Analytical time-dependent distributions for gene expression models with complex promoter switching mechanisms

Chen Jia[1,*], Youming Li[1]

[1] Applied and Computational Mathematics Division, Beijing Computational Science Research Center, Beijing 100193, China.

[*] Correspondence: chenjia@csrc.ac.cn

Abstract

Classical gene expression models assume exponential switching time distributions between the active and inactive promoter states. However, recent experiments have shown that many genes in mammalian cells may produce non-exponential switching time distributions, implying the existence of multiple promoter states and molecular memory in the promoter switching dynamics. Here we analytically solve a gene expression model with random bursting and complex promoter switching, and derive the time-dependent distributions of the mRNA and protein copy numbers, generalizing the steady-state solution obtained in [SIAM J. Appl. Math. 72, 789-818 (2012)] and [SIAM J. Appl. Math. 79, 1007-1029 (2019)]. Using multiscale simplification techniques, we find that molecular memory has no influence on the time-dependent distribution when promoter switching is very fast or very slow, while it significantly affects the distribution when promoter switching is neither too fast nor too slow. By analyzing the dynamical phase diagram of the system, we also find that molecular memory in the inactive gene state weakens the transient and stationary bimodality of the copy number distribution, while molecular memory in the active gene state enhances such bimodality.

1 Introduction

Gene expression in individual cells is an inherently stochastic process due to small copy numbers of biochemical molecules and probabilistic collisions between them [1]. Over the past two decades, numerous strides have been made in the models and theory of single-cell stochastic gene expression dynamics [2], which has a dual representation in terms of its probability distribution and stochastic trajectory. The former is described by a system of chemical master equations (CMEs), while the latter is described by a continuous-time Markov chain that can be simulated via Gillespie's stochastic simulation algorithm. The simplest stochastic gene expression model is the so-called one-state model, where the promoter of the gene of interest is assumed to be always transcriptionally active [3]. This model takes into account synthesis and degradation of mRNA or protein, as well as possible transcriptional or translational bursting with the burst size having a geometric distribution, which was widely observed in experiments [4, 5]. If bursting is not considered, then the steady-state solution of the one-state model is the well-known Poisson distribution; if bursting is considered, then the steady-state solution is given by a negative binomial distribution [3], which was also extensively used in single-cell data analysis [6].

A more detailed gene expression model that is closer to the underlying molecular biology is the classical two-state model, where the promoter of the gene is assumed to switch stochastically between a transcriptionally active state and a transcriptionally inactive state [7–11]. This model has made great success in interpreting many experimental phenomena. Specifically, the two-state model can be used to reveal the biophysical origin of transcriptional bursting [12–14], which is due to a gene that is mostly inactive but transcribes a large number of mRNA molecules when it is active. Moreover, it can account for the experimentally observed bimodal distributions of mRNA and protein copy numbers that cannot be captured by the one-state model [9]. The two-state model is also analytically solvable. If bursting in each gene state is not considered, which usually occurs for the mRNA dynamics,

then the steady-state distribution of the gene product number is given by a confluent hypergeometric function [7] and the time-dependent distribution can also be derived [10]. If bursting in each gene state is considered, which usually occurs for the protein dynamics, then the steady-state solution is given by a Gaussian hypergeometric function [9] and the time-dependent solution was derived in [15]. In addition, the steady-state and time-dependent solutions for the two-state model of an auto-regulatory gene network were also extensively studied [16–20].

In the two-state model, the time spent in the active or inactive gene state has an exponential distribution. This is generally a reasonable assumption for bacteria [21]. However, recent studies in mammalian cells have shown that the inactive periods for many genes may have a non-exponential peaked distribution [22, 23]. This indicates that the gene dynamics in the inactive period may contain multiple exponential steps and exhibit a "refractory" behavior: after leaving the active state, the promoter has to progress through multiple inactive states before becoming active again. In addition, there have been some experimental studies suggesting that the promoter may also have multiple active states, yielding a non-exponential active period [24–28]. In such cases, one could still use a two-state model by accepting the loss of the Markov property [29]. However, adding intermediate states not only provides a convenient way to keep the Markov property, but also has the opportunity to reveal the details of the underlying biological mechanisms behind promoter switching.

In fact, gene expression models with complex multi-state promoter switching mechanisms have been widely studied from the theoretical perspective [30–37]. However, most previous papers focus on the steady-state behavior of the system, while its time-dependent behavior has received comparatively little attention. If there is only one active gene state and multiple inactive gene states, it has been found that the steady-state distribution of the mRNA copy number is given by a generalized hypergeometric function [38–42]. Thus far, there is still a lack of a detailed study about the time-dependent solution of a multi-state gene expression model with transcriptional or translational bursting. In this paper, we fill in this gap by computing the analytical time-dependent distributions of mRNA and protein copy numbers and investigate the influence of molecular memory on such distributions.

The structure of the present paper is organized as follows. In Sec. 2, we describe our multi-state model in detail and provide its CMEs. Our model can simultaneously capture non-exponential active or inactive duration distribution, as well as non-bursty or bursty gene expression pattern, and thus is very general. In Sec. 3, we obtain the analytical time-dependent solution of the multi-state model, discuss how our results reduces to previous results in one-state and two-state models, and reveal three types of possible dynamics behaviors of the system. In Sec. 4, using multiscale simplification techniques, we provide simplified expressions for the time-dependent distributions when the promoter switches rapidly or slowly between the multiple states. In Sec. 5, we investigate the phase diagram characterizing the dynamics of the multi-state model and use it to reveal the influence of molecular memory on the time-dependent distributions. We conclude in Sec. 6.

2 Model

The classical two-state gene expression model assumes that the promoter of a gene can switch between an active and an inactive state with the active state having a higher transcription rate than the inactive one [7]. In the two-state model, the durations of the active and inactive states both have an exponential distribution. However, recent studies have shown that in mammalian cells, the inactive periods for many genes may have a non-exponential distribution with a nonzero peak [22, 23], since the activation of the promoter is a complex multi-step biochemical process due to chromatin remodeling and the binding and release of transcription factors. This indicates that the gene dynamics in the inactive period may contain multiple exponential steps and in sum, the gene would undergo a multiple-state switching process with molecular memory (recall that an exponentially distributed transition time guarantees the Markovian property which is memoryless). Similar multi-state behavior has also been found for the active period [24–28].

Here we consider a gene expression model with complex promoter switching (Fig. 1(a)). Specifically, we assume that the promoter of the gene can exist in $L+1$ states, denoted by $G_0, G_1, \ldots, G_L$. Biophysically, these

states correspond to different conformational states during chromatin remodelling or different binding states with transcription factors [39, 40]. Moreover, we assume that the gene product of interest, i.e. mRNA or protein, is produced in a non-bursty (constitutive) or bursty manner. In each gene state G_i, the synthesis of the mRNA is usually non-bursty, while the synthesis of the protein is often bursty due to rapid translation of protein from a single, short-lived mRNA molecule [13, 14]. Both non-bursty and bursty gene expression are commonly observed in naturally occurring systems [43].

In the bursty case, the synthesis of the gene product is assumed to occur in bursts of random size sampled from a geometric distribution with parameter p, in agreement with experiments [5]. Then the effective reactions describing the gene product dynamics are given by

$$\begin{gathered} G_i \xrightarrow{k_{ij}} G_j, \quad i,j = 0,1,\ldots,L, \quad i \neq j, \\ G_0 \xrightarrow{\rho_0 p^k q} G_0 + kP, \quad G_i \xrightarrow{\rho_1 p^k q} G_i + kP, \quad k \geq 1, \;\; i = 1,\ldots,L, \\ P \xrightarrow{d} \varnothing, \end{gathered} \tag{1}$$

where $q = 1 - p$. Here the first row describes switching of the promoter between all gene states with rates k_{ij}, the second row describes bursty production of the gene product P in all gene states, and the third row describes the decay of the gene product with rate d due to active degradation and dilution during cell division [44, 45]. Following [40], we assume that bursts occur at a rate ρ_0 when the promoter is in state G_0 and occur at a different rate ρ_1 when the promoter is in other states $G_1, \ldots, G_L$. If $\rho_0 > \rho_1$, then G_0 is the active state and $G_1, \ldots, G_L$ are the inactive states. In this case, the active period is exponentially distributed, while the inactive period may have a non-exponential distribution. If $\rho_0 < \rho_1$, then G_0 is the inactive state and $G_1, \ldots, G_L$ are the active states. In this case, the inactive period is exponentially distributed, while the active period is non-exponential distributed.

In our model, the switching dynamics between the $L + 1$ gene states and the switching rates between them can be chosen arbitrarily, and thus the sojourn time in the states $G_1, \ldots, G_L$ can have a very general probability distribution. A special case occurs when promoter progression has an irreversible cyclic structure, as illustrated in Fig. 1(b) [38, 39]. If $\rho_0 > \rho_1$, then the corresponding model is referred to as the multi-OFF model since the promoter must undergo L inactive states one by one before entering the active state. If $\rho_0 < \rho_1$, then the corresponding model is called the multi-ON model. For both the multi-OFF and multi-ON models, the sojourn time in the gene states $G_1, \ldots, G_L$ is the independent sum of L exponential random variables and thus has a hypoexponential distribution (also called a generalized Erlang distribution).

In the non-bursty case, the effective reactions describing the gene product dynamics are given by

$$\begin{gathered} G_i \xrightarrow{k_{ij}} G_j, \quad i,j = 0,1,\ldots,L, \; i \neq j, \\ G_0 \xrightarrow{s_0} G_0 + P, \quad G_i \xrightarrow{s_1} G_i + P, \quad i = 1,\ldots,L, \\ P \xrightarrow{d} \varnothing. \end{gathered} \tag{2}$$

Here we assume that the gene product is produced in a constitutive manner with rate s_0 when the promoter is in state G_0 and is produced with a different rate s_1 when the promoter is in other states $G_1, \ldots, G_L$. Note that when the synthesis rate in the inactive gene state is zero ($s_0 = 0$ or $s_1 = 0$), the non-bursty model introduced above reduces to the model proposed in [40]. Interestingly, the non-bursty model described by Eq. (2) is actually a limiting case of the bursty model described by Eq. (1) [19, 46]. Since the burst size is geometrically distributed, the mean burst size is given by $B = \sum_{k=1}^{\infty} kp^k q = p/q$. It is easy to see that when $\rho_0, \rho_1 \to \infty$ and $B \to 0$, while keeping $\rho_0 B = s_0$ and $\rho_1 B = s_1$ as constant, we have $p \to 0$, $q \to 1$, and

$$\rho_i pq \to s_i, \quad \rho_i p^k q \to 0, \;\; k \geq 2, \;\; i = 0,1.$$

This clearly shows that the bursty model reduces to the non-bursty model in the above limit. Hence in the following, we will first derive the analytical results for the bursty model and then use them to obtain the relevant results for the non-bursty model by taking the above limit.

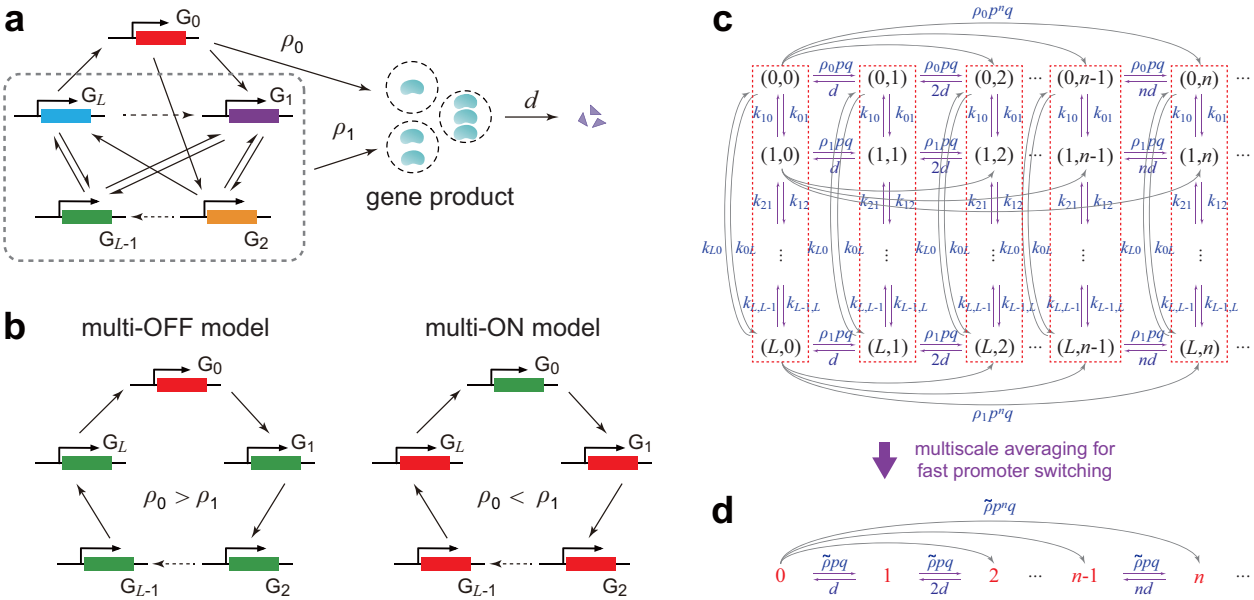

Figure 1: **Multi-state gene expression models.** (a) Schematic diagram of a multi-state model with complex promoter switching and bursting dynamics. The promoter can exist in $L+1$ states $G_0, G_1, \ldots, G_L$ and the connections between them can be arbitrary. The transcription rate is ρ_0 in gene state G_0 and is ρ_1 in all other gene states. The gene product is produced in a bursty manner. (b) Schematic diagrams of the multi-OFF and multi-ON models, where promoter switching has an irreversible cyclic structure. For the multi-OFF model, there is only one active state and multiple inactive states. For the multi-ON model, there is only one inactive state and multiple active states. (c) Transition diagram for the Markovian dynamics associated with the multi-state model. Note that bursting can cause jumps from microstate (i, n) to microstate $(i, n+k)$ for any $k \geq 1$. This this is shown for microstates $(i, 0)$ in the figure but is also true for other microstates. (d) The reduced Markovian model in fast promoter conditions. When promoter switching is fast, the microstates $(0, n), (1, n), \ldots, (L, n)$ in (c), as enclosed by red dashed squares, are in rapid equilibrium and thus can be aggregated into a group that is labeled by group n.

The microstate of the bursty model can be represented by an ordered pair (i, n): the state i of the promoter and the copy number n of the gene product. Let $p_{i,n}(t)$ denote the probability of having n copies of the gene product when the promoter is in state i at time t. Then the stochastic gene expression kinetics can be described by the continuous-time Markov chain illustrated in Fig. 1(c). The evolution of the Markovian model is governed by the following set of CMEs:

$$\begin{cases} \dot{p}_{0,n} = \left[\sum_{k=0}^{n-1} \rho_0 p^{n-k} q p_{0,k} - \sum_{k=1}^{\infty} \rho_0 p^k q p_{0,n}\right] + [(n+1)d p_{0,n+1} - n d p_{0,n}] \\ \qquad + \sum_{j \neq 0} (k_{j0} p_{j,n} - k_{0j} p_{0,n}), \\ \dot{p}_{i,n} = \left[\sum_{k=0}^{n-1} \rho_1 p^{n-k} q p_{i,k} - \sum_{k=1}^{\infty} \rho_1 p^k q p_{i,n}\right] + [(n+1)d p_{i,n+1} - n d p_{i,n}] \\ \qquad + \sum_{j \neq i} (k_{ji} p_{j,n} - k_{ij} p_{i,n}), \quad i = 1, \ldots, L. \end{cases} \tag{3}$$

Here the first term on the right-hand side represents bursty production of the gene product, the second term represents decay of the gene product, and the last term represents promoter switching. In what follows, for convenience, we set $d = 1$ [42], which is equivalent to dividing each parameter by d and rescaling time by $1/d$.

3 Analytical time-dependent solutions

3.1 Analytical solution for the bursty model

Before stating our results, we first introduce some notation. Let $K = (k_{ij})_{(L+1)\times(L+1)}$ denote the generator matrix of promoter switching dynamics, where $k_{ii} = -\sum_{j\neq i} k_{ij}$ for each $0 \leq i \leq L$. Following [42], we assume that K is irreducible, which means that for any pair of gene states G_i and G_j, there exists a transition path

$$G_i = G_{i_1} \to G_{i_2} \to G_{i_3} \to \cdots \to G_{i_n} = G_j$$

from G_i to G_j with positive transition rates, i.e. $k_{i_1 i_2} k_{i_2 i_3} \cdots k_{i_{n-1} i_n} > 0$. Moreover, let $H = (k_{ij})_{1\leq i,j\leq L}$ be the matrix obtained by removing the first row and the first column of the matrix K. Since K is a generator matrix, one eigenvalue of K must be zero. Let $\mu_1, \ldots, \mu_L$ denote all the nonzero eigenvalues of $-K$ and let $\lambda_1, \ldots, \lambda_L$ denote all the eigenvalues of $-H$. It follows from the Perron-Frobenius theorem and the irreducibility of K that λ_i and μ_i all have positive real parts [42].

To proceed, we define the generating function $f_i(t, z) = \sum_{n=0}^{\infty} p_{i,n}(t) z^n$ for each $0 \leq i \leq L$. Then Eq. (3) can be converted into the following system of partial differential equations (PDEs):

$$\begin{cases} \dfrac{\partial f_0}{\partial t} = \dfrac{\rho_0 p(z-1)}{1-pz} f_0 + (1-z)\dfrac{\partial f_0}{\partial z} + \displaystyle\sum_{j\neq 0} k_{j0} f_j - \sum_{j\neq 0} k_{0j} f_0, \\ \dfrac{\partial f_i}{\partial t} = \dfrac{\rho_1 p(z-1)}{1-pz} f_i + (1-z)\dfrac{\partial f_i}{\partial z} + \displaystyle\sum_{j\neq i} k_{ji} f_j - \sum_{j\neq i} k_{ij} f_i, \ i = 1, \ldots, L. \end{cases} \tag{4}$$

Furthermore, let $p_n(t) = \sum_{i=0}^{L} p_{i,n}(t)$ denote the probability of having n copies of the gene product at time t and let $f(t, z) = \sum_{i=0}^{L} f_i(t, z)$ denote its generating function. To simplify the above system of PDEs, we define a new variable $\omega = B(z-1)$. Let $\tilde{f}_i(t, \omega)$ and $\tilde{f}(t, \omega)$ denote the functions with variables t and ω that are associated with $f_i(t, z)$ and $f(t, z)$, i.e.

$$\tilde{f}_i(t, \omega) = f_i(t, \omega/B + 1), \quad \tilde{f}(t, \omega) = f(t, \omega/B + 1). \tag{5}$$

Then Eq. (4) can be rewritten as

$$\begin{cases} \dfrac{\partial \tilde{f}_0}{\partial t} = \dfrac{\rho_0 \omega}{1-\omega} \tilde{f}_0 - \omega \dfrac{\partial \tilde{f}_0}{\partial \omega} + \displaystyle\sum_{j\neq 0} k_{j0} \tilde{f}_j - \sum_{j\neq 0} k_{0j} \tilde{f}_0, \\ \dfrac{\partial \tilde{f}_i}{\partial t} = \dfrac{\rho_1 \omega}{1-\omega} \tilde{f}_i - \omega \dfrac{\partial \tilde{f}_i}{\partial \omega} + \displaystyle\sum_{j\neq i} k_{ji} \tilde{f}_j - \sum_{j\neq i} k_{ij} \tilde{f}_i, \quad i = 1, \ldots, L. \end{cases} \tag{6}$$

For convenience, we set $\rho = \rho_0 - \rho_1$ and

$$g_i(t, \omega) = (1-\omega)^{\rho_1} \tilde{f}_i(t, \omega), \quad i = 0, 1, \ldots, L, \quad g(t, \omega) = (1-\omega)^{\rho_1} \tilde{f}(t, \omega). \tag{7}$$

Moreover, let $\mathscr{A}$ be the differential operator defined by

$$\mathscr{A} = \frac{\partial}{\partial t} + \omega \frac{\partial}{\partial \omega}.$$

It is straightforward to check that g_i satisfies the following system of PDEs:

$$\begin{cases} \mathscr{A} g_0 - \dfrac{\rho\omega}{1-\omega} g_0 + \displaystyle\sum_{j\neq 0} k_{0j} g_0 - \sum_{j\neq 0} k_{j0} g_j = 0, \\ \mathscr{A} g_i + \displaystyle\sum_{j\neq i} k_{ij} g_i - \sum_{j\neq i} k_{ji} g_j = 0, \quad i = 1, \ldots, L. \end{cases} \tag{8}$$

Note that Eq. (8) is much simpler than Eq. (6) since the latter contains both ρ_0 and ρ_1, while the former only contains their difference $\rho = \rho_0 - \rho_1$. Summing up all the equations in Eq. (8) yields

$$g_0(t,\omega) = \frac{1-\omega}{\rho\omega}\mathscr{A}g(t,\omega), \tag{9}$$

which shows that g_0 can be expressed explicitly by g. Complex computations show that $g_0, g_1, \ldots, g_L$ can all be expressed by g as

$$\begin{pmatrix} g_0(t,\omega) \\ g_1(t,\omega) \\ \vdots \\ g_L(t,\omega) \end{pmatrix} = A(\omega) \begin{pmatrix} g(t,\omega) \\ \mathscr{A}g(t,\omega) \\ \vdots \\ \mathscr{A}^L g(t,\omega) \end{pmatrix}, \tag{10}$$

where $A(\omega)$ is a square matrix of order $L+1$ with its entries being functions of ω (see Appendix A for the detailed proof and the explicit expression of $A(\omega)$). Substituting Eqs. (9) and (10) into the relation $g = \sum_{i=0}^{L} g_i$, we obtain a PDE specifically for g, which is given by (see Appendix A for the detailed proof)

$$\left\{ \mathscr{A}\prod_{i=1}^{L}(\mathscr{A} - 1 + \mu_i) - \omega \prod_{i=1}^{L+1}(\mathscr{A} + \gamma_i) \right\} g(t,\omega) = 0, \tag{11}$$

where $\gamma_1, \ldots, \gamma_{L+1}$ are constants satisfying the equations

$$\sigma_k(\gamma_1, \ldots, \gamma_{L+1}) = \sigma_k(\mu_1, \ldots, \mu_L) + \rho\sigma_{k-1}(\lambda_1, \ldots, \lambda_L), \quad k = 1, \ldots, L+1, \tag{12}$$

with $\sigma_k(\mu_1, \ldots, \mu_L)$ being the kth elementary symmetric polynomials defined by

$$\sigma_k(\mu_1, \ldots, \mu_L) = \sum_{1 \le i_1 < \cdots < i_k \le L} \mu_{i_1}\mu_{i_2}\cdots\mu_{i_k}.$$

Here we assume that $\sigma_0(\mu_1, \ldots, \mu_L) = 1$ and $\sigma_{L+1}(\mu_1, \ldots, \mu_L) = 0$. Since Vieta's formulas relate the coefficients of a polynomial to the elementary symmetric polynomials of its roots, we can see that $\gamma_1, \ldots, \gamma_{L+1}$ are actually all the roots of a polynomial of degree $L+1$. The fundamental theorem of algebra guarantees that $\gamma_1, \ldots, \gamma_{L+1}$ are uniquely determined by Eq. (12).

To solve Eq. (11), we use the method proposed in [10, 15] to convert it into a solvable ordinary differential equation (ODE). To this end, we make the variable transformation $\xi = \log\omega - t$ and $\eta = \omega$. Let $\overline{g}(\xi,\eta)$ be the function with variables ξ and η that are associated with $g(t,\omega)$, i.e. $\overline{g}(\xi,\eta) = g(\log\eta - \xi, \eta)$. In terms of the new variables, we have

$$\mathscr{A} = \frac{\partial}{\partial t} + \omega\frac{\partial}{\partial\omega} = \eta\frac{\partial}{\partial\eta}.$$

Then Eq. (11) can be converted into the ODE

$$\left\{ \eta\frac{\partial}{\partial\eta}\prod_{i=1}^{L}\left(\eta\frac{\partial}{\partial\eta} - 1 + \mu_i\right) - \eta\prod_{i=1}^{L+1}\left(\eta\frac{\partial}{\partial\eta} + \gamma_i\right) \right\} \overline{g}(\xi,\eta) = 0. \tag{13}$$

Note that for each given ξ, Eq. (13) is a generalized hypergeometric ODE with respect to η [47, Eq. 16.8.3] and its general solution is given by the linear combination of generalized hypergeometric functions [47, Eq. 16.8.6]

$$\overline{g}(\xi,\eta) = \sum_{i=0}^{L} C_i(\xi) y_i(\eta),$$

where $C_i(\xi)$ are constants depending on ξ and $y_i(\eta)$ are functions defined by

$$\begin{aligned} y_0(\eta) &= {}_{L+1}F_L\left(\gamma_1, \ldots, \gamma_L, \gamma_{L+1}; \mu_1, \ldots, \mu_L; \eta\right), \\ y_i(\eta) &= \eta^{1-\mu_i}\,{}_{L+1}F_L\left(1+\gamma_1-\mu_i, \ldots, 1+\gamma_{L+1}-\mu_i; 2-\mu_i, 1+\mu_1-\mu_i, \ldots, \star, \ldots, 1+\mu_L-\mu_i; \eta\right), \end{aligned} \tag{14}$$

for $i = 1, \ldots, L$. Here ${}_{L+1}F_L(\alpha_1, \ldots, \alpha_{L+1}; \beta_1, \ldots, \beta_L; x)$ denotes the generalized hypergeometric function [47] and "$\star$" represents that the entry $1 + \mu_i - \mu_i$ is removed. In terms of the original variables t and ω, the function $g(t, \omega)$ is given by

$$g(t, \omega) = \sum_{i=0}^{L} \phi_i(\omega e^{-t}) y_i(\omega), \tag{15}$$

where $\phi_i(x) = C_i(\log x)$.

The remaining question is to determine the functions ϕ_i based on the initial conditions, or more precisely, the functions $g_i(0, \omega), i = 0, 1, \ldots, L$ that are associated with the initial distributions. Note that for each $n \geq 0$, it is easy to check that

$$\mathscr{A}^n g(t, \omega) = \sum_{i=0}^{L} \vartheta^n y_i(\omega) \phi_i(\omega e^{-t}), \tag{16}$$

where $\vartheta = \omega \frac{\mathrm{d}}{\mathrm{d}\omega}$. Inserting the above formula into Eq. (10) and setting $t = 0$ yield

$$\begin{pmatrix} g_0(0, \omega) \\ g_1(0, \omega) \\ \vdots \\ g_L(0, \omega) \end{pmatrix} = A(\omega) \begin{pmatrix} y_0(\omega) & y_1(\omega) & \ldots & y_L(\omega) \\ \vartheta y_0(\omega) & \vartheta y_1(\omega) & \ldots & \vartheta y_L(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \vartheta^L y_0(\omega) & \vartheta^L y_1(\omega) & \ldots & \vartheta^L y_L(\omega) \end{pmatrix} \begin{pmatrix} \phi_0(\omega) \\ \phi_1(\omega) \\ \vdots \\ \phi_L(\omega) \end{pmatrix}, \tag{17}$$

from which the functions $\phi_i(\omega)$ can be recovered from the initial conditions $g_i(0, \omega)$. Using the transformations given in Eqs. (5) and (7), in terms of the original variables t and z, the generating function $f(t, z)$ is given by

$$f(t, z) = [1 - B(z - 1)]^{-\rho_1} g(t, B(z - 1)). \tag{18}$$

Taking partial derivatives of $f(t, z)$ with respect to z at $z = 0$, we finally obtain the time-dependent distribution of the gene product number:

$$p_n(t) = \frac{1}{n!} \frac{\partial^n}{\partial z^n} \bigg|_{z=0} f(t, z). \tag{19}$$

We then focus on the steady-state gene product number distribution. At the steady state, the functions $f(t, z)$ and $g(t, \omega)$ are independent of time t and we denote them by $f^{ss}(z)$ and $g^{ss}(\omega)$, respectively. Then Eq. (11) reduces to

$$\left\{ \vartheta \prod_{i=1}^{L} (\vartheta - 1 + \mu_i) - \omega \prod_{i=1}^{L+1} (\vartheta + \gamma_i) \right\} g^{ss}(\omega) = 0.$$

Note that this is a generalized hypergeometric ODE and its solution is given by

$$g^{ss}(\omega) = {}_{L+1}F_L\left(\gamma_1, \ldots, \gamma_L, \gamma_{L+1}; \mu_1, \ldots, \mu_L; \omega\right).$$

In terms of the original variable z, the generating function $f^{ss}(z)$ is given by

$$\begin{aligned} f^{ss}(z) &= [1 - B(z - 1)]^{-\rho_1} g^{ss}(B(z - 1)) \\ &= [1 - B(z - 1)]^{-\rho_1} {}_{L+1}F_L(\gamma_1, \ldots, \gamma_L, \gamma_{L+1}; \mu_1, \ldots, \mu_L; B(z - 1)). \end{aligned} \tag{20}$$

Taking derivatives on both sides of the above equation at $z = 0$ and using the differentiation formula for generalized hypergeometric functions [47, Eq. 16.3.1], we finally obtain the steady-state distribution p_n^{ss} of the gene product number:

$$p_n^{ss} = \frac{q^{\rho_1}}{n!} \sum_{i=0}^{n} \binom{n}{i} \frac{\prod_{j=1}^{L+1} (\gamma_j)_i}{\prod_{j=1}^{L} (\mu_j)_i} {}_{L+1}F_L\left(\gamma_1 + i, \ldots, \gamma_{L+1} + i; \mu_1 + i, \ldots, \mu_L + i; -B\right) p^{n-i} B^i \left(\rho_1\right)_{n-i},$$

where $(x)_i = x(x + 1) \ldots (x + i - 1)$ is the Pochhammer symbol. In the special case of $\rho_1 = 0$, i.e. no gene product molecules are produced during the inactive period, the expression can be simplified as

$$p_n^{ss} = \frac{B^n}{n!} \frac{\prod_{j=1}^{L+1} (\gamma_j)_n}{\prod_{j=1}^{L} (\mu_j)_n} {}_{L+1}F_L\left(\gamma_1 + n, \ldots, \gamma_L + n, \gamma_{L+1} + n; \mu_1 + n, \ldots, \mu_L + n; -B\right).$$

3.2 Analytical solution for the non-bursty model

Next we focus on the non-bursty model. Recall that the non-bursty model given in Eq. (2) is a limiting case of the bursty model given in Eq. (1) when $\rho_0, \rho_1 \to \infty$ and $B \to 0$, while keeping $\rho_0 B = s_0$ and $\rho_1 B = s_1$ as constant. Under this limit, it is clear that

$$|\rho| = |\rho_0 - \rho_1| = |s_0 - s_1|/B \to \infty, \quad \rho B \to s := s_0 - s_1.$$

From Eq. (12), it is easy to see that $\gamma_1, \ldots, \gamma_{L+1}$ are functions of ρ, $\mu_1, \ldots, \mu_L$, and $\lambda_1, \ldots, \lambda_L$. In Appendix B, we have proved that

$$\lim_{|\rho|\to\infty} \gamma_i = \lambda_i, \quad i = 1, \ldots, L, \quad \lim_{|\rho|\to\infty} \frac{\gamma_{L+1}}{\rho} = 1. \tag{21}$$

Recall that when $\alpha_{L+1} \to \infty$ and $z \to 0$, while keeping $\alpha_{L+1} z$ as constant, we have the following limit for generalized hypergeometric functions [47, Eq. 16.8.10]:

$${}_{L+1}F_L(\alpha_1, \ldots, \alpha_L, \alpha_{L+1}; \beta_1, \ldots, \beta_L; z) \to {}_LF_L(\alpha_1, \ldots, \alpha_L; \beta_1, \ldots, \beta_L; \alpha_{L+1} z).$$

Applying this formula and taking $\rho_0, \rho_1 \to \infty$ and $B \to 0$, while keeping $\rho_0 B = s_0$ and $\rho_1 B = s_1$ as constant, we obtain the following limits:

$$\begin{aligned}
&[1 - B(z-1)]^{-\rho_1} \to e^{s_1(z-1)}, \\
&{}_{L+1}F_L(\gamma_1, \ldots, \gamma_L, \gamma_{L+1}; \mu_1, \ldots, \mu_L; B(z-1)) \to {}_LF_L(\lambda_1, \ldots, \lambda_L; \mu_1, \ldots, \mu_L; s(z-1)), \\
&{}_{L+1}F_L(1 + \gamma_1 - \mu_i, \ldots, 1 + \gamma_{L+1} - \mu_i; 2 - \mu_i, 1 + \mu_1 - \mu_i, \ldots, \star, \ldots, 1 + \mu_L - \mu_i; B(z-1)) \\
&\to {}_LF_L(1 + \lambda_1 - \mu_i, \ldots, 1 + \lambda_L - \mu_i; 2 - \mu_i, 1 + \mu_1 - \mu_i, \ldots, \star, \ldots, 1 + \mu_L - \mu_i; s(z-1)).
\end{aligned}$$

Thus from Eqs. (15) and (18), the generating function $f(t, z)$ for the non-bursty model is given by

$$f(t, z) = e^{s_1(z-1)} g(t, s(z-1)),$$

where

$$g(t, \omega) = \sum_{i=0}^{L} \overline{\phi}_i(\omega e^{-t}) \overline{y}_i(\omega),$$

with the functions $\overline{y}_0, \ldots, \overline{y}_L$ being defined by

$$\begin{aligned}
\overline{y}_0(\omega) &= {}_LF_L\left(\lambda_1, \ldots, \lambda_L; \mu_1, \ldots, \mu_L; \omega\right), \\
\overline{y}_i(\omega) &= \omega^{1-\mu_i} {}_LF_L\left(1 + \lambda_1 - \mu_i, \ldots, 1 + \lambda_L - \mu_i; 2 - \mu_i, 1 + \mu_1 - \mu_i, \ldots, \star, \ldots, 1 + \mu_L - \mu_i; \omega\right),
\end{aligned}$$

for $i = 1, \ldots, L$. Here the functions $\overline{\phi}_i$ can be determined by using similar methods. Then the time-dependent copy number distribution can be recovered from $f(t, z)$ based on Eq. (19).

At the steady state, it follows from Eq. (20) that the generating function reduces to

$$f^{ss}(z) = e^{s_1(z-1)} {}_LF_L\left(\lambda_1, \ldots, \lambda_L; \mu_1, \ldots, \mu_L; s(z-1)\right).$$

Taking derivatives on both sides of the above equation at $z = 0$ yields the steady-state copy number distribution

$$p_n^{ss} = \frac{1}{n!} \sum_{i=0}^{n} \binom{n}{i} {}_LF_L(\lambda_1 + i, \ldots, \lambda_L + i; \mu_1 + i, \ldots, \mu_L + i; -s) s_1^{m-i} s^i.$$

In the special case of $s_1 = 0$, i.e. no gene product molecules are produced during the inactive period, the expression can be simplified as

$$p_n^{ss} = \frac{s_0^n}{n!} {}_LF_L(\lambda_1 + n, \ldots, \lambda_L + n; \mu_1 + n, \ldots, \mu_L + n; -s_0).$$

This coincides with the results obtained in [40, 42].

3.3 Special case of $L = 0$

We next focus on two important special cases. In the case of $L = 0$, there is only one gene state G_0 and the promoter is always active. If the gene product is produced in a bursty manner, then our model reduces to the classical bursty gene expression model [3], whose effective reactions are given by

$$G \xrightarrow{\rho_0 p^k q} G + kP, \quad k \geq 1, \quad P \xrightarrow{d} \varnothing,$$

where we assume $d = 1$ for simplicity. Since $L = 0$, it follows from Eq. (15) that

$$g(t, \omega) = \phi_0(\omega e^{-t}) y_0(\omega). \tag{22}$$

If the initial gene product number is zero, then we have $g(0, \omega) = \tilde{f}(0, \omega) = 1$. It the follows from Eq. (22) that

$$\phi_0(\omega) = \frac{1}{y_0(\omega)} = \frac{1}{{}_1F_0\left(\rho_0; ; \omega\right)} = (1 - \omega)^{\rho_0}.$$

Then we obtain

$$g(t, \omega) = \frac{\phi_0(\omega e^{-t})}{\phi_0(\omega)} = \left(\frac{1 - \omega e^{-t}}{1 - \omega}\right)^{\rho_0}.$$

In terms of the original variables t and z, the generating function $f(t, z)$ is given by

$$f(t, z) = g(t, B(z - 1)) = \left[\frac{1 - B(z - 1)e^{-t}}{1 - B(z - 1)}\right]^{\rho_0}. \tag{23}$$

Thus when the promoter is always active, the time-dependent distribution of the gene product number can be recovered from $f(t, z)$ by taking the partial derivatives with respect to z at $z = 0$:

$$p_n(t) = \frac{(\rho_0)_n}{n!}\left(\frac{B}{1 + B}\right)^n \left(\frac{1 + Be^{-t}}{1 + B}\right)^{\rho_0} {}_2F_1\left(-n, -\rho_0; 1 - \rho_0 - n; \frac{1 + B}{e^t + B}\right). \tag{24}$$

This agrees with the results obtained in [9]. Taking $t \to \infty$ in the above equation yields the following steady-state distribution of the gene product number:

$$p_n^{ss} = \frac{(\rho_0)_n}{n!} p^n q^{\rho_0}.$$

This is the well-known negative binomial distribution and agrees with the results obtained in [3].

If the gene product is produced in a non-bursty manner, then our model reduces to the following simple gene expression model:

$$G \xrightarrow{s_0} G + P, \quad P \xrightarrow{d} \varnothing.$$

To solve this model, we only need to apply the limit $\rho_0 \to \infty$ and $B \to 0$, while keeping $\rho_0 B = s_0$ as constant, to the above results. Taking this limit in Eq. (23), we obtain

$$f(t, z) = e^{s_0(1 - e^{-t})(z - 1)}.$$

Taking the derivatives of the generating function $f(t, z)$, we find that the gene product number has a Poisson time-dependent distribution that is given by

$$p_n(t) = \frac{\left[s_0(1 - e^{-t})\right]^n}{n!} e^{-s_0(1 - e^{-t})}.$$

3.4 Special cases of $L = 1$ and $L = 2$

We next consider the special case of $L = 1$, where there are only two gene states G_0 and G_1. If the gene product is produced in a bursty manner, then our model reduces to the classical two-state bursty gene expression model [9, 14], whose effective reactions are given by

$$G_0 \xrightarrow{k_{01}} G_1, \quad G_1 \xrightarrow{k_{10}} G_0,$$
$$G_0 \xrightarrow{\rho_0 p^k q} G_0 + kP, \quad G_1 \xrightarrow{\rho_1 p^k q} G_1 + kP, \quad k \geq 1, \quad P \xrightarrow{d} \varnothing,$$

where we assume $d = 1$ for simplicity. Since $L = 1$, it is easy to check that $\lambda_1 = k_{10}, \mu_1 = k_{01} + k_{10}$. It follows from Eq. (12) that γ_1 and γ_2 are determined by

$$\gamma_1\gamma_2 = k_{10}\rho, \quad \gamma_1 + \gamma_2 = k_{10} + k_{01} + \rho.$$

Based on Eqs. (15) and (17), we have

$$g(t, \omega) = \phi_0(\omega e^{-t})y_0(\omega) + \phi_1(\omega e^{-t})y_1(\omega), \tag{25}$$

where

$$y_0(\omega) = {}_2F_1\left(\gamma_1, \gamma_2; \mu_1; \omega\right), \quad y_1(\omega) = \omega^{1-\mu_1}{}_2F_1\left(1 + \gamma_1 - \mu_1, 1 + \gamma_2 - \mu_1; 2 - \mu_1; \omega\right),$$

and $\phi_0(\omega)$ and $\phi_1(\omega)$ satisfy the following two restrictions:

$$y_0(\omega)\phi_0(\omega) + y_1(\omega)\phi_1(\omega) = g_0(0, \omega) + g_1(0, \omega),$$
$$\frac{1-\omega}{\rho\omega}\vartheta\left(y_0(\omega)\right)\phi_0(\omega) + \frac{1-\omega}{\rho\omega}\vartheta\left(y_1(\omega)\right)\phi_1(\omega) = g_0(0, \omega).$$

Here the first equation follows from Eq. (25) and the second one follows from Eq. (9). If the initial gene product number is zero and the promoter is in the state G_0, then we have $g_0(0, \omega) = (1 - \omega)^{\rho_1}$ and $g_1(0, \omega) = 0$. Solving the above two equations yield

$$\begin{aligned}
\phi_0(\omega) &= (1-\omega)^{\rho_0+1}{}_2F_1\left(1 + \gamma_1 - \mu_1, 1 + \gamma_2 - \mu_1; 1 - \mu_1; \omega\right) \\
&\quad - \frac{\rho}{1-\mu_1}\omega(1-\omega)^{\rho_0}{}_2F_1\left(1 + \gamma_1 - \mu_1, 1 + \gamma_2 - \mu_1; 2 - \mu_1; \omega\right), \\
\phi_1(\omega) &= \frac{k_{01}\rho}{\mu_1(1-\mu_1)}\omega^{\mu_1}(1-\omega)^{\rho_0}{}_2F_1(\gamma_1, \gamma_2; \mu_1 + 1; \omega),
\end{aligned}$$

where we have used [47, Eqs. 15.5.4 and 15.5.21] in the calculations. Inserting the above formulas into Eq. (25) gives the explicit expression of $g(t, \omega)$. Then the generating function in terms of the original variables t and z is given by $f(t, z) = (1 - B(z - 1))^{-\rho_1}g(t, B(z - 1))$. This is in agreement with the result obtained in [15]. The time-dependent copy number distribution can be recovered from $f(t, z)$ by taking derivatives.

If the gene product is produced in a non-bursty manner, then our model reduces to the following classical two-state non-bursty gene expression model [7]:

$$G_0 \xrightarrow{k_{01}} G_1, \quad G_1 \xrightarrow{k_{10}} G_0,$$
$$G_0 \xrightarrow{s_0} G_0 + P, \quad G_1 \xrightarrow{s_1} G_1 + P, \quad k \geq 1, \quad P \xrightarrow{d} \varnothing.$$

To solve this model, we consider the limit $\rho_0, \rho_1 \to \infty$ and $B \to 0$, while keeping $\rho_0 B = s_0$ and $\rho_1 B = s_1$ as constant. Under this limit, it follows from Eq. (21) that $\gamma_1 \to \lambda_1 = k_{10}$ and $\gamma_2/\rho \to 1$. If the initial gene product number is zero and the promoter is in the state G_0, then taking the above limit in Eq. (25) gives the generating function $f(t, z) = e^{s_1(z-1)}g(t, s(z - 1))$, where $s = s_0 - s_1$ and

$$g(t, \omega) = \overline{\phi}_0(\omega e^{-t})\overline{y}_0(\omega) + \overline{\phi}_1(\omega e^{-t})\overline{y}_1(\omega),$$

with

$$\overline{y}_0(\omega) = {}_1F_1\left(k_{10}; k_{01}+k_{10}; \omega\right), \quad \overline{y}_1(\omega) = \omega^{1-k_{01}-k_{10}}{}_1F_1\left(1-k_{01}; 2-k_{01}-k_{10}; \omega\right),$$

and

$$\overline{\phi}_0(\omega) = e^{-s_0\omega/s}{}_1F_1\left(1-k_{01}; 1-k_{01}-k_{10}; \omega\right) - \frac{\omega e^{-s_0\omega/s}}{1-k_{01}-k_{10}}{}_1F_1\left(1-k_{01}; 2-k_{01}-k_{10}; \omega\right),$$
$$\overline{\phi}_1(\omega) = \frac{k_{01}\omega^{k_{01}+k_{10}}e^{-s_0\omega/s}}{(k_{01}+k_{10})(1-k_{01}-k_{10})}{}_1F_1\left(k_{10}; 1+k_{01}+k_{10}; \omega\right).$$

In the special case of $s_0 = 0$, which means G_0 is the inactive state and G_1 is the active state, our result has a slightly different form but is essentially in accordance with the one obtained in [10].

Finally, we consider the case of $L = 2$, where the promoter can switch between three states. In this case, our model reduces to the classical refractory model proposed in [22]. Note that in the original refractory model, G_0 is an active gene state, and the bursting in each gene state is not considered. The analytical solutions for this model are given in Appendix C.

3.5 Three types of dynamic behaviors

To validate our analytical solutions, we compare them with numerical solutions obtained using the finite-state projection algorithm (FSP) [48]. When performing FSP, we truncate the state space at a large integer N and solve the truncated master equation numerically using the MATLAB function ODE45. The truncation size is chosen as $N = 5\max(\rho_0 B/d, \rho_1 B/d)$. Since $\rho_0 B/d$ and $\rho_1 B/d$ are the typical gene product numbers in the active and inactive gene states, respectively, the probability that the copy number is outside the truncated size is very small and practically can always be ignored. From Fig. 2, we can see that our exact solution coincides perfectly with the numerical solution obtained by the FSP at all time points.

In what follows, we assume that the initial gene product number is zero and the promoter initially starts from the steady-state distribution of all gene states $G_0, \ldots, G_L$, unless otherwise stated. Recall that the gene product number distribution can be divided into being unimodal or bimodal according to the number of modes. In general, bimodality implies multiple phenotypic states of the underlying biological system [49]. To further understand the shape of the time-dependent distribution, we classify the dynamic behavior of our model into three different phases, following [20]: (i) the distribution is unimodal at all times (Fig. 2(a)); (ii) the distribution is unimodal at small and large times and becomes bimodal at intermediate times (Fig. 2(b)); and (iii) the distribution is unimodal at small times and becomes bimodal at large times (Fig. 2(c)). Multimodality with more than two modes is not detected over large swaths of the parameter space. To distinguish between them, we refer to (i) as unimodality (U), to (ii) as transient bimodality (TB), and to (iii) as stationary bimodality (SB). It is clear that all the three types of dynamic behaviors can occur in our model.

4 Cases of fast and slow promoter switching

We next focus on two nontrivial special cases. The first case occurs when the promoter switches rapidly between all gene states, i.e. $k_{ij} \gg \rho_0, \rho_1, d$. In this case, the gene product number n is a slow variable and the promoter state i is a fast variable. Hence our model can be greatly simplified using the classical simplification technique of two-time-scale Markov chains called averaging [19, 51, 52]. For simplicity, we only consider the bursty model here; the non-bursty model can be analyzed in a similar way. Since k_{ij} are large, for each $n \geq 0$, the microstates $(0, n), (1, n), \ldots, (L, n)$ are in rapid equilibrium and thus can be aggregated into a group that is labeled by group n, as illustrated in Fig. 1(c). In this way, our model can be simplified to the Markovian model illustrated in Fig. 1(d), whose state space is given by

$$\{\text{group } 0, \text{group } 1, \cdots, \text{group } n, \cdots\},$$

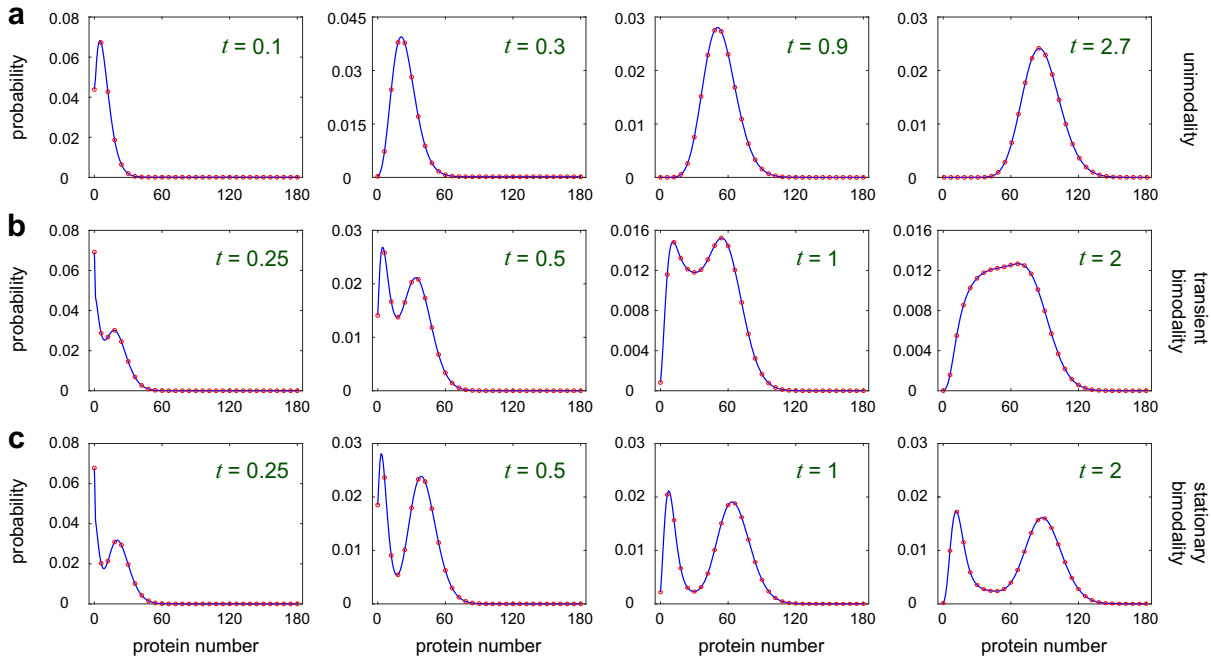

Figure 2: **Three types of dynamic behaviors for the multi-state model.** (a) Unimodality. The blue curve shows the exact solution computed via Eq. (15) and the red circles show the numerical solution computed via FSP. When computing the exact solution, we use the discrete Fourier transform algorithm proposed in [50]. This algorithm can recover the distribution directly from the generating function without calculating the higher order derivatives as in Eq. (19), which is very time-consuming and numerically unstable. Here we have assumed that $k_{01} = u$ and $k_{12} = k_{23} = \cdots = k_{L0} = Lv$. The parameters are chosen as $u = 50, v = 20, \rho_0 = 240, \rho_1 = 20, B = 1, L = 3$. (b) Transient bimodality. The parameters are chosen as $u = 0.65, v = 1.2, \rho_0 = 66, \rho_1 = 10.5, B = 1.5, L = 3$. (c). Stationary bimodality. The parameters are chosen as $u = 0.1, v = 0.25, \rho_0 = 70, \rho_1 = 10, B = 1.5, L = 3$. In (a)-(c), the initial gene product number is zero and the promoter initially starts from the steady-state distribution of all gene states.

The remaining step is to compute the effective transition rates between two groups. In the fast switching regime, for each $n \geq 0$, the microstates (i, n), $i = 0, 1, \ldots, L$ will reach a quasi-steady state. Since the transition rates between gene states are independent of n, the quasi-steady-state distribution is also independent of n. Let p_{G_i} denote the quasi-steady-state probability of being in gene state G_i. It then follows from [42, Lemma 5.3] that

$$p_{G_0} = \frac{\lambda_1 \cdots \lambda_L}{\mu_1 \cdots \mu_L}. \tag{26}$$

For convenience, we define the effective transcription rate as

$$\tilde{\rho} = \frac{\lambda_1 \cdots \lambda_L}{\mu_1 \cdots \mu_L} \rho_0 + \left(1 - \frac{\lambda_1 \cdots \lambda_L}{\mu_1 \cdots \mu_L}\right) \rho_1. \tag{27}$$

Let $q_{(i,n),(i',n')}$ denote the transition rate from microstate (i, n) to microstate (i', n') for the full model. According to the averaging theory [51, 52], the effective transition rate from group n to group $n + k$ is given by

$$\sum_{i=0}^{L} p_{G_i} q_{(i,n),(i,n+k)} = p_{G_0} \rho_0 p^k q + (1 - p_{G_0}) \rho_1 p^k q = \tilde{\rho} p^k q,$$

and the effective transition rate from group n to group $n - 1$ is given by

$$\sum_{i=0}^{L} p_{G_i} q_{(i,n),(i,n-1)} = nd.$$

Thus far, we have obtained all effective transition rates for the group dynamics (Fig. 1(d)). Note that this is equivalent to a one-state model in the case of $L = 0$, where the transcription rate ρ_0 in the gene state G_0 is

replaced by the effective transcription rate $\tilde{\rho}$. Since we have assumed that the promoter initially starts from the quasi-steady-state distribution of all gene states, the averaging theory guarantees that distributions of the full and reduced models agree with each other over the whole time axis [53]. If the initial gene product number is zero, then it follows from Eq. (24) that the approximate time-dependent distribution of the gene product number in the fast switching regime is given by

$$p_n(t) = \frac{(\tilde{\rho})_n}{n!}\left(\frac{B}{1+B}\right)^n \left(\frac{1+Be^{-t}}{1+B}\right)^{\tilde{\rho}} {}_2F_1\left(-n, -\tilde{\rho}; 1-\tilde{\rho}-n; \frac{1+B}{e^t+B}\right). \tag{28}$$

To validate this result, we compare the approximate solution with the exact solution in Fig. 3(a). It can be seen that they coincide with each other perfectly at all time points.

The second case occurs when the promoter switches slowly between all gene states, i.e. $k_{ij} \ll \rho_0, \rho_1, d$. In this case, the gene product number n is a fast variable and the promoter state i is a slow variable. Since k_{ij} are small, for each $0 \le i \le L$, the microstates (i, n), $n \ge 0$ will relax to a quasi-steady state before promoter switching takes places. If the initial gene product number is zero, then it follows from Eq. (24) that the distribution $p_{n|i}(t)$ of the gene product number conditioned on gene state G_i is given by

$$\begin{aligned} p_{n|0}(t) &= \frac{(\rho_0)_n}{n!}\left(\frac{B}{1+B}\right)^n \left(\frac{1+Be^{-t}}{1+B}\right)^{\rho_0} {}_2F_1\left(-n, -\rho_0; 1-\rho_0-n; \frac{1+B}{e^t+B}\right), \\ p_{n|i}(t) &= \frac{(\rho_1)_n}{n!}\left(\frac{B}{1+B}\right)^n \left(\frac{1+Be^{-t}}{1+B}\right)^{\rho_1} {}_2F_1\left(-n, -\rho_1; 1-\rho_1-n; \frac{1+B}{e^t+B}\right), \quad i = 1, \ldots, L. \end{aligned} \tag{29}$$

Let p_{G_i} denote the steady-state probability of being in gene state G_i, respectively. It follows from Eq. (26) that

$$p_{G_0} = \frac{\lambda_1 \cdots \lambda_L}{\mu_1 \cdots \mu_L}.$$

Since we have assumed that the promoter initially starts from the steady-state distribution of all gene states, the approximate time-dependent distribution of the gene product number in the slow switching regime is given by

$$p_n(t) = \sum_{i=1}^{L} p_{G_i} p_{n|i}(t) = \frac{\lambda_1 \cdots \lambda_L}{\mu_1 \cdots \mu_L} p_{n|0}(t) + \left(1 - \frac{\lambda_1 \cdots \lambda_L}{\mu_1 \cdots \mu_L}\right) p_{n|1}(t), \tag{30}$$

where $p_{n|0}(t)$ and $p_{n|1}(t)$ are given in Eq. (29). Taking $t \to \infty$ in the above equation yields the steady-state distribution of the gene product number, which is given by

$$p_n^{ss} = \frac{\lambda_1 \cdots \lambda_L}{\mu_1 \cdots \mu_L} \frac{(\rho_0)_n}{n!} p^n q^{\rho_0} + \left(1 - \frac{\lambda_1 \cdots \lambda_L}{\mu_1 \cdots \mu_L}\right) \frac{(\rho_1)_n}{n!} p^n q^{\rho_1}.$$

This is the mixture of two negative binomial distributions. Since a negative binomial distribution is unimodal, the mixture of two negative binomials can yield a bimodal copy number distribution. To validate this result, we compare the approximate solution with the exact solution in Fig. 3(b), from which we see that they coincide with each other perfectly at all time points.

5 Influence of molecular memory on the time-dependent distribution

5.1 Fast and slow promoter switching

The major difference between our model and the classical one-state or two-state model is the complex promoter switching mechanism, which leads to complex non-exponential distribution of the active or inactive period, i.e. molecular memory. Here we investigate the effect of molecular memory on the time-dependent distribution of the gene product.

To do this, we focus on the multi-ON and multi-OFF models with $L + 1$ gene states (Fig. 1(b)). For the multi-ON (multi-OFF) model, the active (inactive) period has a non-exponential distribution. Here the parameter L

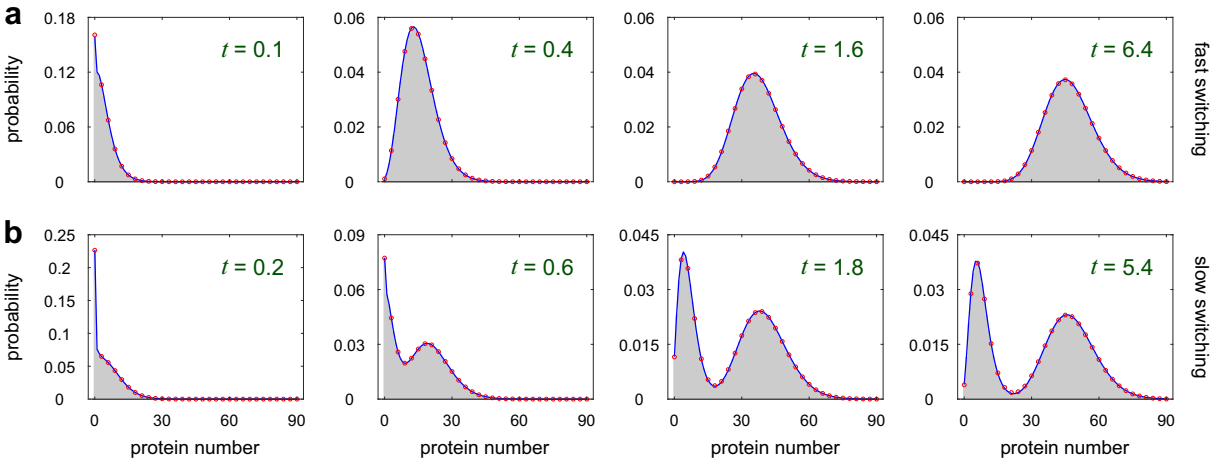

Figure 3: **Comparison between the exact and approximate time-dependent distributions in fast and slow switching conditions for the multi-OFF model**. (a) Fast promoter switching. The parameters are chosen as $u = 100, v = 200, \rho_0 = 42, \rho_1 = 10, B = 1.5$. (b) Slow promoter switching. The parameters are chosen as $u = 0.003, v = 0.005, \rho_0 = 32, \rho_1 = 5, B = 1.5$. In (a),(b), the blue curve shows the exact solution when $L = 2$, the red circles show the approximate solution when $L = 2$, and the grey region shows the exact solutions when $L = 10$. The exact solution is computed via Eq. (18), while the approximate solution is computed via Eq. (28) for fast switching and via Eq. (30) for slow switching. Clearly, molecular memory has no effect on the distributions in fast and slow switching conditions.

can be understood as the strength of molecular memory. A larger L usually gives rise to a larger deviation from the exponential distribution. To make an equitable comparison between models with different values of L, we assume that the promoter switching rates are given by

$$G_0 \xrightarrow{u} G_1 \xrightarrow{Lv} G_2 \xrightarrow{Lv} \cdots \xrightarrow{Lv} G_L \xrightarrow{Lv} G_0.$$

In other words, we assume $k_{01} = u$ and $k_{12} = k_{23} = \cdots = k_{L0} = Lv$. Under this assumption, the transition time from G_0 to G_1, denoted by T_0, has an exponential distribution with mean $1/u$. Moreover, the transition time from G_1 to G_0, denoted by T_1, has an Erlang distribution with shape parameter L and mean $1/v$. Therefore, the parameter v can be interpreted as the effective switching rate from G_1 to G_0. Note that the means of the two transition times are independent of L. The noise, characterized by the coefficient of variation squared, in T_1 is given by $1/L$. As $L \to \infty$, the noise in T_1 tends to zero and thus T_1 converges to a point mass at $1/v$, in which case the molecular memory is the strongest.

To understand the influence of molecular memory on the time-dependent distribution, we first focus on the cases of fast or slow promoter switching. For both the multi-ON and multi-OFF models, the eigenvalues $\mu_1, \ldots, \mu_L$ of the matrix $-K$ satisfy

$$\mu_1 \cdots \mu_L = \det(-K) = (Lv)^L + Lu(Lv)^{L-1},$$

and the eigenvalues $\lambda_1, \ldots, \lambda_L$ of the matrix $-H$ satisfy

$$\lambda_1 \cdots \lambda_L = \det(-H) = (Lv)^L.$$

Combining the above two equations shows that the steady-state probability of being in gene state G_0 is given by

$$p_{G_0} = \frac{\lambda_1 \cdots \lambda_L}{\mu_1 \cdots \mu_L} = \frac{v}{u+v},$$

which does not depend on L. In the fast switching regime, the effective transcription rate given in Eq. (27) and the time-dependent distribution given in Eq. (28) are both independent of L. On the other hand, in the slow switching regime, the time-dependent distribution given in Eq. (30) is also independent of L. This shows that molecular memory has no effect on the time-dependent distribution in the limiting cases of fast and slow promoter switching, as illustrated in Fig. 3.

5.2 General case

We have seen that molecular memory has no influence on the time-dependent distribution when the promoter switches rapidly or slowly between all gene states. A natural question is whether molecular memory will affect the time-dependent distribution when promoter switching is neither too fast nor too slow. To see this, recall that the gene expression system can exhibit three types of dynamical behaviors: unimodality (U), transient bimodality (TB), and stationary bimodality (SB). To gain a deeper insight into a bimodal copy number distribution, following [20], we define the strength of bimodality as

$$\kappa = \frac{H_{\text{low}} - H_{\text{valley}}}{H_{\text{high}}},$$

where H_{low} is the height of the lower peak, H_{high} is the height of the higher peak, and H_{valley} is the height of the valley between them. For unimodal distributions, κ is automatically set to be 0. For bimodal distributions, κ is a quantity between 0 and 1 since $H_{\text{valley}} < H_{\text{low}} \leq H_{\text{high}}$. In general, to display strong bimodality, the following two conditions are necessary: (i) the two peaks should have similar heights and (ii) there should be a deep valley between them. The former ensures that the time periods spent in the low and high expression states are comparable, while the latter guarantees that the two expression levels are distinguishable. Clearly, κ is large if both conditions are satisfied and is small if any one of the two conditions is violated. Hence, κ serves as an effective indicator that characterizes the strength of bimodality.

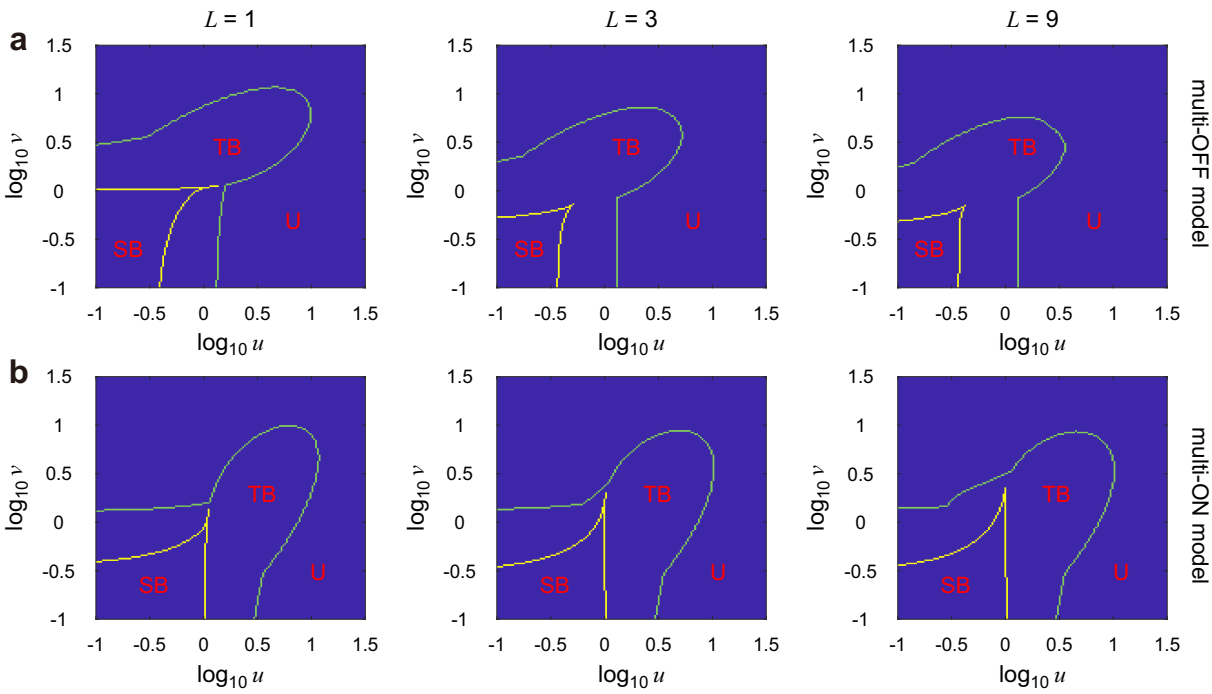

Figure 4: **Dynamic phase diagrams for the multi-OFF and multi-ON models as u and v vary.** The green and yellow curves enclose the regions for TB and SB, respectively. (a) Phase diagram for the multi-OFF model under different choices of L. The parameters are chosen as $\rho_0 = 40, \rho_1 = 0, p = 0.3$. (b) Same as (a) but for the multi-ON model. The parameters are chosen as $\rho_0 = 0, \rho_1 = 40, p = 0.3$.

To understand the influence of molecular memory on the copy number distribution, we illustrate the $u - v$ phase diagrams for the multi-ON and multi-OFF models under different choices of L in Fig. 4. Clearly, for both models, U occurs when promoter switching is very fast, while TB and SB can only occur when promoter switching is relatively slow. This is in sharp contrast to a positive or negative feedback genetic loop, where TB can also occur in fast switching conditions [20]. Compared with TB, the occurrence of SB requires slower promoter switching rates. Previous studies [40] have shown that for both models, the steady-state noise in the gene product number will decrease as the strength L of molecular memory increases. This is because the increase of L leads to a smaller

noise in the transition time from G_1 to G_0 and thus will also reduce noise in the steady-state distribution. Generally, unimodal distributions have smaller noise than bimodal ones. Hence intuitively, we expect that the increase of L should reduce the regions of TB and SB in the phase diagram. For the multi-OFF model, we find that this is indeed the case, where the regions of TB and SB both shrink remarkably as L increases (Fig. 4(a)). However, for the multi-ON model, we find that the intuition is incorrect and the opposite is true — the regions of TB and SB both enlarge with the increase of L (Fig. 4(b)).

To explain this counteractive phenomenon, we illustrate the steady-state copy number distribution under different choices of L in Fig. 5 for both the multi-OFF and multi-ON models. In Fig. 5(a), the steady-state distribution is bimodal when $L = 1$ with two nonzero modes. This usually occurs when ρ_0 and ρ_1 are both nonzero, which means that mRNA can be produced in all gene states. Clearly, increasing L results in a distribution that is more concentrated and has smaller noise, just as expected. Interestingly, the ways that the distribution is concentrated are significantly different for the two models. For the multi-OFF model, the left mode moves to the right and becomes higher as L increases, while the right mode remains almost unchanged and becomes lower. For the multi-ON model, however, the right mode moves to the left and becomes higher as L increases, while the left mode remains almost unchanged and becomes lower.

We emphasize that similar distribution properties also hold when the left mode is located at zero. In Fig. 5(b), the steady-state distribution is bimodal when $L = 1$ with a zero mode and a non-zero mode. This usually takes place when $\rho_0 = 0$ or $\rho_1 = 0$, which means that mRNA can be produced only in the active gene state. For the multi-OFF model, the zero mode moves rightwards as L increases and thus the distribution has two non-zero modes when L is large. Note that the fact that the multi-OFF model may produce two non-zero modes even when $\rho_1 = 0$ has already been reported in [40]. For the multi-ON model, however, the zero mode still exists and the non-zero mode moves leftwards.

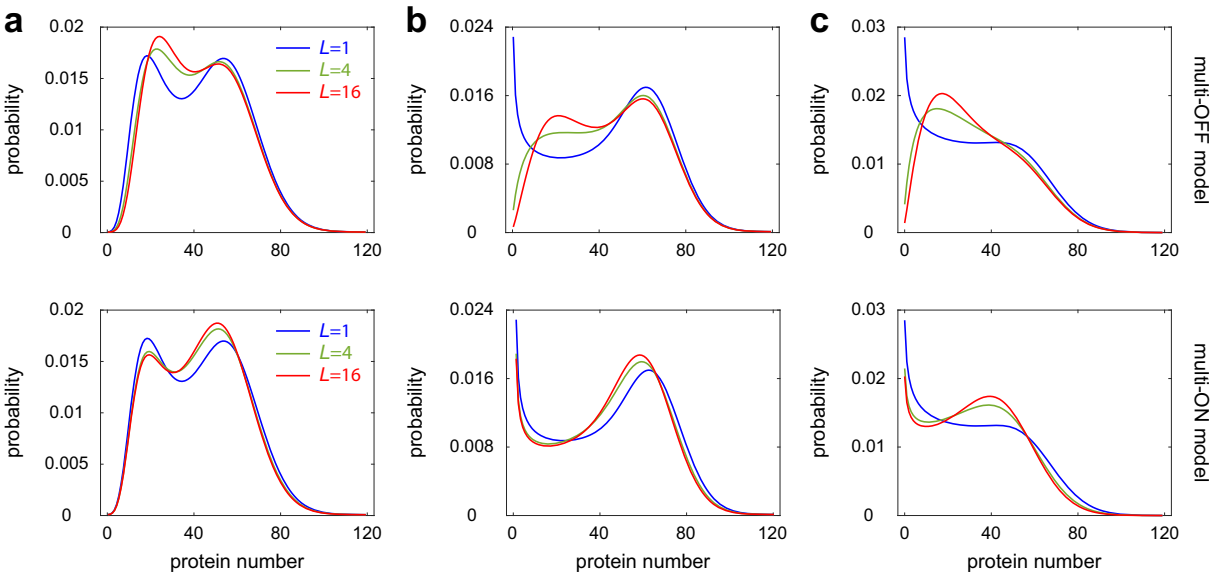

Figure 5: **Influence of molecular memory on the steady-state distributions.** (a) The case where the steady-state distribution is bimodal when $L = 1$ with two non-zero modes. The three curves show the distributions under different choices of L. The parameters are chosen as $u = 0.35, v = 0.46, \rho_0 = 42, \rho_1 = 10, B = 1.5$. (b) The case where the steady-state distribution is bimodal when $L = 1$ with a zero mode and a non-zero mode. The parameters are chosen as $u = 0.4, v = 0.7, \rho_0 = 70, \rho_1 = 0, B = 1$. (c) The case where the steady-state distribution is bimodal when $L = 1$ with an apparent zero mode and an inapparent non-zero mode. The parameters are chosen as $u = 0.8, v = 0.8, \rho_0 = 40, \rho_1 = 0, B = 1.5$.

We next use these distribution properties to explain the counterintuitive enlargement for the TB and SB regions in the phase diagram for the multi-ON model. In Fig. 5(c), the steady-state distribution is bimodal when $L = 1$ with the left mode being dominant and the right mode being inapparent. This usually occurs around the cusp of the SB region in Fig. 4(b). For the multi-OFF model, the left mode moves to the right as L increases and thus U

becomes more apparent when L is large. This is consistent with the observed shrinkage of the TB and SB regions in the phase diagram for the multi-OFF model. In contrast, for the multi-ON model, the right mode moves to the left and becomes higher with the increase of L. Since the left model is dominant, when L is large, the heights of the two modes become closer and bimodality becomes more apparent. This explains why the TB and SB regions enlarge in the phase diagram for the multi-ON model.

To summarize, we find that molecular memory has a marked impact on the dynamical behavior of the system when promoter switching is neither too fast nor too slow. Enhancing molecular memory in the inactive period reduces the TB and SB regions, while enhancing molecular memory in the active period enlarges them. Recent experiments have shown that in mammalian cells, the inactive periods for many genes has a non-exponential distribution [22, 23]. Our results suggest that this non-exponential behavior can significantly reduce transient and stationary bimodality, and thus stabilize both time-dependent and steady-state gene expression levels. This may explain why non-exponential inactive duration distributions are widely observed in nature.

6 Conclusions and discussion

In this paper, we derived the time-dependent distributions of mRNA and protein copy numbers for a multi-state gene expression model with complex promoter switching. Our model allows two different transcriptional rates, one corresponding to the transcriptionally active periods of the promoter and the other corresponding to the transcriptionally inactive periods. Moreover, our model takes into account two possibilities: (i) there are multiple inactive states of the promoter and only one active state; (ii) there are multiple active states and only one inactive state. The former corresponds to the case where the inactive period is non-exponentially distributed and thus has molecular memory, while the latter corresponds to the case where the active period has molecular memory. In each gene state, our model assumes that the gene product is produced in a non-bursty or bursty manner. The non-bursty case usually occurs for the mRNA dynamics while the bursty case usually occurs for the protein dynamics.

Following previous work, we transformed the CMEs satisfied by the probability distribution into the PDEs satisfied by the generating function. The key to our analytical approach is to make nontrivial temporal, spatial, and functional transformations to obtain a generalized hypergeometric differential equation satisfied by the (transformed) generating function. Once this is done, the generating function can be represented as the linear combinations of fundamental solutions of the generalized hypergeometric differential equation. Moreover, we focused on the special cases where the promoter switching dynamics is very fast or very slow. By using multiscale simplification techniques of Markov chains, we obtained the simplified analytical expressions of the time-dependent distributions in the fast and slow switching limiting cases when the gene product number is initially zero and the promoter initially starts from the steady-state distribution of all gene states.

We then investigated the dynamical behavior of the multi-state model. We identified three dynamical phases of the multi-state model that are associated with three distinct types of time-evolution: unimodality, transient bimodality, and stationary bimodality. To understand the influence of molecular memory on the gene product number distribution, we determined the dynamical phase diagram of the multi-state model as the strength L of molecular memory varies. Based on the analytical solution and the phase diagram, we showed that (i) molecular memory has no influence on the time-dependent distribution when the promoter switches rapidly or slowly between all gene states; (ii) molecular memory significantly affects the time-dependent distribution when the promoter switching dynamics is neither too fast nor too slow. In the latter case, we found that molecular memory remarkably enhances the regions of transient bimodality and stationary bimodality if the active period has a non-exponential distribution, while the opposite is true if the inactive period is non-exponentially distributed.

For simplicity and analytical tractability, here we have not included any feedback regulation and cell cycle events. While we have considered an implicit description of gene product dilution due to cell division, via the effective gene product decay rate, it has recently been shown that in some parameter regimes, this type of model cannot capture the stochastic dynamics predicted by models with an explicit description of the cell cycle [54]. We

hence anticipate that a more detailed gene expression model with the inclusion of feedback loops [16–20] and cell cycle events, such as cell growth, cell division, and gene replication [54–59] may display more complex behavior of time-evolution and may even introduce novel dynamical phases hitherto undescribed. These effects are currently under investigation.

Appendix A Derivation of Eqs. (10) and (11)

We first rewrite Eq. (8) in matrix form as

$$\begin{pmatrix} \mathscr{A} - \frac{\rho\omega}{1-\omega} + K_0 & -k_{10} & \cdots & -k_{L0} \\ -k_{01} & \mathscr{A} + K_1 & \cdots & -k_{L1} \\ \vdots & \vdots & \ddots & \vdots \\ -k_{0L} & -k_{1L} & \cdots & \mathscr{A} + K_L \end{pmatrix} \begin{pmatrix} g_0 \\ g_1 \\ \vdots \\ g_L \end{pmatrix} = 0,$$

where $K_i = \sum_{j \neq i} k_{ij}$ for $i = 0, 1, \ldots, L$. This equation, together with Eq. (9), shows that

$$\begin{pmatrix} v_\top \\ \mathscr{A} I - \overline{H} \end{pmatrix} \begin{pmatrix} g_1 \\ \vdots \\ g_L \end{pmatrix} = \begin{pmatrix} \mathscr{A} + K_0 \\ v_\perp^T \end{pmatrix} g_0 - \begin{pmatrix} \mathscr{A} \\ \mathbf{0}^T \end{pmatrix} g, \tag{31}$$

where $v_\top = (k_{10}, \ldots, k_{L0}), v_\perp = (k_{01}, \ldots, k_{0L})$, I is the identity matrix, $\overline{H} = H^T$ is the transpose of H, and $\mathbf{0}$ denotes the row vector whose components are all 0. Let S be an $(L-1) \times L$ operator-valued matrix defined by

$$S = \begin{pmatrix} v_\top \\ v_\top \mathscr{A} \\ v_\top \mathscr{A}^2 \\ \vdots \\ v_\top \mathscr{A}^{L-2} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ v_\top \overline{H} \\ v_\top \mathscr{A} \overline{H} \\ \vdots \\ v_\top \mathscr{A}^{L-3} \overline{H} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ v_\top \overline{H}^2 \\ \vdots \\ v_\top \mathscr{A}^{L-4} \overline{H}^2 \end{pmatrix} + \cdots + \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ v_\top \overline{H}^{L-3} \\ v_\top \mathscr{A} \overline{H}^{L-3} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ v_\top \overline{H}^{L-2} \end{pmatrix}$$
$$:= S_1 + S_2 + S_3 + \cdots + S_{L-2} + S_{L-1},$$

where the vectors $v_\top \mathscr{A}^i$ are defined by $v_\top \mathscr{A}^i = (k_{10}\mathscr{A}^i, \ldots, k_{L0}\mathscr{A}^i)$. Then we have

$$S(\mathscr{A} I - \overline{H}) = \mathscr{A} S_1 - (S_1 \overline{H} - \mathscr{A} S_2) - (S_2 \overline{H} - \mathscr{A} S_3) - \cdots - (S_{L-2}\overline{H} - \mathscr{A} S_{L-1}) - S_{L-1}\overline{H}. \tag{32}$$

Note that

$$S_1 \overline{H} - \mathscr{A} S_2 = \begin{pmatrix} v_\top \overline{H} \\ v_\top \mathscr{A} \overline{H} \\ v_\top \mathscr{A}^2 \overline{H} \\ \vdots \\ v_\top \mathscr{A}^{L-2} \overline{H} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ v_\top \mathscr{A} \overline{H} \\ v_\top \mathscr{A}^2 \overline{H} \\ \vdots \\ v_\top \mathscr{A}^{L-2} \overline{H} \end{pmatrix} = \begin{pmatrix} v_\top \overline{H} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}.$$

Similarly, we have

$$S_{i-1}\overline{H} - \mathscr{A} S_i = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ v_\top \overline{H}^{i-1} \\ v_\top \mathscr{A} \overline{H}^{i-1} \\ \vdots \\ v_\top \mathscr{A}^{L-i} \overline{H}^{i-1} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{0} \\ v_\top \mathscr{A} \overline{H}^{i-1} \\ \vdots \\ v_\top \mathscr{A}^{L-i} \overline{H}^{i-1} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ v_\top \overline{H}^{i-1} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}, \quad 3 \leq i \leq L-1.$$

Inserting the above equations into Eq. (32), we obtain

$$S(\mathscr{A}I-\overline{H})=\begin{pmatrix} v_\top\mathscr{A} \\ v_\top\mathscr{A}^2 \\ \vdots \\ v_\top\mathscr{A}^{L-1} \end{pmatrix}-\begin{pmatrix} v_\top\overline{H} \\ v_\top\overline{H}^2 \\ \vdots \\ v_\top\overline{H}^{L-1} \end{pmatrix}=v(\mathscr{A})^T v_\top-\begin{pmatrix} v_\top\overline{H} \\ v_\top\overline{H}^2 \\ \vdots \\ v_\top\overline{H}^{L-1} \end{pmatrix}, \tag{33}$$

where $v(\mathscr{A})=(\mathscr{A},\ldots,\mathscr{A}^{L-1})$. Next we define an $L\times(L+1)$ operator-valued matrix T as

$$T=\begin{pmatrix} -1 & \mathbf{0} \\ -v(\mathscr{A})^T & S \end{pmatrix}.$$

Then we obtain

$$T\begin{pmatrix} v_\top \\ \mathscr{A}I-\overline{H} \end{pmatrix}=\begin{pmatrix} -v_\top \\ -v(\mathscr{A})^T v_\top+S(\mathscr{A}I-\overline{H}) \end{pmatrix}=M,$$

where

$$M=-\begin{pmatrix} v_\top \\ v_\top\overline{H} \\ \vdots \\ v_\top\overline{H}^{L-1} \end{pmatrix} \tag{34}$$

is a constant matrix which is called the Krylov matrix [60]. Multiplying T on both sides of Eq. (31) then yields

$$\begin{aligned} M\begin{pmatrix} g_1 \\ \vdots \\ g_L \end{pmatrix} &= T\begin{pmatrix} \mathscr{A}+K_0 \\ v_\perp^T \end{pmatrix} g_0-T\begin{pmatrix} \mathscr{A} \\ \mathbf{0}^T \end{pmatrix} g \\ &:= b_1(\mathscr{A})^T g_0-b_2(\mathscr{A})^T g. \end{aligned} \tag{35}$$

Since g_0 can be expressed by g via Eq. (9), the above equation indicates that all $g_1,\ldots,g_L$ can also be expressed by g. This is exactly Eq. (10) in the main text. Clearly, the matrix T only involves terms $\mathscr{A}^i$ with $i\le L-1$. Hence both $b_1(\mathscr{A})$ and $b_2(\mathscr{A})$ involve terms $\mathscr{A}^i$ with $i\le L$. It then follows Eq. (9) that $b_1(\mathscr{A})^T g_0$ involves terms $\mathscr{A}^i g$ with $i\le L+1$. This shows that $b_1(\mathscr{A})^T g_0-b_2(\mathscr{A})^T g$ only involves terms $\mathscr{A}^i g$ with $i\le L+1$. To summarize, we can obtain from Eq. (35) that

$$M\begin{pmatrix} g_1 \\ \vdots \\ g_L \end{pmatrix}=B(\omega)\begin{pmatrix} g(t,\omega) \\ \mathscr{A}g(t,\omega) \\ \vdots \\ \mathscr{A}^{L+1}g(t,\omega) \end{pmatrix}, \tag{36}$$

where $B(\omega)$ is an $L\times(L+2)$ matrix with its entries being functions of ω.

In general, the matrix M defined in Eq. (34) may not be invertible. However, it is easy to check that M is always invertible when $L=1$ and is always invertible when $L=2$ for the multi-OFF and multi-ON models under all model parameters (see Appendix C for the proof). Even if M is not invertible, it can always be approximated by an invertible matrix with an arbitrary degree of accuracy. Thus, without loss of generality, we assume M to be invertible.

By using the Cramer's rule, we obtain

$$g_i=\frac{|M_i|}{|M|},\quad i=1,\ldots,L,$$

where $|M|$ represents the determinant of M, and M_i is the matrix formed by replacing the ith column of M by the vector $b_1(\mathscr{A})^T g_0 - b_2(\mathscr{A})^T g$. Since $g = g_0 + g_1 + \cdots + g_L$, we obtain

$$
\begin{aligned}
|M|g = |M|g_0 + \sum_{i=1}^{L} |M_i| &= \begin{vmatrix} g_0 & \mathbf{1} \\ -b_1(\mathscr{A})^T g_0 + b_2(\mathscr{A})^T g & M \end{vmatrix} \\
&= \begin{vmatrix} 1 & \mathbf{1} \\ T\begin{pmatrix} -\mathscr{A} - K_0 \\ -v_\perp^T \end{pmatrix} & T\begin{pmatrix} v_\top \\ \mathscr{A} I - \overline{H} \end{pmatrix} \end{vmatrix} g_0 - \begin{vmatrix} 0 & \mathbf{1} \\ T\begin{pmatrix} -\mathscr{A} \\ \mathbf{0}^T \end{pmatrix} & T\begin{pmatrix} v_\top \\ \mathscr{A} I - \overline{H} \end{pmatrix} \end{vmatrix} g \\
&:= \mathrm{I} - \mathrm{II},
\end{aligned}
\tag{37}
$$

where $\mathbf{1}$ denotes the row vector whose components are all 1. We first analyze the first part I. Note that

$$
\begin{aligned}
\mathrm{I} &= \begin{vmatrix} \mathbf{1} \\ T\begin{pmatrix} -\mathscr{A} - K_0 & v_\top \\ -v_\perp^T & \mathscr{A} I - \overline{H} \end{pmatrix} \end{vmatrix} g_0 = \begin{vmatrix} \mathbf{1} \\ \begin{pmatrix} -1 & \mathbf{0} \\ -v(\mathscr{A})^T & S \end{pmatrix} \begin{pmatrix} -\mathscr{A} - K_0 & v_\top \\ -v_\perp^T & \mathscr{A} I - \overline{H} \end{pmatrix} \end{vmatrix} g_0 \\
&= \begin{vmatrix} \mathbf{1} \\ \begin{pmatrix} -1 & \mathbf{0} \\ -v(\mathscr{A})^T & S \end{pmatrix} \begin{pmatrix} -1 & \mathbf{0} \\ \mathbf{0}^T & I \end{pmatrix} \begin{pmatrix} \mathscr{A} + K_0 & -v_\top \\ -v_\perp^T & \mathscr{A} I - \overline{H} \end{pmatrix} \end{vmatrix} g_0 \\
&= \begin{vmatrix} \mathbf{1} \\ \begin{pmatrix} 1 & \mathbf{0} \\ v(\mathscr{A})^T & S \end{pmatrix} \begin{pmatrix} \mathscr{A} + K_0 & -v_\top \\ -v_\perp^T & \mathscr{A} I - \overline{H} \end{pmatrix} \end{vmatrix} g_0,
\end{aligned}
\tag{38}
$$

where $\overline{K} = K^T$. Since $\mathbf{1}(\mathscr{A} I - \overline{K}) = \mathbf{1}\mathscr{A}$, we have

$$
\begin{aligned}
\mathscr{A}\mathrm{I} &= \begin{vmatrix} \mathbf{1}\mathscr{A} \\ \begin{pmatrix} 1 & \mathbf{0} \\ v(\mathscr{A})^T & S \end{pmatrix} \begin{pmatrix} \mathscr{A} + K_0 & -v_\top \\ -v_\perp^T & \mathscr{A} I - \overline{H} \end{pmatrix} \end{vmatrix} g_0 = \begin{vmatrix} \begin{pmatrix} 1 & \mathbf{1} \\ 1 & \mathbf{0} \\ v(\mathscr{A})^T & S \end{pmatrix} \left(\mathscr{A} I - \overline{K}\right) \end{vmatrix} g_0 \\
&= -\begin{vmatrix} \mathbf{1} \\ S \end{vmatrix} \left|\mathscr{A} I - \overline{K}\right| g_0.
\end{aligned}
\tag{39}
$$

To proceed, note that

$$
\begin{aligned}
\begin{vmatrix} \mathbf{1} \\ S \end{vmatrix} &= \begin{vmatrix} \mathbf{1} \\ v_\top \\ v_\top \mathscr{A} + v_\top \overline{H} \\ \vdots \\ v_\top \mathscr{A}^{L-2} + v_\top \mathscr{A}^{L-3}\overline{H} + \cdots + v_\top \mathscr{A}\overline{H}^{L-3} + v_\top \overline{H}^{L-2} \end{vmatrix} = - \begin{vmatrix} -\mathbf{1} \\ v_\top \\ v_\top \overline{H} \\ \vdots \\ v_\top \overline{H}^{L-2} \end{vmatrix} \\
&= -\left|-M\overline{H}^{-1}\right| = -|-M| \left|\overline{H}^{-1}\right| = (-1)^{L+1}|M|/|H|,
\end{aligned}
$$

where we have used the fact that $\mathbf{1}\overline{H} = -v_\top$. Since we have assumed that M is invertible, i.e. $|M| \neq 0$, it follows from the above equation that $|\mathbf{1}^T \;\; S^T| \neq 0$. Note that

$$
\left|\mathscr{A} I - \overline{K}\right| = \mathscr{A}\left(\mathscr{A} + \mu_1\right) \ldots \left(\mathscr{A} + \mu_L\right).
$$

Combining this equation with Eqs. (38) and (39), we obtain

$$
\mathrm{I} = -\begin{vmatrix} \mathbf{1} \\ S \end{vmatrix} (\mathscr{A} + \mu_1) \ldots (\mathscr{A} + \mu_L) g_0. \tag{40}
$$

We then analyze the second part II. Since $\mathbf{1}(\mathscr{A}I - \overline{H}) = \mathscr{A}I + v_\top$, we have

$$
\begin{aligned}
\mathrm{II} &= \begin{vmatrix} 0 & \mathbf{1} \\ T\begin{pmatrix} -\mathscr{A} \\ \mathbf{0}^T \end{pmatrix} & T\begin{pmatrix} v_\top \\ \mathscr{A}I - \overline{H} \end{pmatrix} \end{vmatrix} g = \begin{vmatrix} 0 & \mathbf{1} \\ \mathscr{A} & -v_\top \\ \mathscr{A}v(\mathscr{A})^T & -v(\mathscr{A})^T v_\top + S(\mathscr{A}I - \overline{H}) \end{vmatrix} g \\
&= \begin{vmatrix} 0 & \mathbf{1} \\ \mathscr{A} & -v_\top \\ \mathbf{0}^T & S(\mathscr{A}I - \overline{H}) \end{vmatrix} g = - \begin{vmatrix} \mathscr{A}\mathbf{1} + v_\top - v_\top \\ S(\mathscr{A}I - \overline{H}) \end{vmatrix} g \\
&= - \begin{vmatrix} \mathbf{1} \\ S \end{vmatrix} \left| \mathscr{A}I - \overline{H} \right| g + \begin{vmatrix} v_\top \\ S(\mathscr{A}I - \overline{H}) \end{vmatrix} g = - \begin{vmatrix} \mathbf{1} \\ S \end{vmatrix} \prod_{i=1}^{L} (\mathscr{A} + \lambda_i) g + \begin{vmatrix} v_\top \\ S(\mathscr{A}I - \overline{H}) \end{vmatrix} g.
\end{aligned}
\tag{41}
$$

It is easy to see from Eq. (33) that

$$
\begin{vmatrix} v_\top \\ S(\mathscr{A}I - \overline{H}) \end{vmatrix} = \begin{vmatrix} v_\top \\ v_\top \mathscr{A} - v_\top \overline{H} \\ \cdots \\ v_\top \mathscr{A}^{L-1} - v_\top \overline{H}^{L-1} \end{vmatrix} = \begin{vmatrix} v_\top \\ -v_\top \overline{H} \\ \cdots \\ -v_\top \overline{H}^{L-1} \end{vmatrix} = -|M|. \tag{42}
$$

Inserting Eqs. (40), (41), (42) into Eq. (37) and using the fact that $|\mathbf{1}^T \quad S^T| \neq 0$, we obtain

$$
\prod_{i=1}^{L} (\mathscr{A} + \lambda_i) g - \prod_{i=1}^{L} (\mathscr{A} + \mu_i) g_0 = 0. \tag{43}
$$

By induction and using Eq. (9), it is not difficult to prove that

$$
\prod_{i=1}^{L} (\mathscr{A} + \mu_i) g_0 = \left\{ \frac{\mathscr{A} \prod_{i=1}^{L} (\mathscr{A} - 1 + \mu_i)}{\rho\omega} - \frac{\mathscr{A} \prod_{i=1}^{L} (\mathscr{A} + \mu_i)}{\rho} \right\} g.
$$

Inserting the above equation into Eq. (43) yields

$$
\left[\mathscr{A} \prod_{i=1}^{L} (\mathscr{A} - 1 + \mu_i) - \omega \left(\mathscr{A} \prod_{i=1}^{L} (\mathscr{A} + \mu_i) + \rho \prod_{i=1}^{L} (\mathscr{A} + \lambda_i) \right) \right] g = 0.
$$

Note that

$$
\mathscr{A} \prod_{i=1}^{L} (\mathscr{A} + \mu_i) + \rho \prod_{i=1}^{L} (\mathscr{A} + \lambda_i) = \prod_{i=1}^{L+1} (\mathscr{A} + \gamma_i),
$$

where γ_i are determined by Eq. (12). We then obtain Eq. (11) in the main text. From Eq. (11), it is clear that we can express $\mathscr{A}^{L+1} g$ by $g, \mathscr{A} g, \ldots, \mathscr{A}^L g$. Combining Eqs. (9) and (36), we finally obtain Eq. (10) in the main text.

Appendix B Derivation of Eq. (21)

By the definition of elementary symmetric polynomials and Eq. (12), we have

$$
\begin{aligned}
\prod_{k=1}^{L+1} (x - \gamma_k) &= \sum_{k=0}^{L+1} \sigma_k(\gamma_1, \ldots, \gamma_{L+1}) (-1)^k x^{L+1-k} = x^{L+1} + \sum_{k=1}^{L+1} \sigma_k(\gamma_1, \ldots, \gamma_{L+1}) (-1)^k x^{L+1-k} \\
&= x^{L+1} + \sum_{k=1}^{L+1} \left[\sigma_k(\mu_1, \ldots, \mu_L) + \rho \sigma_{k-1}(\lambda_1, \ldots, \lambda_L) \right] (-1)^k x^{L+1-k} \\
&= x^{L+1} + \sum_{k=1}^{L} \sigma_k(\mu_1, \ldots, \mu_L) (-1)^k x^{L+1-k} - \rho \sum_{k=0}^{L} \sigma_k(\lambda_1, \ldots, \lambda_L) (-1)^k x^{L-k},
\end{aligned}
$$

where we have used $\sigma_{L+1}(\mu_1, \ldots, \mu_L) = 0$. Let f be a function defined by

$$f(x) = \frac{x^{L+1} + \sum_{k=1}^{L} \sigma_k(\mu_1, \ldots, \mu_L)(-1)^k x^{L+1-k}}{x^L + \sum_{k=1}^{L} \sigma_k(\lambda_1, \ldots, \lambda_L)(-1)^k x^{L-k}} - \rho.$$

Clearly, the function f has exactly $L+1$ zeros $\gamma_1, \ldots, \gamma_{L+1}$ in the complex plane. Moreover, it is easy to see that

$$\begin{aligned} f(x) &= x - \rho + \frac{\sum_{k=1}^{L} \left[\sigma_k(\mu_1, \ldots, \mu_L) - \sigma_k(\lambda_1, \ldots, \lambda_L)\right](-1)^k x^{L+1-k}}{x^L + \sum_{k=1}^{L} \sigma_k(\lambda_1, \ldots, \lambda_L)(-1)^k x^{L-k}} \\ &= x - \rho + \frac{\sum_{k=0}^{L-1} \left[\sigma_{k+1}(\mu_1, \ldots, \mu_L) - \sigma_{k+1}(\lambda_1, \ldots, \lambda_L)\right](-1)^{k+1} x^{L-k}}{\sum_{k=0}^{L} \sigma_k(\lambda_1, \ldots, \lambda_L)(-1)^k x^{L-k}} \\ &:= x - \rho + h(x). \end{aligned}$$

Since the numerator and denominator of $h(x)$ are polynomials of the same degree, we have

$$\lim_{|x| \to \infty} h(x) < \infty.$$

Hence there exists a constant $C > 0$ such that $|h(x)| \leq C$ when $|x|$ is sufficiently large. Thus for any $\epsilon > 0$, whenever $\rho \geq \rho_0 = C/\epsilon$, we have

$$\begin{aligned} f\left((1+\epsilon)\rho\right) &= \epsilon\rho + h\left((1+\epsilon)\rho\right) \geq \epsilon\rho - C \geq 0, \\ f\left((1-\epsilon)\rho\right) &= -\epsilon\rho + h\left((1+\epsilon)\rho\right) \leq -\epsilon\rho + C \leq 0. \end{aligned}$$

By the mean value theorem, there exists $x_0 \in [(1-\epsilon)\rho, (1+\epsilon)\rho]$ such that $f(x_0) = 0$. Since x_0 is a zero of f, for convenience, we denote it by γ_{L+1}. By the arbitrariness of ϵ, we have

$$\lim_{\rho \to \infty} \frac{\gamma_{L+1}}{\rho} = 1.$$

Along the same line, we can also prove that the above limit holds when $\rho \to -\infty$. Finally we obtain

$$\lim_{|\rho| \to \infty} \frac{\gamma_{L+1}}{\rho} = 1. \tag{44}$$

We then prove by induction that

$$\lim_{|\rho| \to \infty} \sigma_k(\gamma_1, \ldots, \gamma_L) = \sigma_k(\lambda_1, \ldots, \lambda_L), \tag{45}$$

for any $k = 1, \ldots, L$. To this end, we first prove that Eq. (45) holds for $k = L$. Setting $k = L+1$ in Eq. (12), we obtain

$$\prod_{i=1}^{L+1} \gamma_i = \rho \prod_{i=1}^{L} \lambda_i.$$

This equation, together with Eq. (44), shows that Eq. (45) holds for $k = L$. Then we prove the desired result by induction. Suppose that Eq. (45) holds for some $k = 2, \ldots, L$. Then it follows from Eq. (12) that

$$\sigma_k(\gamma_1, \ldots, \gamma_{L+1}) = \sigma_k(\mu_1, \ldots, \mu_L) + \rho\sigma_{k-1}(\lambda_1, \ldots, \lambda_L). \tag{46}$$

By the definition of elementary symmetric polynomials, it is easy to see that

$$\sigma_k(\gamma_1, \ldots, \gamma_L, \gamma_{L+1}) = \sigma_k(\gamma_1, \ldots, \gamma_L) + \gamma_{L+1}\sigma_{k-1}(\gamma_1, \ldots, \gamma_L).$$

Then it follows from Eq. (46) that

$$\sigma_k(\gamma_1, \ldots, \gamma_L) + \gamma_{L+1}\sigma_{k-1}(\gamma_1, \ldots, \gamma_L) = \sigma_k(\mu_1, \ldots, \mu_L) + \rho\sigma_{k-1}(\lambda_1, \ldots, \lambda_L).$$

Since Eq. (45) holds for k, it follows from Eq. (44) that

$$\lim_{|\rho|\to\infty} \sigma_{k-1}(\gamma_1,\ldots,\gamma_L) = \sigma_{k-1}(\lambda_1,\ldots,\lambda_L),$$

which shows that Eq. (45) holds for $k-1$. Therefore, Eq. (45) holds for any $k = 1,\ldots,L$. Finally, by the fundamental theory of algebra, we obtain from Eq. (45) that

$$\lim_{|\rho|\to\infty} \gamma_i = \lambda_i, \quad i = 1,\ldots,L.$$

Appendix C Time-dependent solution for the three-state model

For the three-state model with $L = 2$, we have

$$Q = \begin{pmatrix} -k_{01}-k_{02} & k_{01} & k_{02} \\ k_{10} & -k_{10}-k_{12} & k_{12} \\ k_{20} & k_{21} & -k_{20}-k_{21} \end{pmatrix}, \quad H = \begin{pmatrix} -k_{10}-k_{12} & k_{12} \\ k_{21} & -k_{20}-k_{21} \end{pmatrix}.$$

It follows from Eq. (15) that

$$g(t,\omega) = \phi_0(\omega e^{-t})y_0(\omega) + \phi_1(\omega e^{-t})y_1(\omega) + \phi_2(\omega e^{-t})y_2(\omega), \tag{47}$$

where

$$\begin{aligned} y_0(\omega) &= {}_3F_2(\gamma_1,\gamma_2,\gamma_3;\mu_1,\mu_2;\omega), \\ y_1(\omega) &= {}_3F_2(1+\gamma_1-\mu_1, 1+\gamma_2-\mu_1, 1+\gamma_3-\mu_1; 2-\mu_1, 1+\mu_2-\mu_1;\omega)\omega^{1-\mu_1}, \\ y_2(\omega) &= {}_3F_2(1+\gamma_1-\mu_2, 1+\gamma_2-\mu_2, 1+\gamma_3-\mu_2; 2-\mu_2, 1+\mu_1-\mu_2;\omega)\omega^{1-\mu_2}. \end{aligned}$$

It is easy to check that μ_i, λ_i, and γ_i are determined by the following equations:

$$\mu_1+\mu_2 = -\mathrm{tr}(Q), \quad \mu_1\mu_2 = \frac{1}{2}\left[\mathrm{tr}(Q)^2 - \mathrm{tr}(Q^2)\right], \quad \lambda_1+\lambda_2 = -\mathrm{tr}(H), \quad \lambda_1\lambda_2 = \det(H),$$

$$\gamma_1\gamma_2\gamma_3 = \rho\lambda_1\lambda_2, \quad \gamma_1\gamma_2+\gamma_1\gamma_3+\gamma_2\gamma_3 = \mu_1\mu_2+\rho(\lambda_1+\lambda_2), \quad \gamma_1+\gamma_2+\gamma_3 = \mu_1+\mu_2+\rho.$$

Straight forward computations show that Eq. (35) has the form of

$$M\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} -\mathscr{A}-k_{01}-k_{02} \\ -\mathscr{A}^2-(k_{01}+k_{02})\mathscr{A}+k_{01}k_{10}+k_{02}k_{20} \end{pmatrix} g_0 + \begin{pmatrix} \mathscr{A} \\ \mathscr{A}^2 \end{pmatrix} g, \tag{48}$$

where

$$M = \begin{pmatrix} -k_{10} & -k_{20} \\ k_{10}^2+k_{10}k_{12}-k_{12}k_{20} & k_{20}^2+k_{20}k_{21}-k_{10}k_{21} \end{pmatrix}$$

is a constant matrix. It is easy to check that

$$\det(M) = k_{10}^2k_{21} - k_{12}k_{20}^2 + k_{10}k_{20}(k_{10}+k_{12}-k_{20}-k_{21}).$$

For the multi-ON and multi-OFF models, we have $k_{02} = k_{21} = k_{10} = 0$ and thus $\det(M) = -k_{12}k_{20}^2$, which is always non-zero. Since the analytical solution in the general case is too complicated, we next only focus on the multi-ON and multi-OFF models. From Eq. (9), it is not difficult to prove that

$$\mathscr{A}^i g_0 = \left\{\frac{\mathscr{A}(\mathscr{A}-1)^i}{\rho\omega} - \frac{\mathscr{A}^{i+1}}{\rho}\right\} g, \quad i \geq 1.$$

Applying this equation, Eq. (48) becomes

$$\begin{pmatrix} 0 & -k_{20} \\ -k_{12}k_{20} & k_{20}^2 \end{pmatrix}\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} \frac{k_{01}(\omega-1)}{\rho\omega}+\frac{1}{\rho\omega}+1 & \frac{\omega-1}{\rho\omega} & 0 \\ \frac{k_{01}}{\rho\omega}-\frac{1}{\rho\omega} & \frac{k_{01}(\omega-1)}{\rho\omega}+\frac{2}{\rho\omega}+1 & \frac{\omega-1}{\rho\omega} \end{pmatrix}\begin{pmatrix} \mathscr{A} \\ \mathscr{A}^2 \\ \mathscr{A}^3 \end{pmatrix} g.$$

This shows that

$$g_2(t,\omega) = \left(\frac{k_{01}-1}{k_{20}\rho\omega} - \frac{k_{01}+\rho}{k_{20}\rho}\right)\mathscr{A}g(t,\omega) + \frac{1-\omega}{k_{20}\rho\omega}\mathscr{A}^2 g(t,\omega).$$

This equation, together with Eq. (9), yields

$$\begin{pmatrix} g_0(t,\omega)+g_1(t,\omega)+g_2(t,\omega) \\ g_0(t,\omega) \\ g_2(t,\omega) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1-\omega}{\rho\omega} & 0 \\ 0 & \frac{k_{01}-1}{k_{20}\rho\omega} - \frac{k_{01}+\rho}{k_{20}\rho} & \frac{1-\omega}{k_{20}\rho\omega} \end{pmatrix} \begin{pmatrix} g(t,\omega) \\ \mathscr{A}g(t,\omega) \\ \mathscr{A}^2 g(t,\omega) \end{pmatrix}.$$

Applying Eq. (16) to the above equation and then setting $t=0$, we obtain

$$\begin{pmatrix} y_0(\omega) & y_1(\omega) & y_2(\omega) \\ \frac{1-\omega}{\rho\omega}\vartheta y_0(\omega) & \frac{1-\omega}{\rho\omega}\vartheta y_1(\omega) & \frac{1-\omega}{\rho\omega}\vartheta y_2(\omega) \\ h_0(\omega) & h_1(\omega) & h_2(\omega) \end{pmatrix} \begin{pmatrix} \phi_0(\omega) \\ \phi_1(\omega) \\ \phi_2(\omega) \end{pmatrix} = \begin{pmatrix} g_0(0,\omega)+g_1(0,\omega)+g_2(0,\omega) \\ g_0(0,\omega) \\ g_2(0,\omega) \end{pmatrix}, \quad (49)$$

where

$$h_i(\omega) = \left(\frac{k_{01}-1}{k_{20}\rho\omega} - \frac{k_{01}+\rho}{k_{20}\rho}\right)\vartheta y_i(\omega) + \frac{1-\omega}{k_{20}\rho\omega}\vartheta^2 y_i(\omega), \quad i=0,1,2.$$

We then use Cramer's rule to solve the above system of linear equations. Let $A(\omega)$ denote the coefficient matrix of the above system of linear equations. It is easy to see that

$$\begin{aligned}\det(A(\omega)) &= \begin{vmatrix} y_0(\omega) & y_1(\omega) & y_2(\omega) \\ \frac{1-\omega}{\rho\omega}\vartheta y_0(\omega) & \frac{1-\omega}{\rho\omega}\vartheta y_1(\omega) & \frac{1-\omega}{\rho\omega}\vartheta y_2(\omega) \\ h_0(\omega) & h_1(\omega) & h_2(\omega) \end{vmatrix} = \begin{vmatrix} y_0(\omega) & y_1(\omega) & y_2(\omega) \\ \frac{1-\omega}{\rho}y_0'(\omega) & \frac{1-\omega}{\rho}y_1'(\omega) & \frac{1-\omega}{\rho}y_2'(\omega) \\ \frac{(1-\omega)\omega}{k_{20}\rho}y_0''(\omega) & \frac{(1-\omega)\omega}{k_{20}\rho}y_1''(\omega) & \frac{(1-\omega)\omega}{k_{20}\rho}y_2''(\omega) \end{vmatrix} \\ &= \frac{(1-\omega)^2\omega}{k_{20}\rho^2} W\left(y_0(\omega), y_1(\omega), y_2(\omega)\right).\end{aligned}$$

where $W(y_0(\omega), y_1(\omega), y_2(\omega))$ is the Wronskian of the functions $y_0(\omega)$, $y_1(\omega)$, and $y_2(\omega)$. We emphasize here that the Wronskian is defined with respect to the original derivative, instead of ϑ. The following lemma gives the analytical expression for the Wronskian.

Lemma 1. Let $y_0(\omega), y_1(\omega), \cdots, y_L(\omega)$ be the functions given in Eq. (14). Then for each $L \geq 1$, the Wronskian of these functions is given by

$$W\left(y_0(\omega), y_1(\omega), \ldots, y_L(\omega)\right) = \prod_{i=1}^{L}(1-\mu_i) \prod_{1\leq i<j\leq L}(\mu_i-\mu_j) \frac{(1-\omega)^{\mu_1+\cdots+\mu_L-\gamma_1-\cdots-\gamma_{L+1}-L}}{\omega^{\mu_1+\cdots+\mu_L+\frac{L(L-1)}{2}}}. \quad (50)$$

Proof. It has been shown in [61] that

$$W(y_0(\omega), y_1(\omega), \ldots, y_L(\omega)) = C\frac{(1-\omega)^{\mu_1+\cdots+\mu_L-\gamma_1-\cdots-\gamma_{L+1}-L}}{\omega^{\mu_1+\cdots+\mu_L+\frac{L(L-1)}{2}}}, \quad (51)$$

where C is an undetermined constant. Here we give the explicit expression of C. From Eq. (51), we have

$$C = \lim_{\omega\to 0} W(y_0(\omega), y_1(\omega), \ldots, y_L(\omega))\omega^{\mu_1+\cdots+\mu_L+\frac{L(L-1)}{2}}. \quad (52)$$

It is clear that $y_0(\omega) \to 1$ as $\omega \to 0$. For any $n \geq 1$, it is easy to see that $y_0^{(n)}(\omega) \to C_n$ as $\omega \to 0$ for some constant C_n. For any $i \geq 1$ we write $y_i(\omega) = \omega^{1-\mu_i}Y_i(\omega)$, where $Y_i(\omega)$ is a hypergeometric function which satisfies $Y_i(\omega) \to 1$ as $\omega \to 0$. Note that

$$y_i^{(n)}(\omega) = \sum_{k=0}^{n}\binom{n}{k}(\mu_i-1)_k(-1)^k\omega^{1-\mu_i-k}Y_i^{(n-k)}(\omega), \quad i \geq 1.$$

Thus for each $n \geq 0$, we have

$$\lim_{\omega \to 0} \frac{y_i^{(n)}(\omega)}{(\mu_i - 1)_n \omega^{1-\mu_i - n}(-1)^n} = 1, \quad i = 1, \ldots, L.$$

Using the above limit, we obtain

$$\begin{aligned}
1 &= \lim_{\omega \to 0} \frac{\begin{vmatrix} 1 & \omega^{1-\mu_1} & \omega^{1-\mu_2} & \ldots & \omega^{1-\mu_L} \\ C_1 & -(\mu_1 - 1)_1 \omega^{-\mu_1} & -(\mu_2 - 1)_1 \omega^{-\mu_2} & \ldots & -(\mu_L - 1)_1 \omega^{-\mu_L} \\ C_2 & (-1)^2(\mu_1 - 1)_2 \omega^{-\mu_1 - 1} & (-1)^2(\mu_2 - 1)_2 \omega^{-\mu_2 - 1} & \ldots & (-1)^2(\mu_L - 1)_2 \omega^{-\mu_L - 1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_L & (-1)^L(\mu_1 - 1)_L \omega^{1-\mu_1 - L} & (-1)^L(\mu_2 - 1)_L \omega^{1-\mu_2 - L} & \ldots & (-1)^L(\mu_L - 1)_L \omega^{1-\mu_L - L} \end{vmatrix}}{W(y_0(\omega), y_1(\omega), \ldots, y_L(\omega))} \\
&= \lim_{\omega \to 0} \frac{(-1)^{\frac{L(L+1)}{2}} \begin{vmatrix} (\mu_1 - 1)_1 & (\mu_2 - 1)_1 & \ldots & (\mu_L - 1)_1 \\ (\mu_1 - 1)_2 & (\mu_2 - 1)_2 & \ldots & (\mu_L - 1)_2 \\ \vdots & \vdots & \ddots & \vdots \\ (\mu_1 - 1)_L & (\mu_2 - 1)_L & \ldots & (\mu_L - 1)_L \end{vmatrix}}{W(y_0(\omega), y_1(\omega), \ldots, y_L(\omega)) \omega^{\mu_1 + \cdots + \mu_L + \frac{L(L-1)}{2}}} \\
&= \lim_{\omega \to 0} \frac{(-1)^{\frac{L(L+1)}{2}} \begin{vmatrix} \mu_1 - 1 & \mu_2 - 1 & \ldots & \mu_L - 1 \\ (\mu_1 - 1)^2 & (\mu_2 - 1)^2 & \ldots & (\mu_L - 1)^2 \\ \vdots & \vdots & \ddots & \vdots \\ (\mu_1 - 1)^L & (\mu_2 - 1)^L & \ldots & (\mu_L - 1)^L \end{vmatrix}}{W(y_0(\omega), y_1(\omega), \ldots, y_L(\omega)) \omega^{\mu_1 + \cdots + \mu_L + \frac{L(L-1)}{2}}} \\
&= \frac{(-1)^{\frac{L(L+1)}{2}} \prod_{i=1}^{L} (\mu_i - 1) \prod_{1 \leq j < i \leq L} (\mu_i - \mu_j)}{W(y_0(\omega), y_1(\omega), \ldots, y_L(\omega)) \omega^{\mu_1 + \cdots + \mu_L + \frac{L(L-1)}{2}}},
\end{aligned}$$

where the second equality follows from the expansion of the determinant with respect to the first row, the third equality follows from the fact that $(\mu_i - 1)_k = \sum_{j=1}^{k} C_j (\mu_i - 1)^j$ for some constants C_j that are independent of i, and the fourth equality follows from the expression of the Vandermonde determinant. Combining the above equation with Eq. (52) yields

$$C = (-1)^{\frac{L(L+1)}{2}} \prod_{i=1}^{L} (\mu_i - 1) \prod_{1 \leq j < i \leq L} (\mu_i - \mu_j) = \prod_{i=1}^{L} (1 - \mu_i) \prod_{1 \leq i < j \leq L} (\mu_i - \mu_j).$$

Inserting the above equation into Eq. (51), we finally obtain Eq. (50). □

By using Lemma 1, we immediately see that for $L = 2$,

$$\det(A(\omega)) = \frac{(1-\omega)^2 \omega}{k_{20} \rho^2} W(y_0(\omega), y_1(\omega), y_2(\omega)) = (1 - \mu_1)(1 - \mu_2)(\mu_1 - \mu_2) \frac{(1-\omega)^{-\rho}}{k_{20} \rho^2 \omega^{\mu_1 + \mu_2}}.$$

If the initial protein number is zero and the promoter is in state G_0, then the initial conditions are given by

$$g_1(0, \omega) = g_2(0, \omega) = 0, \quad g_0(0, \omega) = (1 - \omega)^{\rho_1}.$$

It then follows from Eq. (49) and Cramer's rule that

$$\begin{aligned}
\phi_0(\omega) &= \frac{\left[\vartheta y_1(\omega) \vartheta^2 y_2(\omega) - \vartheta y_2(\omega) \vartheta^2 y_1(\omega)\right] (1-\omega)^{\rho_0 + 2}}{(1 - \mu_1)(1 - \mu_2)(\mu_1 - \mu_2) \omega^{2 - \mu_1 - \mu_2}} + \frac{k_{20} \rho^2 [y_2(\omega) h_1(\omega) - y_1(\omega) h_2(\omega)](1-\omega)^{\rho_0}}{(1 - \mu_1)(1 - \mu_2)(\mu_1 - \mu_2) \omega^{-\mu_1 - \mu_2}}, \\
\phi_1(\omega) &= \frac{\left[\vartheta y_2(\omega) \vartheta^2 y_0(\omega) - \vartheta y_0(\omega) \vartheta^2 y_2(\omega)\right] (1-\omega)^{\rho_0 + 2}}{(1 - \mu_1)(1 - \mu_2)(\mu_1 - \mu_2) \omega^{2 - \mu_1 - \mu_2}} + \frac{k_{20} \rho^2 [y_0(\omega) h_2(\omega) - y_2(\omega) h_0(\omega)](1-\omega)^{\rho_0}}{(1 - \mu_1)(1 - \mu_2)(\mu_1 - \mu_2) \omega^{-\mu_1 - \mu_2}}, \\
\phi_2(\omega) &= \frac{\left[\vartheta y_0(\omega) \vartheta^2 y_1(\omega) - \vartheta y_1(\omega) \vartheta^2 y_0(\omega)\right] (1-\omega)^{\rho_0 + 2}}{(1 - \mu_1)(1 - \mu_2)(\mu_1 - \mu_2) \omega^{2 - \mu_1 - \mu_2}} + \frac{k_{20} \rho^2 [y_1(\omega) h_0(\omega) - y_2(\omega) h_0(\omega)](1-\omega)^{\rho_0}}{(1 - \mu_1)(1 - \mu_2)(\mu_1 - \mu_2) \omega^{-\mu_1 - \mu_2}}.
\end{aligned}$$

Inserting these equations into Eq. (47), we obtain the explicit expression $g(t, \omega)$. Finally, in terms of the original variables, the generating function $f(t, z)$ is given by Eq. (18).

Acknowledgements

We are grateful to Prof. Ramon Grima for stimulating discussions. C. J. acknowledges support from the NSAF grant in National Natural Science Foundation of China with grant No. U1930402.

References

[1] Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).

[2] Schnoerr, D., Sanguinetti, G. & Grima, R. Approximation and inference methods for stochastic biochemical kinetics — a tutorial review. *J. Phys. A: Math. Theor.* **50**, 093001 (2017).

[3] Paulsson, J. & Ehrenberg, M. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Phys. Rev. Lett.* **84**, 5447 (2000).

[4] Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).

[5] Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).

[6] Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

[7] Peccoud, J. & Ycart, B. Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.* **48**, 222–234 (1995).

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

[9] Shahrezaei, V. & Swain, P. S. Analytical distributions for stochastic gene expression. *Proc. Natl. Acad. Sci. USA* **105**, 17256–17261 (2008).

[10] Iyer-Biswas, S., Hayot, F. & Jayaprakash, C. Stochasticity of gene products from transcriptional pulsing. *Phys. Rev. E* **79**, 031911 (2009).

[11] Jia, C., Zhang, M. Q. & Qian, H. Emergent Lévy behavior in single-cell stochastic gene expression. *Phys. Rev. E* **96**, 040402 (2017).

[12] Paulsson, J. Models of stochastic gene expression. *Phys. Life Rev.* **2**, 157–175 (2005).

[13] Bokes, P., King, J. R., Wood, A. T. & Loose, M. Multiscale stochastic modelling of gene expression. *J. Math. Biol.* **65**, 493–520 (2012).

[14] Jia, C. Simplification of Markov chains with infinite state space and the mathematical theory of random gene expression bursts. *Phys. Rev. E* **96**, 032402 (2017).

[15] Cao, Z. & Grima, R. Linear mapping approximation of gene regulatory networks with stochastic dynamics. *Nat. Commun.* **9**, 1–15 (2018).

[16] Hornos, J. *et al.* Self-regulating gene: an exact solution. *Phys. Rev. E* **72**, 051907 (2005).

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

[18] Kumar, N., Platini, T. & Kulkarni, R. V. Exact distributions for stochastic gene expression models with bursting and feedback. *Phys. Rev. Lett.* **113**, 268105 (2014).

[19] Jia, C. & Grima, R. Small protein number effects in stochastic models of autoregulated bursty gene expression. *J. Chem. Phys.* **152**, 084115 (2020).

[20] Jia, C. & Grima, R. Dynamical phase diagram of an auto-regulating gene in fast switching conditions. *J. Chem. Phys.* **152**, 174110 (2020).

[21] Chong, S., Chen, C., Ge, H. & Xie, X. S. Mechanism of transcriptional bursting in bacteria. *Cell* **158**, 314–326 (2014).

[22] Suter, D. M. *et al.* Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472–474 (2011).

[23] Harper, C. V. *et al.* Dynamic analysis of stochastic transcription cycles. *PLoS Biol.* **9**, e1000607 (2011).

[24] Neuert, G. *et al.* Systematic identification of signal-activated stochastic gene regulation. *Science* **339**, 584–587 (2013).

[25] Senecal, A. *et al.* Transcription factors modulate c-Fos transcriptional bursts. *Cell reports* **8**, 75–83 (2014).

[26] Corrigan, A. M., Tunnacliffe, E., Cannon, D. & Chubb, J. R. A continuum model of transcriptional bursting. *Elife* **5**, e13051 (2016).

[27] Sepúlveda, L. A., Xu, H., Zhang, J., Wang, M. & Golding, I. Measurement of gene regulation in individual cells reveals rapid switching between promoter states. *Science* **351**, 1218–1222 (2016).

[28] Featherstone, K. *et al.* Spatially coordinated dynamic gene transcription in living pituitary tissue. *Elife* **5**, e08494 (2016).

[29] Shi, C., Jiang, Y. & Zhou, T. Queuing models of gene expression: Analytical distributions and beyond. *Biophysical Journal* **119**, 1606–1616 (2020).

[30] Pedraza, J. M. & Paulsson, J. Effects of molecular memory and bursting on fluctuations in gene expression. *Science* **319**, 339–343 (2008).

[31] Coulon, A., Gandrillon, O. & Beslon, G. On the spontaneous stochastic dynamics of a single gene: complexity of the molecular interplay at the promoter. *BMC Syst. Biol.* **4**, 1–18 (2010).

[32] Innocentini, G. d. C. P., Forger, M., Ramos, A. F., Radulescu, O. & Hornos, J. E. M. Multimodality and flexibility of stochastic gene expression. *Bull. Math. Biol.* **75**, 2600–2630 (2013).

[33] Jenkins, D. J., Finkenstädt, B. & Rand, D. A. A temporal switch model for estimating transcriptional activity in gene expression. *Bioinformatics* **29**, 1158–1165 (2013).

[34] Rieckh, G. & Tkačik, G. Noise and information transmission in promoters with multiple internal states. *Biophys. J.* **106**, 1194–1204 (2014).

[35] Hey, K. L. *et al.* A stochastic transcriptional switch model for single cell imaging data. *Biostatistics* **16**, 655–669 (2015).

[36] Dattani, J. & Barahona, M. Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization. *J. R. Soc. Interface* **14**, 20160833 (2017).

[37] Jia, C. Kinetic foundation of the zero-inflated negative binomial model for single-cell RNA sequencing data. *SIAM J. Appl. Math.* **80**, 1336–1355 (2020).

[38] Zhang, J., Chen, L. & Zhou, T. Analytical distribution and tunability of noise in a model of promoter progress. *Biophysical Journal* **102**, 1247–1257 (2012).

[39] Zhou, T. & Zhang, J. Analytical results for a multistate gene model. *SIAM J. Appl. Math.* **72**, 789–818 (2012).

[40] Zhang, J. & Zhou, T. Promoter-mediated transcriptional dynamics. *Biophys. J.* **106**, 479–488 (2014).

[41] Zoller, B., Nicolas, D., Molina, N. & Naef, F. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Mol. Syst. Biol.* **11**, 823 (2015).

[42] Herbach, U. Stochastic gene expression with a multistate promoter: Breaking down exact distributions. *SIAM J. Appl. Math.* **79**, 1007–1029 (2019).

[43] Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15**, 1263 (2008).

[44] Friedman, N., Cai, L. & Xie, X. S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006).

[45] Jia, C., Qian, H., Chen, M. & Zhang, M. Q. Relaxation rates of gene expression kinetics reveal the feedback signs of autoregulatory gene networks. *J. Chem. Phys.* **148**, 095102 (2018).

[46] Jia, C., Yin, G. G., Zhang, M. Q. *et al.* Single-cell stochastic gene expression kinetics with coupled positive-plus-negative feedback. *Phys. Rev. E* **100**, 052406 (2019).

[47] Olver, F. W., Lozier, D. W., Boisvert, R. F. & Clark, C. W. NIST Digital Library of Mathematical Functions (2017).

[48] Munsky, B. & Khammash, M. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.* **124**, 044104 (2006).

[49] Jia, C., Qian, M., Kang, Y. & Jiang, D. Modeling stochastic phenotype switching and bet-hedging in bacteria: stochastic nonlinear dynamics and critical state identification. *Quant. Biol.* **2**, 110–125 (2014).

[50] Bokes, P., King, J. R., Wood, A. T. & Loose, M. Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *J. Math. Biol.* **64**, 829–854 (2012).

[51] Bo, S. & Celani, A. Multiple-scale stochastic processes: decimation, averaging and beyond. *Phys. Rep.* **670**, 1–59 (2016).

[52] Jia, C. Model simplification and loss of irreversibility. *Phys. Rev. E* **93**, 052149 (2016).

[53] Jia, C. Reduction of Markov chains with two-time-scale state transitions. *Stochastics* **88**, 73–105 (2016).

[54] Beentjes, C. H. L., Perez-Carrasco, R. & Grima, R. Exact solution of stochastic gene expression models with bursting, cell

cycle and replication dynamics. *Phys. Rev. E* **101**, 032403 (2020).

[55] Lin, J. & Amir, A. Homeostasis of protein and mRNA concentrations in growing cells. *Nat. Commun.* **9**, 1–11 (2018).

[56] Cao, Z. & Grima, R. Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. *Proc. Natl. Acad. Sci. USA* **117**, 4682–4692 (2020).

[57] Nieto-Acuña, C., Arias-Castro, J. C., Vargas-García, C., Sánchez, C. & Pedraza, J. M. Correlation between protein concentration and bacterial cell size can reveal mechanisms of gene expression. *Phys. Biol.* **17**, 045002 (2020).

[58] Jia, C. & Grima, R. Frequency domain analysis of fluctuations of mRNA and protein copy numbers within a cell lineage: theory and experimental validation. *Phys. Rev. X* **11**, 021032 (2021).

[59] Jia, C., Singh, A. & Grima, R. Cell size distribution of lineage data: analytic results and parameter inference. *iScience* **24**, 102220 (2021).

[60] Xu, H. Functions of a matrix and Krylov matrices. *Linear algebra and its applications* **434**, 185–200 (2011).

[61] Gorelov, V. A. On algebraic identities between generalized hypergeometric functions. *Mathematical Notes* **88** (2010).