

UC Irvine

UC Irvine Previously Published Works

Title

A moment-convergence method for stochastic analysis of biochemical reaction networks

Permalink

<https://escholarship.org/uc/item/2m25k1qh>

Journal

The Journal of Chemical Physics, 144(19)

ISSN

0021-9606

Authors

Zhang, Jiajun

Nie, Qing

Zhou, Tianshou

Publication Date

2016-05-21

DOI

10.1063/1.4950767

Peer reviewed

A moment-convergence method for stochastic analysis of biochemical reaction networks

Jiajun Zhang,¹ Qing Nie,² and Tianshou Zhou^{1,3,a)}

¹*School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China*

²*Department of Mathematics, University of California at Irvine, Irvine, California 92697, USA*

³*Guangdong Province Key Laboratory of Computational Science and School of Mathematics and Computational Science, Sun Yat-Sen University, Guangzhou 510275, People's Republic of China*

(Received 22 January 2016; accepted 4 May 2016; published online 20 May 2016)

Traditional moment-closure methods need to assume that high-order cumulants of a probability distribution approximate to zero. However, this strong assumption is not satisfied for many biochemical reaction networks. Here, we introduce convergent moments (defined in mathematics as the coefficients in the Taylor expansion of the probability-generating function at some point) to overcome this drawback of the moment-closure methods. As such, we develop a new analysis method for stochastic chemical kinetics. This method provides an accurate approximation for the master probability equation (MPE). In particular, the connection between low-order convergent moments and rate constants can be more easily derived in terms of explicit and analytical forms, allowing insights that would be difficult to obtain through direct simulation or manipulation of the MPE. In addition, it provides an accurate and efficient way to compute steady-state or transient probability distribution, avoiding the algorithmic difficulty associated with stiffness of the MPE due to large differences in sizes of rate constants. Applications of the method to several systems reveal nontrivial stochastic mechanisms of gene expression dynamics, e.g., intrinsic fluctuations can induce transient bimodality and amplify transient signals, and slow switching between promoter states can increase fluctuations in spatially heterogeneous signals. The overall approach has broad applications in modeling, analysis, and computation of complex biochemical networks with intrinsic noise. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4950767>]

I. INTRODUCTION

Intracellular processes are often depicted using biochemical reaction networks.¹ A main task of Systems Biology is to develop mathematical models that faithfully describe the dynamics of these networks. In general, an intracellular reaction network has three common properties: (i) the reactions in the system are stochastic due to discrete nature of chemical species;^{2,3} (ii) the reactions take place far from thermodynamic equilibrium;⁴ and (iii) the system is nonlinear either due to feedback⁵ or the cooperative binding of transcription factors (TFs) to binding sites in the regulatory regions⁶ or due to other nonlinear mechanisms.⁷ Those properties require analyzing the stochastic features of the network, e.g., in terms of the joint probability distribution.⁸ Since Poisson statistics cannot be assumed to be valid for living systems, characterizing the stochastic properties of biochemical reaction networks with intrinsic noise is challenging.

To capture precisely stochastic properties of a biochemical reaction system where intrinsic fluctuations may produce profound effects and even induce extra dynamics,^{9–11} it is needed to analyze the master probability equation (MPE)

of this system, which describes the time evolution of the probability distribution for all the different components in the system.¹² In some cases, the MPE can be analytically solved, often under simplified conditions, and some analytical methods have been developed.^{13–18} In most cases, however, it needs to be numerically solved using methods developed based either on direct integration^{19–22} or on Monte Carlo sampling.^{23–25} The question with these numerical methods is that they are computationally expensive, and in particular, they become infeasible as the number of reactive species is large. This is because the number of equations in the MPE grows exponentially with the number of reactive species, thus making the stochastic analysis difficult even for empirical networks of moderate size.^{26,27}

As an approximation to the MPE, moment equations^{28–34} are gaining increasing attention in modeling stochastic behaviors of reaction systems (see a recent work Ref. 35 and references therein). In moment formulations, however, the equations for low-order moments depend on high-order moments, implying that the resulting moment equations are not a closed system and truncation is required. In many moment-closure schemes, including linear noise approximation (LNA),³⁶ only the first- and second-order moments are included, thus providing only the partial information on a system's stochastic properties.²⁸ Other complex moment-based schemes have been proposed as

^{a)} Author to whom correspondence should be addressed. Electronic mail: mcszhtsh@mail.sysu.edu.cn.

well, e.g., a recently developed moment expansion method,²⁹ which needs to use the lower-order moments to approximate higher-order moments (e.g., a combination of the first- and second-order moments is used to approximate the third-order moment). In a word, how to efficiently truncate the infinite hierarchy of moment equations generated by the MPE is a key challenge in a moment formulation. Typically, there are no efficient methods to control errors in truncation since original or central moments (called ordinary moments) do not converge to zero when their orders tend to infinity. Another main shortcoming of the ordinary moments is that they cannot be used to reconstruct the corresponding joint probability distribution unless it is Poissonian. However, this reconstruction is frequently needed to capture stochastic effects of reaction networks precisely.

Another useful kind of moment formulation is the one based on binomial moments.^{33,34} When used to stochastic analysis of a reaction system, the binomial moment formulation has a remarkable advantage: it can greatly reduce computational complexity from the exponential growth of the master equation to the polynomial growth of the binomial moment equations with regard to the number of reactive species. For some reaction networks, however, binomial moments cannot be guaranteed to converge to zero as their orders go to infinity,³⁷ similar to the case of ordinary moments. To overcome the drawback of this formulation but motivated by this pioneering work, here, we propose convergent moments (CMs) to replace binomial moments for stochastic analysis of reaction networks, which are defined as coefficients in the Taylor expansion of a probability-generating function at some point (see Sec. II for details). A key point of our CMs is that we introduce a parameter vector in binomial moments. As a result, the CMs can be guaranteed to converge to zero as their orders go to infinity (just because of it, they are called as convergent moments). Based on the MPE of a reaction network, we derive a linear set of ordinary differential equations (ODEs) for CMs (called the moment-convergence equations: MCEs). In contrast to Refs. 33 and 34, however, the MCEs are derived in a mathematically rigorous manner. Other important differences between our work and that work include that we elucidate the physical meaning of the first-order MCEs, present an explicit but useful formula for calculating the time-evolutional joint probability distribution using CMs, and discuss sufficient conditions that ensure that MCEs are effectively truncated to become a closed system.

This new formulation, which explicitly captures the reaction network structure in terms of rate laws and stoichiometries, allows for fast and efficient characterization of the stochastic properties of a reaction network. Other advantages include that (i) unlike ordinary moment equations in which the inclusion of high-order moments would be prohibitively complicated, the MCEs can be easily constructed to include moments up to any desired order; (ii) CMs can be conveniently used for fast computation of a time-evolutional joint probability distribution up to a pre-specified accuracy as well as even for analytical derivation of the steady-state distribution; (iii) in contrast to the exponential growth of the MPE with regard to the species number (M), the number of the MCEs increases polynomially with M , thus greatly reducing

computational cost and enabling stochastic simulation of even complex reaction networks involving a large number of reactive species beyond the feasibility limit of any existing method. Accuracy and efficiency are the most important features of this new method in modeling intrinsic noise in biochemical reaction networks.

We will first describe our moment-convergence method, and then analyze several reaction networks of simple structures to demonstrate its applicability. All examples are chosen such that numerical simulations are easily carried out in a personal computer. We will show that this method can be used to determine the effects of parameters on the stochastic properties of each system. In addition, we will uncover several new mechanisms on genetic regulations driven by stochastic effects, which would be difficult to obtain through direct numerical simulation of the MPE. Finally, we will describe how one uses our method for other more complex reaction networks in the Discussion.

II. CONVERGENT MOMENT THEORY

A. Introduction of convergent moments

Before introducing convergent moments, let us analyze two simple examples. The first example is used to show that common moments (raw moments and central moments) do not converge to zero as their orders go to infinity. The second example is used to show the intuitive physical meaning of the first-order convergent moments. First, consider a birth-death process of one single reactive species: $\emptyset \xrightarrow{g} X$ and $2X \xrightarrow{d} \emptyset$. From the corresponding MPE, it is not difficult to show that the steady-state raw moments satisfy $\langle n^{k+1} \rangle > \langle n^k \rangle$ for any $k \in \{1, 2, \dots\}$, implying that the original moment $\langle n^k \rangle$ does not converge to zero as k goes to infinity. Furthermore, according to the expansion formula for the polynomial $(a - b)^k$, we have

$$\langle (n - \langle n \rangle)^k \rangle = \sum_{i=0}^k (-1)^{k-i} \binom{k}{i} \langle n \rangle^{k-i} \langle n^i \rangle, \quad k = 1, 2, \dots \quad (1)$$

Thus, by mathematical induction and using the known fact $\langle n^{i+1} \rangle > \langle n^i \rangle$ for any positive integer i , we can show that the central moment $\langle (n - \langle n \rangle)^k \rangle$ does not tend to zero as k is large. In a word, neither the raw moments nor the central moments for this example tend to zero as their orders go to infinity. Numerical simulation also verifies this fact, referring to the empty circles (central moments) and the triangles (raw moments) in Fig. 1.

Then, consider another toy example, which is mainly used to show how the differential equations for the first-order CMs are written directly based on reactions and how they depend on the higher-order CMs. This example contains the following reactions: $\emptyset \xrightarrow{k_1} A$, $\emptyset \xrightarrow{k_2} B$, $A + A \xrightarrow{k_3} C$, $A + B \xrightarrow{k_4} C$, $C \xrightarrow{k_5} \emptyset$. Note that every reaction rate (k_i) represents the transition probability per time. Thus, the transition rate for reaction $\emptyset \xrightarrow{k_1} A$ is given by k_1 ,

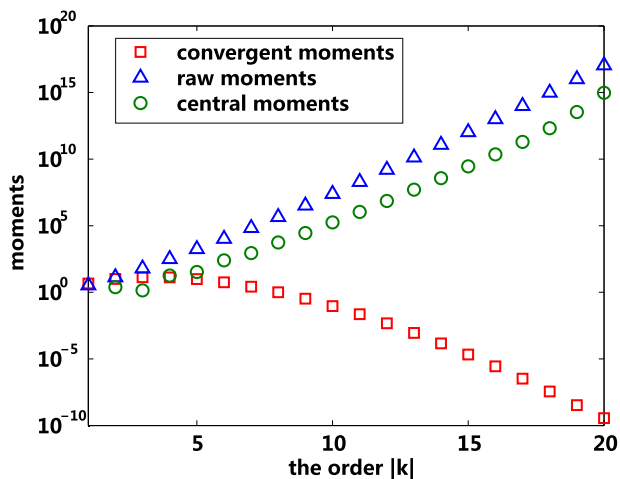


FIG. 1. An example used to show that both raw moments and central moments do not tend to zero but the convergent moments introduced in this paper tend to zero as their orders go to infinity. This example contains only two reactions $\emptyset \xrightarrow{g} X$ and $2X \xrightarrow{d} \emptyset$, where parameter values are set as $g = 20$, $d = 1$.

whereas the transition rate for reaction $A + A \xrightarrow{k_3} C$ by $k_3 A(A-1)$. If the symbol $\binom{Y}{X}$ represents the common binomial coefficient, then the transition rate for $\emptyset \xrightarrow{k_1} A$ can be expressed as $k_1 \binom{A}{0} \binom{B}{0} \binom{C}{0}$, whereas the transition rate for $A + A \xrightarrow{k_3} C$ can be expressed as $k_3 (2!) \binom{A}{2} \binom{B}{0} \binom{C}{0}$. Similarly, we can express transition rates for other reactions. Also note that the deterministic equations for this example can be expressed as

$$\begin{aligned} \frac{d}{dt} \left[\binom{A}{1} \binom{B}{0} \binom{C}{0} \right] &= k_1 \binom{A}{0} \binom{B}{0} \binom{C}{0} - 2k_3 \binom{A}{2} \binom{B}{0} \binom{C}{0} \\ &\quad - k_4 \binom{A}{1} \binom{B}{1} \binom{C}{0} \\ \frac{d}{dt} \left[\binom{A}{0} \binom{B}{1} \binom{C}{0} \right] &= k_2 \binom{A}{0} \binom{B}{0} \binom{C}{0} \\ &\quad - k_4 \binom{A}{1} \binom{B}{1} \binom{C}{0} \\ \frac{d}{dt} \left[\binom{A}{0} \binom{B}{0} \binom{C}{1} \right] &= 2k_3 \binom{A}{2} \binom{B}{0} \binom{C}{0} + k_4 \binom{A}{1} \binom{B}{1} \binom{C}{0} \\ &\quad - k_5 \binom{A}{0} \binom{B}{0} \binom{C}{1}. \end{aligned} \quad (2)$$

We will see that the term, e.g., $\binom{A}{1} \binom{B}{0} \binom{C}{0}$ is nothing but the first-order CM of reactive species A . Other terms can be explained similarly. Thus, the above equations imply that the time change rate of the first-order CM is equal to the algebraic sum of the CMs for all the reactions that the corresponding reactive species involves, where the “+” symbol represents the positive contribution of one reaction to this species, whereas the “-” symbol represents the negative contribution of a different reaction to the species. Similarly, we can write differential equations for higher-order CMs. In this

sense, our CMs have definite physical meanings. Note that the equations for lower-order CMs depend on higher-order CMs, so truncation is required to obtain a closed system. One main task of this paper is to solve this truncation issue.

We can extend the above first-order moment equations for a particular example to a general reaction network (see contents in Subsection II B). In this subsection, however, we focus on introducing convergent moments.

Let $N = (N_1, N_2, \dots, N_M)$ represent the state vector of a reaction network at time t (hereafter bold letters stand for vector or matrix). Then, there is a time-evolutional joint probability distribution for state variable vector N , denoted by $P(N; t)$. As has been seen by example, the cumulants of this distribution may not tend to zero but may be divergent. To understand the essential reason behind it, let us examine the probability-generating function defined by $G(\mathbf{z}; t) = \sum_{\mathbf{k}} \mathbf{z}^{\mathbf{k}} P(\mathbf{k}; t)$, where we define $\mathbf{z}^{\mathbf{k}} \equiv z_1^{k_1} \dots z_M^{k_M}$ for any real vector $\mathbf{z} = (z_1, \dots, z_M)$. Apparently, this function is the sum of a power series with coefficients $P(N; t)$, and the convergence radius of this series, denoted by R , is at least 1 due to $G(\mathbf{1}; t) = \sum_{\mathbf{N}} P(\mathbf{N}; t) \equiv 1$, i.e., $R \geq 1$. Note that $R \geq 1$ does not mean that all the derivatives of $G(\mathbf{z}; t)$ at $\mathbf{z} = \mathbf{1}$ exist. On the other hand, cumulants are actually functions of these derivatives, e.g., in the one-dimensional case, the second-order central moment can be expressed as $\langle (N - \langle N \rangle)^2 \rangle = \partial^2 G / \partial z^2|_{z=1} - (\partial G / \partial z|_{z=1})^2 + \partial G / \partial z|_{z=1}$ and the mean as $\langle N \rangle = \partial G / \partial z|_{z=1}$. The nonexistence of some derivatives would mean that cumulants are divergent. In spite of this, for any point \mathbf{z}_0 in the convergent region $D = \{\mathbf{z} | \|\mathbf{z}\| < R\}$, we know that all the derivatives of $G(\mathbf{z}; t)$ always exist since $G(\mathbf{z}; t)$ can be expanded into a power series at this point. If we denote by $b_{\mathbf{k}}(\mathbf{z}_0; t)$ the coefficients of this series, then simple calculation yields

$$b_{\mathbf{k}}(\mathbf{z}_0; t) = \frac{1}{\mathbf{k}!} \left. \frac{\partial^{\mathbf{k}} G(\mathbf{z}; t)}{\partial \mathbf{z}^{\mathbf{k}}} \right|_{\mathbf{z}=\mathbf{z}_0} = \sum_{\mathbf{N} \geq \mathbf{k}} \mathbf{z}_0^{\mathbf{N}-\mathbf{k}} \binom{\mathbf{N}}{\mathbf{k}} P(\mathbf{N}; t), \quad (3)$$

where we define $\partial^{\mathbf{k}} G / \partial \mathbf{z}^{\mathbf{k}} = \partial^{k_1 + \dots + k_M} G / \partial z_1^{k_1} \dots \partial z_M^{k_M}$ for any configuration $\mathbf{k} = (k_1, \dots, k_M)$, $\mathbf{k}! = k_1! \dots k_M!$, $\mathbf{z}_0^{\mathbf{k}} = z_{10}^{k_1} \dots z_{M0}^{k_M}$, and $\binom{\mathbf{N}}{\mathbf{v}} = \prod_{i=1}^M \binom{N_i}{v_i}$ for two vectors \mathbf{N} and \mathbf{v} , each $\binom{N_i}{v_i}$ representing a common binomial coefficient (owing to this coefficient, $b_{\mathbf{k}}(\mathbf{z}_0; t)$ with $\mathbf{z}_0 = \mathbf{1}$ are sometimes called as binomial moments,^{33,34} which however cannot be guaranteed to converge to zero as their orders go to infinity, for some reaction networks. See the content in Section II C below). Sometimes, we omit \mathbf{z}_0 in $b_{\mathbf{k}}(\mathbf{z}_0; t)$ for convenience, unless the confusion arises.

We emphasize that in Eq. (3), the parameter vector \mathbf{z}_0 is introduced for a rigorous derivation. A more important point is that this introduction is to guarantee that $b_{\mathbf{k}}(\mathbf{z}_0; t)$ converges to zero as its order tends to infinity (in the case of zero convergence, $b_{\mathbf{k}}(\mathbf{z}_0; t)$ is called as a convergent moment (CM) with $|\mathbf{k}| = \sum_{i=1}^M k_i$, representing the order of this moment). Note that a point \mathbf{z}_0 for zero convergence of $b_{\mathbf{k}}(\mathbf{z}_0; t)$ definitely exists in any case. In fact, if $\mathbf{z}_0 = \mathbf{0}$, then $b_{\mathbf{k}}(\mathbf{z}_0; t) = P(\mathbf{k}; t)$, which tends to zero as $|\mathbf{k}| \rightarrow \infty$ due to $\lim_{|\mathbf{N}| \rightarrow \infty} P(\mathbf{N}; t) = 0$. According to Eq. (3), we know that $b_{\mathbf{k}}(\mathbf{z}_0; t)$ is continuous

with regard to z_0 . Therefore, there is a small neighbor of z_0 such that $\lim_{|k| \rightarrow \infty} b_k(z_0; t) = 0$. In a word, we can always select $z_0 \neq \mathbf{0}$ such that $\lim_{|k| \rightarrow \infty} b_k(z_0; t) = 0$. In particular, for many biochemical networks (e.g., gene regulatory networks), we can choose $z_0 = \mathbf{1}$ (note that the case of $z_0 = \mathbf{1}$ corresponds to the binomial moment formulation^{33,34}) but still can guarantee that $b_k(t) \rightarrow 0$ as $|k| \rightarrow \infty$. See the examples shown in Figs. 1 and 2, wherein CMs approach to zero after their orders are large enough. For some peculiar reaction networks, however, $z_0 \neq \mathbf{1}$ should be chosen. See the example analysis shown in Fig. 3. Regarding the choice of z_0 , we will give more discussions in Subsection II C, wherein we also propose a simple strategy to find an “optimal” z_0 .

Apart from the zero convergence, another important advantage of CMs is that they can be conveniently used to calculate the probability distribution. In fact, according to the relationship $P(N; t) = (1/N!) \left[\partial^N G(z; t) / \partial z^N \right]_{z=0}$ combined with the expansion $G(z; t) = \sum_k b_k(z_0; t) (z - z_0)^k$ from which $\partial^N G(z; t) / \partial z^N \big|_{z=0} = \sum_k (-z_0)^{k-N} k(k-1) \cdots (k-N+1) b_k(z_0; t)$ can be derived, we have

$$P(N; t) = \sum_{k \geq N} (-z_0)^{k-N} \binom{k}{N} b_k(z_0; t). \quad (4)$$

This will provide a method for reconstructing a probability distribution based on CMs, where the latter can be easily obtained by a numerical method (see the following content).

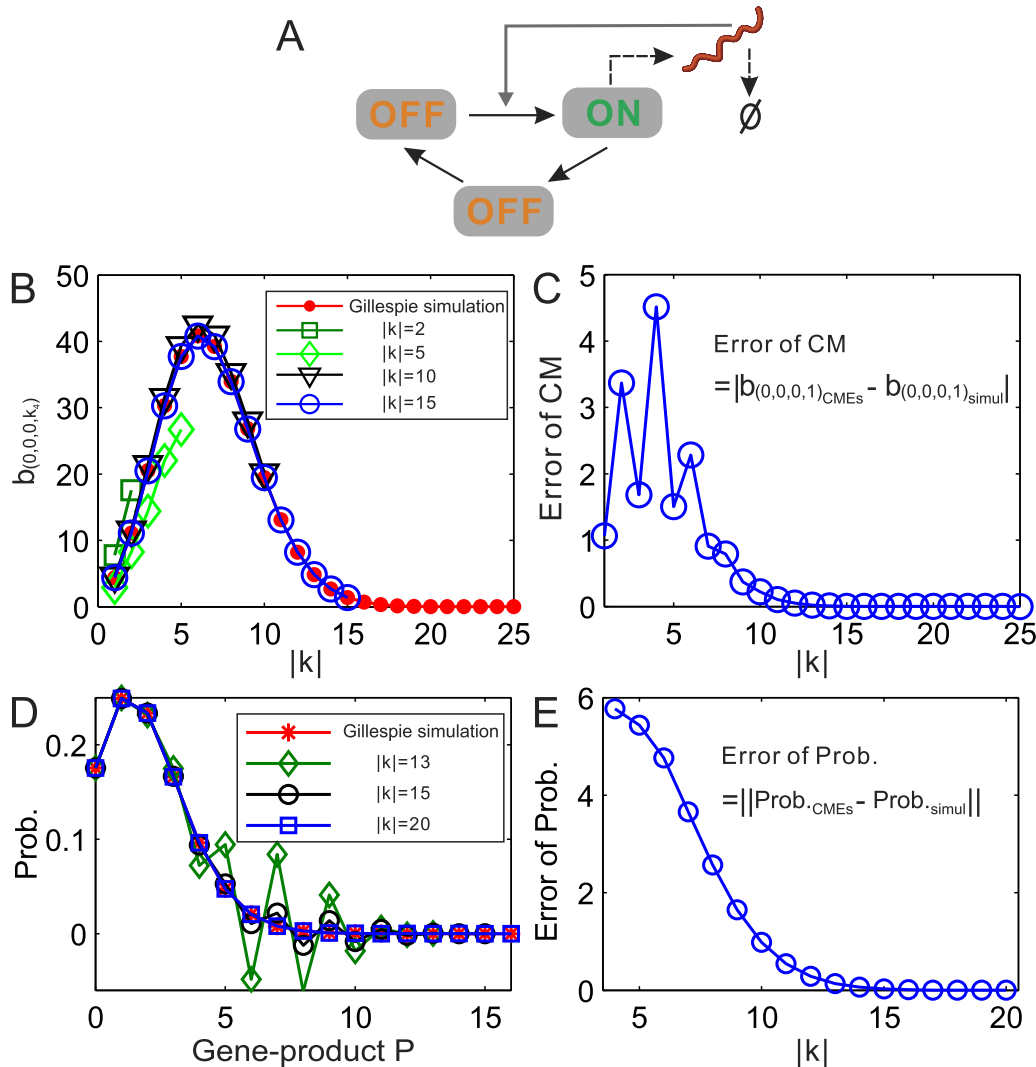


FIG. 2. Zero convergence of convergent moments (CMs) and the distribution reconstructed using CMs. (a) Schematic diagram for a stochastic model of gene expression. All the reactions are listed below: $G_1 \xrightarrow{\lambda_{21}} G_2$, $G_2 \xrightarrow{\lambda_{32}} G_3$, $G_3 \xrightarrow{\lambda_{13}} G_1$, $G_3 + P \xrightarrow{\lambda'_{13}} G_1 + P$, $G_1 \xrightarrow{\mu_1} G_1 + P$, and $P \xrightarrow{\delta} \emptyset$, where G_1 represents the active state of the gene promoter, whereas G_2 and G_3 represent two inactive states; (b) and (c) CMs tend to zero as their orders are large enough, where $b_k = b_{(k_1, k_2, k_3, k_4)}$ represents a CM with k_1, k_2, k_3, k_4 corresponding, respectively, to G_1, G_2, G_3, P , and “error” represents the absolute difference between the first-order CM of the P species obtained by a closed system of CMs (implying that truncation is made), denoted by $b_{(0,0,0,1)_{MCE}}$, and the first-order CM obtained by the Gillespie algorithm denoted by $b_{(0,0,0,1)_{simul}}$. Parameter values are set as $\lambda_{21} = \lambda_{32} = \lambda_{13} = \lambda'_{13} = 10$, $\mu = 5$, $\delta = 1$; (d) and (e) CMs approximate well the probability distribution, where “prob.” represents the gene-product distribution, and “Error of Prob.” represents the Euclidean distance between the distribution reconstructed using the CMs obtained by a closed system of CMs and the distribution obtained by the Gillespie algorithm, i.e., $\text{Error of Prob} = \sqrt{\sum_{i=0}^m [P_{MCE}(i) - P_{simul}(i)]^2}$. Parameter values are set as $\lambda_{21} = \lambda_{32} = \lambda_{13} = \lambda'_{13} = 10$, $\mu = 5$, $\delta = 1$. In panels (b)-(e), $|k| = k_1 + k_2 + k_3 + k_4$ represents the CM order, and initial values are set as $(G_1, G_2, G_3, P) = (1, 0, 0, 0)$.

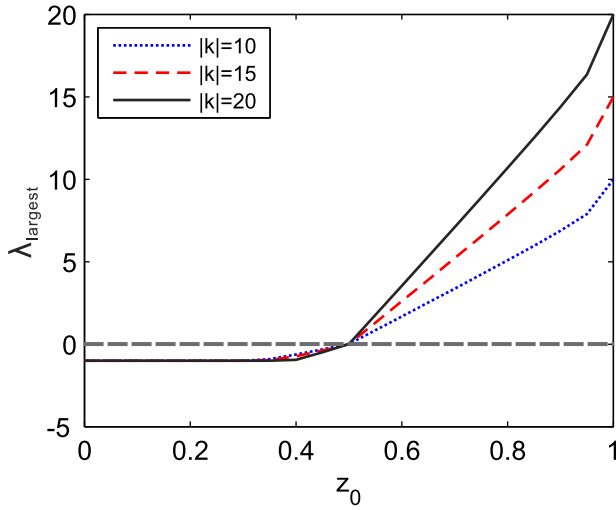


FIG. 3. Dependence of the largest real part of the characteristic values of $A_j(z_0)$ (denoted by λ_{largest}) on z_0 in the microscopic Lotka-Volterra model described by $A + X \xrightarrow{k_1} 2X$, $X + Y \xrightarrow{k_2} 2Y$, $Y \xrightarrow{k_3} P$. Here, $z_0 = z_{10} = z_{20}$ is set and parameter values are set as $ak_1 = 1$ and $k_2 = k_3 = 1$.

As an example, Figs. 2(d) and 2(e) show that CMs can well approximate the distribution as their orders are large. Further advantages of CMs will be demonstrated afterwards.

B. Derivation of moment-convergence equations

Although we have introduced CMs above, an unsolved question is how they are used to stochastic analysis of a general reaction network. Here we derive a linear set of ordinary differential equations (ODEs) for CMs. Assume that this network consists of M interacting species X_i ($1 \leq i \leq M$) that undergo a series of reactions in such a

general form: $\sum_{i=1}^M r_i X_i \rightarrow \sum_{i=1}^M s_i X_i$ with reaction constant rate denoted by c_r^s , where the stoichiometric coefficients r_i and s_i are non-negative integers. For convenience, let $\mathbf{r} \rightarrow \mathbf{s}$ represent this reaction, where $\mathbf{r} = (r_1, \dots, r_M)$ and $\mathbf{s} = (s_1, \dots, s_M)$. More precisely, we let $\mathbf{r} \rightarrow \mathbf{s}$ with reaction rate constant c_r^s represent $\sum_{i=1}^M r_i X_i \rightarrow \sum_{i=1}^M s_i X_i$ with the same rate constant. Let $\mathbf{N} = (N_1, N_2, \dots, N_M)$ represent the state of this system, where N_i is the number of copies of species X_i .

First, we derive a partial differential equation for probability-generating function $G(\mathbf{z}; t)$. Note that if we introduce a shift operator, which is defined as $\Phi^\alpha f(\mathbf{N}) = f(\mathbf{N} + \alpha)$, where f is any function of the state \mathbf{N} and α is a vector of integers, then the MPE can be expressed in the form $\partial P(\mathbf{N}; t) / \partial t = \sum_{\mathbf{r} \rightarrow \mathbf{s}} (c_r^s \cdot \mathbf{r}!) (\Phi^{-\mathbf{s}} \Phi^{\mathbf{r}} - I) \binom{\mathbf{N}}{\mathbf{r}} P(\mathbf{N}; t)$. By differentiating both sides of the equality $G(\mathbf{z}; t) = \sum_{\mathbf{N}} \mathbf{z}^{\mathbf{N}} P(\mathbf{N}; t)$ with regard to t and using this new form of the MPE, we can derive

$$\frac{\partial G(\mathbf{z}; t)}{\partial t} = \sum_{\mathbf{r} \rightarrow \mathbf{s}} c_r^s (\mathbf{z}^{\mathbf{s}} - \mathbf{z}^{\mathbf{r}}) \frac{\partial^{\mathbf{r}} G(\mathbf{z}; t)}{\partial \mathbf{z}^{\mathbf{r}}}, \quad (5)$$

which is a partial differential equation with free boundary conditions, where we have omitted the reaction rate constant c_r^s for the reaction $\mathbf{r} \rightarrow \mathbf{s}$. See the Appendix for derivation of Eq. (5). This equation itself would be very useful since it can be numerically or analytically solved using partial differential equation theories in textbooks. In turn, the resulting $G(\mathbf{z}; t) = \sum_{\mathbf{N}} \mathbf{z}^{\mathbf{N}} P(\mathbf{N}; t)$ can be used to give the distribution according to the formal relationship $P(\mathbf{N}; t) = (1/N!) \partial^{\mathbf{N}} G(\mathbf{z}; t) / \partial \mathbf{z}^{\mathbf{N}}|_{\mathbf{z}=\mathbf{0}}$.

Second, differentiating both sides of Eq. (5) with regard to \mathbf{z} at the $\mathbf{z} = \mathbf{z}_0$ and using Eq. (3), we can further derive the following linear MCEs:

$$\frac{db_{\mathbf{k}}(t)}{dt} = \sum_{\mathbf{r} \rightarrow \mathbf{s}} (c_r^s \cdot \mathbf{r}!) \left[\sum_{\mathbf{i}=0}^{\mathbf{k}} \binom{\mathbf{s}}{\mathbf{i}} \mathbf{z}_0^{\mathbf{s}-\mathbf{i}} - \binom{\mathbf{r}}{\mathbf{i}} \mathbf{z}_0^{\mathbf{r}-\mathbf{i}} \right] \binom{\mathbf{r} + \mathbf{k} - \mathbf{i}}{\mathbf{r}} b_{\mathbf{r} + \mathbf{k} - \mathbf{i}}(t), \quad (6)$$

which holds for any configuration \mathbf{k} , namely, each component k_i in $\mathbf{k} = (k_1, k_2, \dots, k_M)$ may be any non-negative integer. Equation (6) with all the possible \mathbf{k} constitutes a linear ODE group of $b_{\mathbf{k}}(t)$ with constant coefficients. We point out that Eq. (6) is derived through the MPE, but it can be written directly based on a given reaction network since Eq. (6) has captured network structure in terms of rate laws and stoichiometries. In particular, when writing Eq. (6) for a given reaction network, one may first consider individual reactions appearing in this system and then integrate the MCEs for all the reactions. In addition, if the system volume, denoted by Ω , is considered, then Eq. (5) should be rewritten as

$$\frac{db_{\mathbf{k}}(t)}{dt} = \sum_{\mathbf{r} \rightarrow \mathbf{s}} \left(\frac{c_r^s \cdot \mathbf{r}!}{\Omega^{\mathbf{r}}} \right) \left[\sum_{\mathbf{i}=0}^{\mathbf{k}} \binom{\mathbf{s}}{\mathbf{i}} \mathbf{z}_0^{\mathbf{s}-\mathbf{i}} - \binom{\mathbf{r}}{\mathbf{i}} \mathbf{z}_0^{\mathbf{r}-\mathbf{i}} \right] \binom{\mathbf{r} + \mathbf{k} - \mathbf{i}}{\mathbf{r}} b_{\mathbf{r} + \mathbf{k} - \mathbf{i}}(t). \quad (6')$$

To help understand the above general formulation, e.g., Eq. (6), let us consider a simpler reaction system: $\emptyset \xrightarrow{g} X \xrightarrow{d} \emptyset$, which models the birth and death of one single species. Let k represent the number of molecules for species X . Then, the corresponding MCEs (5) become

$$\begin{aligned} \frac{db_k(t)}{dt} &= g [(z_0 - 1) b_k + b_{k-1}] \\ &\quad - d [(z_0 - 1)(1 + k) b_{k+1} + k b_k], \end{aligned} \quad (7)$$

where $k = 0, 1, 2, \dots$. Owing to $b_0 = \sum_{\mathbf{n}} P(\mathbf{n}; t) \equiv 1$ and since Eq. (6) holds for $k = 0$, we have $b_1 = g/d$, which is

model,³⁸ which contains three reactions: $A + X \xrightarrow{k_1} 2X$, $X + Y \xrightarrow{k_2} 2Y$, $Y \xrightarrow{k_3} P$. Here the molecular number of species A , denoted by a , is assumed as a fixed constant. One will see that for this example, choosing $\mathbf{z}_0 = \mathbf{1}$ cannot

guarantee that the real parts of all the characteristic values of matrix $A^{(j)}(\mathbf{z}_0)$ in Eq. (9) are negative. In fact, let n_1 and n_2 represent the molecule number of species X and Y , respectively. Then, the corresponding MCEs take the following form:

$$\begin{aligned} \frac{db_{n_1, n_2}}{dt} = & ak_1 [z_{10}n_1b_{n_1, n_2} + (n_1 - 1)b_{n_1-1, n_2} + (z_{10}^2 - z_{10})(n_1 + 1)b_{n_1+1, n_2}] \\ & + k_2 [(z_{20}^2 - z_{10}z_{20})b_{n_1+1, n_2+1} + (2z_{20} - z_{10})(n_1 + 1)n_2b_{n_1+1, n_2} + (n_1 + 1)(n_2 - 1)b_{n_1+1, n_2-1}] \\ & - k_3 [n_1(n_2 + 1)b_{n_1, n_2+1} - n_1n_2b_{n_1, n_2}] + k_3 [(1 - z_{20})(n_1 + 1)b_{n_1, n_2+1} - n_2b_{n_1, n_2}], \end{aligned} \quad (12)$$

where $b_{0,0} = 1$ and we define $b_{n_1, n_2} = 0$ if $n_1 < 0$ or $n_2 < 0$.

To simplify our analysis, we set $z_0 = z_{10} = z_{20}$. Denote by λ_{largest} , the largest real part of all the characteristic values of coefficient matrix $A^{(j)}(\mathbf{z}_0)$ in Eq. (10). Figure 3 plots the dependence of λ_{largest} on z_0 , showing that there is a critical value of z_0 (denoted by \tilde{z}_0) such that λ_{largest} is negative if z_0 is below \tilde{z}_0 and positive if z_0 is beyond \tilde{z}_0 . A positive real part implies that the MCEs cannot be efficiently truncated. Analysis of this example actually provides a strategy for finding a suitable z_0 . This strategy can be also extended to other reaction systems. As a result, computational complexity can be greatly reduced.

D. Computational cost and statistical quantities

First, recall that the number of equations in the MPE increases exponentially with the reactive species number. This property greatly limits the application range of the MPE. Here, we simply show that the number of MCEs increases only in polynomial with the reactive species number M . In fact, there exists a positive integer M_0 such that $b_{\mathbf{k}} \approx 0$ as $|\mathbf{k}| > M_0$ because of $b_{\mathbf{k}} \rightarrow 0$ as $|\mathbf{k}| \rightarrow \infty$, as mentioned above. Note that the number of different CMs of order $|\mathbf{k}|$ is given by the binomial coefficient $\binom{|\mathbf{k}| + M - 1}{|\mathbf{k}|}$. Therefore, the number of MCEs truncated up to order M_0 is given by $NB = \sum_{|\mathbf{k}|=1}^{M_0} \binom{|\mathbf{k}| + M - 1}{|\mathbf{k}|}$, which is apparently a polynomial of the number of reactive species (M). Owing to this property, MCEs can have broader applications than the MPE.

In general, the choice of M_0 depends on the largest one of the multiplicities of individual reactive species (e.g., the multiplicity of species X in reaction $X + Y \rightarrow Z$ is 1, whereas the multiplicity of species X in reaction $2X + Y \rightarrow Z$ is 2). Moreover, the larger the chosen M_0 is, the higher is the order of the polynomial. For most of common biochemical reaction networks, we may choose $M_0 = 5$ or less, and the approximate results have become very good or have well approached to the exact solutions.

Second, as has been mentioned above, if $\mathbf{z}_0 = \mathbf{0}$ is chosen, then $b_{\mathbf{k}}(t) = P(\mathbf{k}; t)$, implying that the number of the corresponding MCEs grows exponentially with regard

to the number of reactive species. If $\mathbf{z}_0 = \mathbf{1}$ can be chosen, then we know from the above analysis that the MCE number grows polynomially. Thus, we should choose z_0 in such a manner that it approaches to $\mathbf{1}$ as much as possible, so that the computational complexity can be reduced at the most. A general conclusion is that the computational complexity for any reaction network is in between exponential and polynomial growths of the species number. In addition, an ideal z_0 can be quickly found by dichotomy combined with an optimal principle, where by ‘‘ideal’’ we mean that z_0 approaches to $\mathbf{1}$ as much as possible. Some details for finding z_0 are stated as follows. Set $z_0 = z_{10} = \dots = z_{M_0}$, where z_{i0} are components of vector \mathbf{z}_0 . If CMs are convergent for $z_0 = \mathbf{1}$, then we choose $z_0 = \mathbf{1}$. Otherwise, we examine the case of $z_0 = 0.5$. If the CMs are convergent, then we further examine the case of $z_0 = 0.75$. If the CMs are still convergent, then we further examine the case of $z_0 = 0.875$. This process continues until an optimal z_0 is found.

Third, analyzing stochastic properties of a reaction system directly based on the MPE is in general difficult. Therefore, most previous works were focused on analysis of statistic quantities such as mean, variance, noise intensity, Fano factor, skewness, and kurtosis. These statistics can capture stochastic properties of some simple reaction networks such as birth-death processes but would mask useful information on other complex reaction networks. It should be pointed out that the LNA gives only approximate results of the first four statistics for nonlinear reaction networks, and in particular, skewness and kurtosis are zero due to the assumption of Gaussian variables in the LNA. For many other nonlinear reaction systems, however, the latter two indices are not equal to zero. See the following example.

Before analyzing the example, we provide exact formulae for calculating all the above quantities based on CMs. For clarity, let us consider only the case of one variable. The mean and variance are calculated according, respectively, to

$$\langle m \rangle = b_1(t) \sigma^2(t) = 2b_2(t) + b_1(t) - b_1^2(t). \quad (13)$$

With these, the noise strength and the Fano factor are given by

$$\eta^2(t) = \frac{2b_2(t) + b_1(t) - b_1^2(t)}{b_1^2(t)},$$

$$Fano = \frac{2b_2(t) + b_1(t) - b_1^2(t)}{b_1(t)}, \quad (14)$$

which are applicable to both linear and nonlinear reaction networks. These two indices are the most common metrics of noise in, e.g., gene expression³⁷ and will be alternatively used in analysis of our examples below. Since the first- and second-order moments frequently cannot well characterize a distribution, one needs to evaluate the skewness of this distribution over the time denoted by $\gamma_1(t)$, which characterizes the degree of the asymmetry of the distribution around its mean, and the kurtosis, denoted by $\gamma_2(t)$, which measures the degree of peakedness of the distribution. They are calculated according, respectively, to

$$\gamma_1(t) = \frac{\mu_3(t)}{\mu_2^{3/2}(t)}, \quad \gamma_2(t) = \frac{\mu_4(t)}{\mu_2^2(t)} - 3, \quad (15)$$

where the k -order central moment is given by

$$\mu_k(t) = (-b_1(t))^k + \sum_{i=0}^{k-1} \sum_{j=1}^{k-i} R(k, i, j)(j!)(b_1(t))^i b_j(t) \quad (16)$$

$$b_{k+1} = \frac{\langle (n+1)(n+2)\cdots(n+1+k) \rangle}{(k+1)!} = \frac{1}{(k+1)!} \left[\langle n^{k+1} \rangle + \frac{(k+1)(k+2)}{2} \langle n^k \rangle + \cdots + (k+1)! \right], \quad k = 1, 2, \dots, \quad (17)$$

which tend to zero as their orders go to infinity, due to $\langle n^{k+1} \rangle = \frac{1}{k} \sum_{i=0}^{k-1} \left[S\binom{k}{i} \langle n^i \rangle + (-1)^{k-i+1} \binom{k+1}{i} \langle n^{i+1} \rangle \right]$,

$k = 1, 2, \dots$. This indicates that CMs are indeed superior to common moments from the view of convergence.

Finally in this section, we point out that it is known from solving the moment equations in extensions of the linear noise approximation that the equations would become more and more ill-conditioned mainly due to the assumption of Poisson variables that would lead to the ill-conditioned Jacobian matrix. However, our CM method can overcome this drawback mainly since MCEs may incorporate moments of any orders as well as since CMs approach to zero as their orders are large.

III. APPLICATIONS

Since the number of MCEs increases only polynomially in terms of the reactive species number and since truncating MCEs through suitably choosing parameter z_0 is feasible, the above formulation can be in principle applied to stochastic analysis of any complex biochemical reaction network. In this section, we demonstrate its power in stochastic analysis of gene regulatory networks. More applications of MCEs

in which $R(k, i, j) = (-1)^i \binom{k}{i} S(k-i, j)$ with $S(n, k) = \sum_{i=0}^k (-1)^{k-i} \binom{k}{i} i^n$ being the Stirling number of the second kind.³⁹ Equation (16) establishes the relationship between central moments and CMs. Formulae (13)-(16) are easily extended to multivariate cases. We stress that all the above formulae expressed by CMs are exact if the CMs are exactly given and can be applied to any complex reaction networks.

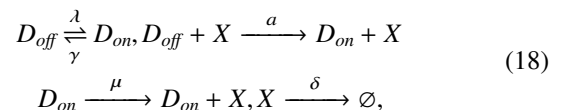
To help understand skewness and kurtosis, let us consider the above system of nonlinear degradation again, i.e., consider the system consisting of two reactions $\emptyset \xrightarrow{g} X$ and $2X \xrightarrow{d} \emptyset$. The MCEs corresponding to the choice of $z_0 = \mathbf{1}$ are given by $db_k/dt = gb_{k-1} - 2dk(k+1)b_{k+1} - dk(k-1)b_k$, where $b_0 = 1$. Note that $b_1 = \sqrt{S}$ is known,³⁴ where $S = g/d$ represents the system size. Therefore, according to the first formula in Eq. (15), it is not difficult to show $\gamma_1(t) = \frac{1}{S}(1 - \sqrt{S}) < 0$. Similarly, we can show that $\gamma_2(t)$ is not equal to zero for any S . This example indicates that skewness and kurtosis needs to be analyzed for better characterizing the corresponding distribution. In addition, for this example, we have rigorously proved that ordinary moments do not tend to zero as their orders are large, seeing Eq. (2). Here we mathematically show that CMs tend to zero as their orders go to infinity. In fact, steady-state CMs can be expressed as

to other biochemical reaction networks will be published elsewhere.

A. Analytical distributions

Here, we demonstrate how MCEs are used to derive analytical distributions as well as how parameter $z_0 \neq \mathbf{1}$ in Eq. (6) is chosen. For this, we analyze a simple example: a gene auto-activating circuit. This example was previously studied.⁵

The corresponding biochemical reactions read



where D_{off} and D_{on} stand for inactive and active states of the gene promoter, respectively, and X for gene product (mRNA or protein). Parameters λ and γ are transition rates between the states of the promoter. The symbol a represents the feedback strength. Parameters μ and δ are transcription and degradation rates, respectively. Let P_0 and P_1 be the probability distribution of the X species molecule number at D_{off} and D_{on} states, respectively. Let G_0 and G_1 be the corresponding probability-generating functions. Denote by $P = P_0 + P_1$, the

total probability distribution and by $G = G_0 + G_1$, the total generating function.

As is previously pointed out, a CM is actually a coefficient in the Taylor series of function G expanded at point z_0 , i.e., $b_k(t) = \frac{1}{k!} G_z^{(k)}(z_0; t)$. Thus, we first write the corresponding partial differential equations for G_0 and G_1

$$\begin{aligned} \frac{\partial G_0}{\partial t} &= -\lambda G_0 + \gamma G_1 - az \frac{\partial G_0}{\partial z} + \delta(1-z) \frac{\partial G_0}{\partial z} \\ \frac{\partial G_1}{\partial t} &= \lambda G_0 - \gamma G_1 + az \frac{\partial G_1}{\partial z} + \mu(z-1)G_1 \\ &\quad + \delta(1-z) \frac{\partial G_1}{\partial z} \end{aligned} \quad (19)$$

and then derive the following ODEs for G at steady state

$$\begin{aligned} \frac{\delta(a+\delta)}{\mu} \left[z - \frac{\delta}{a+\delta} \right] \\ G'' - \left[(a+\delta) \left(z - \frac{\delta}{a+\delta} \right) - \frac{(\lambda+\gamma)\delta}{\mu} \right] G' - \lambda G = 0. \end{aligned} \quad (20)$$

Choosing $z_0 = \frac{\delta}{a+\delta}$ and Taylor expanding $G(z)$ at z_0 , we obtain the following analytical expressions for steady-state CMs:

$$b_k = \frac{b_0}{k!} \left(\frac{\mu}{\delta} \right)^k \frac{(\lambda/(a+\delta))_k}{((\lambda+\gamma)/(a+\delta))_k}, \quad k = 1, 2, 3, \dots, \quad (21)$$

where $(c)_k$ is the Pochhammer symbol defined as $(c)_k = \Gamma(c+k)/\Gamma(c)$, and b_0 is determined by the probability conservation and given by $b_0 = {}_1F_1\left(\frac{\lambda}{a+\delta}; \frac{\lambda+\gamma}{a+\delta}; \frac{a\mu}{\delta(a+\delta)}\right)$, where ${}_1F_1(a; b; z)$ is a confluent hypergeometric function. Using the relationship (3), we thus obtain the analytical steady-state distribution

$$\begin{aligned} P(m) &= \frac{1}{m!} \left(\frac{\mu}{\delta} \right)^m \frac{(\frac{\lambda}{a+\delta})_m}{(\frac{\lambda+\gamma}{a+\delta})_m} \\ &\quad \times \frac{{}_1F_1\left(m + \frac{\lambda}{a+\delta}; m + \frac{\lambda+\gamma}{a+\delta}; -\frac{\mu}{a+\delta}\right)}{{}_1F_1\left(\frac{\lambda}{a+\delta}; \frac{\lambda+\gamma}{a+\delta}; \frac{a\mu}{\delta(a+\delta)}\right)}. \end{aligned} \quad (22)$$

Similarly, we can derive the analytical steady-state distribution in the case of gene auto-repression. Such analytical distributions can be further used to explore the relationship between expression noise and feedback as well as effects of the upstream promoter dynamics on the downstream transcription dynamics including bursting dynamics.^{5,40}

In contrast to the case of the above example, we can always choose $z_0 = 1$ for the examples analyzed in Sec. III B due to special structures of the corresponding reaction networks.

B. Nontrivial effects of intrinsic noise in gene expression models

In this subsection, we analyze two models of gene expression using the above moment-convergence method, showing that intrinsic noise can play a nontrivial role in controlling genetic information flows. The first model is a small system, which contains only 4 reactions. Interestingly, we show that stochastic fluctuations can induce transient bimodality and amplify signals although the steady-state

distribution is unimodal. The second model is a larger system, which contains 32 stochastic variables and 115 reactions. Interestingly, we find that slow switching between promoter states can increase fluctuations in spatially heterogeneous signals. For the first model, traditional methods such as the Gillespie stochastic simulation algorithm²³ and the finite state projection¹⁹ can well solve the corresponding MPE, but for the second model, solving the corresponding MPE using the traditional methods seems difficult and is at least very time-consuming. In contrast, our proposed moment-convergence approach can be easily used to calculate the distributions in these gene models, and the computational cost is comparatively low.

To analyze the models in a united framework using MCEs, we simply introduce a more general model of gene expression at the transcriptional level,⁴¹ which contains the above two models as its particular cases. Assume that the promoter has N states, L states of which are active (denoted by A_i) and the other $K = N - L$ states are inactive (denoted by I_j). Denote by λ_{ij} , the transition rate from state- j to state- i ($\lambda_{ij} = 0$ means that no transition occurs), which altogether constitutes a $N \times N$ matrix $\mathbf{A} = (\lambda_{ij})$ (termed as the promoter's state transition matrix). Let matrix $\mathbf{\Lambda} = \text{diag}(\mu_1, \dots, \mu_N)$ (transcription matrix) describe exits of transcription with μ_i representing the transcription rate of mRNA in state- i ($\mu_i = 0$ means that no transcription takes place). Then, these two matrices \mathbf{A} and $\mathbf{\Lambda}$ altogether characterize the promoter structure. Denote by m , the number of mRNA molecules. Let $P_k(m)$ represent the probability that mRNA has m molecules at the k -th state of promoter, thus having a column vector of probability density states $\mathbf{P} = (P_1, \dots, P_N)^T$. Denote by $\mathbf{b}_k = (b_k^{(1)}, \dots, b_k^{(N)})^T$, where each $b_k^{(i)}$ represents the CM of order k corresponding to P_i . Then, according to Eq. (6), we can write the MCEs for the full system in a straightforward way

$$\frac{d}{dt} \mathbf{b}_k = \mathbf{A} \mathbf{b}_k + \mathbf{\Lambda} \mathbf{b}_{k-1} - k \mathbf{b}_k, \quad (23)$$

where all the parameters are rescaled by δ and $k = 1, 2, \dots$. Let b_k represent the total CM, i.e., $b_k = \mathbf{u}_N \mathbf{b}_k = \sum_{i=1}^N b_k^{(i)}$, where $\mathbf{u}_N = (1, 1, \dots, 1)$ is a row vector. Thus, three terms on the right hand side of Eq. (23) well capture structure of the underlying reaction network, namely, the first term describes promoter kinetics, the second describes transcription, and the third describes degradation.

1. Stochastic fluctuations can induce transient bimodality and amplify signals

Most previous studies on stochastic effects of gene expression have been carried out only through analyzing steady-state distributions or statistical quantities due to technical difficulty.⁴² In many cases, transient dynamics of gene expression may be quite different from the stationary ones.⁴³ Here, we analyze dynamic behaviors of a gene regulatory model using our moment method (Fig. 4).

Assume that the promoter contains only one ON state and one OFF state but a TF dynamically (more precisely, in an enzyme-catalyzed manner) regulates the transition rate

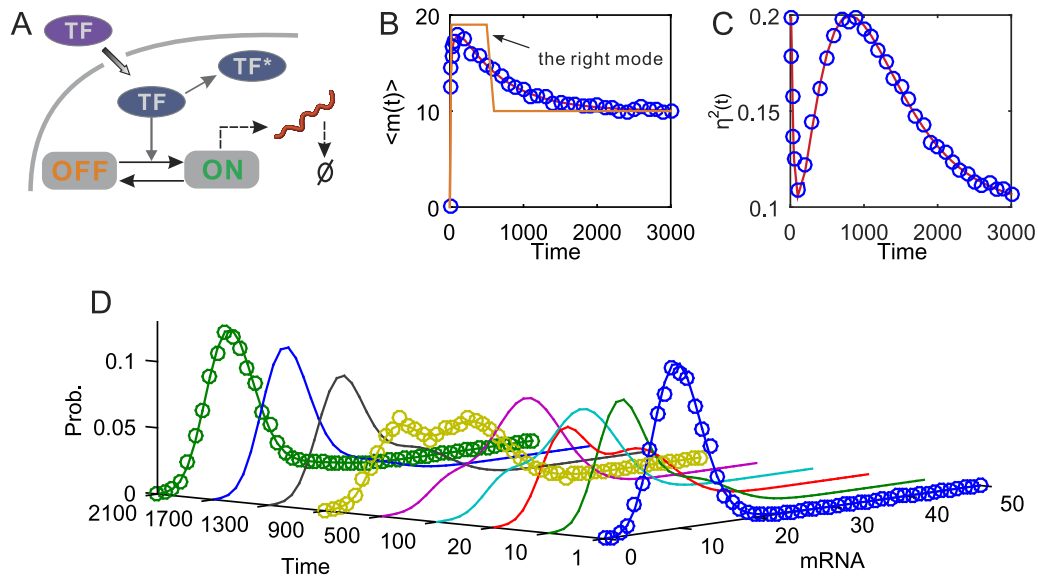


FIG. 4. Stochastic fluctuations can induce transient bimodality and amplify signals. (a) Schematic diagram for a gene model, where the promoter is assumed to have one active state and one inactive state and the transient transcription factor regulates the transition rate from OFF to ON states. The reactions are listed below: $\text{TF}_C \xrightarrow{k_t} \text{TF}_N$, $\text{TF}_C \xrightarrow{\delta_C} \emptyset$, $\text{TF}_N \xrightarrow{\delta_N} \emptyset$, $G_1 + \text{TF}_N \xrightleftharpoons[\lambda_{12}]{\lambda_{21}} G_2$, where TF_C and TF_N represent a regulator protein in the cytoplasm and in the nucleus, respectively, and $G_1 \xrightarrow{\mu_1} G_1 + P$, $G_2 \xrightarrow{\mu_2} G_2 + P$, $P \xrightarrow{\delta_P} \emptyset$; (b) the time evolution of mean expression, where the solid lines describe the mean mRNA number obtained using the distribution corresponding to the right peak along the mRNA axis; (c) the time evolution of expression noise strength; (d) transient mRNA probability distribution, exhibiting different dynamic characteristics (unimodal or bimodal). The parameter values are $k_t = 0.01$, $\delta_C = 0.01$, $\delta_N = 0.01$, $\lambda_{21} = 0.5$, $\lambda_{12} = 0.1$, $\mu_1 = 100$, $\mu_2 = 200$, and $\delta_P = 10$; The initial condition is $(\text{TF}_C, \text{TF}_N, G_1, G_2, P) = (5, 0, 1, 0, 0)$; the order of the truncated CMs is set as $k = 5$. In (b)–(d), empty circles represent the results obtained by the Gillespie algorithm, whereas lines represent the results obtained by our moment-convergence method.

from OFF to ON (Fig. 4(a)). Interestingly, we find that the steady-state distribution is unimodal (see time-evolutional probability distributions in Fig. 4(d)). However, the dynamic distribution may be either unimodal or bimodal (see the curves of moderate times in Fig. 4(d)). From this transient bimodality in which one peak corresponds to a small mRNA number, whereas the other to a large mRNA number, we see that stochastic fluctuations amplify signals (the solid lines shown in Fig. 4(b)). In addition, different from the case of steady state, the mRNA's mean level exhibits one single peak at a finite time and the noise strength demonstrates two different peaks at two different finite times (Figs. 4(b) and 4(c)). Moreover, the moment for the peak of the mean level is different from that for each of the peaks of the noise intensity (Figs. 4(b) and 4(c)). These indicate that different-order moments can exhibit different dynamic characteristics and only the first-order moment cannot capture the complete information on transient stochastic behaviors of this system. This transient bimodality induced by stochastic fluctuations would be important for the activation of signals in the downstream networks since the activation of the downstream signals depends often on the size and frequency of the upstream signals.^{43,44}

2. Slow switching between promoter states increases fluctuations in spatially heterogeneous signals

Here, we consider an application of CM to multiple compartments with interaction among them through diffusion. Transport through diffusion is important for developmental processes,⁴⁵ nanoclusters of signaling proteins,⁴⁶ and cell polarization.⁴⁷ Previously, stochastic branching diffusion

models for gene expression have been developed.⁴⁸ Here, we analyze a reaction-diffusion version of a two-state gene model (Fig. 5(a)) using the CM approach. In particular, we divide the spatial domain into $2^4 = 16$ compartments, leading to 32 stochastic variables and 115 reactions. Traditional computational methods, such as the Gillespie algorithm²³ and the finite state projection method,¹⁸ may become prohibitively expensive for such a case. Direct application of our moment method can efficiently provide the joint distribution for all the reactants and its statistics.

First, we observe that the CM approach gives the similar result as the standard Gillespie algorithm (Fig. 5(b)), and that the steady-state number of protein molecules forms a gradually decreasing gradient with respect to the position of compartments (more precisely, the further the position is, the smaller is the number of synthesized protein molecules).

Then, to investigate the effect of promoter fluctuations on diffusion of signals, we study two cases: fast and slow switching between the promoter's two states, where the ratio of $\lambda_{\text{OFF} \rightarrow \text{ON}} / \lambda_{\text{ON} \rightarrow \text{OFF}} = k_{\text{on}} / k_{\text{off}}$ remains invariant but both k_{on} and k_{off} take very large values for the fast switching whereas very small values for the slow switching. It seems difficult to deal with such a large difference in scale by existing methods (at least they are very time-consuming when handling the case of this slow switching). Our moment method shows that the protein noise intensity (Fig. 5(c)) has an inverse relationship with the spatial gradient (Fig. 5(b)). This is because the fewer proteins generated at a further position imply larger protein fluctuations at this position. In addition, the fast promoter switching (the green line in Fig. 5(b)) leads to smaller fluctuations at the locations where the signal level

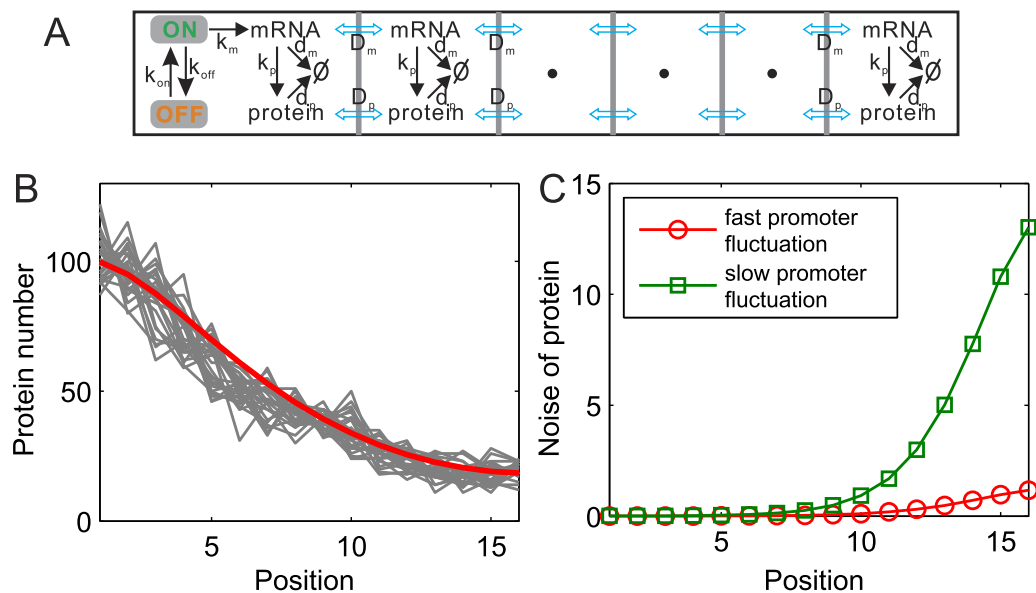


FIG. 5. Application of CM approach to spatial dynamics of gene expressions. (a) Schematic for a gene model with diffusion, where the promoter has one active state and one inactive state, and the gene products, mRNA, and protein diffuse inside a cell; (b) the mean protein number as a function of the spatial compartments shown in (a). The red line is obtained by the CM method, whereas the grey curves are obtained by the Gillespie algorithm; (c) the Fano factor as a function of compartments in space in two cases: fast promoter switching and slow promoter switching. The parameter values are $D_m = 5$, $D_p = 50$, $k_m = 0.1$, $k_p = 0.1$, $\delta_m = 0.005$, $\delta_p = 0.001$; fast promoter fluctuation: $k_{on} = 0.8$, $k_{off} = 0.2$; slow promoter fluctuation: $k_{on} = 0.008$, $k_{off} = 0.002$.

is low than the slow promoter switching (the red line in Fig. 5(b)). This is partially because at the first compartment, the slow switching results in larger fluctuations in protein than the fast switching and partially due to the particular structure of the network shown in Fig. 5(a).

IV. SUMMARY AND CONCLUSION

Biochemical networks in single cells can display large fluctuations in molecule numbers. Mesoscopic approaches, such as MPE, are often used for stochastic analysis of such systems. Here, we have introduced an approximate approach for MPE through MCEs. By using this method, one is able to analyze stochastic properties of biochemical networks in a more convenient way both analytically and computationally. Our CM method is linear and converges as the order of truncation of moment equations increases. In particular, the number of CMEs is substantially reduced compared to the MPE.

Our CMs have a clear physical meaning. For a biochemical reaction system or network, the propensity function of every reaction represents the mean probability that the reactive molecules in this reaction collide. In mathematics, this function is equal to the reaction rate times some CM. Therefore, a CM in essence represents a mean probability. In particular, a high-order CM represents the mean probability that multiple reactive molecules collide. Since this probability tends to zero with the increase of the reactive molecule number, it is not strange that CMs also tend to zero as their orders go to infinity. This would lay a physical foundation for our convergent moment theory.

The moment-convergence method introduced here has more advantages in contrast to the existing methods for stochastic analysis, which can be divided into two classes.

The first class is so-called moment-closure approximations (MCAs), e.g., normal MCAs (also called cumulant neglect MCAs),^{28,31,49–51} the Poisson MCA,⁵² the log-normal MCA,⁵³ and the maximum entropy-based MCA.²⁹ Usually, these approximation methods are obtained by setting some cumulants as zero, and we have established an analytical relationship between cumulants and convergent moments (see Eq. (16)). Therefore, the methods can be viewed as our particular cases. We point out that in general, these approximate methods are efficient only for those reaction systems either with small fluctuations or with particular structures, thus having limited applications. The other class is the direct approximations to the MPE (but they do not belong to moment-closure schemes), e.g., the finite state projection algorithm²³ and the quantized tensor train method.²² In general, however, these two methods can only deal with those reaction systems of small scale. In contrast to the above two classes of methods, our CM method in principle can be applied to stochastic analysis of any reaction networks. In particular, it has no constraints for fluctuations and network structures.

The CM method is mainly designed to work where other moment-closure approaches cannot (perhaps, yet) be applied. Nevertheless, when it is used to address any challenging problem, our method would also be computationally expensive, especially for some reaction networks of complex stochastic behavior (e.g., intricate attractors) or those with very high-order reactions. In addition, the more accurate the reconstructed distributions from CMs, the more challenging the reconstruction would become and models with many reactive species would defy serious analysis by the CM method or, indeed, by any other moment-closure approaches. Although beyond the scope of the current study, more efficient algorithms with improved or guaranteed error involving

distribution reconstruction by CMs can be developed. The CM method may then supplant the Gillespie stochastic simulation algorithm²³ for stochastic chemical simulations, especially for large reaction networks.

The CM method has been used for stochastic analysis of several representative gene regulatory networks, revealing several novel stochastic mechanisms on gene expression dynamics. In general, the CM method can be broadly applied to modeling, analysis, and simulations on biochemical reaction networks of more complex stochastic properties, such as those with bimodal, multimodal probability distributions, or the probability distributions possessing other intricate structures. For such application, our CM method, although still efficient in computing probability distribution, would encounter difficulties associated with, e.g., stiffness of the MPE due to large differences in sizes of rate constants, since in this case the corresponding CM is likely to require higher order of truncations, which leads that the errors become very large due to factorial manipulations involved. For complex networks or the peculiar reaction network analyzed above, one needs to carefully choose the parameter \mathbf{z}_0 such that the truncations of MCEs can be effectively made. We have proposed a strategy for choice of \mathbf{z}_0 by analyzing an example, but the optimal strategy on choice of \mathbf{z}_0 , although existing in theory, needs to be further explored for fast calculation.

Finally, the CM method can also be used for a systematic search for the minimal set of all possible reactions that lead to a particular dynamical property such as bistability, multimodality, oscillation, or excitability. This is because a biochemical reaction network is typically composed of fundamental reaction modules of finite number. This will be very useful in developing design principles of synthetic networks, along with many other interesting and powerful applications.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. 91230204, 91530320, and 11475273 from the Natural Science Foundation No. 2014CB964703, from Science and Technology Department No. 2012J2200017, from the New Star of Science and Technology Project of Zhujiang, Guangdong Province, P. R. China. Q.N. was supported by National Institutes of Health Grant Nos. R01GM107264, R01NS095355, and P50GM76516 and National Science Foundation Grant Nos. DMS1161621 and DMS1562176.

APPENDIX: DERIVATION OF A PARTIAL DIFFERENTIAL EQUATION FOR PROBABILITY-GENERATION FUNCTION

For the master probability equation mentioned in the main text

$$\frac{\partial P(N; t)}{\partial t} = \sum_{r \rightarrow s} (c_r^s \cdot r!) (\Phi^{-s} \Phi^r - I) \binom{N}{r} P(N; t) \quad (\text{A1})$$

we introduce the probability-generating function $G(\mathbf{z}; t)$ for the probability $P(N; t)$: $G(\mathbf{z}; t) = \sum_N \mathbf{z}^N P(N; t)$, where we denote $\mathbf{z} = (z_1, \dots, z_M)$ and define $\mathbf{z}^N \equiv z_1^{N_1} \dots z_M^{N_M}$. Differentiating $G(\mathbf{z}; t)$ with regard to t and using Eq. (A1) yields

$$\frac{\partial G(\mathbf{z}; t)}{\partial t} = \sum_{r \rightarrow s} (c_r^s \cdot r!) \sum_N \mathbf{z}^N [\Phi^{-s} \Phi^r - I] \binom{N}{r} P(N; t). \quad (\text{A2})$$

By changing the order of terms in the summation and switching the signs of the shift operators, then Eq. (A2) will become

$$\frac{\partial G(\mathbf{z}; t)}{\partial t} = \sum_{r \rightarrow s} (c_r^s \cdot r!) \sum_N \binom{N}{r} P(N; t) (\Phi^{-r} \Phi^s - I) \mathbf{z}^N. \quad (\text{A3})$$

Note that $\Phi^{-s} \Phi^r \mathbf{z}^N = \mathbf{z}^{N+r-s}$ and

$$\frac{1}{r!} \frac{\partial^r G(\mathbf{z}; t)}{\partial \mathbf{z}^r} \equiv \frac{1}{r!} \frac{\partial^{r_1 + \dots + r_K} G(\mathbf{z}; t)}{\partial z_1^{r_1} \dots \partial z_K^{r_K}} = \sum_{N \geq r} \binom{N}{r} \mathbf{z}^{N-r} P(N; t). \quad (\text{A4})$$

Thus, Eq. (A3) can be rewritten as Eq. (6) in the main text.

¹J. Gunawardena, *Science* **328**(5978), 581–582 (2010).

²J. Paulsson and M. Ehrenberg, *Q. Rev. Biophys.* **34**(1), 1–59 (2001).

³J. Elf, A. Doncic, and M. Ehrenberg, “Mesoscopic reaction diffusion in intracellular signaling,” in *Fluctuations and Noise in Biological, Biophysical and Biomedical Systems*, edited by S. Bezrukov *et al.* (SPIE, Bellingham, WA, 2003), pp. 114–124.

⁴J. Keizer, *Statistical Thermodynamics of Nonequilibrium Processes* (Springer-Verlag, Berlin, 1987).

⁵J. J. Zhang, Q. Nie, M. He, and T. S. Zhou, *J. Chem. Phys.* **138**(8), 084106 (2013).

⁶M. Ptashne, *A Genetic Switch* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2004).

⁷T. T. Marquez-Lago and J. Stelling, *Biophys. J.* **98**(9), 1742–1750 (2010).

⁸M. Hallen, B. Li, Y. Tanouchi, C. Tan, M. West, and L. You, *PLoS Comput. Biol.* **7**(10), e1002209 (2011).

⁹R. D. Astumian, R. K. Adair, and J. C. Weaver, *Nature* **388**(6643), 632–633 (1997).

¹⁰T. S. Zhou, L. N. Chen, and K. Aihara, *Phys. Rev. Lett.* **95**(17), 178103 (2005).

¹¹L. M. Bishop and H. Qian, *Biophys. J.* **98**(1), 1–11 (2010).

¹²N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 2nd ed. (Elsevier, Amsterdam, 1997).

¹³N. Friedman, L. Cai, and X. S. Xie, *Phys. Rev. Lett.* **97**(16), 168302 (2006).

¹⁴J. J. Zhang, L. N. Chen, and T. S. Zhou, *Biophys. J.* **102**(6), 1247–1257 (2012).

¹⁵V. Shahrezaei and P. S. Swain, *Proc. Natl. Acad. Sci. U. S. A.* **105**(45), 17256–17261 (2008).

¹⁶A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, *PLoS Biol.* **4**(10), e309 (2006).

¹⁷J. J. Zhang and T. S. Zhou, *SIAM J. Appl. Math.* **72**(3), 789–818 (2012).

¹⁸J. J. Zhang and T. S. Zhou, *Biophys. J.* **106**(2), 479–488 (2014).

¹⁹B. Munsky and M. Khammash, *J. Chem. Phys.* **124**(4), 044104 (2006).

²⁰Y. Cao and J. Liang, *BMC Syst. Biol.* **2**, 30 (2008).

²¹Y. Cao, M. Lu, and J. Liang, *Proc. Natl. Acad. Sci. U. S. A.* **107**(17), 18445–18450 (2010).

²²V. Kazeev, M. Khammash, M. Nip, and C. Schwab, *PLoS Comput. Biol.* **10**(3), e1003359 (2014).

²³D. T. Gillespie, *J. Comput. Phys.* **22**(4), 403–434 (1976).

²⁴M. A. Gibson and J. Bruck, *J. Phys. Chem. A* **104**(9), 1876–1889 (2000).

²⁵D. T. Gillespie, *J. Chem. Phys.* **115**(4), 1716–1733 (2001).

²⁶T. Stantcheva and E. Herbst, *Mon. Not. R. Astron. Soc.* **340**(3), 983–988 (2003).

²⁷A. Lipshtat and O. Biham, *Phys. Rev. Lett.* **93**(17), 170601 (2004).

- ²⁸A. Ale, P. Kirk, and M. P. Stumpf, *J. Chem. Phys.* **138**(17), 174101 (2013).
- ²⁹P. Smadbeck and Y. N. Kaznessis, *Proc. Natl. Acad. Sci. U. S. A.* **110**(35), 14261–14265 (2013).
- ³⁰C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koepl, *Proc. Natl. Acad. Sci. U. S. A.* **109**(21), 8340–8345 (2012).
- ³¹R. Grima, *J. Chem. Phys.* **136**(15), 154105 (2012).
- ³²M. S. Sherman and B. A. Cohen, *PLoS Comput. Biol.* **10**(5), e1003596 (2014).
- ³³B. Barzel and O. Biham, *Phys. Rev. Lett.* **106**(15), 150602 (2011).
- ³⁴B. Barzel and O. Biham, *Phys. Rev. E* **86**(3), 031126 (2012).
- ³⁵D. Schnoerr, G. Sanguinetti, and R. Grima, *J. Chem. Phys.* **143**, 185101 (2015).
- ³⁶J. Paulsson and M. Ehrenberg, *Genome Res.* **13**(11), 2475–2484 (2003).
- ³⁷J. J. Zhang, L. F. Huang, and T. S. Zhou, *Phys. Rev. Lett.* **112**(8), 088901 (2014).
- ³⁸A. J. Lotka, *J. Phys. Chem.* **14**(3), 271–274 (1910).
- ³⁹N. Balakrishnan, N. L. Johnson, and S. Kotz, *Stat. Probab. Lett.* **39**(1), 49–54 (1998).
- ⁴⁰L. F. Huang, Z. J. Yuan, P. J. Liu, and T. S. Zhou, *Phys. Rev. E* **90**(5), 052702 (2014).
- ⁴¹J. M. G. Vilar and L. Saiz, *Bioinformatics* **26**(16), 2060–2061 (2010).
- ⁴²A. Sanchez, S. Choubey, and J. Kondev, *Methods* **62**(1), 13–25 (2013).
- ⁴³S. Pelet, F. Rudolf, M. Nadal-Ribelles, E. de Nadal, F. Posas, and M. Peter, *Science* **332**(6030), 732–735 (2011).
- ⁴⁴C. H. Shi, H. X. Li, and T. S. Zhou, *Discrete Contin. Dyn. Syst. Ser. B* **17**(1), 271–282 (2012).
- ⁴⁵A. Turing, *Philos. Trans. R. Soc., B* **237**(641), 37–72 (1952).
- ⁴⁶T. Tian, A. Harding, K. Inder, S. Plowman, R. G. Parton, and J. F. Hancock, *Nat. Cell Biol.* **9**(8), 905–914 (2007).
- ⁴⁷A. S. Howell, N. S. Savage, S. A. Johnson, I. Bose, A. W. Wagner, T. R. Zyla, H. F. Nijhout, M. C. Reed, A. B. Goryachev, and D. J. Lew, *Cell* **139**(4), 731–743 (2009).
- ⁴⁸D. Cottrell, P. S. Swain, and P. F. Tupper, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 9699–9704 (2012).
- ⁴⁹L. Ferm, P. Lötstedt, and A. Hellander, *J. Sci. Comput.* **34**(2), 127–145 (2008).
- ⁵⁰M. Ullah and O. Wolkenhauer, *J. Theor. Biol.* **34**(2), 340–352 (2009).
- ⁵¹C. A. Gomez-Urbe and G. C. Verghese, *J. Chem. Phys.* **126**(2), 024109 (2007).
- ⁵²I. Nasell, *Theor. Popul. Biol.* **64**(2), 233–239 (2003).
- ⁵³M. J. Keeling, *J. Theor. Biol.* **205**(2), 269–281 (2000).