

Generalized master equation for stochastic gene expression with non-Markovian degradation

Changhong Shi ^{*}*School of Public Health, Guangzhou Medical University, Guangzhou, China*

Xiyan Yang

*School of Financial Mathematics and Statistics, Guangdong University of Finance, Guangzhou, China*Tianshou Zhou [†] and Jiajun Zhang [‡]*School of Mathematics, Sun Yat-sen University, Guangzhou, China**and Guangdong Province Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou, China*

(Received 20 August 2025; accepted 17 December 2025; published 7 January 2026)

mRNAs and proteins are believed to degrade in non-Markovian fashions, challenging the common assumption of exponential mRNA and protein lifetimes. Here, we use the queue theory to derive a generalized master equation for a three-stage model of gene expression with nonexponential mRNA and protein degradations. Analytical and numerical results show that the protein degradation memory enlarges the bimodal region and increases the protein noise, whereas the mRNA degradation memory enhances bimodality but reduces the protein noise by lowering the translational burst variance. While these results provide insight into the roles of degradation memories in shaping gene-expression patterns, the generalized master equation is a useful tool for studying gene-expression dynamics in complex cellular environments.

DOI: [10.1103/qbnx-9714](https://doi.org/10.1103/qbnx-9714)

I. INTRODUCTION

Gene expression in both prokaryotes and eukaryotes is inherently stochastic, stemming from the low copy numbers of gene products such as mRNAs and proteins, and further the discrete properties of biochemical reactions underlying transcription, translation, and degradation [1–3]. Common chemical master equations (CMEs) for gene expression, whose derivations rely on the Markovian assumption that the reaction events follow exponentially distributed waiting times, have been widely adopted to study effects of the intrinsic noise on gene-expression levels [4]. The analytical and numerical results obtained by these simplified models have provided insights into the roles of transcriptional regulation and the origins of expression variability in shaping gene-expression patterns [5–10], significantly advancing the understanding of complex gene-expression dynamics. However, gene expression often involves a cascade of intermediate biochemical steps, giving rise to non-Markovian kinetics [11–13]. Consequently, there is growing interest in developing methods that can accurately describe gene-expression systems with non-Markovian properties.

Traditional stochastic models of gene expression often assume single-step degradation for both mRNA and proteins, which imply exponential distributions of mRNA and protein lifetimes. However, recent advances in single-cell technologies have revealed that a substantial portion of mRNA and proteins in eukaryotic cells exhibit nonexponential degradation behaviors. In mammalian eukaryotes, the ubiquitin-proteasome system mediates over 80% of intracellular protein degradation, with more than 10% of proteins displaying nonexponential degradation profiles [14–16]. Similarly, various complex mechanisms have been identified for mRNA degradation. Among these, the most common are decapping pathways [17,18], as well as degradation mediated by microRNAs (miRNAs) and small RNAs [19,20]. These pathways involve multiple biochemical modifications, indicating that mRNA lifetimes are not well described by simple exponential distributions. In a word, it is strongly needed to develop biologically reasonable models to account for nonexponential degradation kinetics of both mRNA and proteins.

Several analytical and numerical approaches have been proposed to handle non-Markovian degradation kinetics [13,21–26]. For example, Xu and Jia developed a gene-expression model incorporating deterministic degradation to describe nascent RNA kinetics and derived an analytical steady-state solution [27]. Based on the queueing theory, Shi *et al.* formulated a generalized master equation (GME) for a two-stage gene-expression model involving promoter switching and transcription with nonexponential mRNA degradation [28]. Li *et al.* proposed an effective Markovian dynamics method for a three-stage gene-expression model involving promoter switching, transcription, and translation process

^{*}Contact author: ashi_248@163.com[†]Contact author: mcszhtsh@mail.sysu.edu.cn[‡]Contact author: zhjiajun@mail.sysu.edu.cn

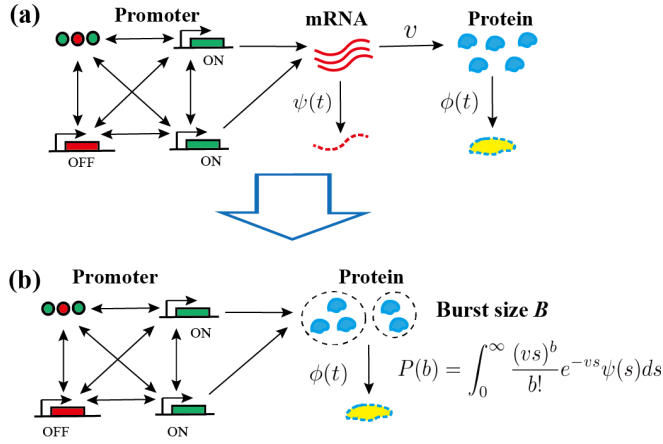


FIG. 1. Schematic for model reduction. (a) A three-stage gene expression model, which involves multistate promoter switching, transcription, translation, mRNA degradation, and protein degradation. The degradation time of both mRNA and proteins conforms to a general probability distribution: mRNA degradation time $t_m \sim \psi(t)$ and protein degradation time $t_p \sim \phi(t)$. (b) A translational bursting model, where B denotes the burst size and follows a probability distribution $P(b)$.

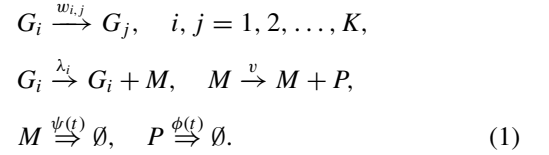
with nonexponential protein degradation [29]. We note that from the viewpoint of the queue theory, all these studied models belong to the $MMPP/G/\infty$ queue. Despite these advances, for stochastic gene expression with nonexponential mRNA and protein degradations, the corresponding CME has not been established, and an efficient method for solving it lacks, either.

In this work, we introduce a three-stage model of gene expression, which assumes that transcription initiation is a Markov process, translation occurs in bursts with a general batch-size distribution, and mRNA and protein degradation times follow nonexponential distributions. First, we map this biologically reasonable model onto a $BMMPP/G/\infty$ queue. Then, we derive a GME by applying the queue theory in combining with a translational bursting approximation. The resulting GME can be solved not only analytically in specific cases, but also numerically using the finite state projection (FSP) method in a general case, thus bringing conveniences for the statistical inference of gene-expression dynamics based on experimental data. In particular, our theoretical framework allows us to investigate how non-Markovian degradation kinetics shape gene-expression dynamics. Through systematic validation against stochastic simulations, we demonstrate that non-Markovianity (or molecular memory) in mRNA or protein degradation can affect the dynamics of gene expression.

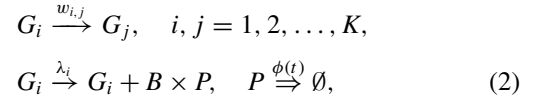
II. GENERALIZED MASTER EQUATION FOR THREE-STAGE GENE EXPRESSION

We introduce a general three-stage model of gene expression that incorporates transcription, translation, and promoter switching, as illustrated in Fig. 1(a). In this model, the gene promoter is assumed to switch among K distinct states, which can be either active or inactive. The kinetics of promoter state transitions are governed by the matrix $\mathbf{W} = [w_{ij}]$, where w_{ij} denotes the transition rate from state i to state

j , satisfying the relationship: $w_{ii} = -\sum_{j \neq i} w_{ij}$. Transcription initiation rates are specified by the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_K)$, where λ_i is the transcription initiation rate when the promoter is in state i . Each mRNA molecule subsequently produces proteins at a constant rate v . We assume that the mRNA and protein lifetimes follow general probability distributions $\psi(t)$ and $\phi(t)$, respectively. With these settings, the full reaction scheme can be written as



In the limit that the mean protein lifetime τ_p is much longer than that of mRNA τ_m , i.e., $\gamma = \tau_p/\tau_m \gg 1$, we apply the translational burst approximation to simplify the system dynamics [Fig. 1(b)]. This approximation assumes that the protein synthesis occurs in instantaneous bursts triggered by mRNA transcription events, allowing the mRNA dynamics to be effectively replaced by a distribution of burst sizes. A detailed justification for this approximation has been provided in earlier work [7]. Under these assumptions, the reaction scheme reduces to



where B is a random variable representing the number of proteins synthesized from a single mRNA before its degradation, commonly referred to as the translational burst size. According to the probability principle, we know that the probability distribution for B is given by [30]

$$h(b) = \int_0^\infty \frac{(vs)^b}{b!} e^{-vs} \psi(s) ds, \quad (3)$$

which is a mixed Poisson distribution with weight $\psi(t)$. Apparently, if $\psi(t)$ is an exponential distribution, then $h(b)$ is a geometric distribution, and this is the case often assumed in previous studies. In the case that the mRNA lifetime follows a gamma distribution, i.e., $\psi(t) = \frac{1}{\Gamma(\alpha)\theta^\alpha} t^{\alpha-1} e^{-t/\theta}$, the probability generating function (PGF) of $h(b)$ takes the form

$$H(z) = (1 + \theta v - \theta v z)^{-\alpha}. \quad (4)$$

This implies that the burst size B follows a negative binomial distribution, denoted as $B \sim \text{NB}(\alpha, 1/(1 + \theta v))$. Note that when $\alpha = 1$, $\psi(t)$ becomes an exponential distribution, and the burst size distribution reduces to a geometric distribution, an assumption frequently adopted in earlier studies [30–32].

Queuing theory has been proven as a useful tool for modeling and analyzing stochastic gene expression [25]. In our previous studies [28], the two-stage gene-expression model (which considered promoter switching and transcription) was mapped onto an $MMPP/G/\infty$ queue. Based on the framework developed by Ramaswami and Neuts [33], we derived a GME for the mRNA distribution. In the current study, we extend this framework by mapping the translational bursting model onto a $BMMPP/G/\infty$ queue. Specifically, we assume that (1) transcription initiation follows a Markov-modulated Poisson

process, (2) translation occurs in bursts with a general batch-size distribution, and (3) all protein degradation times follow the same general distribution. Under these assumptions, and using a similar method as for the *MMPP/G/∞* queue, we derive the generalized master equation for protein distribution.

Let $P_i(n, t)$ denote the probability that the number of protein is n at time t , given that the promoter is i at time 0 and the number of protein is initially 0, i.e.,

$$P_i(n, t) = \Pr\{N(t) = n \mid N(0) = 0, U(0) = i\}, \quad (5)$$

where $P_i(n, t) = \Pr\{N(t) = n \mid N(0) = 0, U(0) = i\}$, where $U(t)$ and $N(t)$ are random variables representing the promoter state and protein number at time t . Let $\mathbf{P}(n, t) = (P_1(n, t), P_2(n, t), \dots, P_K(n, t))^T$; we obtain the following generalized master equation (see Appendix A for derivation):

$$\frac{\partial}{\partial t} \mathbf{P}(n, t) = (\mathbf{W} - \mathbf{\Lambda}) \mathbf{P}(n, t) + \sum_{r=0}^n D_r \mathbf{\Lambda} \mathbf{P}(n-r, t). \quad (6)$$

The initial condition is set as

$$\mathbf{P}(n, 0) = \mathbf{0} \text{ for } n > 0, \quad \mathbf{P}(0, 0) = \mathbf{e}, \quad (7)$$

where $\mathbf{e} = (1, 1, \dots, 1)^T$ is a column vector of ones. The coefficient D_r represents the probability that exactly r surviving proteins are observed at time t since a translational burst took place, and is given by

$$D_r = \sum_{k=r}^{\infty} h(k) \binom{k}{r} [1 - \Phi(t)]^{k-r} \Phi(t)^r, \quad r = 0, 1, 2, \dots, \quad (8)$$

where $\Phi(t) = \int_t^{\infty} \phi(\tau) d\tau$ is the survival function of proteins and $h(k)$ is the probability distribution of the translational burst size B as defined in Eq. (3).

Let $\mathbf{G}(z, t) = (G_1(z, t), G_2(z, t), \dots, G_K(z, t))^T$ denote the probability generating function of $\mathbf{P}(n, t)$, i.e., $\mathbf{G}(z, t) = \sum_{n=0}^{\infty} \mathbf{P}(n, t) z^n$. Then, Eq. (6) can be transformed into (see Appendix A for derivation)

$$\mathbf{G}(z, t) = e^{\mathbf{T}t} \mathbf{e} + \int_0^t \sum_{r=0}^{\infty} D_r z^r e^{\mathbf{T}(t-u)} \mathbf{\Lambda} \mathbf{G}(z, u) du, \quad (9)$$

where $\mathbf{T} = \mathbf{W} - \mathbf{\Lambda}$, and $\mathbf{e} = (1, 1, \dots, 1)^T$. We note that Eq. (9) can also be obtained intuitively by conditioning on the time of the first translational burst. The first term corresponds to the case where no burst occurs during the interval $[0, t]$. The summation in the second term corresponds to the case where the first burst occurs at time $t - u$; in this case, the r surviving proteins from that burst are observed at time t , contributing a factor z^r to the probability generating function.

In the case where the mRNA degradation time $\psi(t)$ follows a Gamma distribution, i.e., $\psi(t) = \frac{1}{\Gamma(\alpha)\theta^\alpha} t^{\alpha-1} e^{-t/\theta}$, we have (see Appendix A for derivation)

$$D_r = \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)\Gamma(r + 1)} \frac{[\theta v \Phi(t)]^r}{[1 + \theta v \Phi(t)]^{r+\alpha}}. \quad (10)$$

When $\alpha = 1$, the mRNA lifetime becomes exponentially distributed, and Eq. (10) reduces to

$$D_r = \frac{[\langle b \rangle \Phi(t)]^r}{[1 + \langle b \rangle \Phi(t)]^{r+1}}, \quad (11)$$

where $\langle b \rangle = \theta v$ represents the mean protein burst size.

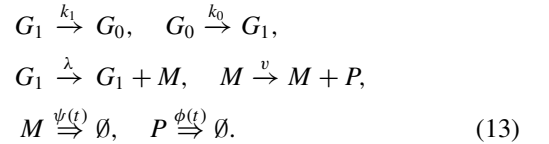
In order to help the reader understand Eq. (11), here we give an intuitive derivation. Note that each mRNA undergoes a competitive process between translation and degradation [34]. The probability that a synthesized protein survives until time t is determined by the survival function $\Phi(t)$. For each translation-degradation event, the probability that a surviving protein is produced is given by

$$P_{\text{surv}} = \frac{v \Phi(t)}{v \Phi(t) + 1/\theta} = \frac{\langle b \rangle \Phi(t)}{1 + \langle b \rangle \Phi(t)}, \quad (12)$$

where v and $1/\theta$ are the translation and degradation rates of the mRNA, respectively. Assuming that proteins have significantly longer lifetimes than mRNA, all proteins translated from a single mRNA can be considered as an instantaneous burst. The probability of producing r surviving proteins, each with a survival time longer than t , corresponds to the probability that the mRNA undergoes r translation events before degradation. This probability follows a geometric distribution, leading to Eq. (11).

III. ANALYSIS OF THE TWO-STATE MODEL OF GENE EXPRESSION

Here, we apply our GME, i.e., Eq. (6), to a two-state gene-expression model in which the promoter switches between an inactive state and an active state with forward and backward rates k_0 and k_1 , respectively. Transcription initiation only occurs in the active state with rate λ . The protein lifetime follows a general probability distribution $\phi(t)$ with corresponding survival function $\Phi(t)$, and the mRNA lifetime follows another general probability distribution $\psi(t)$. The involved biochemical reactions are as follows:



Assuming that the mean time of protein degradation τ_p is much longer than that mRNA degradation time τ_m , Eq. (6) reduces to

$$\begin{aligned} \frac{\partial}{\partial t} P_1(n, t) &= -(k_1 + \lambda) P_1(n, t) + k_1 P_0(n, t) \\ &\quad + \sum_{r=0}^n D_r \lambda P_1(n-r, t), \\ \frac{\partial}{\partial t} P_0(n, t) &= k_0 P_1(n, t) - k_0 P_0(n, t). \end{aligned} \quad (14)$$

Although deriving an exact analytical solution of Eq. (14) is generally difficult, closed-form expressions can be obtained in specific cases. For example, when the protein lifetime is deterministic ($\tau_p = \tau$), or when the promoter remains permanently active ($k_1 = 0$) (see Appendix B for details).

In a general case, the FSP method can be employed to efficiently compute numerical solution of Eq. (14) [35]. This approach allows systematic truncation of the state space while maintaining rigorous error bounds. To validate the accuracy of Eq. (14), we perform stochastic simulations for a

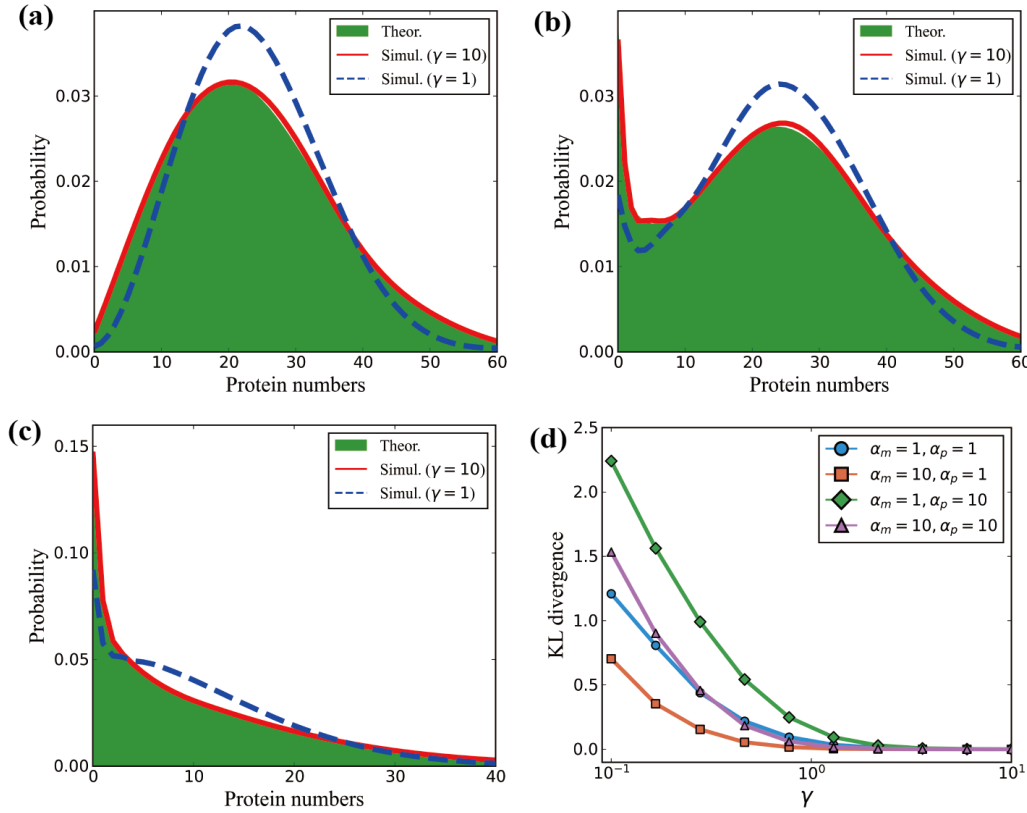


FIG. 2. (a)–(c) Three representative modes of steady-state protein distribution (green) in the translational bursting model and the simulated results (solid or dash line) of the three-stage model for large and small γ values, where $\gamma = \tau_p/\tau_m$. The parameter values are set as (a) $k_0 = 3, k_1 = 1$; (b) $k_0 = 0.9, k_1 = 0.3$; and (c) $k_0 = 1, k_1 = 2$. Degradation times of both mRNA and protein follow Gamma distribution: mRNA lifetime $t_m \sim \Gamma(\alpha_m, 1/(2\gamma))$ and protein lifetime $t_p \sim \Gamma(\alpha_p, 1/2)$, where $\alpha_m = \alpha_p = 2$. Other parameters are set as $\lambda = 8, v = 4\gamma$ (this setting implies that the mean burst size maintains a fixed constant, i.e., $\langle b \rangle = 4$). (d) The Kullback-Leibler divergence between the theoretical and simulated protein distributions is plotted as a function of γ . For $\gamma \gg 1$, the distribution becomes almost indistinguishable. The parameter values are set as $k_0 = 0.6, k_1 = 1, \lambda = 8, v = 4\gamma, t_m \sim \Gamma(\alpha_m, 1/(\alpha_m\gamma))$, and $t_p \sim \Gamma(\alpha_p, 1/\alpha_p)$.

two-state gene-expression model with different parameter sets [Figs. 2(a)–2(c)], where both mRNA and protein lifetimes are modeled as Gamma-distributed random variables. When $\gamma \gg 1$, Eq. (14) provides an excellent approximation for the protein distribution, as confirmed by close agreement with stochastic simulation of the full model. However, for small values of γ , systematic deviation emerges, reflecting the breakdown of the underlying approximations in this parameter regime. To quantify these deviations, we compute the Kullback-Leibler (KL) divergence between the theoretical distribution $P_{\text{theory}}(n)$ and the simulated protein distribution $P_{\text{sim}}(n)$, defined as [36]

$$\text{KL}(P_{\text{theory}} \parallel P_{\text{sim}}) = \sum_n P_{\text{theory}}(n) \log \frac{P_{\text{theory}}(n)}{P_{\text{sim}}(n)}. \quad (15)$$

The KL divergence exhibits a monotonic decrease as γ increases, highlighting the gradual loss of predictive power in the small- γ regime [Fig. 2(d)].

We further investigate the effect of mRNA and protein degradation memory on the stationary protein distribution in the two-state model. Specifically, we assumed that mRNA degradation times follow the Gamma distribution $\Gamma(\alpha_m, \tau_m/\alpha_m)$, and protein degradation times follow the Gamma distribution $\Gamma(\alpha_p, \tau_p/\alpha_p)$, where α_m and α_p denote the memory index of mRNA and protein degradation,

respectively, and τ_m and τ_p are the mean mRNA lifetime and protein lifetime, respectively. Under the condition of $\tau_p \gg \tau_m$, we present a k_0 – k_1 phase diagram for the two-state model in four cases: (1) $\alpha_m = 1$ and $\alpha_p = 1$ [Fig. 3(a)]; (2) $\alpha_m = 1$ and $\alpha_p = 10$ [Fig. 3(b)]; (3) $\alpha_m = 10$ and $\alpha_p = 1$ [Fig. 3(c)]; and (4) $\alpha_m = 10$ and $\alpha_p = 10$ [Fig. 3(d)]. We classify the steady-state protein distribution into three different types following Ref. [7] [Figs. 3(g)–3(i)]: unimodal distribution with zero peak (region I), unimodal distribution with nonzero peak (region II), and bimodal distribution with a zero and a nonzero peak (region III).

As is well known, bimodal distribution has been widely observed experimentally and is known to contribute to phenotypic heterogeneity in genetically identical cell populations, playing a critical role in population survival under fluctuating environmental conditions [37,38]. For the two-state gene-expression model with exponential degradation, bimodality only occurs when the promoter switches slowly between the active and inactive states [39]. Analysis of the k_0 – k_1 phase diagram reveals that increasing protein memory α_p enlarges the bimodal parameter region. In contrast, increasing the mRNA memory α_m also expands the bimodal parameter region, although the effect is less pronounced than that of α_p . Moreover, we observe from Figs. 3(e) and 3(f) that increasing either α_m or α_p can induce a transition from unimodal to bimodal

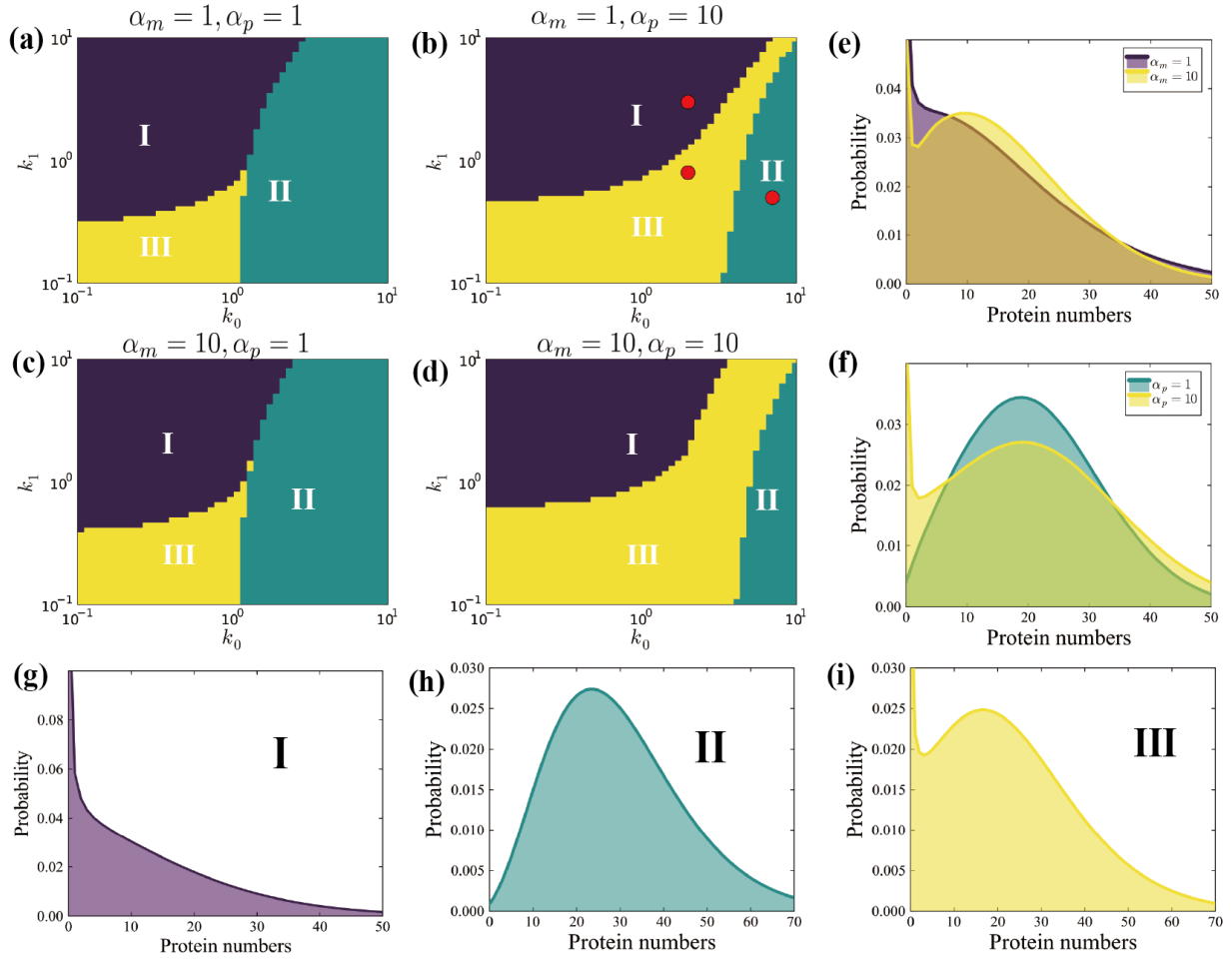


FIG. 3. Phase diagram in the k_0 - k_1 plane for the two-state gene-expression model with different memory indices for mRNA and protein degradation. (a) $\alpha_m = 1, \alpha_p = 1$; (b) $\alpha_m = 1, \alpha_p = 10$; (c) $\alpha_m = 10, \alpha_p = 1$; and (d) $\alpha_m = 10, \alpha_p = 10$. Other parameter values are set as $\lambda = 8$, $v = 4\gamma$, $t_m \sim \Gamma(\alpha_m, 1/(\alpha_m\gamma))$, $t_p \sim \Gamma(\alpha_p, 1/\alpha_p)$, and $\gamma = 10$. (e) Increasing α_m leads to the transition from unimodal to bimodal distributions, where parameter values are set as $k_0 = 3, k_1 = 3$, and $\alpha_p = 1$, and other parameter values are the same as in panel (b). (f) Increasing α_p also leads to the transition from unimodal to bimodal distribution, where parameter values are set as $k_0 = 2, k_1 = 1$, and $\alpha_m = 10$, and other parameter values are the same as in panel (b). (g)–(i) Representative protein distributions corresponding to specific points in different regions of panel (b), where parameter values are set as $k_0 = 2$ and $k_1 = 3$ for type I, $k_0 = 7$ and $k_1 = 0.5$ for type II, and $k_0 = 2$ and $k_1 = 0.8$ for type III, and other parameter values are the same as in panel (b).

distributions. This interesting result indicates that the mRNA or protein degradation memory can significantly alter the behavior of the gene-expression system, and thus may have important biological implications.

IV. BINOMIAL MOMENT OF PROTEIN DISTRIBUTION

In the previous section, we discussed how degradation memory influences the steady-state protein distribution. To investigate the protein noise dynamics further, we now introduce the binomial moment method, which is a widely used moment closure technique [40,41]. This approach allows us to express higher-order moments of the protein distribution in terms of lower-order moments and ultimately derive the mean and noise levels in the stationary state. Thus, it can provide insights into the role of degradation kinetics in shaping protein noise.

We defined the n th binomial moments of the probability $P_i(m, t)$ as

$$\mathbf{a}_n^{(i)}(t) = \sum_{m \geq n} \binom{m}{n} P_i(m, t), \quad n = 0, 1, 2, \dots \quad (16)$$

If we let $\mathbf{a}_n(t) = [a_n^{(1)}(t), \dots, a_n^{(K)}(t)]^T$ be the vector of n th binomial moments across all states, then the time evolution of $\mathbf{a}_n(t)$ is governed by the following equation group (see Appendix C for derivation):

$$\frac{d\mathbf{a}_n(t)}{dt} = \mathbf{W}\mathbf{a}_n(t) + \Lambda \left[\sum_{i=1}^n \mathbf{a}_{n-i}(t) b_i^{\text{burst}} \Phi(t)^i \right], \quad (17)$$

where b_i^{burst} denotes the i th binomial moment of the translational burst size distribution $h(b)$, defined as

$$b_i^{\text{burst}} = \sum_{b=i}^{\infty} \binom{b}{i} h(b) = \int_0^{\infty} \frac{(vs)^i}{i!} \psi(s) ds. \quad (18)$$

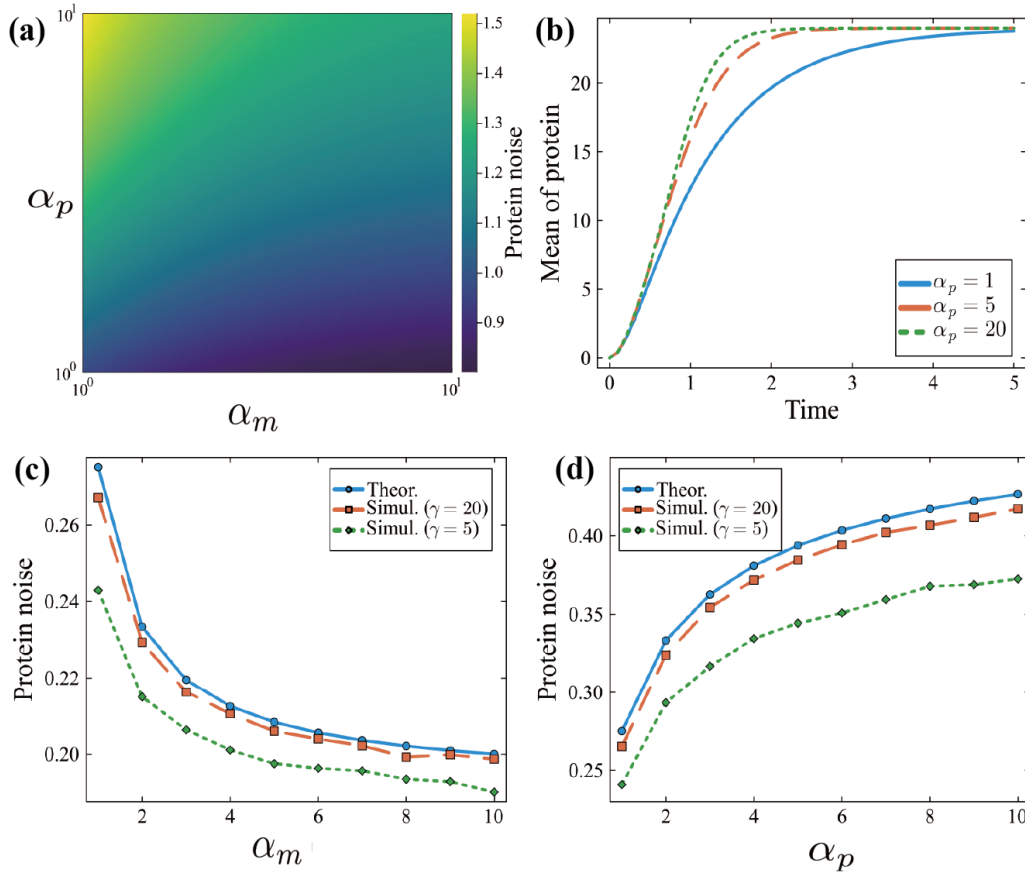


FIG. 4. Effects of degradation memories in mRNA and protein degradations on protein noise. (a) Protein noise as a function of mRNA degradation memory α_m and protein degradation memory α_p . The parameter values are set as $k_0 = 1$, $k_1 = 2$, $\lambda = 8$, $v = 40$, $t_m \sim \Gamma(\alpha_m, 1/(10\alpha_m))$, and $t_p \sim \Gamma(\alpha_p, 1/\alpha_p)$. (b) Time evolution of mean protein after transcription initiation for different memories of protein degradation, where parameter values are set as $t_m \sim \Gamma(2, 1/10)$, and other parameter values are set as those in panel (a). (c), (d) Comparison between the protein noise obtained using Eq. (20) and that obtained by stochastic simulation of the three-stage model for large and small γ values, where parameter values are set as $k_0 = 3$, $k_1 = 1$, $\alpha_p = 1$ for panel (c), $\alpha_m = 1$ for panel (d), and other parameters are the same as in panel (a).

For the above two-state gene-expression model, we can derive the analytical expression for the protein mean $\langle P \rangle$ and noise η_p^2 (defined as the ratio of the variance over the square mean) as follows (see Appendix D for derivation):

$$\langle P \rangle = \frac{k_0 \lambda \langle b \rangle}{k_0 + k_1} \int_0^\infty \Phi(u) du, \quad (19)$$

$$\begin{aligned} \eta_p^2 = \frac{\sigma_p^2}{\langle P \rangle^2} = & \frac{1}{\langle P \rangle} \\ & + \frac{k_1}{k_0 \tau_p^2} \int_0^\infty \int_0^\infty e^{-(k_0 + k_1)|u-v|} \Phi(u) \Phi(v) du dv \\ & + \frac{\langle b \rangle \eta_b^2 + \langle b \rangle - 1}{\tau_p \langle P \rangle} \int_0^\infty \Phi^2(u) du, \end{aligned} \quad (20)$$

where $\langle b \rangle = v \tau_m$ is the mean translational burst size, $\tau_p = \int_0^\infty \Phi(u) du$ is the mean protein lifetime, and $\eta_b^2 = \sigma_b^2 / \langle b \rangle^2$ is the noise strength of translational burst size, taking the form

$$\eta_b^2 = \frac{\text{Var}(b)}{\langle b \rangle^2} = \frac{\langle b \rangle + v^2 \sigma_\psi^2}{\langle b \rangle^2}, \quad (21)$$

where σ_ψ^2 is the variance of the mRNA lifetime. We note that Eq. (20) highlights different source of protein noise.

The first term represents the intrinsic noise of the protein due to birth and death events. The second term describes protein noise resulting from stochastic switching between promoter states. The last term captures noise due to translational burst kinetics. If protein degradation time follows an exponential distribution, i.e., $\Phi(u) = e^{-u/\tau_p}$, then Eq. (20) becomes

$$\eta_p^2 = \frac{1}{\langle P \rangle} + \frac{k_1}{\tau_p k_0 (1 + k_0 + k_1)} + \frac{1}{2} \frac{\langle b \rangle \eta_b^2 + \langle b \rangle - 1}{\langle P \rangle}, \quad (22)$$

which is a known result obtained in a previous study [32].

To investigate the effect of mRNA and protein degradation memory on protein noise, we model the mRNA degradation time with a Gamma distribution $\Gamma(\alpha_m, \tau_m/\alpha_m)$, and the protein degradation time with a Gamma distribution $\Gamma(\alpha_p, \tau_p/\alpha_p)$. As shown in Fig. 4(a), the protein noise exhibits opposing dependencies on the two memory indices: increasing mRNA memory α_m reduces protein noise, whereas increasing protein memory α_p amplifies it. The noise-reducing effect of α_m can be explained through its relationship with translation burst size [i.e., Eq. (20)]: Stronger mRNA memory (a larger α_m) reduces the variance of translational burst size

(B), given by $\sigma_{\psi}^2 = 1/\alpha_m$, thereby decreasing the overall protein noise. In contrast, the noise-amplifying effect of α_p can be qualitatively explained by the fact that strengthening the protein memory (a larger α_p) speeds up the protein response at any given average lifetime [Fig. 4(b)], which in turn prevents the time averaging of fluctuations, leading to the amplification of the protein noise. These theoretical predictions are validated using Gillespie simulations of the full model, as shown in Figs. 4(c) and 4(d). When protein lifetime is greater than mRNA lifetime ($\gamma \gg 1$), simulation results agree closely with theory. Notably, the dependence of protein noise on α_m and α_p persists even when γ is small.

V. CONCLUSION AND DISCUSSION

For a three-stage model of gene expression with non-Markovian kinetics of mRNA and protein degradations, we have developed a GME, which was derived by integrating queuing theory and bursty synthesis approximation, as well as by the assumption that the lifetime of protein is much greater than that of mRNA. This equation extends the classical stochastic gene-expression models since it considers arbitrary mRNA and protein degradation time distributions, which provide an effective way to simplify the modeling of complex processes associated with mRNA and protein degradations [14,17].

Previous theoretical and experimental works have investigated the roles of mRNA and protein degradation kinetics in gene expression [42,43]. Using the GME, we have demonstrated that the molecular memories in mRNA and protein degradation play a nontrivial role that cannot be ignored in gene-expression processes. For example, enhancing the protein degradation memory substantially enlarges the bimodal region of parameters and increases the protein noise by speeding up the protein response to fluctuations. In contrast, the mRNA degradation memory can also enhance bimodality but reduce the protein noise by decreasing the variance of the translational burst size. These findings are consistent with a previous result [11] wherein Pedraza and Paulsson showed that mRNA senescence (i.e., the number of Poisson steps in the degradation pathway) suppresses the protein noise, whereas the protein senescence amplifies it.

While our model focused on non-Markovian kinetics of mRNA and protein degradations, non-Markovianity in gene expression can also arise from other biochemical processes such as transcription initiation [44], promoter switching [45], and feedback regulation [46–48]. Both experimental and theoretical studies have demonstrated that such non-Markovian kinetics can significantly influence gene-expression dynamics. For example, non-Markovian promoter kinetics can alter transcriptional bursting [45], and the time delays (another form of non-Markovianity) in feedback loops can induce stochastic oscillations [49]. While the generalized master equation presented here is specific to the three-stage gene-expression model with nonexponential degradations, the proposed queuing-based framework provides a paradigm for extending the analyses here to more complex case of gene-expression regulation in future work.

A main strength of our GME lies in its flexibility to the settings of arbitrary degradation time distributions, making it applicable to various biomolecular systems with complex regulatory mechanisms. The analytical solutions in special cases (e.g., deterministic protein degradation) and numerical solutions obtained by the FSP method enhance the utility of the GME for both theoretical and experimental studies. However, the computational complexity of the GME poses a limitation, particularly for large-scale systems or high-dimensional regulatory networks [50]. The assumption that protein lifetimes greatly exceed mRNA lifetimes is well supported by both experimental and theoretical studies, particularly in eukaryotes such as yeast [7]. This disparity in timescales justifies the bursty synthesis approximation, whereby protein production occurs in instantaneous bursts following mRNA transcription—a phenomenon observed across both bacterial and eukaryotic systems [51]. Nonetheless, this assumption may not hold in cases where mRNA and protein lifetimes are comparable, such as for regulatory proteins or in fast-turnover systems [7]. Under such conditions, the concurrent modeling of both mRNA and protein fluctuations becomes essential, and the simplified approximations employed here may no longer be valid, potentially leading to deviations of the predicted distributions.

Finally, it is worth pointing out that our GME brings conveniences and even lays a theoretical foundation for statistical inference of molecular mechanisms underlying gene-expression regulations [10,52], based on experimental data such as single-cell transcriptomic data. In fact, data statistical analysis in general involves two aspects: model selection and parameter inference [53]. For the former, our model can be used in inferring hidden biological processes including reaction steps; for the latter, the common maximum likelihood estimation can be used in estimating the values of the parameters in our GME, further determining the reaction network of gene expression. The study of the corresponding statistical inference is underway.

ACKNOWLEDGMENTS

This work was supported by Grants No. 12001129, No. 12171494, No. 62373384, No. 12571551, and No. 12371483 from the Natural Science Foundation of P. R. China, by Grants No. 2019B110233002 and No. 202007030004 from Key-Area Research and Development Program of Guangzhou, P. R. China, and by Grant No. 2022ZDZX2045 from the Special Projects in Key Fields for Colleges and Universities in Guangdong Province, by Grant No. 2014A030310181 from Natural Science Foundation of Guangdong Province, China.

DATA AVAILABILITY

No data were created or analyzed in this study.

APPENDIX A: DERIVATION OF GENERALIZED MASTER EQUATION

We consider a system of gene expression with K promoter states, each producing protein in bursts with a translation

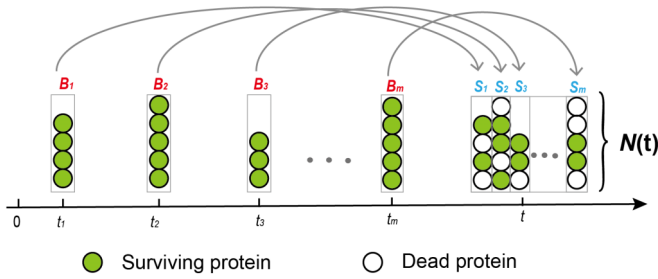


FIG. 5. Illustration of the key idea underlying the derivation of the generalized master equation. Suppose that m bursts occur at times $0 < t_1 < t_2 < \dots < t_m < t$, with the k th burst producing B_k proteins, where B_k is a random variable with PGF $H(z)$. The survival of each protein produced at time t_k but surviving until time t follows a Bernoulli distribution with success probability $\Phi(t - t_k)$. Let S_k denote the total number of surviving proteins from the k th burst, whose PGF is $H[1 + (z - 1)\Phi(t - t_k)]$. The total number of surviving proteins at time t , $N(t)$, has a PGF given by $\prod_{k=1}^m H[1 + (z - 1)\Phi(t - t_k)]$.

rate λ_i , for $i = 1, 2, \dots, K$. The burst size B follows a probability distribution $h(b)$ with probability generating function $H(z) = \mathbb{E}[z^B]$. The protein lifetime follows a generation probability density $\phi(t)$, with survival function $\Phi(t) = \int_t^\infty \phi(u) du$. Let $P_{i,j}(n, t) = \Pr\{N(t) = n, U(t) = j \mid N(0) = 0, U(0) = i\}$ denote the probability that the number of proteins is n and the promoter state is j at time t , given that the promoter state is in state i at $t = 0$ with no initial proteins, where $U(t)$ and $N(t)$ are random variables representing the promoter state and protein number at time t , respectively. Define the probability matrix $\tilde{\mathbf{P}}(n, t) = [P_{i,j}(n, t)]$ and the corresponding matrix PGF

$$\tilde{\mathbf{G}}(z, t) = \sum_{n=0}^{\infty} \tilde{\mathbf{P}}(n, t) z^n. \quad (\text{A1})$$

The promoter state transition is governed by the rate matrix $\mathbf{W} = [w_{ij}]$, where w_{ij} denotes the transition rate from state i to state j , and $w_{ii} = -\sum_{j \neq i} w_{ij}$ are the diagonal elements of the matrix. The burst rates form the diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_K)$.

The key idea underlying the derivation is to condition on the random times of translational bursts (Fig. 5). Given these times, the proteins generated in each burst evolve independently, and their survival at a given time follows a Bernoulli distribution with a success probability given by $\Phi(t)$. Consequently, the total number of surviving proteins at time t can be expressed as a sum of independent Bernoulli random variables, from which the generating function can be obtained. By averaging over all possible realizations of burst arrival sequences, the generalized master equation is finally derived.

The derivation of the generalized master equation includes four main steps.

1. Step 1: PGF of protein number when no burst occurs

Let $M(t)$ denote the number of burst events in $[0, t]$, and $N(t)$ the number of proteins observed at time

t . When $M(t) = 0$, no burst occurs. Therefore, $N(t) = 0$, and the conditional probability is (see Appendix A of [28] for derivation)

$$\Pr\{N(t) = 0, U(t) = j \mid U(0) = i, N(0) = 0\} = [e^{\mathbf{T}t}]_{i,j}, \quad (\text{A2})$$

where $\mathbf{T} = \mathbf{W} - \mathbf{\Lambda}$.

2. Step 2: PGF of protein number when multiple bursts occur

For $M(t) = m > 0$, assume that m bursts occur at times $0 < t_1 < t_2 < \dots < t_m < t$, where the k th burst produces B_k proteins. Each B_k is a random variable with probability generating function $H(z)$.

Let $I_{i,k}(t)$ denote the survival state of the i th protein from the k th burst. The variable $I_{i,k}(t)$ follows a Bernoulli distribution with success probability $\Phi(t - t_k)$, i.e.,

$$I_{i,k}(t) = \begin{cases} 1, & \text{with probability } \Phi(t - t_k), \\ 0, & \text{with probability } 1 - \Phi(t - t_k), \end{cases} \quad (\text{A3})$$

where $\Phi(t - t_k)$ denotes the survival probability of a protein from t_k to t . The corresponding PGF of $I_{i,k}(t)$ is $1 + (z - 1)\Phi(t - t_k)$.

Let $S_k(t)$ denote the total number of surviving proteins from the k th burst at time t . Then,

$$S_k(t) = \sum_{i=1}^{b_k} I_{i,k}(t). \quad (\text{A4})$$

Using the law of total expectation, the PGF is given by

$$\mathbb{E}[z^{S_k(t)}] = H[1 + (z - 1)\Phi(t - t_k)]. \quad (\text{A5})$$

Since the total number of surviving proteins at time t is

$$N(t) = S_1(t) + S_2(t) + \dots + S_m(t) \quad (\text{A6})$$

the conditional PGF of $N(t)$ given $M(t) = m$ and the burst times $\{t_1, t_2, \dots, t_m\}$ is

$$\begin{aligned} \mathbb{E}[z^{N(t)} \mid M(t) = m, t_1, t_2, \dots, t_m] \\ = \prod_{k=1}^m H[1 + (z - 1)\Phi(t - t_k)]. \end{aligned} \quad (\text{A7})$$

3. Step 3: Averaging PGF over all possible realizations of the burst arrival sequence

To obtain the full conditional expectation for $M(t) > 0$, accounting for promoter transitions, we integrate over all possible burst time sequences. This yields

$$\begin{aligned} \mathbb{E}[z^{N(t)} I\{U(t) = j, M(t) > 0\} \mid U(0) = i] \\ = \sum_{m=1}^{\infty} \int_0^t dt_1 \int_{t_1}^t dt_2 \dots \int_{t_{m-1}}^t dt_m \\ \times \prod_{k=1}^m H[1 + (z - 1)\Phi(t - t_k)] \\ \times [e^{\mathbf{T}t_1} \mathbf{\Lambda} e^{\mathbf{T}(t_2 - t_1)} \mathbf{\Lambda} \dots e^{\mathbf{T}(t_m - t_{m-1})} \mathbf{\Lambda} e^{\mathbf{T}(t - t_m)}]_{i,j}. \end{aligned} \quad (\text{A8})$$

Combining Eqs. (A2) and (A8), we have

$$\begin{aligned}\tilde{\mathbf{G}}(z, t) &= e^{\mathbf{T}t} + \sum_{m=1}^{\infty} \int_0^t dt_1 \int_{t_1}^t dt_2 \dots \\ &\times \int_{t_{m-1}}^t dt_m \prod_{k=1}^m [H[1 + (z-1)\Phi(t-t_k)]] \\ &\times [e^{\mathbf{T}t_1} \mathbf{\Lambda} e^{\mathbf{T}(t_2-t_1)} \mathbf{\Lambda} \dots e^{\mathbf{T}(t_m-t_{m-1})} \mathbf{\Lambda} e^{\mathbf{T}(t-t_m)}].\end{aligned}\quad (\text{A9})$$

Next, we introduce the substitution $v_j = t - t_{m+1-j}$ for $j = 1, 2, \dots, m$ and $v_0 = 0$, and define

$$\Psi(v_j) = H[1 + (z-1)\Phi(v_j)] \mathbf{\Lambda} e^{\mathbf{T}(v_j-v_{j-1})}. \quad (\text{A10})$$

Then, Eq. (A9) becomes

$$\begin{aligned}\tilde{\mathbf{G}}(z, t) &= e^{\mathbf{T}t} \sum_{m=1}^{\infty} \int_0^t dv_m \int_0^{v_m} dv_{m-1} \dots \\ &\times \int_0^{v_2} dv_1 e^{\mathbf{T}(t-v_m)} \Psi(v_m) \Psi(v_{m-1}) \dots \Psi(v_1).\end{aligned}\quad (\text{A11})$$

Equation (A11) can be rewritten as

$$\tilde{\mathbf{G}}(z, t) = e^{\mathbf{T}t} + \int_0^t H[1 + (z-1)\Phi(u)] e^{\mathbf{T}(t-u)} \mathbf{\Lambda} \tilde{\mathbf{G}}(z, u) du. \quad (\text{A12})$$

Expand $H[1 + (z-1)\Phi(t)]$ as follows:

$$\begin{aligned}H[1 + \Phi(t)(z-1)] &= \sum_{k=0}^{\infty} h(k) [1 - \Phi(t) + \Phi(t)z]^k \\ &= \sum_{r=0}^{\infty} D_r z^r,\end{aligned}\quad (\text{A13})$$

where

$$D_r = \sum_{k=r}^{\infty} h(k) \binom{k}{r} [1 - \Phi(t)]^{k-r} \Phi(t)^r, \quad r = 0, 1, 2, \dots \quad (\text{A14})$$

Substituting Eq. (A14) into Eq. (A12) yields

$$\tilde{\mathbf{G}}(z, t) = e^{\mathbf{T}t} + \int_0^t \left[\sum_{r=0}^{\infty} D_r z^r \right] e^{\mathbf{T}(t-u)} \mathbf{\Lambda} \tilde{\mathbf{G}}(z, u) du. \quad (\text{A15})$$

4. Step 4: Transformation to the generalized master equation

Left-multiplying both sides of Eq. (A15) by $e^{-\mathbf{T}t}$, differentiating with respect to t , and making simplifications yield

$$e^{-\mathbf{T}t} \frac{\partial}{\partial t} \tilde{\mathbf{G}}(z, t) - e^{-\mathbf{T}t} \mathbf{T} \tilde{\mathbf{G}}(z, t) = \sum_{r=0}^{\infty} D_r z^r e^{-\mathbf{T}t} \mathbf{\Lambda} \tilde{\mathbf{G}}(z, t). \quad (\text{A16})$$

Since $\mathbf{W} = \mathbf{T} + \mathbf{\Lambda}$, we obtain

$$\frac{\partial}{\partial t} \tilde{\mathbf{G}}(z, t) = (\mathbf{W} - \mathbf{\Lambda}) \tilde{\mathbf{G}}(z, t) + \sum_{r=0}^{\infty} D_r z^r \mathbf{\Lambda} \tilde{\mathbf{G}}(z, t), \quad (\text{A17})$$

with the initial condition $\tilde{\mathbf{G}}(z, 0) = \mathbf{I}$. Applying Cauchy's residue formula $P(n, t) = \frac{1}{2\pi i} \oint \frac{\tilde{\mathbf{G}}(z, t)}{z^{n+1}} dz$ to Eq. (A17), we obtain the non-Markovian master equation

$$\frac{\partial}{\partial t} \tilde{\mathbf{P}}(n, t) = (\mathbf{W} - \mathbf{\Lambda}) \tilde{\mathbf{P}}(n, t) + \sum_{r=0}^n D_r \mathbf{\Lambda} \tilde{\mathbf{P}}(n-r, t). \quad (\text{A18})$$

We further denote $\mathbf{P}(n, t) = \tilde{\mathbf{P}}(n, t) \mathbf{e}^T$, where $\mathbf{e}^T = (1, 1, \dots, 1)^T$ is a K -dimensional column vector, then the i th element of $\mathbf{P}(n, t)$ is

$$P_i(n, t) = \Pr\{N(t) = n \mid N(0) = 0, U(0) = i\}. \quad (\text{A19})$$

Then, we obtain the generalized master equation from Eq. (18):

$$\frac{\partial}{\partial t} \mathbf{P}(n, t) = (\mathbf{W} - \mathbf{\Lambda}) \mathbf{P}(n, t) + \sum_{r=0}^n D_r \mathbf{\Lambda} \mathbf{P}(n-r, t), \quad (\text{A20})$$

where the initial condition is $\mathbf{P}(n, 0) = \mathbf{0}$ for $n > 0$ and $\mathbf{P}(0, 0) = \mathbf{e}$.

Let $\mathbf{G}(z, t)$ denote the probability generating function of $\mathbf{P}(n, t)$, i.e., $\mathbf{G}(z, t) = \sum_{n=0}^{\infty} z^n \mathbf{P}(n, t)$. Then, Eq. (A15) can be transformed to

$$\mathbf{G}(z, t) = e^{\mathbf{T}t} \mathbf{e} + \int_0^t \sum_{r=0}^{\infty} D_r z^r e^{\mathbf{T}(t-u)} \mathbf{\Lambda} \mathbf{G}(z, u) du, \quad (\text{A21})$$

where $\mathbf{e} = (1, 1, \dots, 1)^T$, and Eq. (A17) can be transformed to

$$\begin{aligned}\frac{\partial}{\partial t} \mathbf{G}(z, t) &= (\mathbf{W} - \mathbf{\Lambda}) \mathbf{G}(z, t) \\ &+ H[1 + \Phi(t)(z-1)] \mathbf{\Lambda} \mathbf{G}(z, t).\end{aligned}\quad (\text{A22})$$

Next, we consider a special case of degradation time distribution.

5. Gamma-distributed mRNA lifetime

If the mRNA lifetime follows a Gamma distribution, i.e., $\psi(t) = \frac{1}{\Gamma(\alpha)\theta^\alpha} t^{\alpha-1} e^{-t/\theta}$, the burst size PGF is

$$H(z) = (1 + \theta v - \theta v z)^{-\alpha}. \quad (\text{A23})$$

Thus,

$$\begin{aligned}H[1 + \Phi(t)(z-1)] &= [1 + \theta v \Phi(t)]^{-\alpha} \left[1 - \frac{\theta v \Phi(t) z}{1 + \theta v \Phi(t)} \right]^{-\alpha} \\ &= \sum_{r=0}^{\infty} \frac{\Gamma(\alpha + r)}{r! \Gamma(\alpha)} \frac{[\theta v \Phi(t)]^r}{[1 + \theta v \Phi(t)]^{r+\alpha}} z^r.\end{aligned}\quad (\text{A24})$$

The coefficient becomes

$$D_r = \frac{\Gamma(\alpha + r)}{\Gamma(\alpha) \Gamma(r+1)} \frac{[\theta v \Phi(t)]^r}{[1 + \theta v \Phi(t)]^{r+\alpha}}. \quad (\text{A25})$$

APPENDIX B: ANALYTICAL SOLUTION FOR SPECIAL CASES

In general, it is challenging to obtain closed-form solutions for Eq. (A22). However, analytical solutions can be derived

for certain special cases. For instance, when the protein degradation time is deterministic, i.e., $\tau_p = \tau$, the survival function $\Phi(t)$ becomes

$$\Phi(t) = \begin{cases} 1, & t \leq \tau, \\ 0, & t > \tau. \end{cases} \quad (\text{B1})$$

Substituting into Eq. (A22), the time-dependent generating function (PGF) solution is

$$\mathbf{G}(z, t) = \exp[\mathbf{W}(t - \tau)] \exp[(\mathbf{W} + (H(z) - 1)\mathbf{\Lambda})\tau] \mathbf{e}, \quad (\text{B2})$$

As $t \rightarrow \infty$, the matrix exponential $\exp[\mathbf{W}(t - \tau)]$ converges to $\mathbf{e}\boldsymbol{\pi}$ [28], where $\boldsymbol{\pi}$ is the stationary distribution of the promoter states, satisfying $\boldsymbol{\pi}\mathbf{W} = 0$ and $\boldsymbol{\pi}\mathbf{e} = 1$. Since the steady-state distribution of the system is independent of the initial state, we have $G_1(z, \infty) = G_2(z, \infty) = \dots = G_K(z, \infty) = G(z)$. Combining these results with Eq. (B2), we obtain a closed-form solution for the steady-state distribution of protein as follows:

$$G(z) = \boldsymbol{\pi} \exp[(\mathbf{W} + (H(z) - 1)\mathbf{\Lambda})\tau] \mathbf{e}. \quad (\text{B3})$$

We point out that similar results have been obtained in previous studies [29].

We further apply the result given by Eq. (B3) to the two-state gene-expression model [Eq. (14)]. For convenience, we set $\tau = 1$ in the following analysis. Using the Cayley-Hamilton theorem, the steady-state PGF for the protein distribution is

$$G(z) = \frac{r_1 e^{r_2} - r_2 e^{r_1}}{r_1 - r_2} + \frac{e^{r_1} - e^{r_2}}{r_1 - r_2} \pi_1 \lambda (H(z) - 1), \quad (\text{B4})$$

where $\pi_1 = k_0/(k_1 + k_0)$, and r_1 and r_2 are roots of the quadratic equation:

$$(x - r_1)(x - r_2) = x^2 + [k_0 + k_1 - \lambda(H(z) - 1)]x - \lambda k_0(H(z) - 1) = 0. \quad (\text{B5})$$

Note that in the slow switching regime ($k_i \ll \lambda$, $i = 1, 2$), Eq. (B4) can be approximated as

$$G(z) = (1 - \pi_1) + \pi_1 \exp[\lambda(H(z) - 1)], \quad (\text{B6})$$

indicating that the stationary protein distribution in the slow switching regime is a zero-inflated compound Poisson distribution. In contrast, under the fast-switching condition ($k_1, k_0 \rightarrow \infty$), the promoter switching will reach a quasisteady state, and Eq. (B4) can be approximated as

$$G(z) = \exp[\pi_1 \lambda (H(z) - 1)], \quad (\text{B7})$$

indicating that the stationary protein distribution is a compound Poisson distribution.

When $k_1 = 0$, the promoter always stays in the active state, and the two-state model reduces to a one-state model. In this case, Eq. (A22) becomes

$$\frac{\partial}{\partial t} G(z, t) = \lambda(H(1 + \Phi(t)(z - 1)) - 1)G(z, t). \quad (\text{B8})$$

The time-dependent solution of this equation is

$$G(z, t) = \exp\left[\int_0^t \lambda(H(1 + \Phi(u)(z - 1)) - 1) du\right]. \quad (\text{B9})$$

If the mRNA lifetime conforms to a gamma distribution $\Gamma(\alpha, \theta)$, Eq. (B9) becomes

$$G(z, t) = \exp\left[\sum_{k=1}^{\infty} \lambda \beta_k (z^k - 1)\right], \quad (\text{B10})$$

where

$$\beta_k = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\Gamma(k + 1)} \int_0^t \frac{[\theta v \Phi(u)]^k}{[1 + \theta v \Phi(u)]^{(k+\alpha)}} du, \quad (\text{B11})$$

implying that the protein distribution follows a discrete compound Poisson distribution with parameters $(\beta_1 \lambda, \beta_2 \lambda, \dots)$.

APPENDIX C: DERIVATION OF THE MOMENT'S EQUATION

To derive the binomial moments equation for the protein number, we expand the generating function $H[1 + (z - 1)\Phi(t)]$ as

$$\begin{aligned} H[1 + (z - 1)\Phi(t)] &= \sum_{k=0}^{\infty} h(k)[1 + \Phi(t)(z - 1)]^k \\ &= \sum_{k=0}^{\infty} h(k) \sum_{i=0}^{\infty} [\Phi(t)(z - 1)]^i. \end{aligned} \quad (\text{C1})$$

Define $s = z - 1$. Rearranging by collecting terms with the same power of s yields

$$\begin{aligned} H[1 + (z - 1)\Phi(t)] &= \sum_{i=0}^{\infty} \left[\sum_{k=i}^{\infty} h(k) \binom{k}{i} \Phi(t)^i \right] s^i \\ &= 1 + \sum_{i=1}^{\infty} b_i^{\text{burst}} \Phi(t)^i s^i. \end{aligned} \quad (\text{C2})$$

Here, b_i^{burst} denotes the i th binomial moment of the translational burst size distribution $h(b)$, defined as

$$b_i^{\text{burst}} = \sum_{k=i}^{\infty} \binom{k}{i} h(k). \quad (\text{C3})$$

Substituting Eq. (C2) into Eq. (A22), we obtain

$$\begin{aligned} \frac{\partial}{\partial t} \mathbf{G}(s + 1, t) &= \sum_{i=1}^{\infty} b_i^{\text{burst}} \Phi(t)^i s^i \mathbf{\Lambda} \mathbf{G}(s + 1, t) \\ &\quad + \mathbf{W} \mathbf{G}(s + 1, t). \end{aligned} \quad (\text{C4})$$

Expand $\mathbf{G}(s + 1, t)$ in a Taylor series:

$$\mathbf{G}(s + 1, t) = \sum_{n=0}^{\infty} \mathbf{a}_n(t) s^n. \quad (\text{C5})$$

The coefficients $\mathbf{a}_n(t)$ are vectors representing the binomial moments of the protein distribution across promoter states, defined as

$$\mathbf{a}_n(t) = \sum_{m=n}^{\infty} \binom{m}{n} \mathbf{P}(m, t). \quad (\text{C6})$$

Substituting Eq. (C6) into Eq. (C4) and equating coefficients of s^n on both sides yield

$$\sum_{n=0}^{\infty} \frac{d}{dt} \mathbf{a}_n(t) s^n = \sum_{n=0}^{\infty} \mathbf{W} \mathbf{a}_n(t) s^n + \sum_{n=0}^{\infty} \Lambda \left[\sum_{i=1}^{\infty} \mathbf{a}_{n-i}(t) b_i^{\text{burst}} \Phi(t)^i \right] s^n. \quad (\text{C7})$$

For $n \geq 1$, equating coefficients of s^n yields

$$\frac{d \mathbf{a}_n(t)}{dt} = \mathbf{W} \mathbf{a}_n(t) + \Lambda \left[\sum_{i=1}^n \mathbf{a}_{n-i}(t) b_i^{\text{burst}} \Phi(t)^i \right], \quad (\text{C8})$$

with $\mathbf{a}_0(t) = (1, 1, \dots, 1)^T$, and the initial condition $\mathbf{a}_n(0) = \mathbf{0}$ for $n > 0$.

If the mRNA lifetime follows a gamma distribution, i.e., $\psi(t) = \frac{1}{\Gamma(\alpha)\theta^\alpha} t^{\alpha-1} e^{-t/\theta}$, then the expression of b_i^{burst} becomes

$$b_i^{\text{burst}} = \int_0^\infty \frac{(vs)^i}{i!} \psi(s) ds = \frac{\Gamma(\alpha + i)}{i! \Gamma(\alpha)} (v\theta)^i. \quad (\text{C9})$$

In the special case of $\alpha = 1$, the burst size b follows a geometric distribution, and $b_i^{\text{burst}} = \langle b \rangle^i$, where $\langle b \rangle = v\theta$ is the mean burst size.

APPENDIX D: DERIVATION OF THE PROTEIN NOISE

For the two-state model described in the main text, the binomial moment equations corresponding to Eq. (C7) are

$$\frac{d}{dt} a_1^{(0)}(t) = -k_0 a_1^{(0)}(t) + k_0 a_1^{(1)}(t), \quad (\text{D1})$$

$$\frac{d}{dt} a_1^{(1)}(t) = k_1 a_1^{(0)}(t) - k_1 a_1^{(1)}(t) + \lambda b_1^{\text{burst}} \Phi(t), \quad (\text{D2})$$

$$\frac{d}{dt} a_2^{(0)}(t) = -k_0 a_2^{(0)}(t) + k_0 a_2^{(1)}(t), \quad (\text{D3})$$

$$\begin{aligned} \frac{d}{dt} a_2^{(1)}(t) &= k_1 a_2^{(0)}(t) - k_1 a_2^{(1)}(t) \\ &+ \lambda a_1^{(1)}(t) b_1^{\text{burst}} \Phi(t) + \lambda b_2^{\text{burst}} \Phi(t)^2. \end{aligned} \quad (\text{D4})$$

1. Mean protein expression

By $k_1 \times \text{Eq. (D1)}$ + $k_0 \times \text{Eq. (D2)}$, we have

$$\frac{d}{dt} [k_1 a_1^{(0)}(t) + k_0 a_1^{(1)}(t)] = k_0 \lambda b_1^{\text{burst}} \Phi(t). \quad (\text{D5})$$

Integrating with initial condition $k_1 a_1^{(0)}(0) + k_0 a_1^{(1)}(0) = 0$ yields

$$k_1 a_1^{(0)}(t) + k_0 a_1^{(1)}(t) = k_0 \lambda b_1^{\text{burst}} \int_0^t \Phi(u) du. \quad (\text{D6})$$

Combining Eqs. (D2) and (D5), we arrive at

$$\begin{aligned} \frac{d}{dt} a_1^{(1)}(t) &= \lambda b_1^{\text{burst}} \Phi(t) + k_0 \lambda b_1^{\text{burst}} \int_0^t \Phi(u) du \\ &- (k_0 + k_1) a_1^{(1)}(t). \end{aligned} \quad (\text{D7})$$

The solution of this first-order linear differential equation is given by

$$\begin{aligned} a_1^{(1)}(t) &= \frac{k_0 \lambda b_1^{\text{burst}}}{k_0 + k_1} \int_0^t \Phi(u) du \\ &+ \frac{k_1 \lambda b_1^{\text{burst}}}{k_0 + k_1} e^{-(k_0 + k_1)t} \int_0^t e^{(k_0 + k_1)u} \Phi(u) du. \end{aligned} \quad (\text{D8})$$

In the steady state ($t \rightarrow \infty$), the mean protein number is independent of the initial condition; thus,

$$\langle P \rangle = a_1^{(1)}(\infty) = \frac{k_0 \lambda \langle b \rangle}{k_0 + k_1} \int_0^\infty \Phi(u) du, \quad (\text{D9})$$

where $\langle b \rangle = b_1^{\text{burst}}$ representing the mean burst size.

2. Second binomial moments

To derive the second binomial moment, we combine $k_1 \times \text{Eq. (D3)}$ + $k_0 \times \text{Eq. (D4)}$:

$$\begin{aligned} k_1 \frac{d a_2^{(0)}(t)}{dt} + k_0 \frac{d a_2^{(1)}(t)}{dt} &= k_0 \lambda b_2^{\text{burst}} \Phi(t)^2 \\ &+ k_0 \lambda b_1^{\text{burst}} a_1^{(1)}(t). \end{aligned} \quad (\text{D10})$$

Integrating with initial condition $a_2^{(0)}(0) = a_2^{(1)}(0) = 0$ yields

$$\begin{aligned} k_1 a_2^{(0)}(t) + k_0 a_2^{(1)}(t) &= k_0 \lambda b_2^{\text{burst}} \int_0^t \Phi(u)^2 du \\ &+ k_0 \lambda b_1^{\text{burst}} \int_0^t a_1^{(1)}(u) \Phi(u) du. \end{aligned} \quad (\text{D11})$$

In the steady state ($t \rightarrow \infty$), the moments are independent of the initial condition, i.e., $a_2^{(1)}(\infty) = a_2^{(0)}(\infty)$; thus, we have

$$\begin{aligned} a_2^{(1)}(\infty) &= \frac{k_0 \lambda b_2^{\text{burst}}}{k_0 + k_1} \int_0^\infty \Phi(u)^2 du \\ &+ \frac{k_0 \lambda b_1^{\text{burst}}}{k_0 + k_1} \int_0^\infty a_1^{(1)}(u) \Phi(u) du. \end{aligned} \quad (\text{D12})$$

Substituting $a_1^{(1)}(u)$ from Eq. (D8), we obtain

$$a_2^{(1)}(\infty) = \frac{k_0 \lambda b_2^{\text{burst}}}{k_0 + k_1} \int_0^\infty \Phi(u)^2 du + \frac{k_0 k_1}{2} \left(\frac{\lambda b_1^{\text{burst}}}{k_0 + k_1} \right)^2 \left[\left(\int_0^\infty \Phi(u) du \right)^2 + \int_0^\infty \int_0^\infty e^{-(k_0 + k_1)|u-v|} \Phi(u) \Phi(v) du dv \right]. \quad (\text{D13})$$

The variance of protein number in the steady state is

$$\sigma_p^2 = 2a_2^{(1)}(\infty) + \langle P \rangle - \langle P \rangle^2 = \langle P \rangle + k_0 k_1 \left(\frac{\lambda b_1^{\text{burst}}}{k_0 + k_1} \right)^2 \int_0^\infty \int_0^\infty e^{-(k_0+k_1)|u-v|} \Phi(u) \Phi(v) du dv + \frac{2k_0 \lambda b_2^{\text{burst}}}{k_0 + k_1} \int_0^\infty \Phi(u)^2 du. \quad (\text{D14})$$

Given $b_2^{\text{burst}} = \frac{1}{2}(\langle b \rangle \eta_b^2 + \langle b \rangle^2 - \langle b \rangle)$ and $\langle b \rangle = b_1^{\text{burst}}$, the noise strength of protein is given by

$$\eta_p^2 = \frac{\sigma_p^2}{\langle P \rangle^2} = \frac{1}{\langle P \rangle} + \frac{k_1}{k_0 \tau_p^2} \int_0^\infty \int_0^\infty e^{-(k_0+k_1)|u-v|} \Phi(u) \Phi(v) du dv + \frac{\langle b \rangle \eta_b^2 + \langle b \rangle - 1}{\tau_p \langle P \rangle} \int_0^\infty \Phi(u)^2 du, \quad (\text{D15})$$

where $\tau_p = \int_0^\infty \Phi(u) du$ represents the mean protein lifetime and $\eta_b^2 = \sigma_b^2 / \langle b \rangle^2$ denotes the noise strength of translational burst size.

-
- [1] A. Raj and A. van Oudenaarden, Nature, nurture, or chance: Stochastic gene expression and its consequences, *Cell* **135**, 216 (2008).
- [2] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, *Science* **297**, 1183 (2002).
- [3] W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins, Noise in eukaryotic gene expression, *Nature (London)* **422**, 633 (2003).
- [4] D. Kulasiri and R. Kosarwal, *Chemical Master Equation for Large Biological Networks* (Springer, 2021).
- [5] L. Ham, M. A. Coomer, K. Öcal, R. Grima, and M. P. Stumpf, A stochastic vs deterministic perspective on the timing of cellular events, *Nat. Commun.* **15**, 5286 (2024).
- [6] A. G. Nicoll, J. Szavits-Nossan, M. R. Evans, and R. Grima, Transient power-law behaviour following induction distinguishes between competing models of stochastic gene expression, *Nat. Commun.* **16**, 2833 (2025).
- [7] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, *Proc. Natl. Acad. Sci. USA* **105**, 17256 (2008).
- [8] J. Zhang, L. Chen, and T. Zhou, Analytical distribution and tunability of noise in a model of promoter progress, *Biophys. J.* **102**, 1247 (2012).
- [9] J. Zhang and T. Zhou, Promoter-mediated transcriptional dynamics, *Biophys. J.* **106**, 479 (2014).
- [10] S. Luo, Z. Wang, Z. Zhang, T. Zhou, and J. Zhang, Genome-wide inference reveals that feedback regulations constrain promoter-dependent transcriptional burst kinetics, *Nucleic Acids Res.* **51**, 68 (2023).
- [11] J. M. Pedraza and J. Paulsson, Effects of molecular memory and bursting on fluctuations in gene expression, *Science* **319**, 339 (2008).
- [12] 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).
- [13] J. Zhang and T. Zhou, Markovian approaches to modeling intracellular reaction processes with molecular memory, *Proc. Natl. Acad. Sci. USA* **116**, 23542 (2019).
- [14] E. McShane, C. Sin, H. Zaubert, J. N. Wells, N. Donnelly, X. Wang, J. Hou, W. Chen, Z. Storchova, J. A. Marsh, *et al.*, Kinetic analysis of protein stability reveals age-dependent degradation, *Cell* **167**, 803 (2016).
- [15] M. Bazzaro, M. K. Lee, A. Zoso, W. L. Stirling, A. Santillan, I.-M. Shih, and R. B. Roden, Ubiquitin-proteasome system stress sensitizes ovarian cancer to proteasome inhibitor-induced apoptosis, *Cancer Res.* **66**, 3754 (2006).
- [16] J. C. Taggart, H. Zaubert, M. Selbach, G.-W. Li, and E. McShane, Keeping the proportions of protein complex components in check, *Cell Syst.* **10**, 125 (2020).
- [17] C. Deneke, R. Lipowsky, and A. Valleriani, Complex degradation processes lead to non-exponential decay patterns and age-dependent decay rates of messenger RNA, *PLoS One* **8**, e55442 (2013).
- [18] J. G. Belasco, All things must pass: Contrasts and commonalities in eukaryotic and bacterial mRNA decay, *Nat. Rev. Mol. Cell Biol.* **11**, 467 (2010).
- [19] V. Elgart, T. Jia, and R. Kulkarni, Quantifying mRNA synthesis and decay rates using small RNAs, *Biophys. J.* **98**, 2780 (2010).
- [20] E. Huntzinger and E. Izaurralde, Gene silencing by microRNAs: Contributions of translational repression and mRNA decay, *Nat. Rev. Genet.* **12**, 99 (2011).
- [21] C. Shi, X. Yang, T. Zhou, and J. Zhang, Nascent RNA kinetics with complex promoter architecture: Analytic results and parameter inference, *Phys. Rev. E* **110**, 034413 (2024).
- [22] Q. Jiang, X. Fu, S. Yan, R. Li, W. Du, Z. Cao, F. Qian, and R. Grima, Neural network aided approximation and parameter inference of non-Markovian models of gene expression, *Nat. Commun.* **12**, 2618 (2021).
- [23] X. Yang, Z. Wang, C. Shi, T. Zhou, and J. Zhang, Deciphering HIV-1 transcription initiation and elongation from single-molecule imaging data, *Research* **8**, 0645 (2025).
- [24] C. Shi, Y. Jiang, and T. Zhou, Queuing models of gene expression: Analytical distributions and beyond, *Biophys. J.* **119**, 1606 (2020).
- [25] J. Szavits-Nossan and R. Grima, Solving stochastic gene-expression models using queueing theory: A tutorial review, *Biophys. J.* **123**, 1034 (2024).
- [26] Y. Wang, J. Szavits-Nossan, Z. Cao, and R. Grima, Joint distribution of nuclear and cytoplasmic mRNA levels in stochastic models of gene expression: Analytical results and parameter inference, *Phys. Rev. Lett.* **135**, 068401 (2025).
- [27] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Stochastic kinetics of nascent RNA, *Phys. Rev. Lett.* **117**, 128101 (2016).
- [28] C. Shi, X. Yang, J. Zhang, and T. Zhou, Stochastic modeling of the mRNA life process: A generalized master equation, *Biophys. J.* **122**, 4023 (2023).

- [29] Y. Li and C. Jia, Effective Markovian dynamics method for solving non-Markovian dynamics of stochastic gene expression, *Phys. Rev. Lett.* **135**, 048401 (2025).
- [30] N. Kumar, A. Singh, and R. V. Kulkarni, Transcriptional bursting in gene expression: Analytical results for general stochastic models, *PLoS Comput. Biol.* **11**, e1004292 (2015).
- [31] X. Yang, S. Luo, Z. Zhang, Z. Wang, T. Zhou, and J. Zhang, Silent transcription intervals and translational bursting lead to diverse phenotypic switching, *Phys. Chem. Chem. Phys.* **24**, 26600 (2022).
- [32] M. Chen, S. Luo, M. Cao, C. Guo, T. Zhou, and J. Zhang, Exact distributions for stochastic gene expression models with arbitrary promoter architecture and translational bursting, *Phys. Rev. E* **105**, 014405 (2022).
- [33] V. Ramaswami and M. F. Neuts, Some explicit formulas and computational methods for infinite-server queues with phase-type arrivals, *J. Appl. Probab.* **17**, 498 (1980).
- [34] H. H. McAdams and A. Arkin, Stochastic mechanisms in gene expression, *Proc. Natl. Acad. Sci. USA* **94**, 814 (1997).
- [35] B. Munsky and M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, *J. Chem. Phys.* **124**, 044104 (2006).
- [36] S. Kullback and R. A. Leibler, On information and sufficiency, *Ann. Math. Stat.* **22**, 79 (1951).
- [37] D. Fraser and M. Kaern, A chance at survival: Gene expression noise and phenotypic diversification strategies, *Mol. Microbiol.* **71**, 1333 (2009).
- [38] M. Acar, J. T. Mettetal, and A. van Oudenaarden, Stochastic switching as a survival strategy in fluctuating environments, *Nat. Genet.* **40**, 471 (2008).
- [39] F. Jiao, Q. Sun, M. Tang, J. Yu, and B. Zheng, Distribution modes and their corresponding parameter regions in stochastic gene transcription, *SIAM J. Appl. Math.* **75**, 2396 (2015).
- [40] B. Barzel and O. Biham, Binomial moment equations for stochastic reaction systems, *Phys. Rev. Lett.* **106**, 150602 (2011).
- [41] J. Zhang, Q. Nie, and T. Zhou, A moment-convergence method for stochastic analysis of biochemical reaction networks, *J. Chem. Phys.* **144**, 194106 (2016).
- [42] A. Valleriani, Z. Ignatova, A. Nagar, and R. Lipowsky, Turnover of messenger RNA: Polysome statistics beyond the steady state, *Europhys. Lett.* **89**, 58003 (2010).
- [43] C. Chevalier, J. Dornig, Y. Ibrahim, A. Choquet, A. David, J. Ripoll, E. Rivals, F. Geniet, N.-O. Walliser, J. Palmeri, *et al.*, Physical modeling of ribosomes along messenger RNA: Estimating kinetic parameters from ribosome profiling experiments using a ballistic model, *PLoS Comput. Biol.* **19**, e1011522 (2023).
- [44] H. Yin, X. Wen, and H. Wang, Stochastic modeling of the mRNA kinetic process: A delay chemical master-equation model based on a survival probability, *Phys. Rev. E* **112**, 044405 (2025).
- [45] M. Dobrzyński and F. J. Bruggeman, Elongation dynamics shape bursty transcription and translation, *Proc. Natl. Acad. Sci. USA* **106**, 2583 (2009).
- [46] N. A. M. Monk, Oscillatory expression of Hes1, p53, and NF- κ B driven by transcriptional time delays, *Curr. Biol.* **13**, 1409 (2003).
- [47] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, Delay-induced stochastic oscillations in gene regulation, *Proc. Natl. Acad. Sci. USA* **102**, 14593 (2005).
- [48] T. Galla, Intrinsic fluctuations in stochastic delay systems: Theoretical description and application to a simple model of gene regulation, *Phys. Rev. E* **80**, 021909 (2009).
- [49] J. Holehouse, Z. Cao, and R. Grima, Stochastic modeling of autoregulatory genetic feedback loops: A review and comparative study, *Biophys. J.* **118**, 1517 (2020).
- [50] C. Jia and R. Grima, Holimap: An accurate and efficient method for solving stochastic gene network dynamics, *Nat. Commun.* **15**, 6557 (2024).
- [51] R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger, Transcriptional burst frequency and burst size are equally modulated across the human genome, *Proc. Natl. Acad. Sci. USA* **109**, 17454 (2012).
- [52] J. Breda, M. Zavolan, and E. van Nimwegen, Bayesian inference of gene expression states from single-cell RNA-seq data, *Nat. Biotechnol.* **39**, 1008 (2021).
- [53] B. Munsky, G. Li, Z. R. Fox, D. P. Shepherd, and G. Neuert, Distribution shapes govern the discovery of predictive models for gene regulation, *Proc. Natl. Acad. Sci. USA* **115**, 7533 (2018).