}$ and includes the slow noise generated in the birth-death process. To produce the full stochastic trajectory, we have to include the "fast" noise generated in the dimerization process which we assume is in steady state.

We concentrate on the dimerization reaction for fixed N to study the effects of the fast noise. To obtain an analytic expression for the steady state probability distribution we use a generating function formalism, which has been widely used in describing random process in physics and chemistry.[21,47,48] The generating function $\Psi(x)=\Sigma_m P(m)x^m$, where $P(m)$ denotes the probability of having m monomer $\mathcal{M}$, satisfies

$$\frac{\partial\Psi}{\partial t} = \frac{\mu}{2}(1-x^2)\frac{\partial^2\Psi}{\partial x^2} + \frac{\lambda}{2}(x^2-1)\left(N - x\frac{\partial}{\partial x}\right)\Psi, \tag{8}$$

which has the steady state solution

$$\Psi(x) = C\,\mathrm{He}_N(\sqrt{-\lambda/\mu}x), \tag{9}$$

where C is a normalization constant to ensure $\Psi(1)=1$ and He_N is an Nth order Hermite polynomial:

$$\mathrm{He}_N(x) = N!\sum_{n=0}^{[N/2]}\frac{(-1)^n}{n!\,2^n(N-2n)!}x^{N-2n}, \tag{10}$$

where $[N/2]$ denotes the greatest integer not bigger than $N/2$. Thus, explicitly, for even N,

$$\Psi(x) = C_1 N!\sum_{n=0}^{N/2}\frac{x^{N-2n}}{n!\,(N-2n)!}\left(\frac{\mu}{2\lambda}\right)^n, \tag{11}$$

which gives the steady probability distribution

$$P(N-2n) = \frac{C_1 N!}{n!\,(N-2n)!}\left(\frac{\mu}{2\lambda}\right)^n. \tag{12}$$

Using the Sterling approximation

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

and taking the logarithm of Eq. (12), we have

$$\ln P(N-2n) \sim n\ln\frac{\mu}{2\lambda} - n(\ln n - 1) - \frac{1}{2}\ln 2\pi n$$

$$- (N-2n)(\ln(N2n) - 1) - \frac{1}{2}\ln 2\pi(N-2n). \tag{13}$$

In the limit of large particle numbers, we may use a Gaussian approximation for the probability distribution and, thus, we calculate the first and second derivatives of Eq. (13),

$$\frac{d\ln P(N-2n)}{dn} = \ln\frac{\mu}{2\lambda} + 2\ln(N-2n) + \frac{1}{N-2n} - \ln n - \frac{1}{2n}, \tag{14}$$

$$\frac{d^2\ln P(N-2n)}{dn^2} = \frac{2}{(N-2n)^2} + \frac{1}{2n^2} - \frac{1}{n} - \frac{4}{N-2n}. \tag{15}$$

The condition $d\ln P(N-2n)/dn=0$ then gives an equation for the average,

$$\mu(N-2n)(N-2n+1) = \lambda(2n+1), \tag{16}$$

to leading order. If we write $m=N-2n$, then Eq. (16) is almost equivalent to Eq. (4). Therefore, the average of the stationary distribution is well approximated by the deterministic result for $\bar{m}$, which stays close to the slow manifold. If we define $b=-d^2\ln P(N-2n)/dn^2|_{m=\bar{m}}>0$, then we get a Gaussian approximation of the probability distribution

$$P(m) = \sqrt{\frac{b}{8\pi}}e^{-(b/8)(m-\bar{m})^2}, \tag{17}$$

which describes the fast noise distribution conditional on N.

Consequently, the random variable m is well approximated by

$$m = \bar{m} + \xi_N, \tag{18}$$

where ξ_N is a random variable with the distribution given by Eq. (17). In this way, we achieve the desired reduction. To get the stochastic trajectory of $m(t)$, we first solve the slow Langevin equation (7) to get $\bar{m}(t)$ and hence $N(t)$ by Eq. (4). Next, this $N(t)$ is used to generate a random number ξ_N according to the probability distribution equation (17) and $m(t)$ is obtained through Eq. (18). In practice, computational resources can be directed to evolve the slow variable, considering the full stochasticity of $m(t)$ only when necessary.

III. NUMERICAL IMPLEMENTATION AND COMPARISON WITH THE GILLESPIE ALGORITHM

In this section, we examine the validity of our technique using two examples. First, we compare the simulation results from Eqs. (7) and (18) with exact discrete stochastic simulations. Next a gene regulatory network is introduced and a similar reduction is carried out. This example requires a modification to the reduction method to accommodate the discrete dynamics of the gene promoter. Numerical solutions of the simplified model are also compared with discrete stochastic simulations. The Gillespie algorithm[3] is used to produce exact results for both systems. A second-order forward integration scheme[49–53] is invoked to evolve the chemical Langevin equations. In both simulations we used 4×10^4 trajectories to compute the averages. The probability distribution functions (PDFs) at particular times and time series for the average and variance of the monomer number are plotted. As above, we use $\bar{m}$ to denote the conditional average of m for fixed N and introduce the angular bracket notation $\langle \cdot \rangle$ to denote the overall average of a variable.

A. The birth-death-dimerization problem

To satisfy the fast dimerization condition, we choose $g = 0.2$, $k = 0.001$, $\mu = 0.002$, and $\lambda = 30$. The initial condition is $(m,n) = (0,0)$. Figure 1 shows a comparison of the distributions for $\mathcal{M}$ at $t = 200$ and $t = 500$ obtained from our reduced Langevin approximation (solid line) and Gillespie simulations (circles), indicating excellent agreement. The reduced equation captures the characteristic features of the distribution. Although the figure only shows the distribution at two specific snapshots, the distributions are in good agreement throughout the time evolution.

Figure 2 displays the average and the variance of protein number m in the time interval $[0,500]$. The reduced equation produced a result in perfect agreement with the Gillespie result. Both the average and the variance grow monotonically with time. One interesting observation is that the average is approximately equal to the variance, which is a characteristic feature of the Poisson process and can be explained by Eq. (7). With the current parameter values, the first factor on the right hand side of Eq. (7) is very close to 1 since $\mu \bar{m}/\lambda \ll 1$. Thus, to a very good approximation, the conditional average $\bar{m}$ satisfies a usual birth-death equation which describes a Poisson process. The dimerization in Eq. (1) rarely happens in the interval $t = [0,500]$ since $\langle n \rangle \approx 0.2$ when $\langle m \rangle \approx 80$ at $t = 500$.

B. Positive autoregulation

We now turn to a more complex system involving a genetic switch.[17,54–61] In addition to the birth-death-dimerization process, the switch involves a positive feedback loop in which the dimer form of the protein acts a transcriptional activator by binding to the promoter region of the gene. As the time scales become even more entangled[60,62] in this model, we test the validity of the reduction scheme in this new context.

The reactions for this system are as follows:[63]
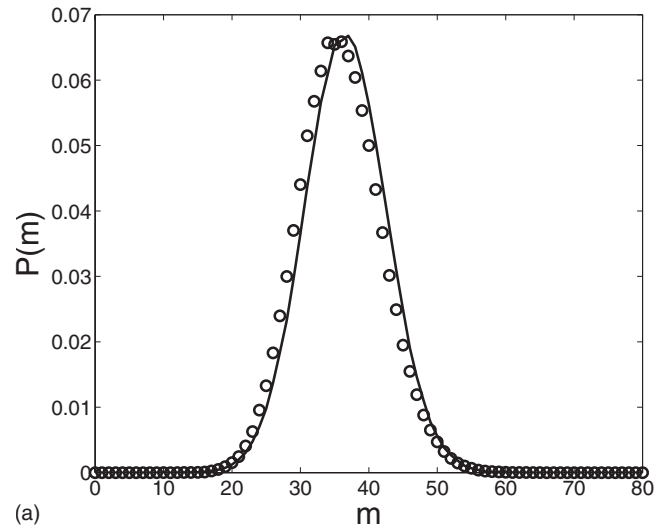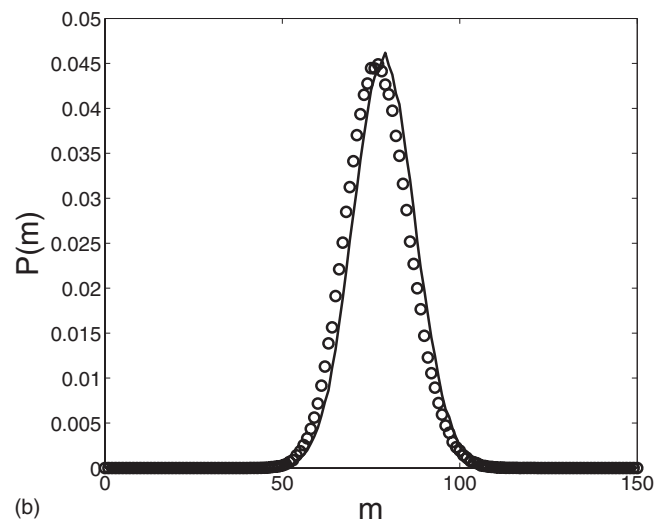
FIG. 1. Comparison of the m distribution for the two-step cascade at times (a) $t = 200$ and (b) $t = 500$: Gillespie simulation (circles) and reduced Langevin approximation (solid line). $g = 0.2$, $k = 0.001$, $\mu = 0.002$, and $\lambda = 30$ with the initial condition $(m,n) = (0,0)$.

$$\mathcal{M} \xrightarrow{\delta} \varnothing, \tag{19}$$

$$\varnothing \xrightarrow{\alpha_s} \mathcal{M}, \tag{20}$$

$$\mathcal{M} + \mathcal{M} \underset{\theta\Lambda}{\overset{\Lambda}{\rightleftharpoons}} \mathcal{D}, \tag{21}$$

$$\mathcal{O}_0 + \mathcal{D} \underset{\beta k}{\overset{k}{\rightleftharpoons}} \mathcal{O}_1. \tag{22}$$

The above equations describe a simplified version of an autoregulated gene. The promoter is denoted by $\mathcal{O}_s$, where $s \in \{0,1\}$ denotes the states of the promoter, unoccupied and dimer bound, respectively. The constitutive production rate of the promoter is very small compared to the production rate of the activated gene when a dimer bounds to the promoter. The transitions in the promoter state are represented by Eq. (22), where k is the forward rate and β is the dimensionless dissociation constant. For simplicity, we lump the
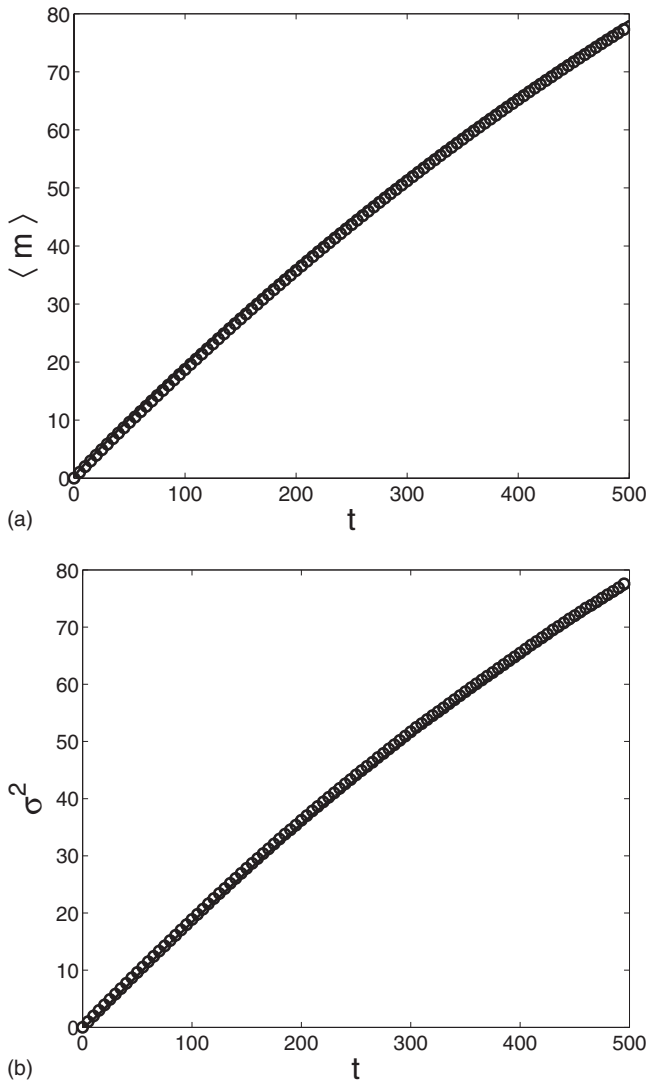
FIG. 2. Comparison of the m average and variance computed for the two-step cascade in the time interval $t=[0,500]$: Gillespie simulation (circle) and reduced Langevin approximation (solid line). $g=0.2$, $k=0.001$, $\mu=0.002$, and $\lambda=30$ with the initial condition $(m,n)=(0,0)$.

transcription step and the translation step into one production rate (20) with a rate α_s that depends on the state of the promoter. When $s=0$, α_0 is small and marks the constitutive production rate of the monomer; when $s=1$, α_1 is much larger than α_0 and represents the active state of the gene. The protein $\mathcal{M}$ decays at a constant rate δ in Eq. (19) while it dimerizes to $\mathcal{D}$ at a rate Λ in Eq. (21). The dissociation rate is $\theta\Lambda$.

This system is more complicated than the previous one because now the number n of the dimer $\mathcal{D}$ affects the production rate of the monomer $\mathcal{M}$ by binding to the gene promoter. Furthermore, the promoter has only two discrete states $\mathcal{O}_0$ and $\mathcal{O}_1$, which cannot be approximated by a continuously varying random number. Thus, the Gaussian noise approximation is not applicable here, and we have to keep the original discrete stochastic process to model transitions in the state of the promoter.

If the dimerization is much faster than other reactions, we may use the steady state approximation as in the previous example; specifically,

$$\bar{n} = \frac{\bar{m}(\bar{m}-1)}{2\theta}, \tag{23}$$

where $\bar{m}$ and $\bar{n}$ denote the conditional averages of the monomer number m and the dimer number n, respectively, under the fast reaction (21). Let $N=2n+m$ denote the slow variable; from

$$N = \bar{m} + \frac{\bar{m}(\bar{m}-1)}{\theta}, \tag{24}$$

we derive

$$\frac{d\bar{m}}{dt} = (1+(2\bar{m}-1)/\theta)^{-1}\frac{dN}{dt}, \tag{25}$$

which shows how $\bar{m}$ changes with N. For the slow variable N, we may write the corresponding chemical Langevin equation,

$$\frac{dN}{dt} = -\delta\bar{m} + \alpha_s - \Gamma_1\sqrt{\delta\bar{m}} + \Gamma_2\sqrt{\alpha_s}. \tag{26}$$

The primary difference between Eqs. (6) and (26) is that here α_s is a two-state random variable with a Poisson switching probability determined by the equations

$$\frac{dP_0}{dt} = -k\bar{n}P_0 + \beta k P_1,$$

$$\frac{dP_1}{dt} = k\bar{n}P_0 - \beta k P_1, \tag{27}$$

where P_0 and P_1 denote the probabilities for the state variable s having a value of 0 or 1.

We use a fixed time-step method to integrate Eq. (26) and simulate the random switching event of the promoter. For sufficiently small time step dt, the transition probability from $\mathcal{O}_0$ to $\mathcal{O}_1$ in the time interval dt is $k\bar{n}P_0dt$ and $\beta k P_1 dt$ is the probability for the reverse transition. In each time interval, we fix the promoter state and allow it to change only at the end of each step according to the transition probability. As long as dt is much smaller than the average switching time, this numerical scheme is relatively accurate. Within each time interval, we used a second-order stochastic integrator to update $\bar{m}$ as in the previous example. Based on this scheme, we carried out simulations with two sets of parameter values and calculated the averages, variances, and PDFs. We also carried out Gillespie simulations with identical parameters to compare with our reduced system. In the plots below, the Gillespie results are depicted with open circles and the chemical Langevin results with solid lines (or dashed lines). All results consist of averages over 40 000 trajectory realizations.

Simulation results shown in Fig. 3, were obtained in the time interval $t \in [0,2000]$ with parameter values $\delta=0.01$, $\alpha_0=0.1$, $\alpha_1=1.0$, $\lambda=0.2$, $\theta=15$, $k=0.05$, and $\beta=3$ and the initial condition $(m,N_{\phi_0},N_{\phi_1},n)=(0,1,0,0)$. We see that the evolution of the average and the variance of two different calculations agree extremely well. The average is a monotonically increasing function of time, saturating at long times, while the variance shows a fast initial rise to a peak
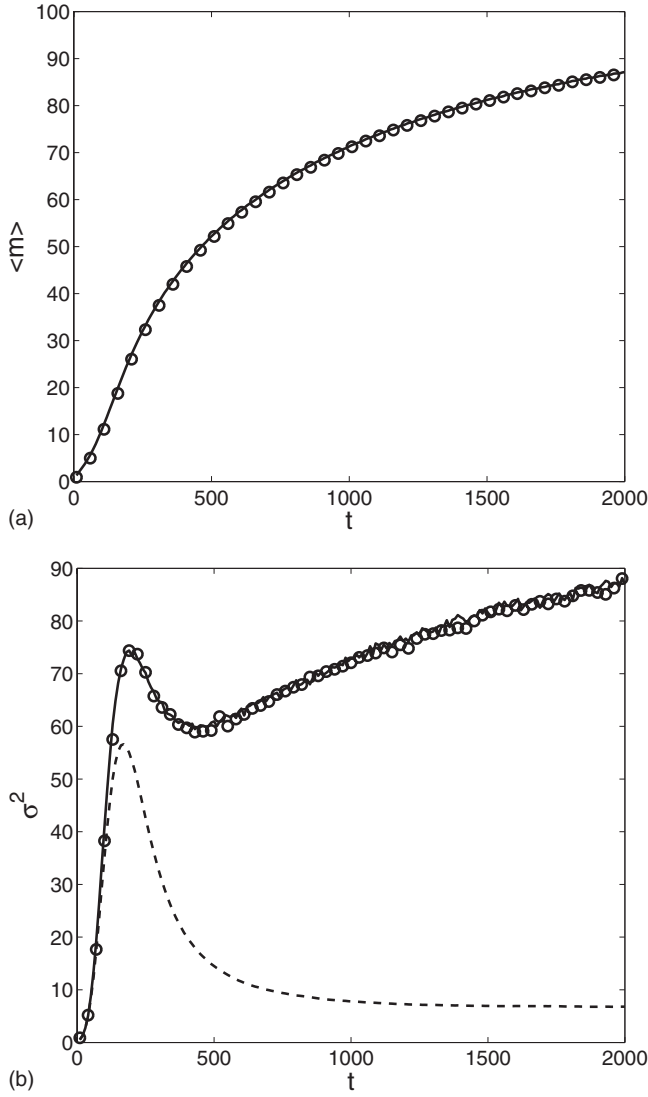
FIG. 3. Comparison of (a) the average and (b) the variance of the monomer $\mathcal{M}$ number computed for the gene regulation cascade in the time interval $t = [0, 2000]$: the Gillespie simulation (circle), the reduced Langevin approximation (solid line), and without fast noise (dashed line). $\delta = 0.01$, $\alpha_0 = 0.1$, $\alpha_1 = 1.0$, $\lambda = 0.2$, $\theta = 15$, $k = 0.05$, and $\beta = 3$ with the initial condition $(m, N_{\phi_0}, N_{\phi_1}, n) = (0, 1, 0, 0)$.

FIG. 4. Comparison of the autocorrelation function $\text{Cor}(t, \tau)$ at (a) $t = 400$ and (b) $t = 1000$ of the monomer $\mathcal{M}$ number computed for the gene regulation cascade: the Gillespie simulation (circle) and the reduced Langevin approximation (solid line). $\delta = 0.01$, $\alpha_0 = 0.1$, $\alpha_1 = 1.0$, $\lambda = 0.2$, $\theta = 15$, $k = 0.05$, and $\beta = 3$ with the initial condition $(m, N_{\phi_0}, N_{\phi_1}, n) = (0, 1, 0, 0)$.

and after a small drop starts a second slowly growing stage. The peak in the variance arises due to the discreteness of the dimers at the initial stage. The dashed lines in Fig. 3 display the results for the reduced model without fast noise, i.e., with only $m = \bar{m}$ in Eq. (18). The average overlaps with the full noise approximation as expected while the variance shows large discrepancies at later times. Initially, dimerization rarely happens, and thus contributes little to the variance. After that, however, the monomers dimerize more frequently and the dimerization process starts to dominate the fluctuations. At $t = 2000$, taking $\bar{m} = 87$, we get $\bar{n} \approx 250$ by Eq. (23). According to Eqs. (16) and (17), the contribution of the fast noise to the variance is

$$\frac{4}{b} \approx \frac{4}{1/250 + 4/87} \approx 80,$$

which accounts for the difference between the solid and the dashed line in Fig. 3(b).
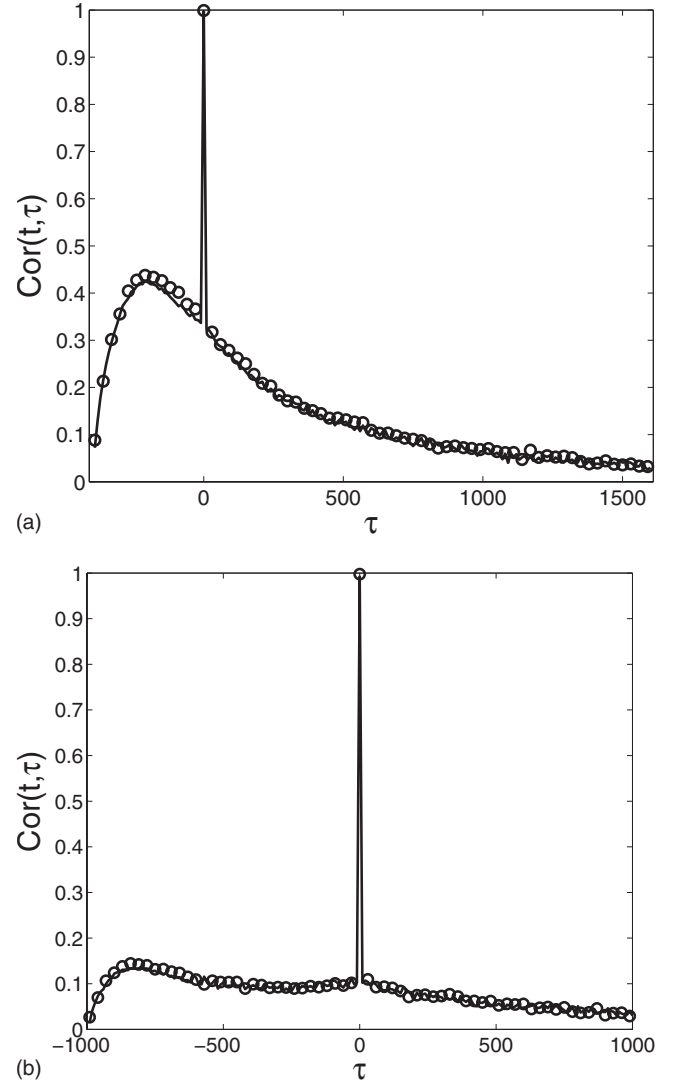
To further demonstrate that the reduced method captures the dynamic behavior of the full system, in Fig. 4 we compare the temporal autocorrelation function

$$\text{Cor}(t, \tau) \equiv \frac{\langle (m(t) - \langle m \rangle(t))(m(t + \tau) - \langle m \rangle(t + \tau)) \rangle}{\sigma_t \sigma_{t+\tau}},$$

which is a function of two variables t and τ representing two time instants t and $t + \tau$ (where σ_t^2 indicates variance at time t). When the system reaches equilibrium, $\text{Cor}(t, \tau)$ depends only on the time difference τ. In the transient phase, however, the correlation function also has a strong dependence on the variable t. This is clearly exhibited in Fig. 4(a) for $t = 400$ and Fig. 4(b) for $t = 1000$. The graphs look similar and both have a sharp protrusion near $\tau = 0$ due to the short decorrelation time. For $t = 1000$ (Fig. 4), the profile becomes more symmetric around $\tau = 0$ than for $t = 400$ since the system approaches equilibrium with increasing time. In both figures, the reduced method (solid lines) agrees with the Gillespie result (circles) extremely well.
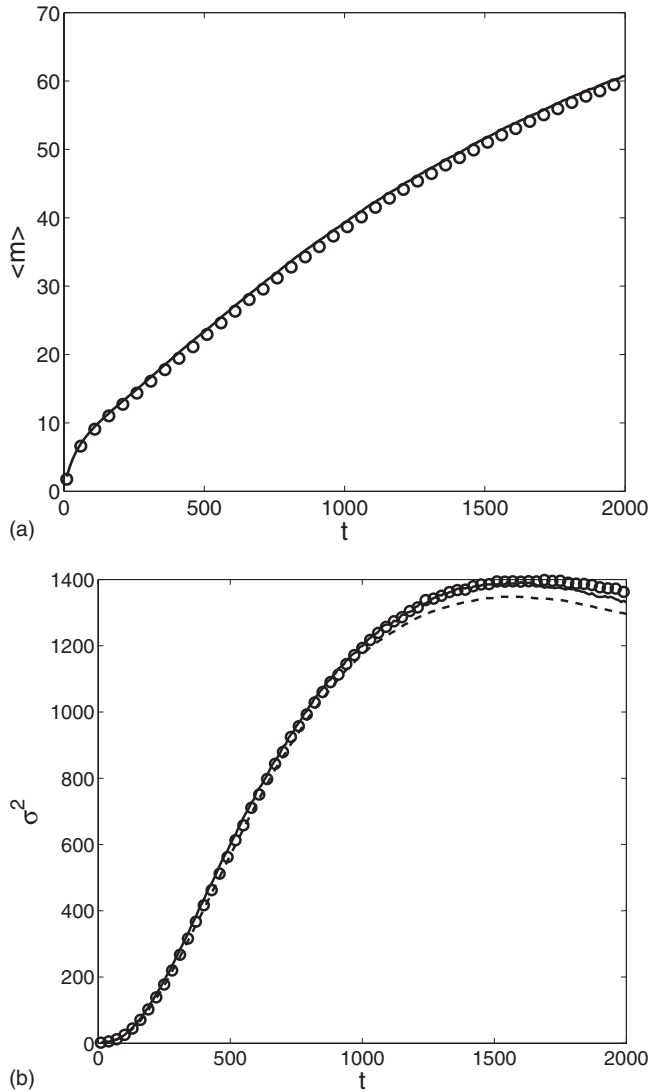
FIG. 5. Comparison of (a) the average and (b) the variance of the monomer $\mathcal{M}$ number computed for the gene regulation cascade in the time interval $t \in [0, 2000]$: the Gillespie simulation (circle), the reduced Langevin approximation (solid line), and without fast noise (dashed line). $\delta = 0.02$, $\alpha_0 = 0.2$, $\alpha_1 = 2.0$, $\lambda = 0.4$, $\theta = 50$, $k = 0.001$, and $\beta = 10$ with the initial condition $(m, N_{\phi_0}, N_{\phi_1}, n) = (0, 1, 0, 0)$.

FIG. 6. Comparison of the autocorrelation function $\text{Cor}(t, \tau)$ at (a) $t = 400$ and (b) $t = 1000$ of the monomer $\mathcal{M}$ number computed for the gene regulation cascade: the Gillespie simulation (circle) and the reduced Langevin approximation (solid line). $\delta = 0.02$, $\alpha_0 = 0.2$, $\alpha_1 = 2.0$, $\lambda = 0.4$, $\theta = 50$, $k = 0.001$, and $\beta = 10$ with the initial condition $(m, N_{\phi_0}, N_{\phi_1}, n) = (0, 1, 0, 0)$.

Figure 5 shows similar calculations but with a different set of parameter values, $\delta = 0.02$, $\alpha_0 = 0.2$, $\alpha_1 = 2.0$, $\lambda = 0.4$, $\theta = 50$, $k = 0.001$, and $\beta = 10$, where we significantly increased the backward reaction rate θ and dissociation constant β but decreased the promoter switching rate k. The agreement with Gillespie calculation is still very good, but small discrepancies exist, especially in the variance near the end of the simulation. The average is still a monotonic function, but the variance displays a broad peak which is very different from the one depicted in Fig. 3(b). The computation without fast noise now agrees quite well with the full noise approximation in both averages and variances. In the current case, the birth-death fluctuations resulting from switching in the promoter state dominates. The dimerization noise becomes relatively insignificant. Also note the large values of the variance, which indicate a strong deviation from the Poisson process. Thus, the dynamics for the two chosen sets of parameters differ significantly. The PDFs shown below reveal additional differences.
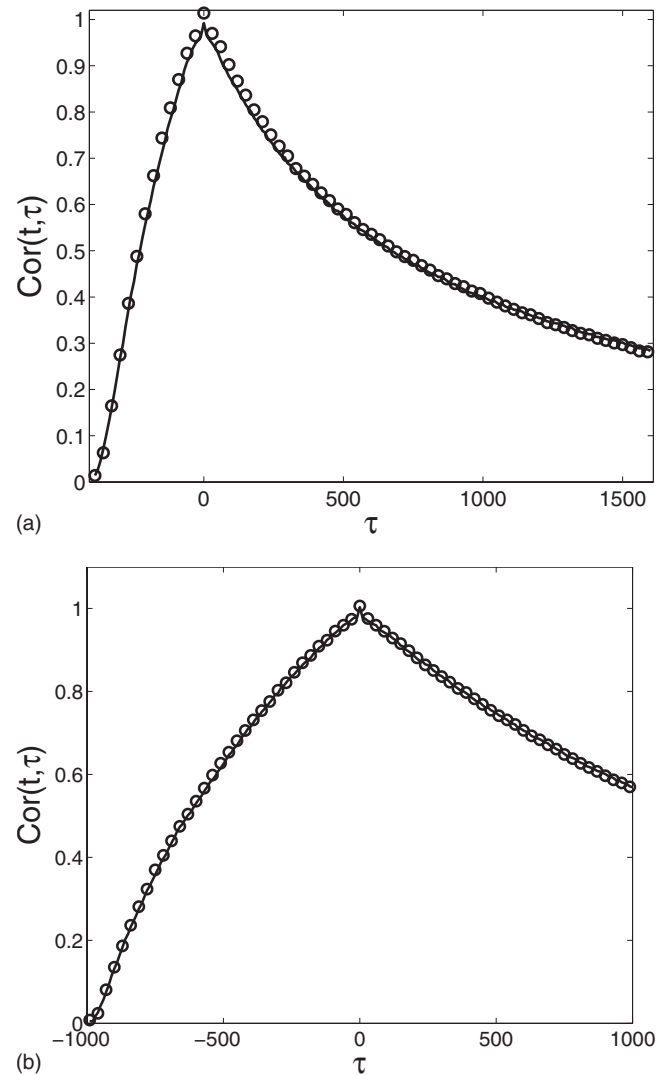
We compare the time correlation function $\text{Cor}(t, \tau)$ in Fig. 6 again for $t = 400$ and $t = 1000$ with the new set of parameter values. They continue to show the t dependence of $\text{Cor}(t, \tau)$ out of equilibrium. The profile around the peak in both (a) and (b) seems to display a cusp at the peak. The sharp protrusion at the origin in Fig. 4 is not obvious here because the fast noise is subordinate to the fluctuations induced by the slow changes in the state of the promoter. Still, the agreement between the reduced method and the exact Gillespie simulations is nearly perfect, demonstrating that the temporal dynamics of the original system is very well captured by the reduced equations.

Figure 7 shows the probability distributions at time $t = 400$ calculated with the two sets of parameter values. Our reduced Langevin approximation captures the distribution function very well. In Fig. 7(a), the distribution is close to a Gaussian centered around $m = 50$ while in Fig. 7(b) the distribution has a peak at $m = 10$ plus a broad elevation near $m = 70$. For both sets of parameters, when $s = 0$, the monomer
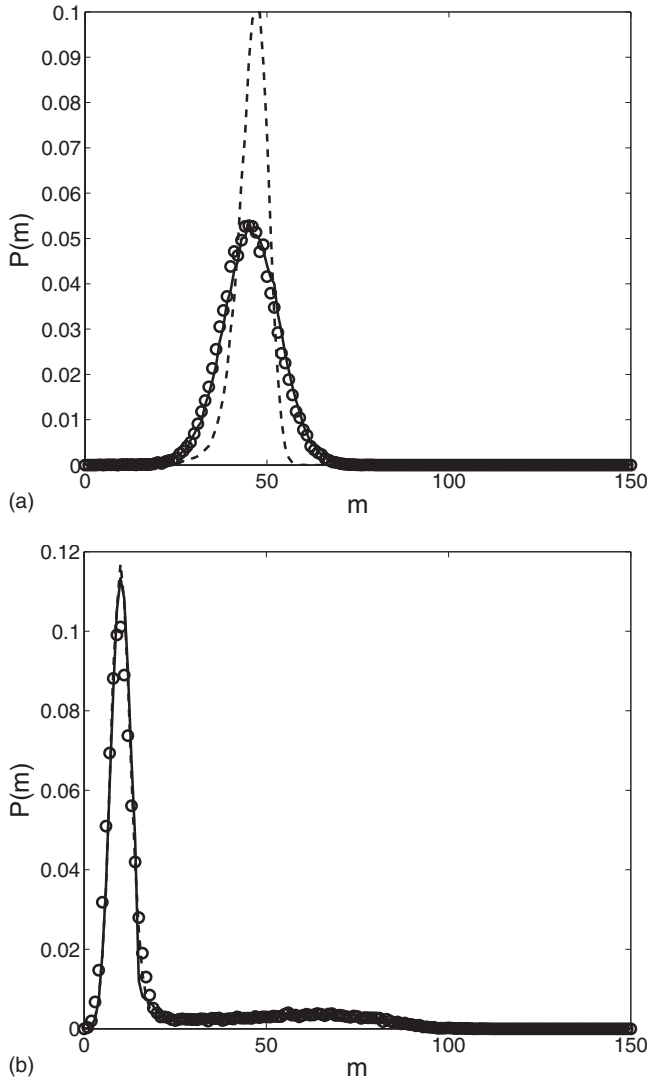
FIG. 7. Comparison of the probability distributions for the gene regulation cascade at $t=400$: the Gillespie simulation (circle), the reduced Langevin approximation (solid line), and without fast noise (dashed line) for parameter values (a) $\delta=0.01$, $\alpha_0=0.1$, $\alpha_1=1.0$, $\lambda=0.2$, $\theta=15$, $k=0.05$, and $\beta=3$ and (b) $\delta=0.02$, $\alpha_0=0.2$, $\alpha_1=2.0$, $\lambda=0.4$, $\theta=50$, $k=0.001$, and $\beta=10$, with the initial condition $(m,N_{\phi_0},N_{\phi_1},n)=(0,1,0,0)$.
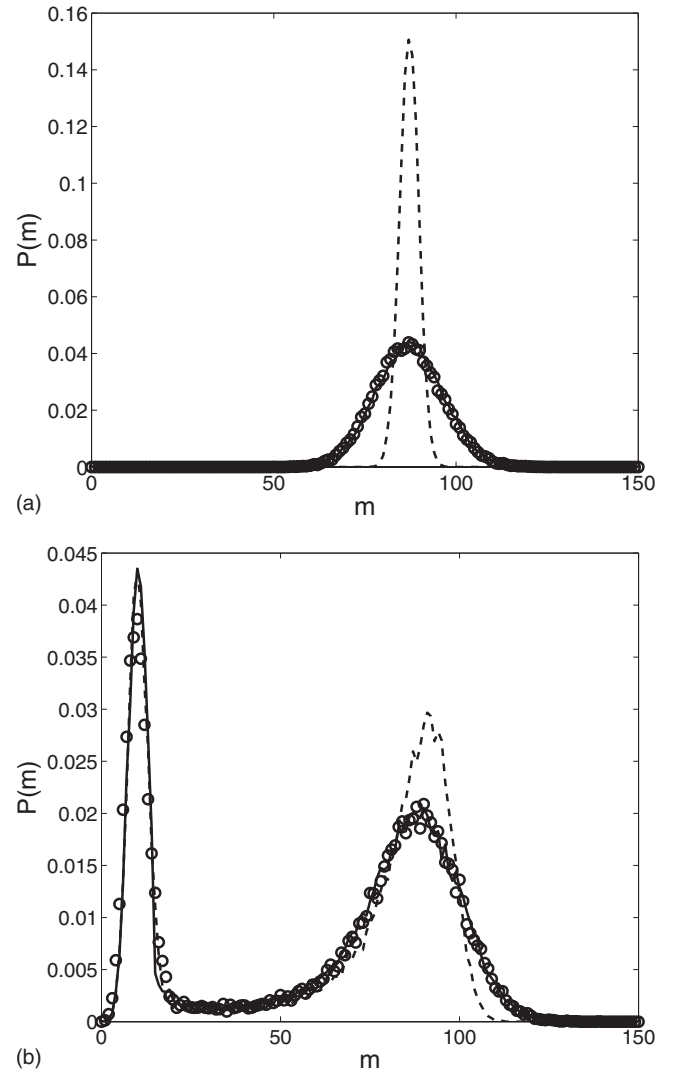

FIG. 8. Comparison of the probability distributions for the gene regulation cascade at $t=2000$: the Gillespie simulation (circle), the reduced Langevin approximation (solid line), and without fast noise (dashed line) for parameter values (a) $\delta=0.01$, $\alpha_0=0.1$, $\alpha_1=1.0$, $\lambda=0.2$, $\theta=15$, $k=0.05$, and $\beta=3$ and (b) $\delta=0.02$, $\alpha_0=0.2$, $\alpha_1=2.0$, $\lambda=0.4$, $\theta=50$, $k=0.001$, and $\beta=10$, with the initial condition $(m,N_{\phi_0},N_{\phi_1},n)=(0,1,0,0)$.

number is approximately 10 but is approximately 100 when $s=1$. The values of θ and β in Fig. 7(a) are considerably smaller than those in Fig. 7(b); thus, it is a lot easier for the first system to retain the $s=1$ state than the second.

We let the two systems further evolve to $t=2000$ resulting in the probability distributions depicted in Fig. 8. Our reduced model still perfectly matches the exact simulations. The Gaussian profile of the first distribution is retained but its center moved to $m \approx 90$ which means that most of the trajectories reach the $s=1$ state. In Fig. 8(b), a large second peak appears near $m=100$ while the first peak is considerably lower when compared to Fig. 7(b). The second peak corresponds to the $s=1$ state while the trajectories in the first peak mainly stay at the $s=0$ state.

The dashed lines in Figs. 7 and 8 depict the results of the reduced model without fast noise. For the first set of parameter values, the PDF profiles are considerably narrower than the exact ones while there are only small discrepancies for

the second set of parameter values. In fact, in Fig. 3(b) the dashed line shows a variance of $\sigma_d^2 \approx 7$, which is much smaller than the value produced by a pure birth-death process with the corresponding average value $\langle m \rangle \approx 87$. The dimers act as a reservoir to buffer fluctuations in the monomer number. When monomers decay, some dimers subsequently dissociate to keep the monomer number from decreasing; when monomers are born, dimers are subsequently formed to reduce the monomer number. Thus, the dimer reservoir has the effect of damping fluctuations generated in the birth-death process. Of course, at short time scales, the fluctuations caused by dimerization could be very large. For the second parameter set, the fluctuations are mainly due to the slow promoter state switching, so the dashed lines in Figs. 7(b) and 8(b) are closer to the exact profiles.

IV. DISCUSSION

Since the introduction of the Gillespie exact stochastic simulation algorithm,[3] much effort has been devoted to ac-

celerate its implementation. The first reaction method[4] was improved by the next reaction method,[5] which also is exact and requires only one random number for each reaction event. The τ-leap method[24] determines the number of transitions that occur in a fixed time interval with the approximation that the propensities of all the reactions do not change significantly during that time. In order to avoid negative protein numbers in the τ-leap method with large step size, binomial sampling scheme can be invoked.[32] If all species exist in large quantities, the chemical Langevin equations provide a continuous approximation to the underlying discrete process. In real reaction networks, however, there often coexist species with large and small copy numbers, as well as reactions with very different rates. The approximations needed in these schemes do not hold uniformly across the whole system and may not significantly reduce the computation load in many practical situations. Consequently, further reduction methods have been developed that partition the species or the reactions into different groups.[7,8,29–38] Fast reactions are assumed to reach partial equilibrium[30,31,35,38] such that effective reaction rates for the slow reactions can be estimated conditional on the averages of the fast variables. Alternatively, if the fast species assume large populations, deterministic equations, chemical Langevin equations, or the τ-leap method[7,8,29,33,37] can be used to evolve these species while the original or slightly modified Gillespie algorithm is used to advance the slow reactions.

The reduction scheme introduced here is related to the acceleration schemes mentioned above. Our contribution resides in treating the chemical Langevin equation instead of the chemical master equation as all the above schemes did. Therefore, our scheme deals with reaction species with large population. The chemical Langevin equations are a reduction in the master equation and the treatment here can be regarded as a second reduction. It turns out that even with the chemical Langevin equation, the partial equilibrium assumption[30,31,35,41] is still applicable as long as the reaction rates are well separated. In this sense, the current scheme is an extension of the partition scheme for the master equation. Many techniques such as the hybrid integration method developed for the master equation are also applicable here. In our second example, the promoter switches between two discrete states. We retained this discreteness in our reduced system and utilized a simple hybrid scheme. The excellent agreement of the results from this method with exact Gillespie simulations convincingly verifies the validity of the scheme.

In the examples considered here, our reduction scheme takes only 2% or 3% of the computational time as the exact Gillespie algorithm, demonstrating considerable efficiency. The main advantage of the reduction is that it removes the stiffness[64] embedded in the chemical Langevin equations. The fast dimerization process contributes stiff terms to both the monomer and dimer equations such that very small steps need to be taken in order to avoid instabilities when integrating these equations. Our reduction scheme preprocesses the fast reactions, and thus retains only the slow variables. This allows for much larger time steps. The increase in the integration step size is approximately proportional to the ratio of the fast and the slow reaction rates to maintain the integration stability $kdt \sim 1$, where k is a reaction rate and dt is the integration step size. Note that only a forward integration scheme[64] was considered because such methods are the most popular and convenient schemes for integrating the chemical Langevin equations.

An interesting question is whether it is possible to incorporate the fast reaction distribution equation (17) into the slow Langevin equation (7) to forge a single Langevin equation for m. The answer seems to be negative. Equation (7) is the core of the whole reduction but is only valid for the conditional average $\bar{m}$ as we argued in Sec. II. Thus, to write an equation purely in terms of m, we have to deduce the average $\bar{m}$ from the known full random m at each time instant in order to make use of the slow manifold property. This is impossible without simulating the fast noise which breaks our original goal of reducing computational load. In the literature discussed above, the slowest and the fastest motions including fluctuations are always treated separately. If there are more than two time scales, the fastest reactions should reach steady state first, followed by the second fastest, and then the remaining reactions. Our reduction scheme evolves only the slowest set of reactions and is able to sequentially generate the distributions of fast variables based on the values of the slower variables.

V. SUMMARY

In this paper, we proposed a new scheme for performing model reduction using a separation of time scales in the chemical Langevin equation. In the reduced model, a Langevin equation is used to advance the slow variables while assuming the fast variables remain in a quasisteady state. A stochastic numerical simulation technique is devised to implement the proposed scheme. Two examples are used to demonstrate the effectiveness of our method. The first one is a birth-death-dimerization process and the second is a simple gene regulation network. In these examples, reactions involved in the dimerization reactions are assumed to be faster than other processes, and thus represent fast degrees of freedom. However, the monomer participates in both the slow and the fast reactions, thus mixing time scales. We introduced an auxiliary variable to successfully separate reaction components with different time scales and achieved significant computation efficiency in the resulting reduced model. The approximate solutions agree extremely well with the results computed from exact Gillespie simulations.

Although more sophisticated integration techniques are available that would further improve the efficiency of the current scheme, the main goal here was to demonstrate our scheme's validity and usefulness. We chose to use a simple method to integrate Eq. (26). The time step dt is required to be much smaller than the average switching time between promoter states. This restriction can be removed by using hybrid integration schemes like those employed in treating the master equation.[7,8,33,37] Furthermore, these integration schemes are able to handle time-dependent noise and may be used in our reduction technique to treat nonautonomous systems.

The reduction scheme has been shown to be very effective in the two examples in which the network structures are relatively simple and the slow and fast variables are easy to identify. In more realistic situations, the cellular network could be much more complicated and the time scales are more entangled. Defining slow variables and separating different time scales in a systematic way is essential for the implementation of the current technique. Furthermore, often in realistic large-scale networks, there exists a span of time scales, where there is no clear-cut distinction between the slow and fast processes. Systematically extending our technique to model systems with multiple time scales is a challenging and interesting problem. Recent work by Ball *et al.*[36] provides a possible avenue for pursuing this problem.

ACKNOWLEDGMENTS

[1] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, Nature (London) **402**, C47 (1999).
[2] *Foundations of Systems Biology*, edited by H. Kitano (MIT Press, Cambridge, 2001).
[3] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[4] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[5] M. A. Gibson and J. Bruck, J. Phys. Chem. A **104**, 1876 (2000).
[6] C. V. Rao, D. M. Wolf, and A. P. Arkin, Nature (London) **420**, 231 (2002).
[7] T. R. Kiehl, R. M. Mattheyses, and M. K. Simmons, Bioinformatics **20**, 316 (2004).
[8] J. Pucha_lka and A. M. Kierzek, Biophys. J. **86**, 1357 (2004).
[9] Y. Lan and G. Papoian, Biophys. J. **94**, 3839 (2008).
[10] Y. Lan and G. Papoian, Phys. Rev. Lett. **98**, 228301 (2007).
[11] N. Barkai and S. Leibler, Nature (London) **387**, 913 (1997).
[12] F. Rieke and D. A. Baylor, Biophys. J. **75**, 1836 (1998).
[13] J. M. G. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler, Proc. Natl. Acad. Sci. U.S.A. **99**, 5988 (2002).
[14] N. Barkai and S. Leibler, Nature (London) **403**, 267 (2000).
[15] P. S. Swain, J. Mol. Biol. **344**, 965 (2004).
[16] L. S. Weinberger, J. C. Burnett, J. E. Toettcher, A. P. Arkin, and D. V. Schaffer, Cell **122**, 169 (2005).
[17] M. Thattai and A. van Oudenaarden, Genetics **167**, 523 (2004).
[18] E. Korobkova, T. Emonet, J. M. G. Vilar, T. S. Shimizu, and P. Cluzel, Nature (London) **428**, 574 (2004).
[19] P. W. Zandstra, D. A. Lauffenburger, and C. J. Eaves, Blood **96**, 1215 (2000).
[20] G. M. Süel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M. B. Elowitz, Science **315**, 1716 (2007).
[21] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland Personal Library, Amsterdam, 1992).
[22] L. E. Reichl, *A Modern Course in Statistical Physics* (Wiley, New York, 1998).
[23] D. T. Gillespie, J. Chem. Phys. **113**, 297 (2000).
[24] D. T. Gillespie, J. Chem. Phys. **115**, 1716 (2001).
[25] T. C. Meng, S. Somani, and P. Dhar, In Silico Biology **4**, 0024 (2004).
[26] J. Hasty and J. J. Collins, Nat. Genet. **31**, 13 (2002).
[27] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, New York, 2002).
[28] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1984).
[29] E. L. Haseltine and J. B. Rawlings, J. Chem. Phys. **117**, 6959 (2002).
[30] C. V. Rao and A. P. Arkin, J. Chem. Phys. **118**, 4999 (2003).
[31] R. Bundschuh, F. Hayot, and C. Jayaprakash, Biophys. J. **84**, 1606 (2003).
[32] T. Tian and K. Burrage, J. Chem. Phys. **121**, 10356 (2004).
[33] H. Salis and Y. Kaznessis, J. Chem. Phys. **122**, 054103 (2005).
[34] R. Erban, I. G. Kevrekidis, D. Adalsteinsson, and T. C. Elston, J. Chem. Phys. **124**, 084106 (2005).
[35] Y. Cao, D. T. Gillespie, and L. Petzold, J. Comput. Phys. **206**, 395 (2005).
[36] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala, J. Appl. Probab. **16**, 1925 (2006).
[37] A. Hellander and P. Lötstedt, J. Comput. Phys. **227**, 100 (2007).
[38] M. J. Morelli, R. J. Allen, S. Tănase-Nicola, and P. R. ten Wolde, J. Chem. Phys. **128**, 045105 (2008).
[39] J. J. Hornberg, B. Binder, F. J. Bruggeman, B. Schoeberl, R. Heinrich, and H. V. Westerhoff, Oncogene **24**, 5533 (2005).
[40] J. Ross and M. O. Vlad, Annu. Rev. Phys. Chem. **50**, 51 (1999).
[41] G. K. Ackers, A. D. Johnson, and M. A. Shea, Proc. Natl. Acad. Sci. U.S.A. **79**, 1129 (1982).
[42] R. Temam, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, Appl. Math. Sci. No. 68 (Springer-Verlag, New York, 1988).
[43] C. Foias, M. S. Jolly, I. G. Kevrekidis, G. R. Sell, and E. S. Titi, Phys. Lett. A **131**, 433 (1988).
[44] I. G. Kevrekidis, C. W. Gear, and G. Hummer, AIChE J. **50**, 1346 (2004).
[45] M. Valorani, F. Creta, D. A. Goussis, J. C. Lee, and H. N. Najm, Combust. Flame **146**, 29 (2006).
[46] S. H. Lam and D. A. Goussis, Int. J. Chem. Kinet. **26**, 461 (1994).
[47] Y. Lan and G. A. Papoian, J. Chem. Phys. **125**, 154901 (2006).
[48] Y. Lan, P. Wolynes, and G. A. Papoian, J. Chem. Phys. **125**, 124106 (2006).
[49] R. L. Honeycutt, Phys. Rev. A **45**, 600 (1992).
[50] J. Lehn, A. Rößler, and O. Schein, J. Comput. Appl. Math. **138**, 297 (2002).
[51] A. C. Brańka and D. M. Heyes, Phys. Rev. E **58**, 2611 (1998).
[52] J. Wilkie, Phys. Rev. E **70**, 017701 (2004).
[53] R. F. Fox, J. Stat. Phys. **54**, 1353 (1989).
[54] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[55] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, Nat. Genet. **31**, 69 (2002).
[56] M. Kaern, W. J. Blake, and J. J. Collins, Annu. Rev. Biomed. Eng. **5**, 179 (2003).
[57] M. Sasai and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **100**, 2374 (2003).
[58] W. J. Blake, M. Kaern, C. R. Cantor, and J. J. Collins, Nature (London) **422**, 633 (2003).
[59] A. M. Walczak, M. Sasai, and P. G. Wolynes, Biophys. J. **88**, 828 (2004).
[60] A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **102**, 18926 (2005).
[61] G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, Nature (London) **440**, 545 (2006).
[62] Y. Okabe, Y. Yagi, and M. Sasai, J. Chem. Phys. **127**, 105107 (2007).
[63] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).
[64] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, England, 1992).