Analytical results for non-Markovian models of bursty gene expression

Zihao Wang, Zhenquan Zhang, and Tianshou Zhou*

Guangdong Province Key Laboratory of Computational Science School of Mathematics, Sun Yat-sen University, Guangzhou 510275, China

(Received 11 April 2019; revised manuscript received 5 May 2019; accepted 24 March 2020; published 22 May 2020)

Modeling stochastic gene expression has long relied on Markovian hypothesis. In recent years, however, this hypothesis is challenged by the increasing availability of time-resolved data. Correspondingly, there is considerable interest in understanding how non-Markovian reaction kinetics of gene expression impact protein variations across a population of genetically identical cells. Here, we analyze a stochastic model of gene expression with arbitrary waiting-time distributions, which includes existing gene models as its special cases. We find that stationary probabilistic behavior of this non-Markovian system is exactly the same as that of an equivalent Markovian system with the same substrates. Based on this fact, we derive analytical results, which provide insight into the roles of feedback regulation and molecular memory in controlling the protein noise and properties of the steady states, which are inaccessible via existing methodology. Our results also provide quantitative insight into diverse cellular processes involving stochastic sources of gene expression and molecular memory.

DOI: 10.1103/PhysRevE.101.052406

I. INTRODUCTION

Gene expression in single cells is inherently stochastic. This stochasticity is critical for the maintenance of cellular functions as well as for the generation of phenotypic variability among genetically identical cells. Quantifying the contributions of different sources of the gene expression noise using stochastic models [1–3] is an essential step towards understanding fundamental intracellular processes and cell-to-cell variability in gene expression.

Traditionally, modeling stochastic gene expression is based on the Markovian assumption, i.e., the stochastic motion of messenger (m)RNA or protein is uninfluenced by previous states, only by the current state. This memoryless property implies that the reaction kinetics can be described by Poissonian processes with constant rates, which are characterized by exponential waiting time distributions [4]. However, gene products (mRNA or proteins) observed in experiments are, in general, synthesized not in a single-step manner but in a multistep manner, creating a memory between individual events and leading to non-Markovian kinetics [5–8]. More generally, the dynamics of a given reactant resulting from its interactions with the environment cannot be described as a Markovian process [9]. Indeed, molecular memory (MM) in gene expression has been verified by the increasing availability of time-resolved data [10–16].

Given the above experimental facts, an important yet unsolved question is: How does MM affect stochastic gene expression? This issue was previously addressed in terms of statistics [6], but in general a random variable is best characterized by its distribution. Since the Markov theory cannot translate directly to modeling and analysis of non-Markovian processes, this leads to many significant challenges. Here we develop a technique to analyze a stochastic model of gene expression with arbitrary intrinsic event waiting-time distributions (a popular way to characterize MM in the physical field [17], but different from queuing waiting-time distributions in queue theory [6]). The key point of this technique is to introduce an effective transition rate for each reaction involved and prove that stationary probabilistic behavior of the original non-Markovian system is exactly the same as that of an equivalent Markovian system with the same substrates. As such, we derive analytical results for statistics and stationary distributions, which provide quantitative insights into diverse cellular processes involving stochastic sources of gene expression and MM.

II. MODEL DESCRIPTION

A gene model to be studied is schematically shown in Fig. 1(a), where different from the previous descriptions of Markovian models of gene expression [18–24], we adopt intrinsic event waiting-time distributions [17] to characterize reaction kinetics. First, assume that the gene promoter has one active (ON) and one inactive (OFF) states, and each mRNA degrades instantaneously after producing a protein molecule. Second, since the chromatin template accumulates over time until the promoter becomes active [25], the waiting-time distribution from OFF to ON is, in general, nonexponential and is denoted by $\psi_1(t;n)$, where n represents the number of protein molecules. Self-regulation exists if $\psi_1(t;n)$ depends on n and does not exist otherwise. The promoter switching from ON to OFF has been reported to occur with essentially a single rate-limiting step, and thus can be modeled by a constant switching rate [26], implying that the corresponding waiting-time distribution is exponential, i.e., $\psi_2(t;n) = \beta e^{-\beta t}$, where β represents the mean switching rate. Third,

*Corresponding author: mcszhtsh@mail.sysu.edu.cn

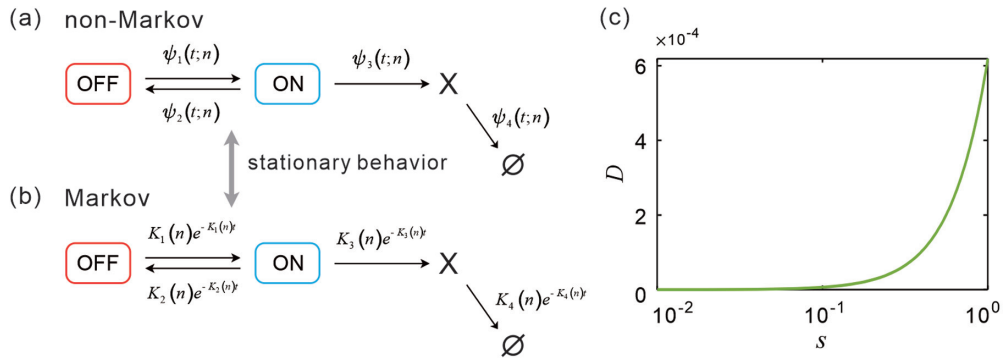

FIG. 1. This paper proves that the stationary behavior of a non-Markovian gene expression system with general waiting-time distributions $\psi_i(t;n)$ (a) is exactly the same as that of an equivalent Markovian gene expression system with exponential waiting-time distributions $K_i(n)e^{-K_i(n)t}$ (b), where $1 \leqslant i \leqslant 4$, X stands for protein and n for its molecular number, and $K_i(n)$ represents the effective transition rate (see the main text). (c) It is demonstrated that dynamic behavior of a non-Markovian gene expression system described by $DNA \xrightarrow{(\mu^k/\Gamma(k))t^{k-1}e^{-\mu t}} DNA + B \cdot X$ and $X \xrightarrow{\delta e^{-\delta t}} \emptyset$ can be well approximated by that of an equivalent Markovian gene expression system described by $DNA \xrightarrow{K_1(n)} DNA + B \cdot X$ and $X \xrightarrow{K_2(n)=n\delta e^{-n\delta t}} \emptyset$, where B represents burst size that is assumed to follow a geometric distribution: $prob\{B = i\} = b^i/(1+b)^{i+1}$, and $\Gamma(\cdot)$ is the common Gamma function. In (c), D represents the Kullback-Leibler divergence between two distributions in the sense of Laplace transform. Parameter values are set as $k = 2$, $\mu = 10$, $\delta = 1$, and $b = 2$.

assume that proteins are generated in bursts with burst size (B) following a distribution, and degrade in a linear manner with constant rate δ (implying that the corresponding waiting-time distribution is also exponential, i.e., $\psi_4(t;n) = n\delta e^{-n\delta t}$, where δ represents the mean decay rate and will be set as 1 without loss of generality). Finally, we assume that the intrinsic event waiting-time distribution for protein production, $\psi_3(t;n)$, is general. If $\psi_3(t;n)$ depends on n, this implies the occurrence of posttranscriptional or posttranslational regulation. We emphasize that the gene model described here includes almost the existing gene models [19,20,24] (e.g., the common ON-OFF model of gene expression where all the reaction rates are assumed to be constants) as its special cases. For convenience, we denote by R_i $(1 \leqslant i \leqslant 4)$ four reactions for transitions from OFF to ON and vice versa, synthesis, and degradation of proteins, respectively.

Let $P_0(n;t)$ and $P_1(n;t)$ represent the probabilities that protein has n molecules in OFF and ON states at time t, respectively. Let $M_i(t;n)$ be the memory function for R_i. According to continuous time random walk theory [27,28], the chemical master equation (CME) for the gene model depicted in Fig. 1(a) can be described as

$$\begin{aligned} s\tilde{P}_0(n;s) - P_0(n;0) &= -\tilde{M}_1(s;n)\tilde{P}_0(n;s) + \tilde{M}_2(s;n)\tilde{P}_1(n;s) + (\mathbb{E} - \mathbb{I})[\tilde{M}_4(s;n)\tilde{P}_0(n;s)], \\ s\tilde{P}_1(n;s) - P_1(n;0) &= \tilde{M}_1(s;n)\tilde{P}_0(n;s) - \tilde{M}_2(s;n)\tilde{P}_1(n;s) + \sum_{i=0}^{n} g_{n-i}\tilde{M}_3(s;i)\tilde{P}_1(i;s) - \tilde{M}_3(s;n)\tilde{P}_1(n;s) \\ &\quad + (\mathbb{E} - \mathbb{I})[\tilde{M}_4(s;n)\tilde{P}_1(n;s)], \end{aligned} \tag{1}$$

where $g_n \equiv \text{prob}\{B = n\}$ is the burst size distribution that is assumed to follow a geometric distribution given by $\text{prob}\{B = i\} = b^i/(1+b)^{i+1}$ with b representing the mean burst size, $\mathbb{E}$ is the step operator, and $\mathbb{I}$ is the unit operator. In Eq. (1), function $\tilde{M}_i(s;n)$ is the Laplace transform of memory function $M_i(t;n)$, which is defined as $M_i(t;n) = L^{-1}(s\tilde{\varphi}_i(s;n)/(1 - \sum_{k=1}^{4}\tilde{\varphi}_k(s;n)))$, where L^{-1} represents the inverse of the Laplace operator, and $\tilde{\varphi}_i(s;n)$ is the Laplace transform of function $\varphi_i(t;n)$ where $\varphi_i(t;n)dt$ represents the probability of both the ith reaction happening and the reaction waiting time being in the inetrval $[t, t+dt]$. Using the inverse of Laplace transform, we can obtain differential equations for $P_i(n;t)$ $(i = 0, 1)$, but directly solving these equations is, in general, very difficult since memory functions are implicitly expressed by waiting-time distributions [17]. Only for some special cases (e.g., Ref. [29]), can the analytical protein distributions be found.

Below, we assume that stationary protein distribution exists (in fact, numerical simulation has verified this point, referring to Fig. 3(a) and is denoted by $P(n)$. Thus, two factorial stationary distributions, $P_0(n;t)$ and $P_1(n;t)$, also exist due to $P(n) = P_0(n) + P_1(n)$.

III. MAIN RESULTS

Before presenting main results, let us define an effective transition rate for each of four reactions R_i $(1 \leqslant i \leqslant 4)$ and derive the analytical formula expressed in terms of waiting-time distributions. Recall that function $\tilde{M}_i(s;n)$, which is the Laplace transform of memory function $M_i(t;n)$, is defined according to [5,17]

$$\tilde{M}_i(s;n) = \frac{s\tilde{\varphi}_i(s;n)}{1 - \sum_{i=1}^{4}\tilde{\varphi}_i(s;n)}, \quad 1 \leqslant i \leqslant 4, \tag{2}$$

where $\tilde{\varphi}_i(s;n)$ is the Lapace transform of function $\varphi_i(t;n)$. Note that if the gene is in the OFF state, reactions R_2 and R_3 will not happen, implying that $\psi_2(t;n)=\psi_3(t;n)=0$, and if the gene is in the ON state, reaction R_1 will not happen, implying that $\psi_1(t;n)=0$. Thus, if we appropriately define $\bar{\psi}_i(t;n)$ when considering ON and OFF states, then

$$\begin{aligned}\varphi_i(t;n) &= \bar{\psi}_i(t;n)\prod_{j\neq i}\left[1-\int_0^t \bar{\psi}_j(t';n)dt'\right]\\ &= \bar{\psi}_i(t;n)\prod_{j\neq i}\int_t^\infty \bar{\psi}_j(t';n)dt'. \end{aligned} \tag{3}$$

Also note that $\int_t^\infty \bar{\psi}_j(t';n)dt' = 1-\int_0^t \bar{\psi}_j(t';n)dt'$ represents the probability that reaction R_j does not happen in the time interval $[0,t]$, so the product $\prod_{j\neq i}[1-\int_0^t \bar{\psi}_j(t';n)dt']$ represents the probability that these reactions does not occur in this interval. Since.

$$\begin{aligned}\sum_{i=1}^4 \varphi_i(t;n) &= \sum_{i=1}^4 \bar{\psi}_i(t;n)\prod_{j\neq i}\left[1-\int_0^t \bar{\psi}_j(t';n)dt'\right]\\ &= -\frac{\partial}{\partial t}\prod_{j=1}^4\int_t^\infty \bar{\psi}_j(t';n)dt',\end{aligned}$$

we have

$$\begin{aligned}&1-\sum_{i=1}^4\int_0^\infty e^{-st}\varphi_i(t;n)dt\\ &= 1+\int_0^\infty e^{-st}\left(\frac{\partial}{\partial t}\prod_{j=1}^4\int_t^\infty \bar{\psi}_j(t';n)dt'\right)dt\\ &= s\int_0^\infty e^{-st}\left(\prod_{j=1}^4\int_t^\infty \bar{\psi}_j(t';n)dt'\right)dt.\end{aligned}$$

Thus, if we define $K_i(n)=\lim_{s\to 0}\tilde{M}_i(s;n)$, then

$$\begin{aligned}K_i(n) &= \lim_{s\to 0}\frac{s\tilde{\varphi}_i(s;n)}{1-\sum_{i=1}^4\tilde{\varphi}_i(s;n)}\\ &= \lim_{s\to 0}\frac{s\int_0^\infty e^{-st}\varphi_i(t;n)dt}{s\int_0^\infty e^{-st}\left(\prod_{j=1}^4\int_t^\infty \bar{\psi}_j(t';n)dt'\right)dt}\\ &= \frac{\int_0^\infty \varphi_i(t;n)dt}{\int_0^\infty\left(\prod_{j=1}^4\int_t^\infty \bar{\psi}_j(t';n)dt'\right)dt}.\end{aligned}$$

Using Eq. (3), $K_i(n)$ is explicitly expressed as

$$K_i(n)=\frac{\int_0^{+\infty}\bar{\psi}_i(t;n)\left[\prod_{j\neq i}\int_t^\infty \bar{\psi}_j(t';n)dt'\right]dt}{\int_0^{+\infty}\left[\prod_{j=1}^4\int_t^\infty \bar{\psi}_j(t';n)dt'\right]dt},\quad 1\leqslant i\leqslant 4. \tag{4}$$

Function $K_i(n)$ will be called the effective transition rate for reaction R_i $(1\leqslant i\leqslant 4)$. Specifically, we have $\bar{\psi}_i(t;n)=\psi_i(t;n)$ $(i=1,4)$ and $\bar{\psi}_k(t;n)=0$ $(k=2,3)$ when calculating $K_1(n)$ according to $K_1(n)=\lim_{s\to 0}\tilde{M}_1(s;n)$; $\bar{\psi}_1(t;n)=0$, $\bar{\psi}_i(t;n)=\psi_i(t;n)$ $(i=2,3,4)$ when calculating $K_2(n)$ or $K_3(n)$ according to $K_k(n)=\lim_{s\to 0}\tilde{M}_k(s;n)(i=2,3)$; and $(1\leqslant i\leqslant 4)$ when calculating $K_2(n)$ according to $K_4(n)=\lim_{s\to 0}\tilde{M}_4(s;n)$.

Interestingly, if $\psi_i(t;n)$ is an exponential distribution of the form $\psi_i(t;n)=\lambda_i(n)e^{-\lambda_i(n)t}$, where $\lambda_i(n)$ should be understood as the reaction propensity function for reaction R_i, we have

$$K_i(n)=\frac{\int_0^{+\infty}\lambda_i(n)e^{-\lambda_i(n)t}\left[\prod_{j\neq i}\int_t^\infty \psi_j(t';n)dt'\right]dt}{\int_0^{+\infty}e^{-\lambda_i(n)t}\left[\prod_{j\neq i}\int_t^\infty \psi_j(t';n)dt'\right]dt}=\lambda_i(n).$$

This indicates that the effective transition rate for reaction R_i is equal to the reaction propensity function for this reaction in this case. Therefore, effective transition rates are extensions of common reaction propensity functions. If all waiting-time distributions are exponential, which corresponds to the Markovian case, Eq. (1) reduces to the common ON-OFF model in the sense of Laplace transform. We point out that the introduction of effective transition rates will be a key to analyze the behavior of the original non-Markovian gene system.

Based on $K_i(n)$, it is natural to construct an equivalent reaction network with the same substrates but without MM, referring to Fig. 1(b). For this network, the reaction process is Markovian since the corresponding waiting-time distributions are exponential: $\psi_i(t;n)=K_i(n)e^{-K_i(n)t}$. Importantly, there is a close relationship between the behaviors of this artificial system and the original non-Markovian system. First, we can show that the stationary probabilistic behaviors of these two systems are exactly the same. In fact, according to the final value theorem [30], we know $\lim_{s\to 0}s\tilde{P}_i(n;s)=P_i(n)$, where $i=0,1$. First multiplying s on both sides of Eq. (1) and then taking the limt with regard to s yield the following common stationary generalized CME:

$$\begin{aligned}&-K_1(n)P_0(n)+K_2(n)P_1(n)+(\mathbb{E}-\mathbb{I})[nP_0(n)]=0,\\ &K_1(n)P_0(n)-K_2(n)P_1(n)+(\mathbb{E}-\mathbb{I})[nP_1(n)]\\ &\quad+\sum_{i=0}^n g_{n-i}K_3(i)P_1(i)-K_3(n)P_1(n)=0,\end{aligned} \tag{5}$$

where we have used the assumption of $\delta=1$. This fact not only reveals the essential characteristic of non-Markovian reaction kinetics but also lays a solid foundation for further analyzing the effect of MM. Second, there would be differences between dynamic behaviors of the two systems. In fact, if we denote by $Q_0(n;t)$ and $Q_1(n;t)$ the probabilities that protein has n molecules in OFF and ON states at time t in the constructed gene system, respectively, we can easily write the CME for $Q_0(n;t)$ and $Q_1(n;t)$ in the sense of Laplace transform, whose form is similar to Eq. (1) if $\tilde{M}_i(s;n)$ is replaced with $K_i(n)$. Define $Q=Q_0+Q_1$, which represents the total probability. Let $D(s)$ represent the Kullback-Leibler divergence between $\tilde{Q}(n;s)$ and $\tilde{P}(n;s)$, that is,

$$D(s)=\sum_{n=0}^\infty\frac{\tilde{Q}(n;s)}{\sum_{i=0}^\infty\tilde{Q}(i;s)}\ln\frac{\tilde{Q}(n;s)/\sum_{i=0}^\infty\tilde{Q}(i;s)}{\tilde{P}(n;s)/\sum_{i=0}^\infty\tilde{P}(i;s)}.$$

Note that $t\to+\infty$ corresponds to $s\to 0$. From Fig. 1(c), we observe that $D(s)\to 0$ as $s\to 0$. This indicates that the

dynamic distribution in the Markovian case can well approximate that in the non-Markovian case after the time is long enough. Nevertheless, we only consider stationary behavior in this paper.

A. Molecular memory can adjust the strength of feedback

Since an Erlang distribution is the convolution of exponential distributions [31], it can be used to model a multistep process. In the following, we analyze two special cases to show the explicit effect of MM characterized by Erlang distribution.

Case 1: $\psi_1(t;n) = [(\lambda_1(n))^{I_1}/\Gamma(I_1)]t^{I_1-1}e^{-\lambda_1(n)t}$, $\psi_2(t;n) = \beta e^{-\beta t}$ and $\psi_3(t;n) = \mu e^{-\mu t}$, where $\lambda_1(n)$ is a feedback regulation function [if $\lambda_1(n)$ is independent of n, this implies the absence of feedback] and is set as $\lambda_1(n) = \alpha + f_1 n^{h_1}/(K_1^{h_1} + n^{h_1})$ with f_1 representing feedback strength, h_1 being the Hill coefficient and K_1 being a threshold constant. Positive constants α, β, and μ represent average switching rates from OFF to ON and vice versa, and the average transcriptional or translational rate, respectively. Note that $I_1 = 1$ corresponds to the Markovian case whereas $I_1 > 1$ to the non-Markovian case. Therefore, I_1 is called memory index. According to Eq. (4), we can show $K_2(n) = \beta$, $K_3(n) = \mu$, $K_1(n) = n(\lambda_1(n))^{I_1}/[(\xi_1(n))^{I_1} - (\lambda_1(n))^{I_1}]$, where $\xi_1(n) = n + \lambda_1(n)$ and $K_1(0) = \lambda_1(0)/I_1$. If $I_1 = 1$, then $K_1(n) = \lambda_1(n)$, indicating that feedback is kept unchanged if $f_1 \neq 0$. However, if $I_1 > 1$ then $K_1(n)$ will not be equal to $\lambda_1(n)$. Furthermore, if $f_1 = 0$ (i.e., if the original non-Markovian system has no feedback), then $K_1(n) = n(n+\alpha)^{I_1}/[(n+\alpha)^{I_1} - \alpha^{I_1}]$, implying that the effect of MM is equivalent to the introduction of negative feedback if $I_1 > 1$, e.g., $K_1(n) = \alpha^2/(n+2\alpha)$ if $I_1 = 2$. If $f_1 \neq 0$ (i.e., if the original non-Markovian system has feedback), then $K_1(n)$ is a monotonically decreasing function of n if $I_1 > 1$. Moreover, $K_1(n)$ with $I_1 > 1$ is always less than $K_1(n)$ with $I_1 = 1$, implying that MM always reduces the feedback strength. Extremely, feedback regulation disappears if I_1 tends to infinity.

Case 2: $\psi_1(t;n) = \alpha e^{-\alpha t}$, $\psi_2(t;n) = \beta e^{-\beta t}$, and $\psi_3(t;n) = [(\lambda_3(n))^{I_3}/\Gamma(I_3)]t^{I_3-1}e^{-\lambda_3(n)t}$, where I_3 is a memory index, and the meaning of $\lambda_3(n)$ is similar to that of $\lambda_1(n)$ in Case 1. Function $\lambda_3(n)$ will be set as $\lambda_3(n) = \mu + f_3 n^{h_3}/(K_3^{h_3} + n^{h_3})$ with f_3 representing feedback strength, h_3 being the Hill coefficient and K_3 being a threshold constant. In this case, we can show $K_3(n) = (\lambda_3(n))^{I_3}\frac{\xi_3(n)-\lambda_3(n)}{(\xi_3(n))^{I_3}-(\lambda_3(n))^{I_3}}$ with $\xi_3(n) = \alpha + n + \lambda_3(n)$, $K_1(n) = \alpha$, and $K_2(n) = \beta$. Similar to Case 1, MM can adjust the strength of posttranscriptional or posttranslational regulation and its effect is equivalent to the introduction of a negative feedback even if $\lambda_3(n)$ does not depend on n (or is a constant).

In addition, we can analyze the combination of the above two cases. In a word, effective transition rates introduced above not only explicitly decode the memory effect but also can give us useful information on the underlying system. MM can adjust the feedback strength as shown above, so it would not be strange that MM can induce additional dynamics.

B. Stationary protein distributions

Equation (5) is essentially an iterating system, but solving it is still difficult since $K_i(n)$ is in general a nonlinear function of n and in particular, initial values $P_i(0)$ $(i = 0, 1)$ are not known. Here, we develop a technique to derive the explicit expression of $P(n)$ from Eq. (5). Simply speaking, this technique first expresses $P_0(n)$ and $P_1(n)$ as well as $P(n)$ in terms of $P_0(0)$ and $P_1(0)$ by applying the mathematical induction, then shows $P_0(0) = CP(0)$ with C being a positive constant between 0 and 1, and finally determines C by the conservative condition of probability combined with

$$-\sum_{n=0}^{N} K_1(n)P_0(n) + \sum_{n=0}^{N} K_2(n)P_1(n) + (N+1)P_0(N+1) = 0, \quad \text{for any } N,$$

which can be derived from Eq. (5). Interestingly, we find that $P(n)$ can be formally expressed as (see Appendix A)

$$P(0) = \lim_{N\to+\infty} \frac{1}{1 + \sum_{i=1}^{N}(a_i - Cb_i)/i!}, \tag{6}$$

$$P(n) = \frac{1}{n!}[a_n P(0) - b_n P_0(0)] = \frac{1}{n!}\frac{a_n - Cb_n}{1 + \lim_{N\to+\infty}\sum_{i=1}^{N}(a_i - Cb_i)/i!}, \quad n = 1, 2, \cdots, \tag{7}$$

where C is given by

$$C = \lim_{N\to+\infty} \frac{K_2(0) + \sum_{i=1}^{N}[(a_i + c_i)K_2(i) + c_i K_1(i)]/i!}{K_1(0) + K_2(0) + \sum_{i=1}^{N}[(b_i + d_i)K_2(i) + d_i K_1(i)]/i!}. \tag{7a}$$

In Eq. (6), we have $a_0 = 1$, $b_0 = 0$ and $a_1 = b_1 = \frac{b}{1+b}K_3(0)$, and the other a_n and b_n are expressed as

$$a_n = \frac{1}{n}\sum_{k=0}^{n-1}\frac{K_3(k)}{k!}\left(\frac{b}{1+b}\right)^{n-k}\left[a_k + \sum_{i=1}^{k-1}K_2(i)a_i\prod_{j=i+1}^{k-1}(j + K_1(j) + K_2(j))\right] + \frac{K_2(0)}{n}\sum_{k=1}^{n-1}\frac{K_3(k)}{k!}\left(\frac{b}{1+b}\right)^{n-k}$$
$$\times\prod_{i=1}^{k-1}(i + K_1(i) + K_2(i)) \tag{7b}$$

and

$$b_n = \frac{1}{n}\sum_{k=0}^{n-1}\frac{K_3(k)}{k!}\left(\frac{b}{1+b}\right)^{n-k}\left[b_k + \sum_{i=1}^{k-1}K_2(i)b_i\prod_{j=i+1}^{k-1}(j+K_1(j)+K_2(j))\right] + \frac{1}{n}\sum_{k=1}^{n-1}\frac{K_3(k)}{k!}\left(\frac{b}{1+b}\right)^{n-k} \times \prod_{i=0}^{k-1}(i+K_1(i)+K_2(i)), \tag{7c}$$

implying that all a_n and b_n can be iteratively given. After having determined a_n and b_n, we can further determine c_n and d_n in Eq. (7a), according to the following relations:

$$c_n = \sum_{i=1}^{n-1}K_2(i)a_i\prod_{j=i+1}^{n-1}(j+K_1(j)+K_2(j)) + K_2(0)\prod_{i=1}^{n-1}(i+K_1(i)+K_2(i)), \tag{7d}$$

$$d_n = \sum_{i=1}^{n-1}K_2(i)b_i\prod_{j=i+1}^{n-1}(j+K_1(j)+K_2(j)) + \prod_{i=0}^{n-1}(i+K_1(i)+K_2(i)). \tag{7e}$$

We emphasize that the formulas above are exact although they look complicated.

It is worth pointing out that the above method (or see the Supplemental Material of this paper wherein the complete mathematical details are given and more numerical results are demonstrated) can be used to derive the formal expression for gene-product distribution in the common ON-OFF model of gene expression with feedback of arbitrary forms (including linear or nonlinear feedback, posttranscriptional or posttranslational regulation, etc.). This is because $K_i(n)$ $(i = 1, 2, 3)$ may be arbitrary functions of n. In addition, known distributions for gene expression models in the existing literature can be reproduced via Eq. (6). In fact, for the common ON-OFF model for which the waiting-time distributions are given by $\psi_1(t;n) = \alpha e^{-\alpha t}$, $\psi_2(t;n) = \beta e^{-\beta t}$, $\psi_3(t;n) = \mu e^{-\mu t}$, we can show $P(0) = {}_1F_1(\alpha;\alpha+\beta;-\mu)$, $P_0(0) = \frac{\beta}{\alpha+\beta}{}_1F_1(\alpha;\alpha+\beta+1;-\mu)$, and $C = \frac{\beta}{\alpha+\beta}\frac{{}_1F_1(\alpha,\alpha+\beta+1;-\mu)}{{}_1F_1(\alpha,\alpha+\beta;-\mu)}$. According to Eq. (7), we find that the stationary protein distribution is given by (see Supplemental Material [37])

$$P(n) = \frac{\mu^n}{n!}\frac{(\alpha)_n}{(\alpha+\beta)_n}{}_1F_1(\alpha+n,\alpha+\beta+n;-\mu),$$

where ${}_1F_1(a,b;z)$ is the Kummer confluent hypergeometric function. Similarly, for this gene model, if feedback is considered, e.g., if $\psi_1(t;n) = (\alpha+nf)e^{-(\alpha+nf)t}$, $\psi_2(t;n) = (\beta+ng)e^{-(\beta+ng)t}$, and $\psi_3(t;n) = \mu e^{-\mu t}$ are set, then we can show that the stationary protein distribution is given by

$$P(n) = \frac{P(0)}{(n)!}\left(\frac{\mu\xi}{\eta^2}\right)^n\frac{(a)_n}{(b)_n}{}_1F_1\left(n+a,n+b;-\frac{\mu}{\eta^2}\right),$$

where $P(0) = [{}_1F_1(a,b;-\mu/\eta^2)]^{-1}$, $\xi = 1+f$, $\eta = 1+f+g$, $a = \frac{\alpha}{\xi}$, $b = \frac{\beta}{\xi}+\frac{\mu g}{\eta^2}$. These are all known distributions for gene expression [24,32].

More importantly, based on the equations or relations above, we can establish an effective algorithm to calculate stationary protein distributions. Here we list the main steps for this algorithm below.

Step-0: Input parameter values and N (a larger positive integer, e.g., $N = 200$), and calculate $a_1 = b_1 = K_3(0)$, and $K_1(0)$, $K_2(0)$;

Step-1: Set $n = 1$;

Step-2: Calculate $K_i(n)$ $(1 \leqslant i \leqslant 3)$, a_n and b_n according to Eqs. (7b) and (7c), c_n and d_n according to Eq. (7d) and (7e);

Step-3: Update $n+1 \to n$. If $n \leqslant N$, go to Step-2 and turn to the next step elsewhere;

Step-4: Calculate C according to Eq. (7a), and $P(n)$ according to Eqs. (6) and (7), where $n = 0, 1, 2, \cdots, N$;

Step-5: Output $P(n)$.

If burst size is not considered, then by a series of matematical operations, we can show that Eqs. (7a) and (7b) reduce, respectively, to (see Appendix B for derivation)

$$a_{n+1} = \frac{K_3(n)(n-1+K_1(n-1)+K_2(n-1)+K_3(n-1))}{K_3(n-1)} \times a_n - K_3(n)(n-1+K_1(n-1))a_{n-1} \tag{8a}$$

and

$$b_{n+1} = \frac{K_3(n)(n-1+K_1(n-1)+K_2(n-1)+K_3(n-1))}{K_3(n-1)} \times b_n - K_3(n)(n-1+K_1(n-1))b_{n-1}, \tag{8b}$$

where $n \geqslant 2$. Note that $a_1 = K_3(0)$ and $a_2 = K_3(1)(K_3(0)+K_2(0))$, $b_1 = K_3(0)$ and $b_2 = K_3(1)(K_3(0)+K_2(0)+K_1(0))$. These iterarive formulas are much simpler than Eqs. (7b) and (7c).

We point out that since positive series $\sum_{n=0}^{\infty}(a_n - Cb_n)/n!$ is convergent, implying that $(a_n - Cb_n)/n!$ tends to zero as $n \to +\infty$, we can have the approximation $C \approx a_n/b_n$ as n is sufficiently large. Numerical simulation verifies that the rate a_n/b_n rapidly approaches a stable value with the increase of n, referring to Fig. 2. In other words, the value of C is easily obtained by numerical calculation. After having determined C in such a manner, we can adopt the following method to calculate the stationary distribution. First, we set $y_n = a_n - Cb_n$, where $n = 2, 3, \cdots$, $y_2 = K_3(1)(K_3(0)+K_2(0))(1-C) - CK_3(1)K_1(0)$, and $y_1 =$

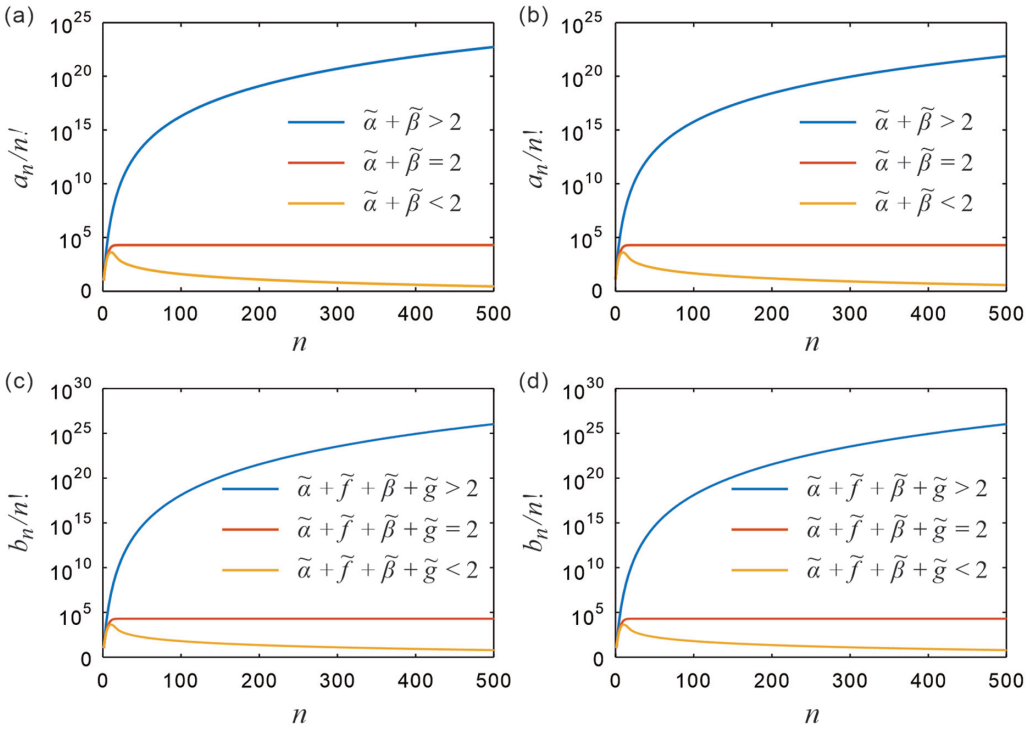

FIG. 2. Convergence of series $\sum_{n=1}^{\infty} a_n/n!$ and $\sum_{n=1}^{\infty} b_n/n!$, where bursts are not considered. (a), (b) $\tilde{K}_1(n)=\tilde{\alpha}$, $\tilde{K}_2(n)=\tilde{\beta}$, and $\tilde{K}_3(n)=\tilde{\mu}$, where $\tilde{\alpha}=1$, $\tilde{\beta}=10$, $\tilde{\mu}=10$ for $\alpha+\beta>2\delta$; $\tilde{\alpha}=1$, $\tilde{\beta}=1$, $\tilde{\mu}=10$ for $\tilde{\alpha}+\tilde{\beta}=2$; and $\tilde{\alpha}=0.1$, $\tilde{\beta}=0.5$, $\tilde{\mu}=10$ for $\alpha+\beta<2\delta$. (c), (d) $\tilde{K}_1(n)=\tilde{\alpha}+\tilde{f}\frac{n^h}{K^h+n^h}$, $\tilde{K}_2(n)=\tilde{\beta}+\tilde{g}\frac{n^h}{K^h+n^h}$, $\tilde{K}_3(n)=\tilde{\mu}$, where $\tilde{\alpha}=1$, $\tilde{\beta}=10$, $\tilde{\mu}=10$, $\tilde{f}=1.2$, $\tilde{g}=1$, $h_1=h_2=2$, $K_1=K_2=\sqrt{10}$ for $\tilde{\alpha}+\tilde{f}+\tilde{\beta}+\tilde{g}>2$; $\tilde{\alpha}=0.5$, $\tilde{\beta}=0.5$, $\tilde{\mu}=10$, $\tilde{f}=0.5$, $\tilde{g}=0.5$, $h_1=h_2=2$, $K_1=K_2=\sqrt{10}$ for $\tilde{\alpha}+\tilde{f}+\tilde{\beta}+\tilde{g}=2$; and $\tilde{\alpha}=0.1$, $\tilde{\beta}=0.1$, $\tilde{\mu}=10$, $\tilde{f}=0.2$, $\tilde{g}=0.2$, $h_1=h_2=2$, $K_1=K_2=\sqrt{10}$ for $\tilde{\alpha}+\tilde{f}+\tilde{\beta}+\tilde{g}<2$. The tilde bar represents the normalization by δ, e.g., $\tilde{K}_1(n)=K_1(n)/\delta$. Note that the barred parameters are actually the same as the corresponding unbarred parameters since $\delta=1$ has been set.

$K_3(0)(1-C)$. Then, according to Eqs. (7a) and (7b), we have

$$y_{n+1}=\frac{K_3(n)(n-1+K_1(n-1)+K_2(n-1)+K_3(n-1))}{K_3(n-1)}\times y_n-K_3(n)(n-1+K_1(n-1))y_{n-1}, \tag{9}$$

which is an iterative system, where $n=2,3,\cdots$. Note that $P(n)$ can be expressed directly by y_n due to Eq. (7).

Finally, we point out that the above method actually gives the formal expressions of stationary gene-product distributions in general gene models with feedback of arbitrary (linear or nonlinear) forms including positive or negative feedback of arbitrary forms as well as posttranscriptional or posttranslational feedback of arbitrary forms (i.e., $K_1(n)$, $K_2(n)$ and $K_3(n)$ may be arbitrary functions of n). The corresponding result itself is interesting since the distributions in models of gene expression with nonlinear feedbacks have not been derived.

C. Noise tunability

While dynamics of a stochastic variable (e.g., the protein number in our case) is best characterized by the probability mass function, statistics such as noise intensity (defined as the ratio of the variance over the squared mean) can give more intuitive characteristics of randomness. Noise tunability has been extensively studied [1,2,6–8,10,19,20], but how memory index fine tunes the protein noise remains elusive.

Although Eq. (4) gives the stationary protein distribution through which we can, in principle, calculate the mean protein level and the protein noise, one cannot clearly see how MM affects the protein expression since the resulting expressions for the statistics are rather formal. Based on Eqs. (6) and (7), however, we can derive an effective approximation for the mean protein level, which is given by

$$\langle X\rangle=b\frac{K_1(0)K_3(0)}{K_1(0)+K_2(0)}. \tag{10}$$

Apparently, this expression is exact in the absence of MM due to $K_1(0)=\alpha$, $K_2(0)=\beta$ and $K_3(0)=\mu$ in this case. However, $\langle X\rangle$ is in general a function of memory index in the presence of MM. In fact, if we consider Case 1 with $\lambda_1(n)=\alpha$, then $K_1(0)=\alpha/I_1$, $K_2(0)=\beta$ and $K_3(0)=\mu$ in this case. Therefore, $\langle X\rangle=b\mu\frac{\alpha/I_1}{\alpha/I_1+\beta}$, which is a monotonically decreasing function of I_1, implying that MM reduces the mean protein level. Similarly, for Case 2 with $\lambda_3(n)=\mu$, we have $K_1(n)=\alpha$, $K_3(0)=\frac{\alpha\mu^{I_3}}{(\alpha+\mu)^{I_3}-\mu^{I_3}}$, and $K_2(n)=\beta$. Therefore, $\langle X\rangle=\frac{\alpha b}{(1+\alpha/\mu)^{I_3}-1}$, which is a monotonically decreasing function of I_3, implying that MM also reduces the mean protein level.

Furthermore, if the protein noise is denoted by η_X, then it can be approximated as

$$\eta_X=\frac{1+b}{\langle X\rangle}+\frac{K_2(0)}{K_1(0)}\frac{1}{1+K_1(0)+K_2(0)}, \tag{11}$$

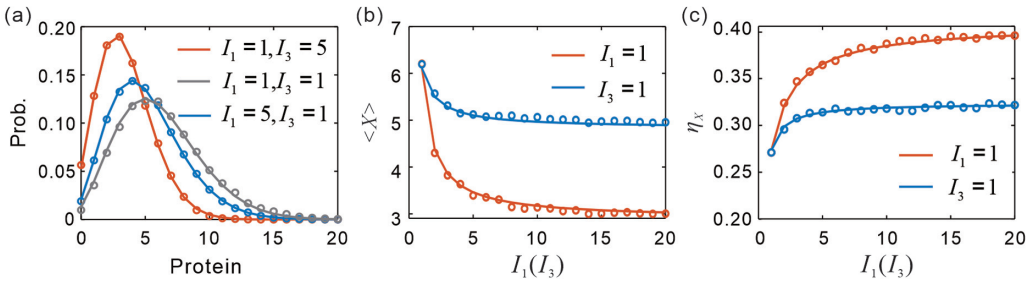

FIG. 3. Effects of molecular memory on protein expression. This is the case where burst is not considered, where solid lines represent the results obtained by theoretical prediction, whereas circles represent the results obtained by the Gillespie algorithm [34]. Parameter values are set as: $\beta = 5$, $b = 1$ (i.e., burst is not considered), and $\delta = 1$. Other parameters are determined by relations: $\alpha = 5I_1$, $\mu = 10I_3$, $f_1 = 0.2I_1$, $f_3 = 0.2I_3$. Here we set $\lambda_1(n) = \alpha + nf_1$ and $\lambda_3(n) = \mu + nf_3$.

where b represents the mean burst size. Equation (11) actually gives a decomposition formula for the protein noise, which includes two parts: the internal noise and the promoter noise [33], which correspond to the first and second terms in η_X, respectively. If we consider Case 1, then $K_1(0)$ is a monotonically decreasing function of I_1. Therefore, MM reduces the mean protein level but amplifies the protein noise if the mean protein level is fixed, referring to Figs. 3(b) and 3(c). Similarly, if $\psi_3(t;n)$ is an Erlange distribution of order I_3 and $\psi_1(t;n)$ is an exponential distribution, then I_3 reduces the mean protein level but amplifies the protein noise, also referring to Figs. 3(b) and 3(c).

D. Memory-induced bimodality

Bimodality has important biological implication [35], but whether molecular memory can induce bimodal protein distributions is unclear. Here we address this issue by modeling the time evolution of the protein concentration x as a hybrid switching ordinary differential equation, a special class of the so-called piecewise deterministic Markov process [36]. With this approximation, the time evolution of the constructed Markovian system [referring to Fig. 1(a)] is governed by the Kolmogorov forward equation (see Appendix C for derivation):

$$\begin{aligned}\frac{\partial}{\partial t}p_0(x;t) &= -K_1(x)p_0(x;t)+K_2(x)p_1(x;t)+\frac{\partial}{\partial x}[xp_0(x;t)],\\ \frac{\partial}{\partial t}p_1(x;t) &= K_1(x)p_0(x;t)-K_2(x)p_1(x;t)\\ &\quad -b\frac{\partial}{\partial x}[K_3(x)p_1(x;t)]+\frac{\partial}{\partial x}[xp_1(x;t)],\end{aligned} \tag{12}$$

where the assumption of $\delta = 1$ has been used. In Eq. (12), $K_i(x)$ $(1 \leqslant i \leqslant 3)$ and b all depend, in principle, on the system size [36], but we omit for notation simplicity. Note that the stationary probabilistic behavior of the constructed Markovian reaction network is exactly the same as that of the original non-Markovian reaction network, where effective transition rates in the former system has incorporated the effect of molecular memory in the latter system. This implies that the stationary distributions of the two systems are also the same. In the following analysis, we will therefore focus on Eq. (12) at steady state. In order to numerically solve the corresponding stationary equations, we adopt a simple yet effective approximate approach, the core idea of which is to use a piecewise deterministic equation group to the stochastic behavior of the system described by Eq. (12) (see Ref. [36] for details).

Set $p(x) = p_0(x) + p_1(x)$, which represents the total protein probability. By adding the two equations of Eq. (12) and considering the steady state, we can obtain the following relationship:

$$p_1(x)=\frac{x}{bK_3(x)}p(x). \tag{13}$$

Furthermore, we have

$$p_0(x) = p(x) - p_1(x) = \left[1 - \frac{x}{bK_3(x)}\right]p(x). \tag{14}$$

Substituting Eqs. (13) and (14) to the second equation of Eq. (12) at steady state yields

$$\begin{aligned}&\left[K_1(x) - \frac{xK_1(x)+xK_2(x)}{bK_3(x)}\right]p(x) - \frac{d}{dx}\left[xp(x)\right]\\ &\quad + \frac{1}{b}\frac{d}{dx}\left[\frac{x^2}{K_3(x)}p(x)\right] = 0,\end{aligned}$$

from which we can obtain the following expression of steady-state protein distribution:

$$\begin{aligned}p(x) = N\exp\Bigg\{\int^x &\frac{K_1(y)-1}{y} + \frac{K_2(y)-1}{y-bK_3(y)}\\ &+ \frac{y\left[\ln K_3(y)\right]'}{y-bK_3(y)}dy\Bigg\},\end{aligned} \tag{15}$$

where N is a normalized factor, and $[\ln K_3(y)]'$ is the derivative of the argument with respect to y. If $K_3(x) = \mu$, then

$$p(x) = N\exp\left\{\int^x\left[\frac{K_1(y)-1}{y} + \frac{K_2(y)-1}{y-b\mu}\right]dy\right\}. \tag{16}$$

Furthermore, if $K_1(x) = \alpha$ and $K_2(x) = \beta$, then

$$p(x) = \frac{\Gamma(\alpha+\beta)w^{1-\alpha-\beta}}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(w-x)^{\beta-1}, \tag{17}$$

where $0 < x < b\mu$ with $w = b\mu$ being the maximum protein concentration. Eq. (17) indicates that the steady-state protein number follows a β distribution.

In the following, we separately consider two special non-Markovian cases.

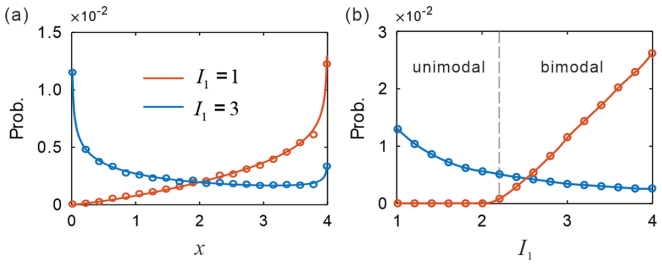

FIG. 4. (a) Molecular memory can induce bimodal protein expressions: **Case 1**, where empty circles represent the results obtained by a numerical method [36] whereas the solid curves represent the results obtained by theoretical prediction. (b) Dependence of the peaks of stationary distribution on memory index I_1, where blue symbols correspond to the left peak (i.e., the peak at $x = 0$) whereas orange symbols to the right peak [(i.e., the peak at $x = b\mu$)]. In (b), the vertical dashed line represents the boundary of unimodality and bimodality, and the solid curves conduct numerical results (since analytical results cannot be given for some I_1). In (a) and (b), we set $\lambda_1(n) = \alpha$, $\alpha = 2.2$, $\beta = 0.8$, $\mu = 2$, $b = 2$, $\delta = 1$.

(1) Assume that $K_1(x) = \frac{x\alpha^{I_1}}{(\alpha+x)^{I_1}-\alpha^{I_1}}$ with $K_1(0) = \frac{\alpha}{I_1}$ (this corresponds to the switching reaction from OFF to ON is non-Markovian), $K_2(x) = \beta$ and $K_3(x) = \mu$. In this case, the steady-state protein distribution can be bimodal but cannot be trimodal. For example, if $I_1 = 2$, then the steady-state protein distribution takes the form

$$p(x) \sim x^{\alpha/2-1}(w-x)^{\beta-1}(2\alpha+x)^{-\alpha/2}, \tag{18}$$

which is bimodal.

(2) Assume that $K_1(x) = \alpha$, $K_2(x) = \beta$ and $K_3(x) = \frac{\mu^{I_3}(\alpha+x)}{(\alpha+\mu+x)^{I_3}-\mu^{I_3}}$. In this case, the steady-state protein distribution can be bimodal. In fact, we have

$$y - bK_3(y) = \frac{H(y)}{(\alpha+\mu+y)^{I_3}-\mu^{I_3}},$$

where $H(y) = y(\alpha+\mu+y)^{I_3} - y\mu^{I_3} - b\mu^{I_3}(\alpha+y)$. Note that the algebraic equation $H(y) = 0$ has one positive root if $I_3 > 1$. Therefore, the stationary protein distribution takes the form

$$p(x) \sim x^A(x-x_1)^B(x+x_2)^C, \tag{19}$$

where A, B, and C are constants depending on the system parameters, x_1 and x_2 are positive constants.

Next, we perform numerical calculations using the numerical method mentioned above. For clarity, we only consider Case 1. In this case, numerical simulation shows that the stationary protein distribution is bimodal for an appropriate I_1, although the distribution corresponding to $I_1 = 1$ is unimodal, referring to examples shown in Fig. 4(a). Note that in the case of the occurrence of bimodality, one peak corresponds to $x = 0$ whereas the other peak to $x = b\mu$. For example, if we set $\lambda_1(x) = \alpha$ and $I_1 = 2$, then $p(x)$ takes the form $p(x) = Nx^{r-1}(w-x)^{\beta-1}(2\alpha+x)^{-r}$, which is apparently bimodal, where $r = \alpha/2$ and $0 < x < w$ with $w = b\mu$ being the maximum protein concentration.

IV. CONCLUSION AND DISCUSSION

In this paper, we have developed a stochastic gene model that integrates key features of gene expression regulation such as molecular memory, bursting, promoter switching and feedback regulation, and derived the analytical expression of the stationary protein distribution. The analytical results derived, which generalize or extend previous results obtained in the literature, provide insights into the effect of MM in inducing bimodal protein distributions and fine-tuning noise in gene expression as well as the role of feedback regulation (including posttranscriptional or posttranslational regulation) in controlling gene expression noise. Furthermore, the results obtained here can provide guidelines for synthetic biologists to design functional modules of gene expression.

We emphasize that most previously obtained results (in particular, analytical stationary gene-product distributions obtained) for ON-OFF models of gene expression (including those with or without linear or nonlinear autoregulation) can be reproduced via our method stated above. This is mainly because effective transition rates $K_i(n)$ in the Markovian reaction case reduce to common reaction propensity functions. In addition, the Supplemental Material [37] of this paper provides more analytical and numerical results, which further verify the qualitative conclusions obtained above.

It is worth noting that the effective transition rates introduced above, which explicitly decode MM, can be easily extended to other more complex biological reaction systems with MM. As such, the stationary probabilistic behavior of these non-Markovian systems can be converted to that of the corresponding Markovian systems as done above. This will open a direction for studying various non-Markovian reaction processes on networks, and would lead to discovery of new biological knowledge as shown above.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science foundation of P.R. China through Grants No. 11775314 and No. 11931019. We thank Dr. J. Zhang for useful discussions.

APPENDIX A: DERIVATION OF EQ. (7)

For Eq. (5) in the main text, summing up two equations yields

$$(n+1)P(n+1) - nP(n) = K_3(n)P_1(n) - \sum_{i=0}^{n} g_{n-i}K_3(i)P_1(i), \quad n = 0, 1, 2, \cdots.$$

Further, summing up both sides of the equation over n yields

$$P(n) = \frac{1}{n}\sum_{k=0}^{n-1} K_3(k)P_1(k) - \frac{1}{n}\sum_{k=0}^{n-1}\sum_{i=0}^{k} g_{k-i}K_3(i)P_1(i),$$

which can be rewritten as

$$P(n)=\frac{1}{n}K_3(0)P_1(0)\left(1-\sum_{i=0}^{n-1}g_i\right)+\frac{1}{n}\sum_{k=1}^{n-1}K_3(k)P_1(k)\left(1-\sum_{i=0}^{n-k-1}g_i\right), \tag{A1}$$

where $n\geqslant 1$. From the first equation of Eq. (5) in the main text in combination with $P_1(n)=P(n)-P_0(n)$, we have

$$P_0(n)=\frac{n-1+K_1(n-1)+K_2(n-1)}{n}P_0(n-1)-\frac{K_2(n-1)}{n}P(n-1). \tag{A2}$$

In the following, we will take $P(0)$ and $P_0(0)$ as parameters. Then, $P_0(n)$ can be explicitly expressed by $P(0)$ and $P_0(0)$. For this, we first establish the following lemma:

Lemma 1. If $x_n=a_nx_{n-1}+b_n$, $n=1,2,\cdots$, then $x_n=x_0\prod_{i=1}^{n}a_i+\sum_{i=1}^{n}b_i\prod_{j=i}^{n-1}a_{j+1}$, where we define $\prod_{j=n}^{n-1}a_{j+1}\equiv 1$.

Using this lemma, $P_0(n)$ can be expressed as

$$P_0(n)=\frac{1}{n!}P_0(0)\prod_{i=0}^{n-1}(i+K_1(i)+K_2(i))-\sum_{i=0}^{n-1}\frac{K_2(i)}{i+1}P(i)\prod_{j=i+1}^{n-1}\frac{j+K_1(j)+K_2(j)}{j+1}, \tag{A3}$$

where we define $\prod_{i=n}^{n-1}f(i)=1$ for function $f(i)$, and $n=1,2,\cdots$. Therefore, we have

$$P_1(n)=P(n)-P_0(n)=P(n)+\sum_{i=0}^{n-1}\tfrac{K_2(i)}{i+1}P(i)\prod_{j=i+1}^{n-1}\tfrac{j+K_1(j)+K_2(j)}{j+1}-\tfrac{1}{n!}P_0(0)\prod_{i=0}^{n-1}(i+K_1(i)+K_2(i)).$$

Simultaneously, Eq. (A1) becomes

$$\begin{aligned}P(n)=&\frac{1}{n}K_3(0)P_1(0)\left(1-\sum_{i=0}^{n-1}g_i\right)+\frac{1}{n}\sum_{k=1}^{n-1}K_3(k)P(k)\left(1-\sum_{i=0}^{n-k-1}g_i\right)+\frac{1}{n}\sum_{k=1}^{n-1}K_3(k)\sum_{i=0}^{k-1}\frac{K_2(i)}{i+1}P(i)\\&\times\prod_{j=i+1}^{k-1}\frac{j+K_1(j)+K_2(j)}{j+1}\left(1-\sum_{i=0}^{n-k-1}g_i\right)-\frac{1}{n}\sum_{k=1}^{n-1}K_3(k)\frac{1}{k!}P_0(0)\prod_{i=0}^{k-1}(i+K_1(i)+K_2(i))\left(1-\sum_{i=0}^{n-k-1}g_i\right).\end{aligned} \tag{A4}$$

Then, by the mathematical induction, we can prove Eq. (7) in the main text.

Using the expression of $P(n)$ given by Eq. (7) in the main text, $P_0(n)$ can formally be expressed as

$$P_0(n)=\frac{1}{n!}[-c_nP(0)+d_nP_0(0)], \tag{A5}$$

where $n\geqslant 1$, and

$$c_n=\sum_{i=1}^{n-1}K_2(i)a_i\prod_{j=i+1}^{n-1}(j+K_1(j)+K_2(j))+K_2(0)\prod_{i=1}^{n-1}(i+K_1(i)+K_2(i)), \tag{A5a}$$

$$d_n=\sum_{i=1}^{n-1}K_2(i)b_i\prod_{j=i+1}^{n-1}(j+K_1(j)+K_2(j))+\prod_{i=0}^{n-1}(i+K_1(i)+K_2(i)). \tag{A5b}$$

Meanwhile, $P_1(n)$ can formally be expressed as

$$P_1(n)=\frac{1}{n!}[(a_n+c_n)P(0)-(b_n+d_n)P_0(0)], \tag{A6}$$

where $n=1,2,\cdots$.

The left question is how we determine $P(0)$ and $P_0(0)$. Summing up the first equation of Eq. (5) in the main text over all n from 0 to N yields

$$-\sum_{n=0}^{N}K_1(n)P_0(n)+\sum_{n=0}^{N}K_2(n)P_1(n)+(n+1)P_0(n+1)=0.$$

Using the expressions of $P_0(n)$ and $P_1(n)$ given above and assuming $\lim_{n\to\infty}nP_0(n)=0$, we know that there is a relationship between $P_0(0)$ and $P(0)$

$$P_0(0)=CP(0), \tag{A7}$$

where

$$C = \lim_{N\to+\infty} \frac{K_2(0) + \sum_{i=1}^{N} [(a_i + c_i)K_2(i) + c_i K_1(i)]/i!}{K_1(0) + K_2(0) + \sum_{i=1}^{N} [(b_i + d_i)K_2(i) + d_i K_1(i)]/i!} \tag{A7a}$$

is a positive constant. In combination with the probability conservative condition,

$$1 = \sum_{m=0}^{\infty} P(m) = P(0) + \sum_{m=1}^{\infty} \frac{1}{m!}[a_m P(0) - b_m P_0(0)] = P(0) + P(0) \sum_{m=1}^{\infty} \frac{1}{m!}(a_m - Cb_m),$$

we can thus determine $P(0)$ and $P_0(0)$, which are formally given by

$$P(0) = \lim_{N\to+\infty} \frac{1}{1 + \sum_{i=1}^{N} (a_i - Cb_i)/i!}, \quad P_0(0) = \lim_{N\to+\infty} \frac{C}{1 + \sum_{i=1}^{N} (a_i - Cb_i)/i!}. \tag{A8}$$

To that end, the stationary probability distribution is formally expressed as

$$P(n) = \frac{1}{n!}[a_n P(0) - b_n P_0(0)] = \frac{1}{n!} \frac{a_n - Cb_n}{1 + \lim_{N\to+\infty} \sum_{i=1}^{N} (a_i - Cb_i)/i!}, \quad n = 1, 2, \cdots. \tag{A9}$$

APPENDIX B: DERIVATION OF EQS. (8A) AND (8B)

First, note that $a_1 = b_1 = K_3(0)$, $a_2 = K_3(1)(K_3(0) + K_2(0))$, and $b_2 = K_3(1)(K_3(0) + K_2(0) + K_1(0))$. Then, note that for $n \geqslant 3$, we have

$$\frac{K_2(n)a_n}{\prod_{i=1}^{n}(i + K_1(i) + K_2(i))} = \frac{K_3(n-1)K_2(n)}{(n + K_1(n) + K_2(n))K_2(n-1)} \frac{K_2(n-1)a_{n-1}}{\prod_{i=1}^{n-1}(i + K_1(i) + K_2(i))} + \frac{K_3(n-1)K_2(n)\prod_{i=1}^{n-2}(i + K_1(i) + K_2(i))}{\prod_{i=1}^{n}(i + K_1(i) + K_2(i))}$$

$$\times \sum_{i=1}^{n-2} \frac{K_2(i)a_i}{\prod_{j=1}^{i}(j + K_1(j) + K_2(j))} + \frac{K_3(n-1)K_2(0)K_2(n)\prod_{i=1}^{n-2}(i + K_1(i) + K_2(i))}{\prod_{i=1}^{n}(i + K_1(i) + K_2(i))}.$$

That is,

$$\frac{K_2(n)a_n}{\prod_{i=1}^{n}(i + K_1(i) + K_2(i))} = \frac{K_3(n-1)K_2(n)}{(n + K_1(n) + K_2(n))K_2(n-1)} \frac{K_2(n-1)a_{n-1}}{\prod_{i=1}^{n-1}(i + K_1(i) + K_2(i))}$$

$$+ \frac{K_3(n-1)K_2(n)}{(n + K_1(n) + K_2(n))(n - 1 + K_1(n-1) + K_2(n-1))} \sum_{i=1}^{n-2} \frac{K_2(i)a_i}{\prod_{j=1}^{i}(j + K_1(j) + K_2(j))}$$

$$+ \frac{K_3(n-1)K_2(n)K_2(0)}{(n + K_1(n) + K_2(n))(n - 1 + K_1(n-1) + K_2(n-1))}.$$

If we denote $A_n = \frac{K_2(n)a_n}{\sum_{i=1}^{n}(i+K_1(i)+K_2(i))}$ and $S_n = \sum_{i=1}^{n} A_n$, then

$$A_n = \frac{K_3(n-1)K_2(n)}{(n + K_1(n) + K_2(n))K_2(n-1)} A_{n-1} + \frac{K_3(n-1)K_2(n)}{(n + K_1(n) + K_2(n))(n - 1 + K_1(n-1) + K_2(n-1))} S_{n-2}$$

$$+ \frac{K_3(n-1)K_2(n)K_2(0)}{(n + K_1(n) + K_2(n))(n - 1 + K_1(n-1) + K_2(n-1))}.$$

Therefore

$$S_{n-2} = \frac{(n + K_1(n) + K_2(n))(n - 1 + K_1(n-1) + K_2(n-1))}{K_3(n-1)K_2(n)} A_n - \frac{(n - 1 + K_1(n-1) + K_2(n-1))}{K_2(n-1)} A_{n-1} - K_2(0)$$

and

$$A_{n+1} = \frac{K_3(n)K_2(n+1)K_2(0)}{(n+K_1(n)+K_2(n))(n+1+K_1(n+1)+K_2(n+1))} + \frac{K_3(n)K_2(n+1)}{(n+1+K_1(n+1)+K_2(n+1))K_2(n)}A_n + \frac{K_3(n)K_2(n+1)}{(n+K_1(n)+K_2(n))(n+1+K_1(n+1)+K_2(n+1))}(A_{n-1}+S_{n-2}).$$

Furthermore, we have

$$\begin{aligned} A_{n+1} &= \left[\frac{K_3(n)K_2(n+1)}{(n+1+K_1(n+1)+K_2(n+1))K_2(n)} + \frac{K_3(n)K_2(n+1)(n-1+K_1(n-1)+K_2(n-1))}{K_3(n-1)K_2(n)(n+1+K_1(n+1)+K_2(n+1))}\right]A_n \\ &+ \frac{K_3(n)K_2(n+1)}{(n+K_1(n)+K_2(n))(n+1+K_1(n+1)+K_2(n+1))}\left[1-\frac{(n-1+K_1(n-1)+K_2(n-1))}{K_2(n-1)}\right]A_{n-1} \\ &- \frac{K_3(n)\tilde{K}_2(n+1)K_2(0)}{(n+K_1(n)+K_2(n))(n+1+K_1(n+1)+K_2(n+1))} + \frac{K_3(n)K_2(n+1)K_2(0)}{(n+K_1(n)+K_2(n))(n+1+K_1(n+1)+K_2(n+1))} \\ &= \frac{K_3(n)K_2(n+1)(n-1+K_1(n-1)+K_2(n-1)+K_3(n-1))}{K_3(n-1)K_2(n)(n+1+K_1(n+1)+K_2(n+1))}A_n \\ &- \frac{K_3(n)K_2(n+1)(n-1+K_1(n-1))}{(n+K_1(n)+K_2(n))(n+1+K_1(n+1)+K_2(n+1))K_2(n-1)}A_{n-1}. \end{aligned}$$

Also note that

$$\begin{aligned} \frac{K_2(n+1)a_{n+1}}{\prod_{i=1}^{n+1}(i+K_1(i)+K_2(i))} &= \frac{K_3(n)K_2(n+1)(n-1+K_1(n-1)+K_2(n-1)+K_3(n-1))}{K_3(n-1)K_2(n)(n+1+K_1(n+1)+K_2(n+1))}\frac{K_2(n)a_n}{\prod_{i=1}^{n}(i+K_1(i)+K_2(i))} \\ &- \frac{K_3(n)K_2(n+1)(n-1+K_1(n-1))}{(n+K_1(n)+K_2(n))(n+1+K_1(n+1)+K_2(n+1))K_2(n-1)}\frac{K_2(n-1)a_{n-1}}{\prod_{i=1}^{n-1}(i+K_1(i)+K_2(i))}. \end{aligned}$$

That is, Eq. (8a) in the main text holds. In a similar way, we can derive Eq. (8b) in the main text.

APPENDIX C: DERIVATION OF EQ. (12)

Let Ω represent the system size. If Ω is large enough, then we can introduce continuous variable $x \approx n/\Omega$. Denote $q_i(x,t) = p_i(x\cdot\Omega,t)$, $(i=0,1)$, $\bar{K}_j(x) = K_j(x\cdot\Omega)$, $(j=1,2,3)$. Then, the CME for the constructed Markovian reaction network can be expressed as

$$\begin{aligned} \frac{\partial q_0(x,t)}{\partial t} &= -\bar{K}_1(x)q_0(x,t) + \bar{K}_2(x)p_1(x,t) + (x\cdot\Omega+1)q_0\left(x+\frac{1}{\Omega},t\right) - x\cdot\Omega q_0(x,t), \\ \frac{\partial q_1(x,t)}{\partial t} &= \bar{K}_1(x)q_0(x,t) - \bar{K}_2(x)q_1(x,t) + (x\cdot\Omega+1)q_1\left(x+\frac{1}{\Omega},t\right) - x\cdot\Omega q_1(x,t) + \sum_{i=0}^{n} g_i\bar{K}_3\left(x-\frac{i}{\Omega}\right)q_1\left(x-\frac{i}{\Omega},t\right) \\ &- \bar{K}_3(x)q_1(x,t), \end{aligned} \tag{C1}$$

where we have used the assumption of $\delta = 1$. Using the approximations of

$$f\left(x\pm\frac{i}{\Omega},t\right) \approx f(x,t) \pm \frac{i}{\Omega}\frac{\partial f(x,t)}{\partial x},$$

and neglecting higher-order small quantities as done in Ref. [36], we can have

$$\begin{aligned} \frac{\partial q_0(x,t)}{\partial t} &\approx -\bar{K}_1(x)q_0(x,t) + \bar{K}_2(x)q_1(x,t) + \frac{\partial}{\partial x}[xq_0(x,t)], \\ \frac{\partial q_1(x,t)}{\partial t} &\approx -\bar{K}_1(x)q_0(x,t) + \bar{K}_2(x)q_1(x,t) + \frac{\partial}{\partial x}[xq_1(x,t)] + \left(\sum_{i=0}^{n} g_i - 1\right)\bar{K}_3(x)q_1(x,t) - \left(\sum_{i=0}^{n} ig_i\right)\frac{1}{\Omega}\frac{\partial}{\partial x}\left[\bar{K}_3(x)q_1(x,t)\right]. \end{aligned} \tag{C2}$$

Furthermore, using approximations $\sum_{i=0}^{n} g_i \approx 1$ and $\sum_{i=0}^{n} i g_i \approx b$, which are reasonable for sufficiently large n, and normalizing b by the system size, we can immediately obtain Eq. (12) in the main text except for notations.

[1] J. M. Paulsson, Models of stochastic gene expression, Phys. Life Rev. **2**, 157 (2005).

[2] S. Azaele, J. R. Banavar, and A. Maritan, Probing noise in gene expression and protein production, Phys. Rev. E **80**, 031916 (2009).

[3] B. Munsky, B. Trinh, and M. Khammash, Listening to the noise: Random fluctuations reveal gene network parameters, Mol. Syst. Biol. **5**, 318 (2009).

[4] C. Gardiner, *Stochastic Methods-A Handbook for the Natural and Social Sciences* (Springer, New York, 2009).

[5] J.J. Zhang and T.S. Zhou, Markovian approaches to modeling intracellular reaction processes with molecular memory, Proc. Natl. Acad. Sci. U.S.A. **116**, 23542 (2019).

[6] J. M. Pedraza and J. Paulsson, Effects of molecular memory and bursting on fluctuations in gene expression, Science **319**, 339 (2008).

[7] T. Jia and R. V. Kulkarni, Intrinsic Noise in Stochastic Models of Gene Expression with Molecular Memory and Bursting, Phys. Rev. Lett. **106**, 058102 (2011).

[8] A. Schwabe, K. N. Rybakova, and F. J. Bruggeman. Transcription stochasticity of complex gene regulation models, Biophys. J. **103**, 1152 (2012).

[9] T. Guérin, N. Levernier, O. Bénichou, and R. Voituriez, Mean first-passage times of non-Markovian random walkers in confinement, Nature (London) **534**, 356 (2016).

[10] C.V. Harper, B. Finkenstadt, D.J. Woodcock, S. Friedrichsen, S. Semprini *et al.*, Dynamic analysis of stochastic transcription cycles, PLoS Bio **l9**, e1000607 (2011).

[11] M. Salathé, M. Kazandjieva, J.W. Lee, P. Levis, M.W. Feldman, and J.H. Jones, A high resolution human contact network for infectious disease transmission, Proc. Natl. Acad. Sci. USA. **107**, 22020 (2010).

[12] A. Corral, Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes, Phys. Rev. Lett. **92**, 108501 (2004).

[13] P.S. Stumpf, R.C. Smith, M. Lenz, A. Schuppert, F.J. Müller, A. Babtie *et al.*, Stem cell differentiation as a non-Markov stochastic process, Cell Syst. **5**, 268 (2017).

[14] D. M. Suter, N. Molina, and F. Naef, Mammalian genes are transcribed with widely different bursting kinetics, Science, **332**, 472 (2011).

[15] T. Guérin, O. Bénichou, and R. Voituriez, Non-Markovian polymer reaction kinetics, Nat. Chem. **4**, 568 (2012).

[16] A.L. Barabasi, The origin of bursts and heavy tails in human dynamics, Nature (London) **435**, 207 (2005).

[17] T. Aquino and M. Dentz, Chemical Continuous Time Random Walks, Phys. Rev. Lett. **119**, 230601 (2017).

[18] P. Visco, R. J. Allen, and M. R. Evans, Exact Solution of a Model DNA-Inversion Genetic Switch with Orientational Control, Phys. Rev. Lett. **101**, 118104 (2008).

[19] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, Proc. Natl. Acad. Sci. U.S.A. **105**, 17256 (2008).

[20] J. J. Zhang and T. S. Zhou, Promoter architecture-mediated transcriptional dynamics, Biophys. J. **106**, 479 (2014).

[21] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, Self-regulating gene: An exact solution, Phys. Rev. E **72**, 051907 (2005).

[22] R. Grima, D. R. Schmidt, and T. J. Newman, Steady-state fluctuations of a genetic feedback loop: An exact solution, J. Chem. Phys. **137**, 035104 (2012).

[23] Y. Vandecan and R. Blossey, Self-regulatory gene: An exact solution for the gene gate model, Phys. Rev. E **87**, 042705 (2013).

[24] N. Kumar, T. Platini, and R. V. Kulkarni, Exact Distributions for Stochastic Gene Expression Models with Bursting and Feedback, Phys. Rev. Lett. **113**, 268105 (2014).

[25] Hager G., C. Elbi, and S. John, Chromatin dynamics and the evolution of alternate promoter states, Chromosome Res. **14**, 107 (2006).

[26] N. Kumar, Singh A, and R.V. Kulkarni, Transcriptional bursting in gene expression: Analytical results for general stochastic models, PLoS Comput. Biol. **11**, e1004292 (2015).

[27] M. Silver and L. Cohen, Adiabatic bond charge model for the phonons in diamond, Si, Ge, and $\alpha-$Sn, Phys. Rev. B **15**, 3276 (1977).

[28] B. Berkowitz, A. Cortis, M. Dentz, and H. Scher, Modeling non-Fickian transport in geological formations as a continuous time random walk, Rev. Geophys. **44**, RG2003 (2006).

[29] H.H. Qiu, B.G. Zhang, and T.S. Zhou, Analytical results for a generalized model of bursty gene expression with molecular memory, Phys. Rev. E **100**, 012128 (2019).

[30] J. C. Jaeger and G. Newstead, *An Introduction to the Laplace Transformation with Engineering Applications* (Methuen, London, UK, 1969).

[31] W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 2008), Vol. 2.

[32] A. Raj, C. S. Peskin, D. Tranchina, D.Y. Vargas, and S. Tyagi, Stochastic mRNA synthesis in mammalian cells, PLoS Biol. **4**, e309 (2006).

[33] P.J. Liu, Z.J. Yuan, L.F. Huang, and T.S. Zhou, Roles of factorial noise in inducing bimodal gene expression, Phys. Rev. E **91**, 062706 (2015).

[34] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J Comput Phys **22**(4), 403 (1976).

[35] T.-L. To and N. Maheshri, Noise can induce bimodality in positive transcriptional feedback loops without bistability, Science **327**, 1142 (2010).

[36] H. Ge, H. Qian, and X. S. Xie, Stochastic Phenotype Transition of a Single Cell in an Intermediate Region of Gene-State Switching, Phys. Rev. Lett. **114**, 078101 (2015).

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.101.052406 for more mathematical details and more numerical results.