Steady-state joint distribution for first-order stochastic reaction kinetics

Youming Li [1,2] Da-Quan Jiang,[1,3,*] and Chen Jia [2,†]

[1]*LMAM, School of Mathematical Sciences, Peking University, Beijing 100871, China*
[2]*Applied and Computational Mathematics Division, Beijing Computational Science Research Center, Beijing 100193, China*
[3]*Center for Statistical Science, Peking University, Beijing 100871, China*

While the analytical solution for the marginal distribution of a stochastic chemical reaction network has been extensively studied, its joint distribution, i.e., the solution of a high-dimensional chemical master equation, has received much less attention. Here we develop an alternative method of computing the exact joint distributions of a wide class of first-order stochastic reaction systems in steady-state conditions. The effectiveness of our method is validated by applying it to four gene expression models of biological significance, including models with 2A peptides, nascent mRNA, gene regulation, translational bursting, and alternative splicing.

I. INTRODUCTION

Stochastic modeling of chemical reaction networks has attracted massive attention in recent years due to its wide applications in biology, chemistry, ecology, and epidemics [1]. If a reaction system is well mixed and the number of molecules is very large, then random fluctuations can be ignored and the evolution of concentrations of all chemical species can be modeled deterministically as a set of ordinary differential equations (ODEs) based on the law of mass action. If the chemical species are present in low numbers, however, then random fluctuations can no longer be ignored and the evolution of copies numbers of all species is usually modeled stochastically as a Markov jump process whose dynamics is governed by the well-known chemical master equation (CME). Thus far, stochastic chemical reaction networks have become a fundamental model for single-molecule enzymology [2,3] and single-cell gene expression dynamics [4]. Over the past two decades, the marginal distributions of stochastic reaction systems, such as Michalies-Menten enzyme kinetics [5,6], gene expression dynamics [7–9], and gene regulatory networks [10–16], have been studied extensively by solving the CME exactly or approximately based on various methods. These approaches include the generating function method [7], method of characteristics [8], multiscale techniques [17], moment closure approximation [18], moment convergence method [19], linear noise approximation [20], linear mapping approximation [21], etc.

The joint distribution of all chemical species for stochastic chemical reaction kinetics has received relatively little attention. Mathematically, the steady-state distribution of a reaction system corresponds to the eigenvector associated with the zero eigenvalue of the rate matrix of the underlying Markovian model. It can always be solved analytically when the rate matrix is finite-dimensional. However, for most reaction systems, the rate matrix is infinite-dimensional since the numbers of reactants are not bounded. In this case, simple approaches like diagonalization of the rate matrix usually fail. Due to the limitation of techniques, the joint distribution can only be solved for some particular systems. It has long been known [22,23] that (i) the steady-state joint distribution of a closed monomolecular system, which only includes reactions of the form $S_i \to S_j$, must be a multinomial distribution and (ii) the joint distribution of a detailed balanced reaction network is given by a product of Poissons [24]. Here detailed balance means that there is no net flux between any pair of reversible reactions. The CME for an open monomolecular system, which consists of synthesis reactions $\varnothing \to S_i$, degradation reactions $S_i \to \varnothing$, and conversion reactions $S_i \to S_j$, has also been solved exactly and the steady-state joint distribution is given by a product-form Poisson distribution [25–28]. Recently, this result has been extended to general stochastic reaction networks that are complex balanced. Here complex balance means that the flux flowing into each complex (see Ref. [29] for definition) is precisely balanced by the flux flowing out of that complex [29]. In fact, the steady-state joint distribution of a complex balanced reaction network is also given by a product-form Poisson-like distribution [30,31].

However, the condition of complex balance is very restrictive and not applicable to most systems of biological relevance. If complex balance is not satisfied, then the joint distribution has been analytically derived for hierarchic first-order reaction networks [32]. In the context of stochastic gene expression, the joint distribution for the copy numbers of mRNA and protein has been exactly solved for the two-stage model involving transcription and translation [33,34] and the joint distribution for the copy numbers of two mRNA isoforms has also been analytically derived in the presence of alternative splicing [35]. In addition, the joint distributions of gene expression models have also been studied using the linear noise approximation in the limit of large system size [20]. In most previous papers, the closed-form solution of the

*jiangdq@math.pku.edu.cn
†chenjia@csrc.ac.cn

joint distribution is computed by first converting the CME into a system of partial differential equations (PDEs) satisfied by the generating function and then solving the system of PDEs using the method of characteristics. However, this method is often very difficult to apply because of the tedious computations involved. Thus far, there is still a lack of a simple and effective approach that can be applied to a wide class of first-order reaction networks.
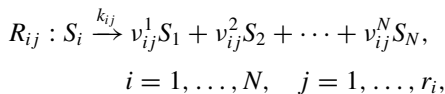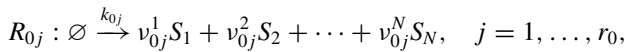
In this article, we propose an effective method of computing the joint distribution of a first-order reaction system in steady-state conditions. The key idea is to simplify the Markovian model of stochastic reaction kinetics to a modified Markovian model by allowing all zero-order reactions to occur only when all chemical species have zero copies. It turns out that the modified model has a much simpler state space and thus its joint distribution is much easier to solve. Once the modified model is solved analytically, the joint distribution of the original model is automatically obtained by making a simple transformation. Compared with the classical method of characteristics, our approach greatly reduces the theoretical complexity. The paper is organized as follows. In Sec. II, we describe the stochastic model of first-order reaction networks and introduce our method in detail. In Sec. III, we validate the effectiveness of our approach by applying it to four gene expression models of biological significance. These models include (i) a gene expression model involving 2A self-cleaving peptides, (ii) a multistep gene expression model involving nascent mRNA, (iii) a gene regulatory model involving translational bursting, and (iv) a multistep gene expression model involving alternative splicing. We conclude in Sec. IV.

II. MODEL AND METHODS

A chemical reaction involving a set of chemical species $S_1, \ldots, S_N$ can be written in the following general form:

$$\mu^1 S_1 + \mu^2 S_2 + \cdots + \mu^N S_N \xrightarrow{k} \nu^1 S_1 + \nu^2 S_2 + \cdots + \nu^N S_N,$$
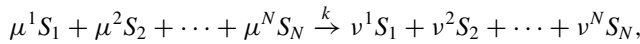
where μ^i and ν^i are nonnegative integers and k is the rate constant. The order of this reaction is the sum of coefficients of all the reactants, i.e., $\mu^1 + \mu^2 + \cdots + \mu^N$. Following the definition in Ref. [1], a reaction system is said to be *first-order* if it only consists of zero-order and first-order reactions. By definition, a first-order reaction system can be written in the following general form:

$$R_{0j} : \varnothing \xrightarrow{k_{0j}} \nu_{0j}^1 S_1 + \nu_{0j}^2 S_2 + \cdots + \nu_{0j}^N S_N, \quad j = 1, \ldots, r_0,$$

$$R_{ij} : S_i \xrightarrow{k_{ij}} \nu_{ij}^1 S_1 + \nu_{ij}^2 S_2 + \cdots + \nu_{ij}^N S_N,$$

$$i = 1, \ldots, N, \quad j = 1, \ldots, r_i,$$

where $R_{0j}, j = 1, \ldots, r_0$ are all zero-order reactions involved in the system and $R_{ij}, j = 1, \ldots, r_i$ are all first-order reactions associated with the reactant S_i. For convenience, we write $\nu_{ij} = (\nu_{ij}^1, \ldots, \nu_{ij}^N)$ for each $i = 0, 1, \ldots, N$ and $j = 1, \ldots, r_i$. A first-order reaction system can include synthesis reactions $\varnothing \to S_i$, degradation reactions $S_i \to \varnothing$, conversion reactions $S_i \to S_j$, catalytic reactions $S_i \to S_i + S_j$, and splitting reactions $S_i \to S_j + S_k$; hence, it can be widely applied

to model various naturally occurring systems in biology and physics.

We next focus on the stochastic dynamics of a first-order reaction network. The microstate of the system can be described by an ordered N-tuple $n = (n_1, \ldots, n_N)$, where n_i denotes the molecule number of S_i. Based on the law of mass action, the stochastic dynamics of the system can be described by a Markov jump process whose transition rates are given by

$$q_{n,n+\nu_{0j}} = k_{0j}, \quad 1 \leqslant j \leqslant r_0,$$
$$q_{n,n+\nu_{ij}-e_i} = k_{ij}n_i, \quad 1 \leqslant i \leqslant N, \quad 1 \leqslant j \leqslant r_i, \quad (1)$$

where $q_{n,n'}$ denotes the transition rate from microstate n to microstate n', ν_{0j} is the reaction vector of the zero-order reaction R_{0j}, i.e., the vector indicating the species change after the reaction, and $\nu_{ij} - e_i$ is the reaction vector of the first-order reaction R_{ij} with e_i being the vector whose ith component is 1 and all other components are zero.

Throughout this paper, we assume that the reaction system is ergodic, which guarantees that the system has a unique steady-state distribution. Let $p_n = p_{n_1, \cdots, n_N}$ denote the probability of observing microstate n. Then the evolution of the Markovian system is governed by the CME

$$\dot{p}_n = \sum_{i=1}^{N} \sum_{j=1}^{r_i} k_{ij} \big[(n_i + 1 - \nu_{ij}^i) p_{n+e_i-\nu_{ij}} - n_i p_n \big]$$
$$+ \sum_{j=1}^{r_0} k_{0j} \big[p_{n-\nu_{0j}} - p_n \big], \quad (2)$$

where the first term on the right-hand side corresponds to the occurrence of first-order reactions and the second term corresponds to the occurrence of zero-order reactions. To proceed, let

$$F(x_1, \cdots, x_n) = \sum_{n_1, \cdots, n_N} p_{n_1, \ldots, n_N} x_1^{n_1} \ldots x_N^{n_N}$$

denote the generating function associated with the joint distribution $p_{n_1, \ldots, n_N}$. Then F satisfies the following PDE [32]:

$$\frac{\partial F}{\partial t} = \sum_{i=1}^{N} \sum_{j=1}^{r_i} k_{ij} (x^{\nu_{ij}} - x_i) \frac{\partial F}{\partial x_i} + \sum_{j=1}^{r_0} k_{0j} (x^{\nu_{0j}} - 1) F. \quad (3)$$

A classical method of solving the CME is to first solve Eq. (3) to obtain the closed form of the generating function F and then recover the joint distribution p_n by taking the derivatives of F at zero. However, it is remarkably difficult to solve Eq. (3) analytically in most cases, even at the steady state.

Here we propose an alternative method of solving the CME in steady-state conditions. To this end, we construct a simpler Markov jump process called the *modified Markovian model*. The microstate of the modified model is still described by an ordered N-tuple $n = (n_1, n_2, \ldots, n_N)$. Note that for the original model, the zero-order reaction R_{0j} can lead to a transition from any microstate n to microstate $n + \nu_{0j}$; in other words, the zero-order reactions can occur at any microstate of the original model. However, for the modified model, we only allow the zero-order reactions to occur at the microstate $\mathbf{0} = (0, \ldots, 0)$, which is called the *zero microstate*, while the

first-order reactions follow the same transition rule as the original model. To summarize, the transition rates for the modified model are given as follows:

$$\tilde{q}_{n,n+v_{0j}} = \begin{cases} 0, & n \neq \mathbf{0}, \\ k_{0j}, & n = \mathbf{0}, \end{cases} \quad 1 \leqslant j \leqslant r_0,$$

$$\tilde{q}_{n,n+v_{ij}-e_i} = k_{ij}n_i, \quad 1 \leqslant i \leqslant N, \ 1 \leqslant j \leqslant r_i. \quad (4)$$

Comparing Eqs. (1) and (4), we can see that the modified model can be easily derived from the original one by eliminating those transitions from n to $n + v_{0j}$ for $n \neq \mathbf{0}$. Let $\pi_n = \pi_{n_1,n_2,\ldots,n_N}$ denote the probability of observing microstate n for the modified model and let

$$H(x_1, \ldots, x_N) = \sum_n \pi_{n_1,\ldots,n_N} x_1^{n_1} \ldots x_N^{n_N}$$

denote its generating function. Then the evolution of the modified model is then governed by the master equation

$$\dot{\pi}_n = \sum_{i=1}^{N} \sum_{j=1}^{r_i} \tilde{q}_{n+e_i-v_{ij},n} \pi_{n+e_i-v_{ij}} - \sum_{i=1}^{N} \sum_{j=1}^{r_i} \tilde{q}_{n,n+v_{ij}-e_i} \pi_n$$

$$+ \sum_{j=1}^{r_0} \tilde{q}_{n-v_{0j},n} \pi_{n-v_{0j}} - \sum_{j=1}^{r_0} \tilde{q}_{n,n+v_{0j}} \pi_n, \quad (5)$$

where the first two terms on the right-hand side correspond to first-order reactions and the last two terms correspond to zero-order reactions. We next make a crucial observation that if the system contains at least one zero-order reaction and both the original and modified models have reached the steady state, then the two generating functions F and H are related by (see Appendix A for the proof)

$$F(x_1, \ldots, x_N) = e^{\frac{H(x_1,\ldots,x_N)-1}{\pi_{\mathbf{0}}}}, \quad (6)$$

where $\pi_{\mathbf{0}}$ is the probability of observing the zero microstate for the modified model. In general, the modified model has a simpler transition diagram than the original model and thus the master equation for the former is much easier to solve. Once we have obtained the generating function H of the modified model, we can use Eq. (6) to compute the generating function F of the original model. Finally, the steady-state joint distribution for the copy numbers of all chemical species can be recovered by taking the derivatives of F at zero, i.e.,

$$p_n = \frac{1}{n_1! \cdots n_N!} \frac{\partial^{n_1+\cdots+n_N} F}{\partial x_1^{n_1} \ldots \partial x_N^{n_N}}(0, \ldots, 0).$$

We summarize the above method as follows: first, we construct the modified model (which is usually much simpler than the original model) and compute its steady-state joint distribution π_n; next, we calculate the generating function H of the modified model and use Eq. (6) to compute the generating function F of the original model; finally, we recover the steady-state joint distribution of the original model by taking the derivatives of F. We emphasize that Eq. (6) does not hold if the two models have not reached the steady state. In fact, the proof of Eq. (6) relies on the close relationship between the partial derivatives of F and H with respect to x_i in steady-state conditions, while for the time-dependent case, we need to take the partial derivatives with respect to t into consideration, which invalidates our approach (see Appendix A for details).

We next focus on the transition diagrams of the two models. Recall that a microstate n is called recurrent if there exists a path in the transition diagram that starts from n and returns to itself; otherwise it is called transient. Actually, transient microstates contribute nothing to the steady-state probabilities and thus the steady-state distribution is only concentrated on the collection of all recurrent microstates, which is called the irreducible state space [36,37]. Hence, for both models, we only need to focus on the irreducible state space, instead of the whole state space. We now use a simple example to show the relationship between the two models. Consider the following open monomolecular system:

$$\varnothing \underset{d_1}{\overset{k_1}{\rightleftharpoons}} P_1 \overset{k_2}{\rightarrow} P_2 \overset{d_2}{\rightarrow} \varnothing. \quad (7)$$
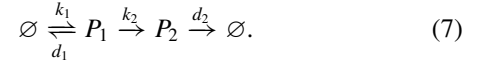
If we regard P_1 and P_2 as two conformational states of a protein, then this reaction scheme describes the synthesis, degradation, and conformational changes for the protein. The transition diagrams for the original and modified models associated with this reaction scheme are depicted in Fig. 1. The difference between them is that the zero-order reaction $\varnothing \rightarrow P_1$ (blue arrows) can occur at any microstate for the original model, but it can only occur at only the zero microstate for the modified model. It can be seen from Fig. 1(b) that the modified model has many transient microstates. The irreducible state space (the collection of all recurrent microstates) for the original model is the whole two-dimensional nonnegative integer lattice, while the irreducible state space for the modified model is simply the collection of the following three microstates:

$$\{(0, 0), (1, 0), (0, 1)\},$$

which is much simpler than that of the original model. Once the modified model enters the irreducible state space, it can never leave it anymore. Since the steady-state distribution of the modified model is only concentrated on the irreducible state space which contains only three microstates, we immediately obtain

$$\pi_{0,0} = \frac{1}{1+a+b}, \quad \pi_{1,0} = \frac{a}{1+a+b}, \quad \pi_{0,1} = \frac{b}{1+a+b},$$

where $a = k_1/(k_2 + d_1)$ and $b = k_1 k_2/(k_2 + d_1)d_2$. Thus, the generating function of the modified model is given by $H(x_1, x_2) = \pi_{0,0} + \pi_{1,0}x_1 + \pi_{0,1}x_2$. It then follows from Eq. (6) that the generating function of the original model is given by

$$F(x_1, x_2) = e^{\frac{\pi_{0,0}+\pi_{1,0}x_1+\pi_{0,1}x_2-1}{\pi_{0,0}}} = e^{a(x_1-1)+b(x_2-1)}.$$

Then the steady-state joint distribution for the copy numbers of P_1 and P_2 can be recovered by taking the derivatives of F at zero, which finally gives

$$p_{n_1,n_2} = \frac{a^{n_1}b^{n_2}}{n_1!n_2!}e^{-(a+b)}.$$

Note that this is the product of two Poisson distributions. In fact, it has been shown in Ref. [26] that the joint distribution of an open monomolecular system must be a product of Poissons, which is consistent with our result. However, compared with the derivation in Ref. [26], our method is much simpler.
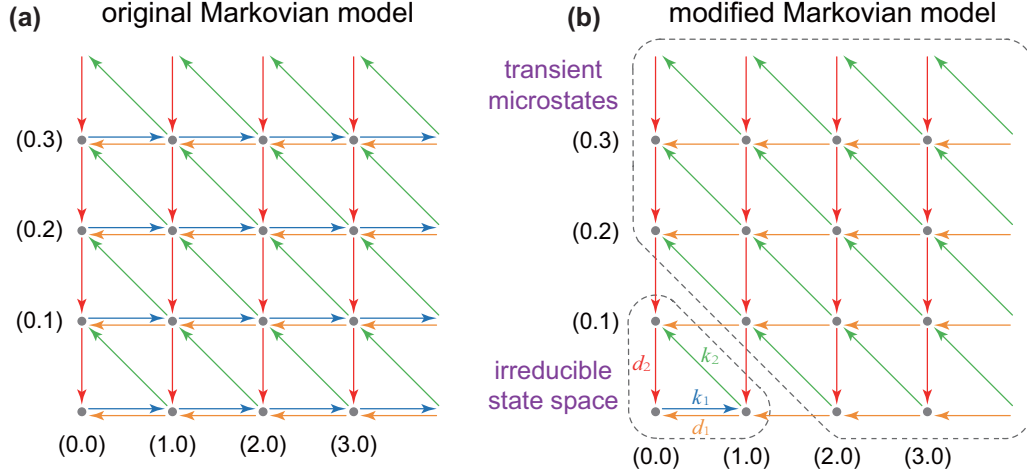
FIG. 1. Transition diagrams of the original and modified Markovian models for the reaction scheme given in Eq. (7). The red arrows correspond to the reaction $P_2 \rightarrow \varnothing$, the green arrows correspond to $P_1 \rightarrow P_2$, the blue arrows correspond to $\varnothing \rightarrow P_1$, and the orange arrows correspond to $P_1 \rightarrow \varnothing$. For the modified model, we only allow the zero-order reaction (blue arrows) to occur at the zero microstate. The irreducible state space of the modified model only consists of three microstates: (0,0), (1,0), and (0,1).

Our method can also be used to compute many other quantities of interest. First, the steady-state marginal distribution for the copy number of any chemical species can be easily computed. To see this, let $p_{n_i}^i$ denote the steady-state probability of having n_i copies of S_i. Then the marginal distribution can be recovered from the generating function F as

$$p_{n_i}^i = \frac{1}{n_i!} \frac{\partial^{n_i} F}{\partial x_i^{n_i}} (1, \cdots, 0, \cdots, 1), \qquad (8)$$

where $(1, \cdots, 0, \cdots, 1)$ is the vector whose ith component is 0 and other components are all 1. Note that the generating function given in Eq. (6) is a composite function. The following Faà di Bruno's formula [38] gives the explicit expression for the higher-order derivatives of a composite function:

$$\frac{d^n}{dx^n} f[g(x)]$$
$$= \sum_{k=1}^{n} f^{(k)}[g(x)] B_{n,k}[g'(x), g''(x), \ldots, g^{(n-k+1)}(x)],$$

where $B_{n,k}(x_1, \ldots, x_{n-k+1})$ is the incomplete Bell polynomial [39]. The above two equations, together with Eq. (6), give the following analytical expression for the marginal distributions of all species:

$$p_{n_i}^i = \frac{B_{n_i}(g_{i,1}, g_{i,2}, \cdots, g_{i,n_i})}{n_i!} e^{\frac{H(1,\cdots,0,\cdots,1)-1}{\pi_0}}, \qquad (9)$$

where

$$B_n(x_1, \cdots, x_n) = \sum_{k=1}^{n} B_{n,k}(x_1, \cdots, x_{n-k+1})$$

is the complete Bell polynomial [39], and

$$g_{i,k} = \frac{1}{\pi_0} \frac{\partial^k H}{\partial x_i^k} (1, \cdots, 0, \cdots, 1), \quad k = 1, \ldots, n_i.$$

In addition, the steady-state mean and variance for the copy number of S_i can be obtained as

$$\langle n_i \rangle = \frac{\partial F}{\partial x_i} (1, \ldots, 1),$$

$$\sigma_{n_i}^2 = \left[\frac{\partial^2 F}{\partial x_i^2} + \frac{\partial F}{\partial x_i} - \left(\frac{\partial F}{\partial x_i} \right)^2 \right] (1, \ldots, 1),$$

where $\sigma_{n_i}^2 = \langle n_i^2 \rangle - \langle n_i \rangle^2$ denotes the copy number variance of S_i. Finally, the steady-state covariance for the copy numbers of any pair of chemical species S_i and S_j can be computed as

$$\mathrm{Cov}(n_i, n_j) = \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle$$

$$= \left[\frac{\partial^2 F}{\partial x_i \partial x_j} - \frac{\partial F}{\partial x_i} \frac{\partial F}{\partial x_j} \right] (1, \ldots, 1).$$

In particular, the correlation coefficient between the copy numbers of S_i and S_j is given by

$$\rho_{n_i, n_j} = \frac{\mathrm{Cov}(n_i, n_j)}{\sigma_{n_i} \sigma_{n_j}}. \qquad (10)$$

These formulas will be used to analyze the dynamic properties of some important gene expression models in what follows.

We have seen from the previous example that our method is particularly effective when the modified model has a finite irreducible state space. A natural question is when this occurs. To answer this, we recall that a family of reactions

$$R_i : \mu_i^1 S_1 + \cdots + \mu_i^N S_N \xrightarrow{k_i} \nu_i^1 S_1 + \cdots + \nu_i^N S_N, \quad i = 1, \ldots, r$$
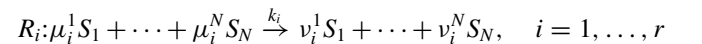
has a conservation law, if there exists a nonzero vector $\omega = (\omega_1, \ldots, \omega_N)$ such that

$$\omega_1 \mu_i^1 + \omega_2 \mu_i^2 + \cdots + \omega_N \mu_i^N = \omega_1 \nu_i^1 + \omega_2 \nu_i^2 + \cdots + \omega_N \nu_i^N$$

for all $i = 1, \ldots, r$. In Appendix B, we prove that if all the first-order reactions except degradation reactions have a conservation law with positive coefficients $\omega_1, \cdots, \omega_N > 0$, then the modified model must have a finite irreducible state space. To verify this criterion, we apply it to the reaction scheme
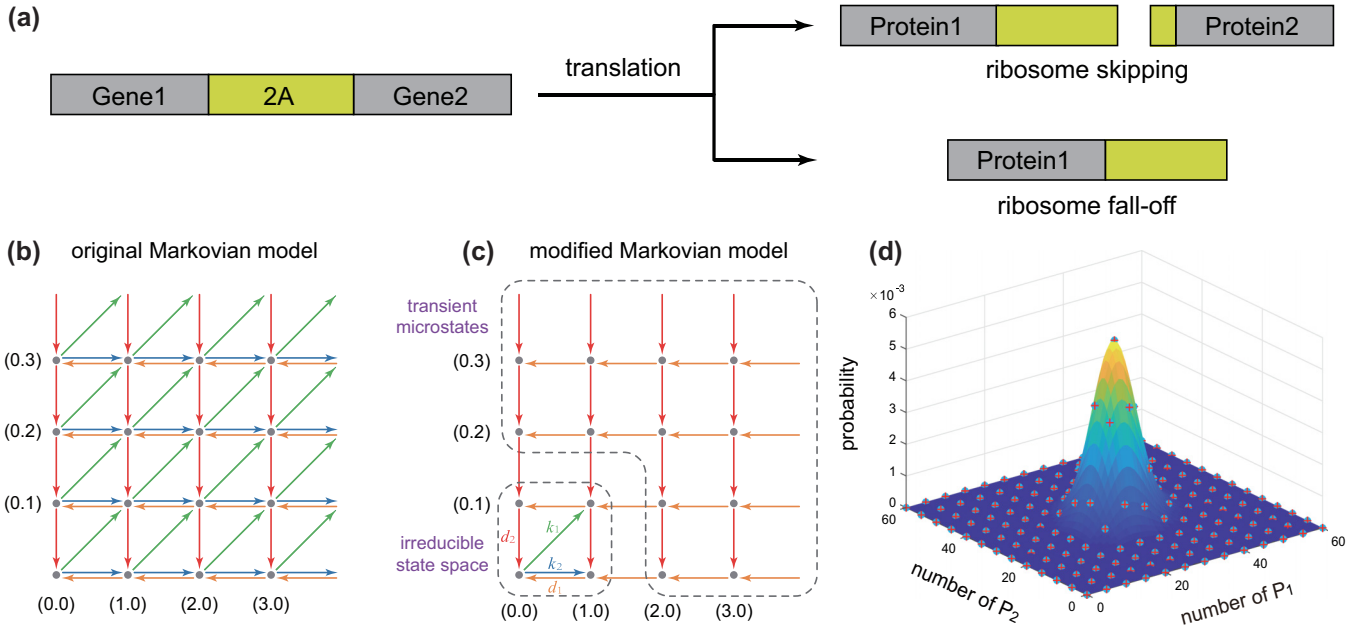
FIG. 2. A gene expression model involving 2A peptides. (a) Translation mechanism of two genes with the coding region of a 2A peptide (2A) inserted in between. There are two possibilities: ribosome skipping results in two cleaved proteins and ribosome fall-off results in only the protein upstream of the 2A [42]. (b, c) Transition diagrams for the original and modified models. The green arrows correspond to the reaction $G \to G + P_1 + P_2$, the blue arrows correspond to $G \to G + P_1$, the red arrows correspond to $P_2 \to \varnothing$, and the orange arrows correspond to $P_1 \to \varnothing$. (d) Comparison of the analytical steady-state joint distribution for the numbers of the two proteins given in Eq. (13) (colored surface) with the numerical simulations obtained using FSP (red plus signs) and stochastic simulations obtained using SSA (light blue dots). Here SSA is performed by generating 80 000 stochastic trajectories. The model parameters are chosen as $k_1 = 30$, $k_2 = 30$, $d_1 = 2$, $d_2 = 1$.

given in Eq. (7). For this reaction system, there are three first-order reactions:
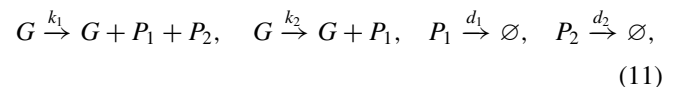
$$P_1 \to P_2, \quad P_1 \to \varnothing, \quad P_2 \to \varnothing.$$
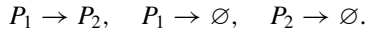
Among these reactions, only $P_1 \to P_2$ is not a degradation reaction and obviously, it has a conservation law with positive coefficients $\omega_1 = \omega_2 = 1$ since the total number of P_1 and P_2 is invariant. It then follows from the above criterion that the corresponding modified model has a finite irreducible state space, which is consistent with the previous discussion. Before leaving this section, we emphasize that if a family of reactions contains a first-order catalytic reaction such as $S_i \to S_i + S_j$, which appears in many biochemical systems, then the family of reactions can never have a conservation law with positive coefficients. In this case, the modified model may have an infinite irreducible state space. Fortunately, for many biochemical systems involving catalytic reactions, our method is still applicable, although the computation will be more complicated than the case of finite irreducible state space. In the next section, we shall apply our method to compute the steady-state joint distributions of mRNAs and/or proteins in four gene expression models of biological significance.

III. APPLICATIONS

A. Gene expression model with 2A self-cleaving peptides

As the first application, we consider a gene expression system involving 2A self-cleaving peptides, also called 2A peptides. Biologically, 2A peptides are 18–22 amino-acid-long oligopeptides derived from a wide range of viral families

[40,41] that mediate cleavage of polypeptides during translation in eukaryotic cells [42], and therefore enable the synthesis of several gene products (proteins) from a single transcript. For this reason, 2A peptides are widely used in genetic engineering to cleave a long peptide into two shorter peptides. Specifically, the coding region of a 2A peptide (2A) is inserted between the coding regions of two proteins [Fig. 2(a)]. The mechanism of 2A-mediated self-cleavage was recently discovered to be ribosome skipping the formation of a peptide bond at the C-terminus of the 2A [43,44]. There are two possibilities for a 2A-mediated skipping event: (i) successful skipping and recommencement of translation results in two cleaved proteins: the protein upstream of the 2A is attached to the complete 2A peptide except for the C-terminal proline, while the protein downstream of the 2A is attached to one proline at the N-terminus; (ii) successful skipping but ribosome fall-off and discontinued translation results in only the protein upstream of the 2A [42]. Then the effective reactions describing the gene expression system are given by

$$G \xrightarrow{k_1} G + P_1 + P_2, \quad G \xrightarrow{k_2} G + P_1, \quad P_1 \xrightarrow{d_1} \varnothing, \quad P_2 \xrightarrow{d_2} \varnothing,$$
$$(11)$$

where G is the coding region illustrated in Fig. 2(a) and P_1 and P_2 are two proteins. The first reaction describes ribosome skipping, the second reaction describes ribosome fall-off, and the remaining two reactions describe the degradation of the two proteins. The microstate of the system can be represented by an ordered pair (n_1, n_2), where n_i denotes the copy number of P_i. Let p_{n_1, n_2} denote the probability of observing microstate

(n_1, n_2) and let

$$F(x_1, x_2) = \sum_{n_1, n_2} p_{n_1, n_2} x_1^{n_1} x_2^{n_2}$$

denote the corresponding generating function. Then the stochastic gene expression dynamics can be described by a Markov jump process with transition diagram illustrated in Fig. 2(b). The evolution of the Markovian system is governed by the CME

$$\dot{p}_{n_1, n_2} = k_1 p_{n_1-1, n_2-1} + k_2 p_{n_1-1, n_2} + d_1(n_1 + 1) p_{n_1+1, n_2}$$
$$+ d_2(n_2 + 1) p_{n_1, n_2+1} - (k_1 + k_2 + d_1 n_1 + d_2 n_2) p_{n_1, n_2}.$$

To solve this CME, we consider the modified Markovian model. We emphasize here that we do not take copy number variation of the gene into account and hence the first two reactions in Eq. (11) can be regarded as zero-order reactions. The reason why we explicitly write out the gene G, instead of using $\varnothing$, in the first two reactions is to stress that proteins are produced from genes. Since the zero-order reactions can only occur at the zero microstate, the modified model has the transition diagram illustrated in Fig. 2(c). While the transition diagram of the modified model is complicated, the irreducible state space is actually finite and only contains the following four microstates:

$$\{(0, 0), (1, 1), (1, 0), (0, 1)\}.$$

Since the steady-state distribution of the modified model is concentrated on the irreducible state space which contains only four microstates, it can be easily computed as

$$\pi_{0,0} = \frac{\alpha_0}{\alpha}, \quad \pi_{1,0} = \frac{\alpha_1}{\alpha}, \quad \pi_{0,1} = \frac{\alpha_2}{\alpha}, \quad \pi_{1,1} = \frac{\alpha_{12}}{\alpha},$$

where

$$\alpha_0 = 1, \quad \alpha_1 = \frac{k_2(d_1 + d_2) + k_1 d_2}{d_1(d_1 + d_2)}, \quad \alpha_2 = \frac{k_1 d_1}{d_2(d_1 + d_2)},$$

$$\alpha_{12} = \frac{k_1}{d_1 + d_2},$$

and $\alpha = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_{12}$. Then the generating function of the modified model is given by

$$H(x_1, x_2) = \frac{1}{\alpha}(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_{12} x_1 x_2).$$

It then follows from Eq. (6) that the generating function of the original model is given by

$$F(x_1, x_2) = e^{\frac{H(x_1, x_2)-1}{\pi_{0,0}}} = e^{\alpha_1(x_1-1)+\alpha_2(x_2-1)+\alpha_{12}(x_1 x_2-1)}. \quad (12)$$

This shows that the copy numbers of the two proteins have a bivariate Poisson distribution [45], which can be recovered from F by taking the derivatives:

$$p_{n_1, n_2} = \frac{1}{n_1! n_2!} \frac{\partial^{n_1+n_2} F}{\partial x_1^{n_1} \partial x_2^{n_2}}(0, 0)$$
$$= \sum_{i=0}^{n_1 \wedge n_2} \frac{\alpha_1^{n_1-i} \alpha_2^{n_2-i} \alpha_{12}^i}{i!(n_1-i)!(n_2-i)!} e^{-(\alpha_1+\alpha_2+\alpha_{12})}, \quad (13)$$

where $n_1 \wedge n_2$ denotes the smaller one of n_1 and n_2. Taking $x_2 = 1$ and $x_1 = 1$ in Eq. (12), we obtain

$$F(x_1, 1) = e^{(\alpha_1+\alpha_{12})(x_1-1)}, \quad F(1, x_2) = e^{(\alpha_2+\alpha_{12})(x_2-1)}.$$

It then follows from Eq. (8) that the steady-state marginal distributions for the two proteins are given by

$$p_{n_1}^1 = \frac{(\alpha_1 + \alpha_{12})^{n_1}}{n_1!} e^{-(\alpha_1+\alpha_{12})}, \quad p_{n_2}^2 = \frac{(\alpha_2 + \alpha_{12})^{n_2}}{n_2!} e^{-(\alpha_2+\alpha_{12})}. \quad (14)$$

This shows that both proteins have a marginal Poisson distribution but their joint distribution is not the product of two Poisson distributions (note that a bivariate Poisson distribution may not be the product of two Poissons). This reaction system should be compared with complex balanced networks. In fact, it was shown in Ref. [30] that if a reaction system is complex balanced, then the copy numbers of all chemical species must have a product-form Poisson distribution in steady-state conditions, which is very different from the nonproduct-form Poisson distribution studied here.

To validate our analytical solution, we compare it with the numerical solutions obtained using the finite-state projection algorithm (FSP) [46] and the stochastic simulation algorithm (SSA), as illustrated in Fig. 2(d). When using FSP, we truncate the state space at large enough N_1 and N_2, with N_1 and N_2 being the truncation sizes for n_1 and n_2, respectively, and then solve the normalized eigenvector of the truncated rate matrix corresponding to the zero eigenvalue numerically using MATLAB. The truncation sizes are chosen to be $N_1 = 5(k_1 + k_2)/d_1$ and $N_2 = 5k_1/d_2$. Since $(k_1 + k_2)/d_1$ and k_1/d_2 are the typical copy numbers for proteins P_1 and P_2, respectively, the probability that the protein numbers are outside the truncation region is very small and practically can always be ignored. It can be seen that the analytical solution coincides perfectly with both FSP and SSA. Our analytical results can also be used to analyze the correlation between the two proteins. It follows from Eqs. (10) and (12) that the correlation coefficient between the numbers of P_1 and P_2 is given by

$$\rho_{P_1, P_2} = \frac{\alpha_{12}}{\sqrt{(\alpha_1 + \alpha_{12})(\alpha_2 + \alpha_{12})}}$$
$$= \frac{1}{\sqrt{\left(1 + \frac{k_2}{k_1}\right)\left(1 + \frac{d_1}{d_2}\right)\left(1 + \frac{d_2}{d_1}\right)}}.$$

Clearly, the numbers of the two proteins are always positively correlated and their correlation coefficient has the upper bound

$$\rho_{P_1, P_2} \leqslant \frac{1}{2\sqrt{1 + \frac{k_2}{k_1}}},$$

where the equality holds if and only if $d_1 = d_2$. This means that the correlation is the strongest when the degradation rates of the two proteins are equal. In addition, we can see that the correlation coefficient is always smaller than 0.5 and is comparatively large when the degradation rates of the two proteins are close to each other, i.e., $d_1 \approx d_2$, and when the translation rate due to ribosome skipping is much larger compared to the translation rate due to ribosome fall-off, i.e., $k_1 \gg k_2$. Before leaving this section, we point out that biologically, it is possible to generate three or more cleaved proteins from a single transcript using coding sequences of multiple 2A peptides [42]. In this case, our method can still be used to compute the
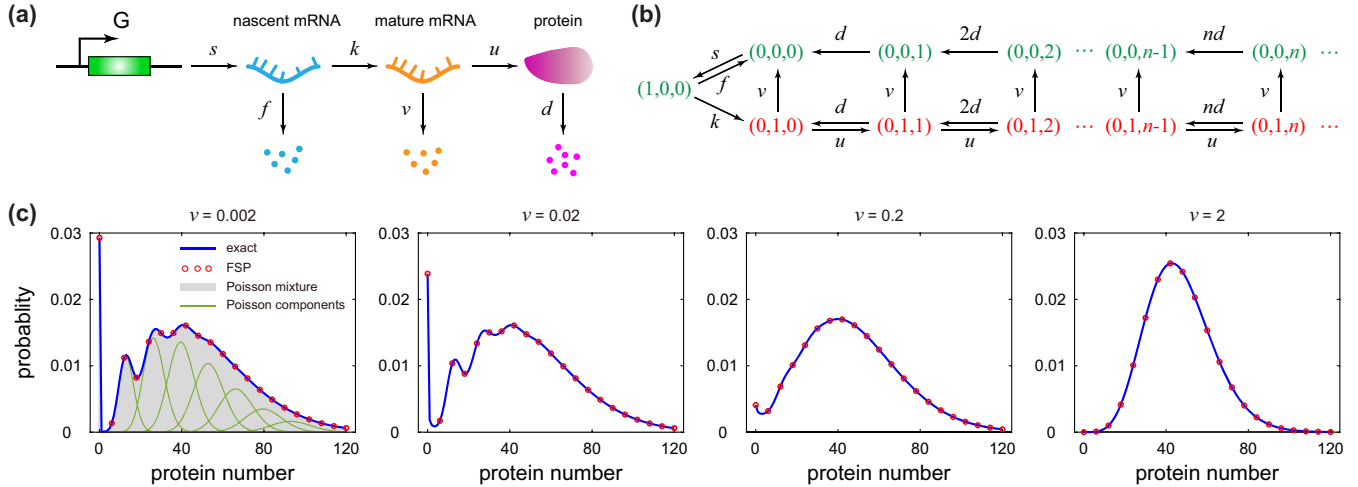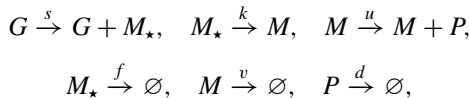
FIG. 3. A multistep gene expression model. (a) Schematic of a multistep gene expression model, which includes transcription, translation, and the production of mature mRNA from nascent mRNA. (b) Transition diagram for the modified model restricted to the irreducible state space. (c) Comparison between the exact steady-state distribution for the protein number given in Eq. (17) (blue curve) and FSP simulations (red circles) as $\nu = v/d$ varies while keeping $b = ks/(k + f)v$ as constant. The left panel also compares the exact distribution (blue curve) with the mixed Poisson approximation given in Eq. (18) (gray region). The Poisson components of the mixed distribution are shown by the green curves. The model parameters are chosen as $u = 40/3$, $d = 1$, $k = 0.002$, $s = 10$ and the parameter f is chosen so that $b = 3.52$.

joint copy number distributions for these proteins since the irreducible state space of the modified model is always finite.

B. Gene expression model with nascent mRNA

Based on the central dogma of molecular biology, the gene expression dynamics in an individual cell has a standard two-stage representation involving transcription and translation [8]. In the literature, the transcription step is usually modeled as the elementary reaction $G \rightarrow G + M$, where G is the gene of interest and M is the corresponding mRNA. However, in living cells, the realistic transcription process is much more complicated: first the gene is transcribed to produce the so-called nascent mRNA and then several steps such as 5′ capping, 3′ polyadenylylation, and mRNA splicing to remove the introns are necessary for the nascent mRNA to become the mature mRNA [47]. Only the mature mRNA can undergo translation to produce the protein. Recent studies about single-cell RNA-sequencing data analysis have highlighted the need to incorporate the nascent mRNA dynamics into the model to introduce the key concept of RNA velocity [48,49].

Here we consider a more realistic gene expression model depicted in Fig. 3(a). Let G denote the gene of interest, let $M_\star$ denote the nascent mRNA, let M denote the mature mRNA, and let P denote the protein. Then the effective reactions for the gene expression model are given by

$$G \xrightarrow{s} G + M_\star, \quad M_\star \xrightarrow{k} M, \quad M \xrightarrow{u} M + P,$$

$$M_\star \xrightarrow{f} \varnothing, \quad M \xrightarrow{v} \varnothing, \quad P \xrightarrow{d} \varnothing,$$

where the first reaction represents transcription, the second reaction represents the conversion of nascent mRNA into mature mRNA, the third reaction represents translation, and the remaining three reactions represent the degradation of all gene products. If the dynamics of nascent mRNA is ignored, then the steady-state joint distribution of mRNA and protein

numbers has been derived in Ref. [33]. Here we consider a more complicated model involving nascent mRNA (a similar model has been solved in Ref. [34] using a different method). The microstate of the gene can be represented by an ordered triple $(m_\star, m, n)$: the copy number m_* of nascent mRNA, the copy number m of mature mRNA, and the copy number n of protein. Let $p_{m_\star,m,n}$ denote the probability of observing microstate $(m_\star, m, n)$ and let

$$F(x_\star, x, y) = \sum_{m_\star, m, n} p_{m_\star,m,n} x_\star^{m_\star} x^m y^n$$

denote the corresponding generating function. Then the evolution of the gene expression model is governed by the CME

$$
\begin{aligned}
\dot{p}_{m_\star,m,n} = {} & s p_{m_\star-1,m,n} + k(m_\star + 1) p_{m_\star+1,m-1,n} + um p_{m_\star,m,n-1} \\
& + f(m_\star + 1) p_{m_\star+1,m,n} + v(m + 1) p_{m_\star,m+1,n} \\
& + d(n + 1) p_{m_\star,m,n+1} \\
& - (s + km_\star + fm_\star + um + vm + dn) p_{m_\star,m,n}.
\end{aligned}
$$

To solve this CME, we consider the modified Markovian model. We emphasize again that we do not take copy number variation of the gene into consideration and thus the reaction $G \rightarrow G + M_\star$ can be viewed as a zero-order reaction. Since the zero-order reaction can only occur at the zero microstate, it is easy to see that the irreducible state space of the modified model is given by

$$\{(1, 0, 0)\} \cup \{(0, 0, n), (0, 1, n) : n \geqslant 0\},$$

and the transition diagram restricted to the irreducible state space is illustrated in Fig. 3(b), which has a ladder-shaped structure. In fact, ladder-shaped models arise in many gene expression models and have been extensively studied in the literature [7–15]. Such models are usually analytically tractable with their solutions being represented by hypergeometric functions (see Ref. [17] for a detailed discussion on

the analytical theory of ladder-shaped models). Note that the irreducible state space of the original model is the whole three-dimensional nonnegative integer lattice since $m_\star$, m, and n can take all nonnegative integer values. Using the method proposed in this paper, we simplify a three-dimensional problem for the original model to a coupled one-dimensional problem for the modified model (here "coupled" means that

m can only take the values of 0 and 1 and "one-dimensional" means that n ranges over all nonnegative integers), which greatly reduces the theoretical complexity.

Since the modified model is essentially one-dimensional, its generating function H can be easily computed in steady-state conditions, which is given by (see Appendix C for details)

$$H(x_\star, x, y) = a\pi_0(x_\star - 1) + b\pi_0\left[(x-1)\,_1F_1(1; 1+v; \mu(y-1)) + \mu\int_1^y {}_1F_1(1; 1+v; \mu(z-1))dz\right] + 1,$$

where $_1F_1$ denotes the confluent hypergeometric function and

$$a = \frac{s}{k+f}, \quad b = \frac{ks}{(k+f)v}, \quad \mu = \frac{u}{d}, \quad v = \frac{v}{d}.$$

It then follows from Eq. (6) that the generating function F of the original model is given by

$$F(x_\star, x, y) = e^{a(x_\star-1)+b\left[(x-1)\,_1F_1(1;1+v;\mu(y-1))+\mu\int_1^y {}_1F_1(1;1+v;\mu(z-1))dz\right]}. \tag{15}$$

This implies that the number of nascent mRNA is independent of the numbers of mature mRNA and protein, while the numbers of mature mRNA and protein are correlated. Taking the derivatives of F gives the joint distributions for the nascent mRNA, mature mRNA, and protein numbers. In particular, taking $x = y = 1$ and $x_\star = y = 1$, we obtain

$$F(x_\star, 1, 1) = e^{a(x_\star-1)}, \quad F(1, x, 1) = e^{b(x-1)}.$$

This shows that the numbers of nascent and mature mRNAs both have a Poisson distribution:

$$p_{m_\star}^{M^\star} = \frac{a^{m_\star}}{m_\star!}e^{-a}, \quad p_m^M = \frac{b^m}{m!}e^{-b}.$$

Taking $x_\star = x = 1$, we obtain

$$F(1, 1, y) = e^{b\mu\int_1^y {}_1F_1(1;1+v;\mu(z-1))dz}. \tag{16}$$

It then follows from Eq. (9) that the marginal distribution for the protein number is given by

$$p_n^P = \frac{B_n(g_1, \ldots, g_n)}{n!}e^{-b\mu\int_0^1 {}_1F_1(1;1+v;\mu(z-1))dz}, \tag{17}$$

where B_n is the complete Bell polynomial and

$$g_i = \frac{b\mu^i(i-1)!}{(1+v)_{i-1}}\,_1F_1(i; i+v; -\mu), \quad i = 1, \ldots, n.$$

Our analytical results can also be used to compute the correlation coefficient $\rho_{M,P}$ between the mature mRNA and protein numbers. Combining Eqs. (10) and (15), it is easy to obtain

$$\rho_{M,P} = \sqrt{\frac{\mu}{(1+v)(1+\mu+v)}}.$$

This shows that the mature mRNA and protein numbers are always positively correlated; the correlation is strong when the translation rate u is large and the mature mRNA degradation rate v is small compared to the protein decay rate d.

We next take a deeper look at the marginal protein distribution. It is a classical result that if mRNA decays much faster than protein, then the protein number has a negative binomial distribution [8]. In fact, this assumption holds for

the majority of genes in bacteria and yeast, but it fails for many genes in higher prokaryotes, where mRNA and protein often decay at the same time scale (see Table S1 in Ref. [50] for such time scales in various cell types). Here we consider another important case where mature mRNA decays much slower compared to protein. Specifically, we consider the limiting case of $v = v/d \ll 1$, while keeping $b = ks/(k+f)v$ as constant. In this limit, the synthesis and degradation of mature mRNA are both very slow and thus the mature mRNA is a slow variable. Actually, a similar limit has been considered in Ref. [33] where the nascent mRNA is not modeled explicitly; here we take a deeper look at this limit. Since $v \ll 1$, we have $_1F_1[1; 1+v; \mu(z-1)] \approx {}_1F_1[1; 1; \mu(z-1)] = e^{\mu(z-1)}$. Then the generating function in Eq. (16) can be simplified as

$$F(1, 1, y) = e^{b\left[e^{\mu(y-1)}-1\right]}.$$

Taking the derivatives of the generating function $F(1, 1, y)$ at $y = 0$, we find that the protein number has the following mixed Poisson distribution with Poissonian weights:

$$p_n^P = e^{-b}\delta_0(n) + \sum_{k=1}^{\infty}\frac{b^j e^{-b}}{k!}\left[\frac{(k\mu)^n e^{-k\mu}}{n!}\right], \tag{18}$$

where $\delta_0(n)$ is Kronecker's δ function which takes the value of 1 when $n = 0$ and the value of 0 otherwise. This can be understood intuitively as follows. We have seen that the mature mRNA number has the Poisson distribution $\mathbb{P}(N_M = k) = b^k e^{-b}/k!$, where N_M denotes the number of M. Since the mature mRNA is a slow variable, given that k copies of mature mRNA has been produced, the total synthesis rate of protein is given by ku and thus the conditional distribution of the protein number is also Poissonian:

$$\mathbb{P}(N_P = n|N_M = k) = \frac{(k\mu)^n e^{-k\mu}}{n!},$$

where N_P denotes the number of P. Hence, the mixed Poisson distribution given in Eq. (18) is nothing but the formula of
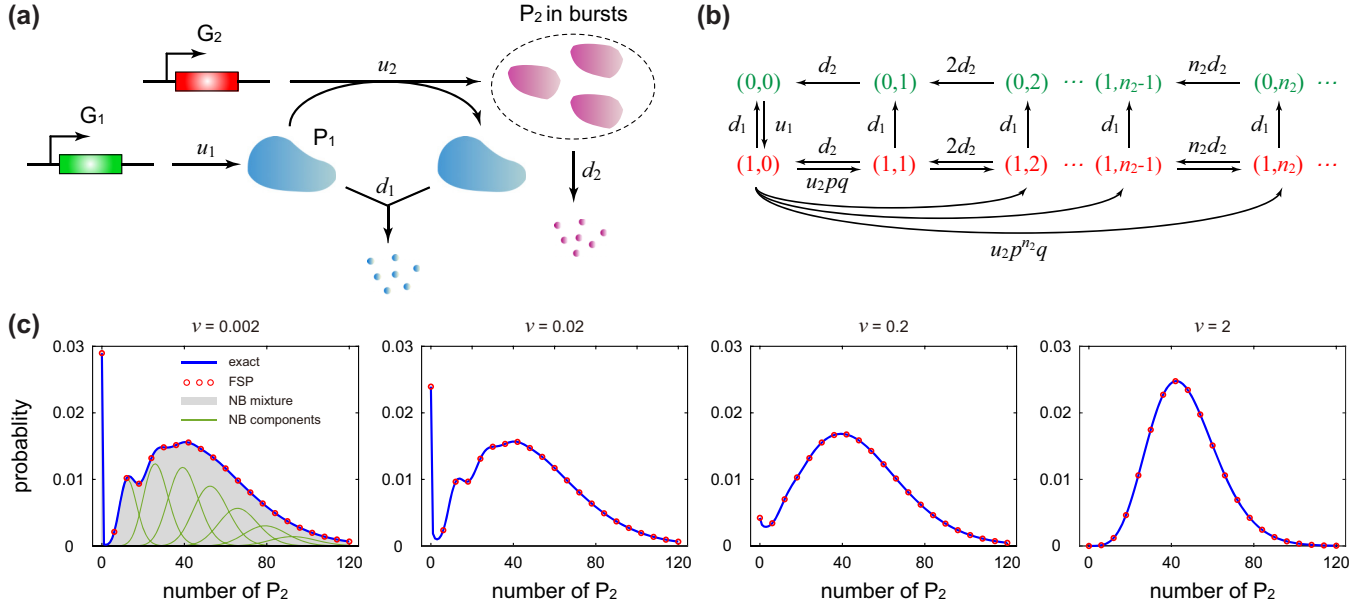
FIG. 4. A gene regulatory model with translational bursting. (a) Schematic of a simple gene regulatory model where the product of gene G_1 activates the expression of gene G_2. The protein synthesis of gene G_2 occurs in bursts. (b) Transition diagram for the modified model restricted to the irreducible state space. Note that translational bursting can cause jumps from microstate $(1, n_2)$ to $(1, n_2')$ with $n_2' > n_2$. This is shown for microstate $(1,0)$ in the figure but is also true for other microstates. (c) Comparison between the exact steady-state distribution for the number of protein P_2 given in Eq. (21) (blue curve) with FSP simulations (red circles) as $\nu = d_1/d_2$ varies while keeping $\mu_1 = u_1/d_1$ as constant. The left panel also compares the exact solution (blue curve) with mixed negative binomial approximation given in Eq. (24) (gray region). The negative binomial components of the mixed distribution are shown by the green curves. The model parameters are chosen as $u_2 = 40, d_2 = 1, \ p = 0.25$ and the parameters u_1 and d_1 are chosen so that $\mu_1 = 3.52$.
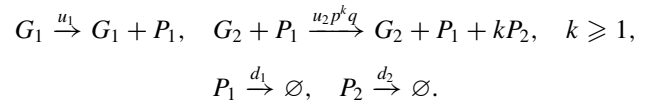
total probability:

$$\mathbb{P}(N_P = n) = \sum_{k=0}^{\infty} \mathbb{P}(N_P = n | N_M = k) \mathbb{P}(N_M = k).$$

Figure 3(c) shows the comparison between the exact solution given in Eq. (17), the approximate solution given in Eq. (18), and FSP simulations under different values of ν. It can be seen that the exact and approximation solutions coincide perfectly with each other for small ν, but they fail as expected for large ν. When $\nu \ll 1$, the protein distribution is a mixture of Poisson distributions and thus is capable of producing multiple peaks that are located around $k\mu$, $k = 0, 1, 2...$ with μ being the averaged amount of protein produced by a single mature mRNA molecule. Note that only the first several Poisson components contribute to the multiple peaks of the protein distribution since the Poisson components become lower and flatter as k increases [Fig. 3(c)]. In the literature [15], it is widely believed that bimodality of the protein distribution has two major origins—it can occur either when there is a positive feedback loop involved in the system or when the switching between promoter states are slow. Here we show that multimodality can also be caused by slow synthesis and degradation of mature mRNA, even when the gene is constitutively expressed (no promoter switching). As ν increases, multimodality disappears and the protein distribution becomes closer to a negative binomial distribution [Fig. 3(c)].

C. Gene regulatory model with translational bursting

As the third application, we consider a simple gene regulatory system where the product of a gene, as a transcription factor, regulates the expression of another gene in a bursty manner [Fig. 4(a)]. Let G_1 and G_2 denote the two genes and let P_1 and P_2 denote the corresponding gene products. The effective reactions describing the gene regulatory system are given by

$$G_1 \xrightarrow{u_1} G_1 + P_1, \quad G_2 + P_1 \xrightarrow{u_2 p^k q} G_2 + P_1 + kP_2, \quad k \geqslant 1,$$

$$P_1 \xrightarrow{d_1} \varnothing, \quad P_2 \xrightarrow{d_2} \varnothing.$$

Here the first reaction describes the expression of gene G_1 with effective translation rate u_1, the second reaction describes the expression of gene G_2 which is activated by protein P_1, and the last two reactions describe the degradation of the two proteins. In agreement with experiments [51], the production of protein P_2 is assumed to occur in bursts of random size sampled from a geometric distribution with parameter p. Each burst is due to rapid synthesis of protein from a single, short-lived mRNA molecule; thus the effective translation rate of gene G_2 is the product of the corresponding transcription rate u_2 and the geometric distribution $p^k q$, where $q = 1 - p$ [52]. The microstate of the system can be represented by an ordered pair (n_1, n_2), where n_i denotes the copy number of protein P_i. Let p_{n_1, n_2} denote the probability of observing microstate

(n_1, n_2) and let

$$F(y_1, y_2) = \sum_{n_1, n_2} p_{n_1, n_2} y_1^{n_1} y_2^{n_2}$$

denote the corresponding generating function. Then the evolution of the gene regulatory system is governed by the CME

$$\dot{p}_{n_1, n_2} = u_1 p_{n_1-1, n_2} + \sum_{i=0}^{n_2-1} u_2 p^{n_2-i} q n_1 p_{n_1, i} + d_1(n_1 + 1) p_{n_1+1, n_2}$$

$$+ d_2(n_2 + 1) p_{n_1, n_2+1}$$

$$- (u_1 + u_2 p n_1 + d_1 n_1 + d_2 n_2) p_{n_1, n_2},$$

where $u_2 p n_1 = \sum_{k=1}^{\infty} u_2 p^k q n_1$ in the bracket is the sum of transition rates from microstate (n_1, n_2) to other microstates due to translational bursting.

To solve this CME, we next consider the modified Markovian model. Similarly, we do not take the copy number variation of the gene into account and thus the reaction

$G_1 \rightarrow G_1 + P_1$ can be viewed as a zero-order reaction. Since the zero-order reaction can only occur at the zero microstate, it is easy to see that the irreducible state space of the modified model is given by

$$\{(1, n_2), (0, n_2) : n_2 \geqslant 0\},$$

and the transition diagram restricted to the irreducible state space is illustrated in Fig. 4(b). Note that the irreducible state space of the original model is the two-dimensional lattice since n_1 and n_2 can take all nonnegative integer values. Thus, the method proposed in this paper reduces a two-dimensional problem for the original model to a coupled one-dimensional problem for the modified model (here "coupled" means that n_1 can only take the values of 0 and 1), which greatly reduces the theoretical complexity.

Since the modified model is essentially one-dimensional, its generating function H can be easily computed in steady-state conditions, which is given by (see Appendix D for details)

$$H(y_1, y_2) = \pi_0 \mu_1 \left\{ {}_2F_1[-\mu_2, 1; 1+\nu; \omega(y_2)](y_1 - 1) + \mu_2 B \int_1^{y_2} {}_2F_1[1+\mu_2+\nu, 1; 1+\nu; B(z-1)]dz \right\} + 1,$$

where ${}_2F_1$ denotes the Gaussian hypergeometric function, $B = p/q = \sum_{n_2=1}^{\infty} n_2 p^{n_2} q$ is the mean burst size of protein P_2, and

$$\mu_1 = \frac{u_1}{d_1}, \quad \mu_2 = \frac{u_2}{d_2}, \quad \nu = \frac{d_1}{d_2}, \quad \omega(y_2) = \frac{p(y_2 - 1)}{py_2 - 1}.$$

It then follows from Eq. (6) that the generating function F for the original model is given by

$$F(y_1, y_2) = e^{\mu_1 \left[{}_2F_1(-\mu_2, 1; 1+\nu; \omega(y_2))(y_1-1) + \mu_2 B \int_1^{y_2} {}_2F_1(1+\mu_2+\nu, 1; 1+\nu; B(z-1))dz \right]}. \tag{19}$$

Taking the derivatives of F at zero yields the steady-state joint distribution for the numbers of the two proteins. In particular, taking $y_2 = 1$, we obtain $F(y_1, 1) = e^{\mu_1(y_1 - 1)}$. This shows that the number of protein P_1 has the Poisson distribution

$$p_{n_1}^{P_1} = \frac{\mu_1^{n_1}}{n_1!} e^{-\mu_1}.$$

Moreover, taking $y_1 = 1$, we obtain

$$F(1, y_2) = e^{\mu_1 \mu_2 B \int_1^{y_2} {}_2F_1(1+\nu+\mu_2, 1; 1+\nu; B(z-1))dz}. \tag{20}$$

It then follows from Eq. (9) that the number of protein P_2 has the following distribution:

$$p_{n_2}^{P_2} = \frac{B_{n_2}(g_1, \ldots, g_{n_2})}{n_2!} e^{-\mu_1 \mu_2 B \int_0^1 {}_2F_1(1+\nu+\mu_2, 1; 1+\nu; B(z-1))dz}, \tag{21}$$

where B_n is the complete Bell polynomial and

$$g_i = \frac{\mu_1 \mu_2 B^i (1 + \nu + \mu_2)_{i-1}(i-1)!}{(1+\nu)_{i-1}} {}_2F_1(i + \nu + \mu_2, i; i + \nu; -B), \quad i = 1, \ldots, n_2.$$

We next focus on two limiting cases. The first case occurs when protein P_1 decays much faster than protein P_2, i.e., $\nu = d_1/d_2 \gg 1$, and the constant $\mu_1 = u_1/d_1$ is strictly positive and bounded. In this case, both the synthesis and degradation of protein P_1 are very fast and thus it can be viewed as a fast variable. When $\nu \gg 1$, we have

$${}_2F_1[1 + \nu + \mu_2, 1; 1 + \nu; B(z - 1)]$$

$$\approx {}_1F_0[1; B(z - 1)] = [1 - B(z - 1)]^{-1}.$$

It then follows from Eq. (20) that

$$F(1, y_2) = \left(\frac{q}{1 - py_2} \right)^{\mu_1 \mu_2}. \tag{22}$$

This shows that the number of protein P_2 has the negative binomial distribution

$$p_{n_2}^{P_2} = \frac{(\mu_1 \mu_2)_{n_2}}{n_2!} p^{n_2} q^{\mu_1 \mu_2}. \tag{23}$$

The second case occurs when protein P_1 decays much slower than protein P_2, i.e., $\nu = d_1/d_2 \ll 1$, and the constant

$\mu_1 = u_1/d_1$ is strictly positive and bounded. In this case, both the synthesis and degradation of protein P_1 are very slow and thus it can be viewed as a slow variable. When $\nu \ll 1$, we have

$$_2F_1[1 + \nu + \mu_2, 1; 1 + \nu; B(z - 1)]$$
$$\approx {}_1F_0[1 + \mu_2; B(z - 1)] = [1 - B(z - 1)]^{-(1+\mu_2)}.$$

Then the generating function in Eq. (20) can be simplified as

$$F(1, y_2) = e^{\mu_1[(\frac{q}{1-py_2})^{\mu_2} - 1]}.$$

Taking the derivatives of $F(1, y_2)$ at $y_2 = 0$, we find that the number of protein P_2 has the following mixed negative binomial distribution with Poissonian weights:

$$p_{n_2}^{P_2} = e^{-\mu_1}\delta_0(n_2) + \sum_{k=1}^{\infty} \frac{\mu_1^k e^{-\mu_1}}{k!}\left[\frac{(k\mu_2)_{n_2}}{n_2!}p^{n_2}q^{k\mu_2}\right]. \quad (24)$$

This can be explained intuitively as follows. We have seen that the number of protein P_1 has the Poisson distribution $\mathbb{P}(N_{P_1} = k) = \mu_1^k e^{-\mu_1}/k!$, where N_{P_1} denotes the number of P_1. Since protein P_1 is a slow variable, given that k copies of P_1 has been produced, the effective transcription rate of gene G_2 is given by ku_2 and thus the conditional distribution for the number of protein P_2 is negative binomial:

$$\mathbb{P}(N_{P_2} = n_2|N_{P_1} = k) = \frac{(k\mu_2)_{n_2}}{n_2!}p^{n_2}q^{k\mu_2},$$

where N_{P_2} denotes the number of P_2. Thus, the mixed negative binomial distribution given in Eq. (24) is nothing but the formula of total probability:

$$\mathbb{P}(N_{P_2} = n_2) = \sum_{k=0}^{\infty} \mathbb{P}(N_{P_2} = n_2|N_{P_1} = k)\mathbb{P}(N_{P_1} = k).$$

Figure 4(c) shows the comparison between our exact solution given in Eq. (21), the approximate solution given in Eq. (24), and FSP simulations under different values of ν. Clearly, the exact and approximation solutions coincide perfectly with each other for small ν, but deviate significantly from each other for large ν. When $\nu \ll 1$, the copy number distribution for protein P_2 is a mixture of negative binomials and thus can produce multiple peaks around $k\mu_2 B$, $k = 0, 1, 2...$ with $\mu_2 B$ being the averaged amount of protein P_2 produced by a single protein P_1 molecule. As ν increases, multimodality disappears and the protein distribution becomes closer to a negative binomial distribution [Fig. 4(c)].

We finally examine the correlation between the two proteins using our analytical results. It follows from Eqs. (10) and (19) that the correlation coefficient between the numbers of P_1 and P_2 is given by

$$\rho_{P_1,P_2} = \sqrt{\frac{\mu_2 B}{(1 + \nu)[1 + \nu + (1 + \nu + \mu_2)B]}}.$$
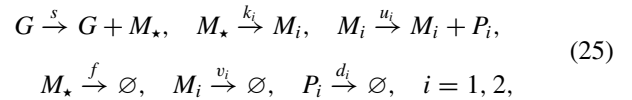
Clearly, the numbers of the two proteins are always positively correlated; the correlation is strong when the burstiness of

protein P_2 is large, the translation rate of protein P_2 is large, and the degradation rate of protein P_1 is small.

D. Gene expression model with alternative splicing

Alternative splicing is a process during gene expression that results in a single gene coding for multiple proteins [47]. In this process, particular exons of a gene may be included within or excluded from the final processed mRNA that are produced from that gene. Consequently, the proteins translated from different spliced mRNAs will be different [see Fig. 5(a) for an illustration]. A gene expression model involving alternative splicing has been solved in Ref. [35], which considers the expression of mRNAs but not proteins. Here we take proteins into consideration.

Let G denote the gene of interest, $M_\star$ denote the nascent mRNA, M_1 and M_2 denote two mRNA isoforms, and P_1 and P_2 denote the corresponding protein isoforms. Based on the central dogma of molecular biology, the effective reactions involved in the gene expression system are listed as follows:

$$G \xrightarrow{s} G + M_\star, \quad M_\star \xrightarrow{k_i} M_i, \quad M_i \xrightarrow{u_i} M_i + P_i,$$
$$M_\star \xrightarrow{f} \varnothing, \quad M_i \xrightarrow{v_i} \varnothing, \quad P_i \xrightarrow{d_i} \varnothing, \quad i = 1, 2, \quad (25)$$

where s is the transcription rate, k_i are the rates of alternative splicing, u_i are the translation rates of the two mRNA isoforms, and f, v_i, and d_i are the degradation rates of all gene products. Recently, it has been found that alternative splicing can be regulated by a system of proteins (regulators) binding to a nascent transcript that in turn direct the splicing machinery to include or skip specific exons [53,54]; moreover, the regulators usually exert distinct effects on exon inclusion or exclusion depending on the position of its binding [53] and thus different binding positions lead to different mRNA isoforms. Here we take this effect into account by assuming that there is a regulator P which activates the formation of two mRNA isoforms M_1 and M_2 (via exon inclusion and/or exclusion). Hence, the copy number of the regulator P, which is denoted by n, will influence the splicing rates $k_1 = k_1(n)$ and $k_2 = k_2(n)$. For simplicity, we further assume that the number of regulator has a fixed distribution that is independent of the numbers of gene products.

The microstate of the system can be represented by an ordered five-tuple (m, m_1, n_1, m_2, n_2): the copy number m of nascent mRNA, the copy numbers m_1 and m_2 of the two mRNA isoforms, and the copy numbers n_1 and n_2 of the two protein isoforms. Let p_{m,m_1,n_1,m_2,n_2} denote the probability of observing microstate (m, m_1, n_1, m_2, n_2) and let

$$F(x, x_1, y_1, x_2, y_2) = \sum_{m,m_1,n_1,m_2,n_2} p_{m,m_1,n_1,m_2,n_2} x^m x_1^{m_1} y_1^{n_1} x_2^{m_2} y_2^{n_2}$$

denote the corresponding generating function. Given that there are n copies of regulator P, we can treat the splicing rates $k_1 = k_1(n)$ and $k_2 = k_2(n)$ as constants and the evolution of the gene expression model is governed by
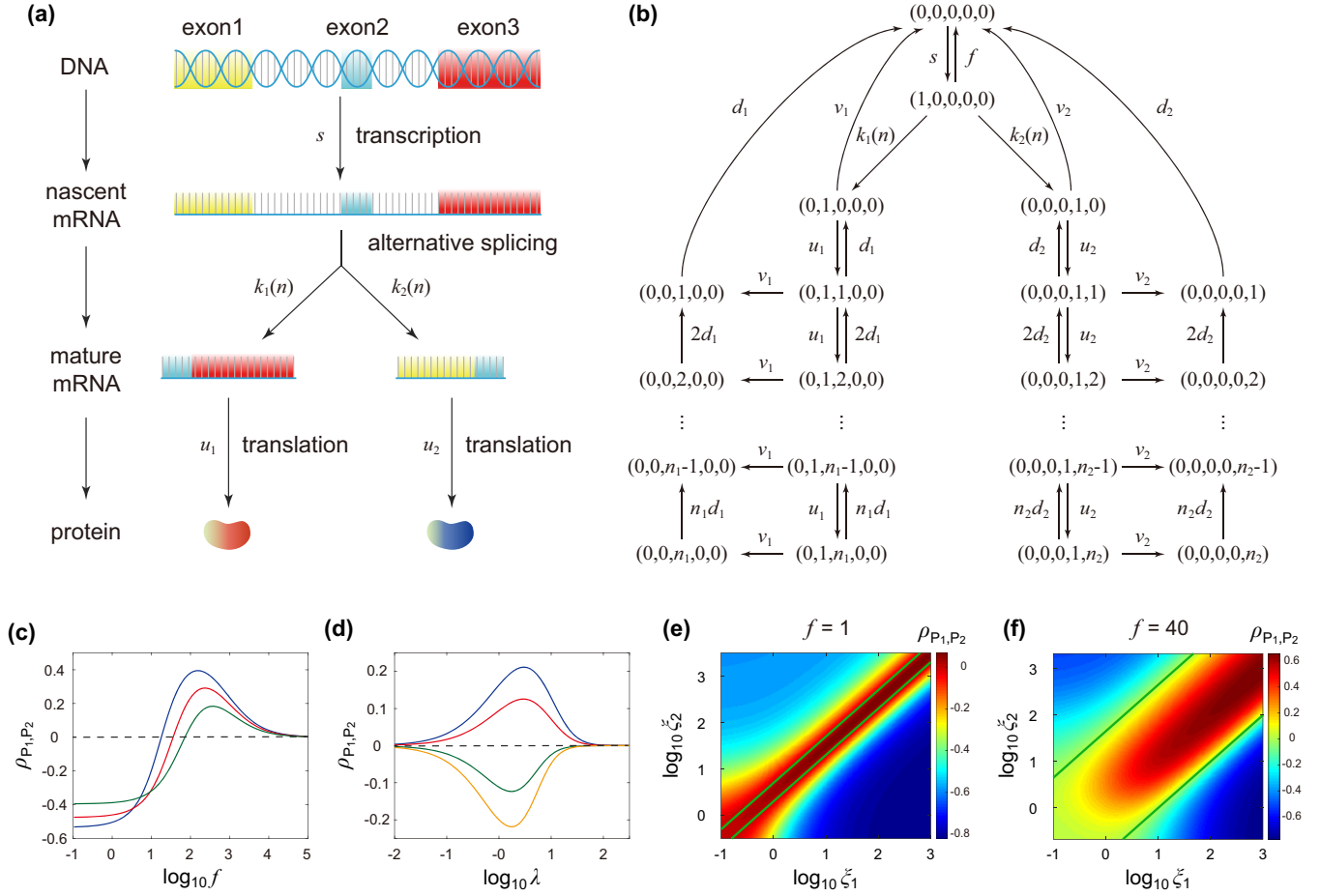
FIG. 5. A multistep gene expression model with alternative splicing. (a) Schematic of a multistep gene expression model involving transcription, translation, and alternative splicing. Due to alternative splicing, the nascent mRNA is spliced in two different ways to produce two mature mRNA isoforms. (b) Transition diagram for the modified model restricted to the irreducible state space. Note that the transition diagram has two branches (left and right), corresponding to the production of two mRNA/protein isoforms. Each branch has a ladder-shaped structure. (c) Correlation coefficient ρ_{P_1,P_2} between the numbers of the two protein isoforms versus the degradation rate f of nascent mRNA. The model parameters are chosen as $s = 100$, $u_1 = 30$, $u_2 = 20$, $v_1 = 15$, $v_2 = 4$, $d_1 = 3$, $d_2 = 2$, $\xi_1 = 10$, $\xi_2 = 20$, $\lambda = 2$. The remaining parameters are chosen as $\eta_1 = 10$, $\eta_2 = 3$ (blue curve), $\eta_1 = 20$, $\eta_2 = 6$ (red curve), and $\eta_1 = 40$, $\eta_2 = 12$ (green curve). (d) Correlation coefficient ρ_{P_1,P_2} versus the mean λ of the regulator number. The model parameters are chosen as $u_1 = 30$, $u_2 = 20$, $v_1 = 3$, $v_2 = 4$, $d_1 = 3$, $d_2 = 4$, $\xi_1 = 7$, $\xi_2 = 5$, $\eta_1 = 15$, $\eta_2 = 28$. The remaining parameters are chosen as $s = 200$, $f = 80$ (blue curve), $s = 100$, $f = 80$ (red curve), $s = 100$, $f = 0$ (green curve), and $s = 200$, $f = 0$ (orange curve). (e), (f) Correlation coefficient ρ_{P_1,P_2} versus the regulation strengths ξ_1 and ξ_2. (e) Slow degradation of nascent mRNA with $f = 1$. (f) Fast degradation of nascent mRNA with $f = 40$. The model parameters are chosen as $s = 100$, $u_1 = 30$, $u_2 = 20$, $v_1 = 15$, $v_2 = 4$, $d_1 = 3$, $d_2 = 2$, $\eta_1 = 1$, $\eta_2 = 4$, $\lambda = 2$. The two green lines separate the region with positive correlation and the region with negative correlation.

the CME

$$
\begin{aligned}
\dot{p}_{m,m_1,n_1,m_2,n_2} =\; & s p_{m-1,m_1,n_1,m_2,n_2} + k_1(m+1)p_{m+1,m_1-1,n_1,m_2,n_2} + k_2(m+1)p_{m+1,m_1,n_1,m_2-1,n_2} + u_1 m_1 p_{m,m_1,n_1-1,m_2,n_2} \\
& + u_2 m_2 p_{m,m_1,n_1,m_2,n_2-1} + f(m+1)p_{m+1,m_1,n_1,m_2,n_2} + v_1(m_1+1)p_{m,m_1+1,n_1,m_2,n_2} \\
& + v_2(m_2+1)p_{m,m_1,n_1,m_2+1,n_2} + d_1(n_1+1)p_{m,m_1,n_1+1,m_2,n_2} + d_2(n_2+1)p_{m,m_1,n_1,m_2,n_2+1} \\
& - [s + (k_1 + k_2 + f)m + (u_1 + v_1)m_1 + (u_2 + v_2)m_2 + d_1 n_1 + d_2 n_2]p_{m,m_1,n_1,m_2,n_2}.
\end{aligned}
$$

To solve this CME, we next consider the modified Markovian model, which has only one zero-order reaction. Since the zero-order reaction $G \rightarrow G + M_\star$ can only occur at the zero microstate, it is easy to see that the irreducible state space of the modified model is given by

$$\{(1,0,0,0,0), (0,1,n_1,0,0), (0,0,n_1,0,0), (0,0,0,1,n_2), (0,0,0,0,n_2) : n_1, n_2 \geqslant 0\}.$$

The transition diagram restricted to the irreducible state space is illustrated in Fig. 5(b). Clearly, the zero microstate can only transition to microstate (1,0,0,0,0). If the nascent transcript $M_\star$ produces the mRNA isoform M_1, then the modified model enters the left branch in Fig. 5(b); if $M_\star$ produces M_2, then the modified model enters the right branch. Hence, our method reduces a five-dimensional problem for the original model to a coupled one-dimensional problem for the modified model. In analogy to the derivation in Sec. III B, given that there are n copies of regulator P in a single cell, the generating function of the original model is given by (see Appendix E for details)

$$F(x, x_1, y_1, x_2, y_2 | n) = e^{a(n)(x-1) + \sum_{i=1}^{2} K_i(n) b_i [(x_i-1) \, _1F_1(1;1+\nu_i;\mu_i(y_i-1)) + \mu_i \int_1^{y_i} \, _1F_1(1;1+\nu_i;\mu_i(z-1))dz]}, \tag{26}$$

where

$$a(n) = \frac{s}{k_1(n) + k_2(n) + f}, \quad K_1(n) = \frac{k_1(n)}{k_1(n) + k_2(n) + f},$$

$$K_2(n) = \frac{k_2(n)}{k_1(n) + k_2(n) + f},$$

$$b_1 = \frac{s}{v_1}, \quad b_2 = \frac{s}{v_2}, \quad \mu_1 = \frac{u_1}{d_1}, \quad \mu_2 = \frac{u_2}{d_2},$$

$$\nu_1 = \frac{v_1}{d_1}, \quad \nu_2 = \frac{v_2}{d_2}.$$

Finally, when taking into account the copy number variation of regulator P, it follows from the total probability formula that the generating function F is given by

$$F(x, x_1, y_1, x_2, y_2) = \sum_{n=0}^{\infty} p_n^P F(x, x_1, y_1, x_2, y_2 | n),$$

where p_n^P is the probability of observing n copies of regulator in a cell. Finally, the joint distribution of all gene products can be recovered by taking the derivatives of the generating function. It is easy to see that the marginal distributions for nascent mRNA and the two mRNA isoforms are all mixed Poisson distributions with the weights being the distribution of the regulator number; however, the marginal distributions for the two protein isoforms are much more complicated.

In recent years, the correlation between different mRNA and protein species produced from a single gene by means of alternative splicing has attracted increasing attention [53,55]. It has been shown that the numbers of two mRNA isoforms are independent of each other if they are not controlled by the regulator [35]; moreover, transcriptional bursting (which is not considered in our current model) may lead to positive correlation between two mRNA isoforms [35]. Here we analyze such correlation when the two mRNA isoforms are controlled by the same regulator. To do this, we assume that the splicing rates depend on the regulator number linearly as

$$k_1(n) = \xi_1 n + \eta_1, \quad k_2(n) = \xi_2 n + \eta_2,$$

where $\eta_i > 0$, $i = 1, 2$ are the spontaneous splicing rates and $\xi_i \geqslant 0$ characterize the strengths of regulation. Such linear dependence has been widely used in the modeling of stochastic gene regulatory networks [13,56–58]. In addition, we assume that the number of regulator has a Poisson distribution with mean λ. Under these assumptions, the correlation coefficient between (the numbers of) the two mRNA isoforms is given by (see Appendix E for details)

$$\rho_{M_1, M_2} = \frac{\alpha_1 \alpha_2}{\sqrt{(\alpha_1^2 + \beta_1)(\alpha_2^2 + \beta_2)}}, \tag{27}$$

and the correlation coefficient between the two protein isoforms is given by (see Appendix E for details)

$$\rho_{P_1, P_2} = \frac{\alpha_1 \alpha_2}{\sqrt{(\alpha_1^2 + \beta_1 L_1)(\alpha_2^2 + \beta_2 L_2)}}, \tag{28}$$

where

$$\alpha_1 = \frac{\xi_2 \eta_1 - \xi_1 \eta_2 - \xi_1 f}{\eta_1 + \eta_2 + f}, \quad \alpha_2 = \frac{\xi_1 \eta_2 - \xi_2 \eta_1 - \xi_2 f}{\eta_1 + \eta_2 + f},$$

$$L_1 = \frac{1 + \mu_1 + \nu_1}{\mu_1(1 + \nu_1)}, \quad L_2 = \frac{1 + \mu_2 + \nu_2}{\mu_2(1 + \nu_2)}, \quad \gamma = \frac{\eta_1 + \eta_2 + f}{\xi_1 + \xi_2},$$

$$h_1 = \, _1F_1(1; \gamma + 1; -\lambda), \quad h_2 = \, _2F_2(\gamma, \gamma; \gamma + 1, \gamma + 1; \lambda)e^{-\lambda},$$

$$\beta_1 = \frac{(\xi_1 + \xi_2)(\xi_1 + \alpha_1 h_1)}{b_1(h_2 - h_1^2)}, \quad \beta_2 = \frac{(\xi_1 + \xi_2)(\xi_2 + \alpha_2 h_1)}{b_2(h_2 - h_1^2)},$$

where $_{12}F$ denotes the generalized hypergeometric function. In the above formulas, the parameters β_1 and β_2 depend on the parameters h_1 and h_2, which further depend on the parameter γ. In Appendix E, we have proved that the parameters β_1 and β_2, together with $h_2 - h_1^2$, must be positive. Therefore, the correlation coefficients ρ_{M_1, M_2} and ρ_{P_1, P_2} must have the same sign and the sign is determined by the sign of $\alpha_1 \alpha_2$. In particular, when the nascent mRNA decays very slowly, i.e., $f \ll 1$, we have

$$\alpha_1 \alpha_2 \approx -\frac{(\xi_2 \eta_1 - \xi_1 \eta_2)^2}{(\eta_1 + \eta_2)^2}.$$

In this case, the numbers of the two mRNA/protein isoforms are negatively correlated. However, when the nascent mRNA decays very fast, i.e., $f \gg 1$, we have $\alpha_1 \alpha_2 \approx \xi_1 \xi_2 > 0$. In this case, the numbers of the two mRNA/protein isoforms are positively correlated.

These results can be understood intuitively as follows. When the nascent mRNA decays very slowly, once a nascent transcript is synthesized, it can either produce an M_1 or an M_2 molecule. Thus, there is strong competition between the two isoforms; the more one isoform, the less the other isoform. This results in negative correlation between them. However, when the nascent mRNA decays very fast, its molecule number relaxes to the steady-state value rapidly [59] and thus there is an ample supply of nascent mRNA. In this case, there is little competition between the two isoforms; the more (less) the regulator, the more (less) the two isoforms. This results in positive correlation between them.

Our results indicate that the degradation rate f of nascent mRNA has a critical value

$$f_c = \begin{cases} |\xi_2 \eta_1 - \xi_1 \eta_2|/\xi_1, & \text{if } \xi_2 \eta_1 - \xi_1 \eta_2 \geqslant 0, \\ |\xi_2 \eta_1 - \xi_1 \eta_2|/\xi_2, & \text{if } \xi_2 \eta_1 - \xi_1 \eta_2 < 0, \end{cases}$$

and the system undergoes a stochastic bifurcation as f varies. When $f < f_c$, we have $\alpha_1 \alpha_2 < 0$ and thus the levels of the two isoforms are negatively correlated; when $f = f_c$, we have $\alpha_1 \alpha_2 = 0$ and thus they are not correlated; when $f > f_c$, we have $\alpha_1 \alpha_2 > 0$ and thus they are positively correlated. Note that the size of the critical value f_c depends on the sizes of η_1 and η_2. As η_1 and η_2 increase, the critical value f_c becomes larger. These observations coincide with stochastic simulations in Fig. 5(c), which illustrates the correlation coefficient ρ_{P_1,P_2} as a function of f.

The correlation between the two mRNA/protein isoforms is also influenced by the abundance of regulator. Figure 5(d) depicts the correlation coefficient ρ_{P_1,P_2} as a function of the regulator mean λ. It can be seen that the correlation is weak when λ is very small or very large. Interestingly, there is an optimal λ such that $|\rho_{P_1,P_2}|$ attains its maximum. This shows that the correlation is the strongest when the regulator mean is neither too small nor too large. This can be understood intuitively as follows. It follows from Eq. (26) that the regulator number n affects the joint distribution by adjusting the three parameters $a(n)$, $K_1(n)$, and $K_2(n)$. When $\lambda \ll 1$ or $\lambda \gg 1$, the three parameters are almost invariant and thus the gene expression model under consideration behaves like a system with no regulator. This explains the weak correlation observed when $\lambda \ll 1$ or $\lambda \gg 1$. Figure 5(d) also shows that a larger transcription rate s will enhance the correlation between the two isoforms. This is consistent with our analytical result in Eq. (28) since a larger value of s results in smaller values of β_1 and β_2 and thus results in stronger correlation.

Furthermore, the correlation is also influenced by the regulation strengths ξ_1 and ξ_2. Figures 5(e) and 5(f) illustrate the correlation coefficient ρ_{P_1,P_2} as a function of ξ_1 and ξ_2 under different values of f, where the two green lines in each figure separate the region with positive correlation (inside the two green lines) and the region with negative correlation (outside the two green lines). One of the two green lines corresponds to the case of $\alpha_1 = 0$ and the other corresponds to the case of $\alpha_2 = 0$. Recall that the two isoforms are positively correlated when $\alpha_1 \alpha_2 > 0$, i.e.,

$$\frac{\eta_2}{\eta_1 + f} < \frac{\xi_2}{\xi_1} < \frac{\eta_2 + f}{\eta_1}.$$

Therefore, to observe positive correlation, $\log \xi_2 - \log \xi_1$ must be controlled within a belt-shaped region that becomes wider as f increases [Figs. 5(e) and 5(f)]. In the absence of regulator ($\xi_1 = \xi_2 = 0$), we have $\alpha_1 = \alpha_2 = 0$ and thus there is no correlation between the two isoforms [35]. If only one of the two isoforms is controlled by the regulator ($\xi_1 > 0$ and $\xi_2 = 0$), then we have $\alpha_1 < 0$ and $\alpha_2 > 0$ and thus they are negatively correlated. If both isoforms are controlled by the regulator ($\xi_1, \xi_2 > 0$), then the correlation coefficient can be either positive or negative, depending on whether the degradation rate of nascent mRNA is above or below its critical value.

Finally, we make a crucial observation that the correlation between the two protein isoforms can be either weaker or stronger than that between the two mRNA isoforms, depending on the values of the parameters L_1 and L_2. Comparing Eq. (27) with Eq. (28), we can see that the protein correlation is less than the mRNA correlation when $L_1, L_2 > 1$. However, when $L_1, L_2 < 1$, i.e., when the translation rates u_i and degradation rates v_i of mRNA isoforms are large compared to the degradation rates d_i of protein isoforms, the protein correlation can be even greater than the mRNA correlation, which means that the translation step may even enhance the correlation between the two isoforms of the gene product.

IV. DISCUSSION

In this paper, we propose an alternative method of computing the joint distribution for a wide class of first-order stochastic reaction networks in steady-state conditions. By allowing all zero-order reactions to occur only at the zero microstate, we simplify the Markovian model of stochastic reaction kinetics to a modified Markovian model whose transition diagram is usually much simpler than that of the original one. In many models of biological relevance, the joint distribution of the modified model can be computed analytically. Finally, the joint generating function of the original model can be recovered from that of the modified model by taking a simple exponential transformation.

While the modified model is generally simpler than the original one, it may not be analytically tractable. However, we show its analytical tractability in two special cases: (i) its irreducible state space is finite and (ii) its irreducible state space has a ladder-shaped topological structure. We provide an easily verifiable criterion for the case (i), which states that if all the first-order reactions except degradation reactions have a conservation law with positive coefficients, then the modified model must have a finite irreducible state space. We also show that the case (ii) is satisfied in many gene expression models of biological interest. Here the ladder-shaped structure results from the fact that for the modified model, we only allow zero-order reactions to occur at the zero microstate. For example, if we allow $\varnothing \rightarrow P$ to occur only at the zero microstate, then the number of P can only vary between 0 and 1, which correspond to the two branches of the ladder-shaped structure. In fact, ladder-shaped models have been extensively studied in the literature and their generating functions are always represented by various kinds of hypergeometric functions [17]. Hence, for the case (ii), the generating function of the original model is given by the exponential of hypergeometric functions since an exponential transformation needs to be taken in our approach.

In most previous papers, the exact joint distribution is computed by first converting the CME into a system of PDEs satisfied by the joint generating function and then solving the system of PDEs using the method of characteristics. Compared with this method which often involves tedious computations, our approach greatly reduces the theoretical complexity. We then validate the effectiveness of our method by applying it to four gene expression models of biological significance. The analytical results obtained reveal some interesting biological phenomena: (i) multimodality can be caused by slow synthesis and degradation of some gene product, even when the gene is constitutively expressed; (ii) in the presence of alternative splicing, the numbers of two mRNA-protein isoforms are negatively regulated if one isoform is controlled by the regulator and the other isoform is not; (iii) if both mRNA-protein isoforms are controlled by the regulator, then their abundances can be either positively or

negatively correlated, depending on whether the degradation rate of nascent mRNA is above or below its critical value; (iv) the protein isoform correlation may be even greater than the mRNA isoform correlation when the translation rates and degradation rates of mRNA isoforms are large compared to the degradation rates of protein isoforms.

We emphasize that we construct the modified model by allowing all zero-order reactions to occur only at the zero microstate. Hence, to apply our method, the reaction system must have at least one zero-order reaction. However, in some biological systems, there may not be a zero-order reaction involved in the system. For example, consider the following gene expression model with promoter switching [7]:

$$G \xrightarrow{a} G^*, \quad G^* \xrightarrow{b} G, \quad G^* \xrightarrow{\rho} G^* + P, \quad P \xrightarrow{d} \varnothing,$$

where G and G^* denote the inactive and active states of the promoter, respectively, and P denotes the corresponding protein. Note that in this model, while the total number of genes in the two promoter states is constant, the number of genes in the active (inactive) state is not constant. Therefore, the two switching reactions, $G \to G^*$ and $G^* \to G$, as well as the synthesis reaction $G^* \to G^* + P$, are actually first-order reactions and cannot be regarded as zero-order reactions. In this case, there are no zero-order reactions involved in the system and thus our approach can no longer be applied. This is the major limitation of our method. In the presence of promoter switching, it has been shown that the analytical solution of a gene expression model is usually represented by hypergeometric functions [7–15]. In our paper, we do not take promoter switching into account and show that the joint distributions for a class of gene expression models can be represented by the exponential of hypergeometric functions. The reason for this discrepancy is that promoter switching is considered for the former but is not considered for the latter.

The current method is aimed to compute the exact solution of the steady-state joint distribution of first-order reaction kinetics. If a system contains higher-order reactions, then the PDEs satisfied by the generating function involve higher-order partial derivatives and hence it is very difficult to solve these PDEs analytically. Current research work aims to develop novel methods of computing the joint distribution of higher-order stochastic reaction kinetics. We anticipate that the method developed in this paper can be combined with various approximate techniques developed recently [18–21] to solve the joint distribution of complex biochemical reaction networks and gene regulatory networks.

APPENDIX A: RELATIONSHIP BETWEEN THE GENERATING FUNCTIONS OF THE TWO MODELS

Here we uncover the relationship between the generating functions of the original and modified models. Multiplying $x^n = x_1^{n_1} \dots x_N^{n_N}$ on both sides of Eq. (5) and then summing over all microstates, we obtain

$$\frac{\partial H}{\partial t} = \sum_n \left(\sum_{i=1}^{N} \sum_{j=1}^{r_i} \tilde{q}_{n+e_i-v_{ij},n} \pi_{n+e_i-v_{ij}} - \sum_{i=1}^{N} \sum_{j=1}^{r_i} \tilde{q}_{n,n+v_{ij}-e_i} \pi_n \right) x^n + \sum_n \left(\sum_{j=1}^{r_0} \tilde{q}_{n-v_{0j},n} \pi_{n-v_{0j}} \right) x^n - \sum_n \left(\sum_{j=1}^{r_0} \tilde{q}_{n,n+v_{0j}} \pi_n \right) x^n$$

$$:= \mathrm{I} + \mathrm{II} - \mathrm{III}. \tag{A1}$$

Recall that first-order reactions lead to the same transitions for the two models. It then follows from the classical result about first-order reaction systems (see Appendix A.2 in Ref. [32]) that

$$\mathrm{I} = \sum_{i=1}^{N} \sum_{j=1}^{r_i} k_{ij} (x^{v_{ij}} - x_i) \frac{\partial H}{\partial x_i}.$$

Moreover, since $\tilde{q}_{n,n+v_{0j}}$ is nonzero only when $n = \mathbf{0}$, it is easy to see that

$$\mathrm{II} = \sum_{j=1}^{r_0} k_{0j} \pi_{\mathbf{0}} x^{v_{0j}}, \quad \mathrm{III} = \sum_{j=1}^{r_0} k_{0j} \pi_{\mathbf{0}}.$$

Inserting the above two equations into Eq. (A1) yields

$$\frac{\partial H}{\partial t} = \sum_{i=1}^{N} \sum_{j=1}^{r_i} k_{ij} (x^{v_{ij}} - x_i) \frac{\partial H}{\partial x_i} + \sum_{j=1}^{r_0} k_{0j} \pi_{\mathbf{0}} (x^{v_{0j}} - 1). \tag{A2}$$

We next prove that Eq. (6) holds. If both the original and modified models are at the steady state, then it follows from Eq. (A2) that

$$\sum_{i=1}^{N}\sum_{j=1}^{r_i} k_{ij}(x^{\nu_{ij}} - x_i)\frac{\partial}{\partial x_i}\left(e^{\frac{H-1}{\pi_0}}\right) + \sum_{j=1}^{r_0} k_{0j}e^{\frac{H-1}{\pi_0}}(x^{\nu_{0j}} - 1)$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{r_i} k_{ij}(x^{\nu_{ij}} - x_i)e^{\frac{H-1}{\pi_0}}\frac{1}{\pi_0}\frac{\partial H}{\partial x_i} + \sum_{j=1}^{r_0} k_{0j}e^{\frac{H-1}{\pi_0}}(x^{\nu_{0j}} - 1) = e^{\frac{H-1}{\pi_0}}\frac{1}{\pi_0}\left[\sum_{i=1}^{N}\sum_{j=1}^{r_i} k_{ij}(x^{\nu_{ij}} - x_i)\frac{\partial H}{\partial x_i} + \sum_{j=1}^{r_0} k_{0j}\pi_0(x^{\nu_{0j}} - 1)\right] = 0.$$

Thus, we have

$$\sum_{i=1}^{N}\sum_{j=1}^{r_i} k_{ij}(x^{\nu_{ij}} - x_i)\frac{\partial}{\partial x_i}\left(e^{\frac{H-1}{\pi_0}}\right) + \sum_{j=1}^{r_0} k_{0j}e^{\frac{H-1}{\pi_0}}(x^{\nu_{0j}} - 1) = 0.$$

(A3)

Comparing Eq. (A3) with Eq. (3), we finally conclude that $F = e^{\frac{H-1}{\pi_0}}$ in steady-state conditions.

APPENDIX B: FINITENESS OF THE IRREDUCIBLE STATE SPACE OF THE MODIFIED MODEL

Here we prove the following criterion: if all the first-order reactions except degradation reactions has a conservation law with positive coefficients, then the modified model must have a finite irreducible state space. To prove this criterion, we need the following lemma.

Lemma IV.1. Suppose that a family of reactions

$$R_i : \mu_i^1 S_1 + \cdots + \mu_i^N S_N \xrightarrow{k_i} \nu_i^1 S_1 + \cdots + \nu_i^N S_N, \quad i = 1, \ldots, r,$$
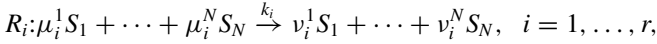
has the conservation law

$$\omega_1\mu_i^1 + \omega_2\mu_i^2 + \cdots + \omega_N\mu_i^N = \omega_1\nu_i^1 + \omega_2\nu_i^2 + \cdots + \omega_N\nu_i^N,$$

for all $i = 1, \ldots, r$. If the coefficients $\omega_1, \cdots, \omega_N$ are all positive, then for any microstate n, the family of reactions can only lead microstate n to a finite number of microstates.

Proof. For simplicity, we write $\mu_i = (\mu_i^1, \cdots, \mu_i^N)$ and $\nu_i = (\nu_i^1, \cdots, \nu_i^N)$. Suppose that the family of reactions lead microstate n to microstate $\bar{n}$. Then there exists nonnegative integers $\xi_1, \cdots, \xi_r$ such that

$$\bar{n} = n + \xi_1(\nu_1 - \mu_1) + \cdots + \xi_r(\nu_r - \mu_r),$$

with ξ_i being the number of occurrence of the ith reaction. Then we have

$$\omega \cdot \bar{n} = \omega \cdot n + \xi_1\omega \cdot (\nu_1 - \mu_1) + \cdots + \xi_r\omega \cdot (\nu_r - \mu_r)$$

$$= \omega \cdot n,$$

where $\omega \cdot n = \omega_1 n_1 + \omega_2 n_2 + \cdots + \omega_N n_N$ denotes the usual scalar product of vectors. This clearly shows that $\omega \cdot (\bar{n} - n) = 0$, which implies that all the microstates accessible from n must lie in some hyperplane H with normal vector ω. Since the normal vector ω has positive components, it always points into the first orthant and thus the hyperplane H can only contain a finite number of microstates within the first orthant (see Fig. 6 for an illustration). This completes the proof. ∎

We are now in a position to prove the above criterion. Since the original model is ergodic, all nonzero microstates can lead

to the zero microstate via a series of first-order reactions. Since first-order reactions result in the same transitions for the original and modified models, for the modified model, all nonzero microstates can also lead to the zero microstate via a series of first-order reactions. This shows that the zero microstate is contained in the irreducible state space of the modified model. Therefore, to identify the irreducible state space of the modified model, we only need to determine which microstates are accessible from the zero microstate. First, since zero-order reactions can only occur at the zero microstate for the modified model, all zero-order reactions can only lead the zero microstate to a finite number of microstates, denoted by $n_1, \cdots, n_k$. Next, since the family of first-order reactions except degradation reactions has a conservation law with positive coefficients, it follows from Lemma IV.1 that all first-order reactions can only lead microstates $n_1, \cdots, n_k$ to a finite number of microstates. This completes the proof of the criterion.

APPENDIX C: JOINT DISTRIBUTION FOR THE GENE EXPRESSION MODEL WITH NASCENT mRNA

Let $\pi_{m_*, m, n}$ denote the steady-state probability of observing microstate (m_*, m, n) for the modified model. From the transition diagram in Fig. 3(b), these steady-state probabilities
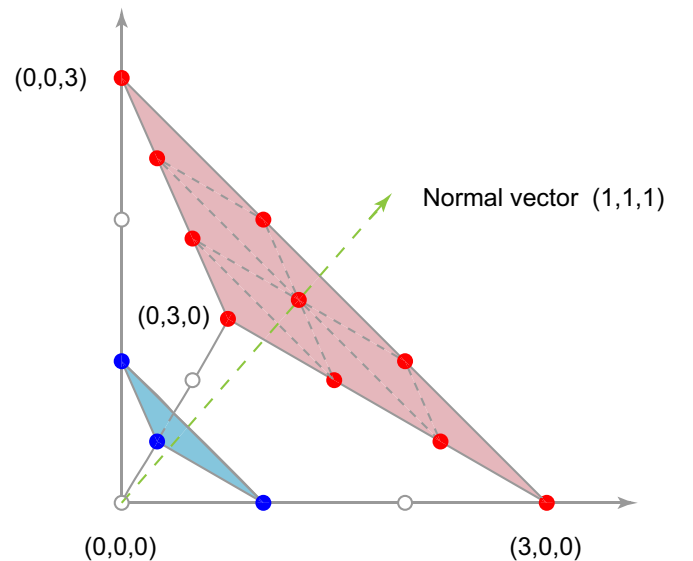

FIG. 6. Two hyperplanes with the same normal vector $(1,1,1)$. The blue hyperplane contains three points in the first orthant of the lattice space and the red hyperplane contains ten points.

satisfy the following equations:

$$f\pi_{1,0,0} + v\pi_{0,1,0} + d\pi_{0,0,1} - s\pi_{0,0,0} = 0,$$
$$s\pi_{0,0,0} - (k+f)\pi_{1,0,0} = 0,$$
$$k\pi_{1,0,0} + d\pi_{0,1,1} - (u+v)\pi_{0,1,0} = 0,$$
$$v\pi_{0,1,n} + (n+1)d\pi_{0,0,n+1} - nd\pi_{0,0,n} = 0, \quad n \geqslant 1,$$
$$u\pi_{0,1,n-1} + (n+1)d\pi_{0,1,n+1} - (nd+u+v)\pi_{0,1,n} = 0, \quad n \geqslant 1. \tag{C1}$$

To proceed, we define the following two generating functions:

$$\phi(y) = \sum_{n=0}^{\infty} \pi_{0,0,n} y^n, \quad \psi(y) = \sum_{n=0}^{\infty} \pi_{0,1,n} y^n.$$

Then the generating function of the modified model is given by

$$H(x_\star, x, y) = \pi_{1,0,0} x_\star + \phi(y) + x\psi(y). \tag{C2}$$

Note that Eq. (C1) can be converted into the following system of ODEs:

$$k\pi_{1,0,0} + (uy - u - v)\psi(y) + d(1-y)\psi'(y) = 0, \tag{C3}$$

$$-k\pi_{1,0,0} + v\psi(y) + d(1-y)\phi'(y) = 0. \tag{C4}$$

By the second equation in Eq. (C1) we obtain

$$\pi_{1,0,0} = a\pi_{0,0,0},$$

where $a = s/(k+f)$. Taking the derivative on both sides of Eq. (C3) yields

$$d(1-y)\psi''(y) + (uy - u - v - d)\psi'(y) + u\psi(y) = 0.$$

This is a confluent hypergeometric differential equation [60, Eq. (13.2.1)] and its solution is given by

$$\psi(y) = K\,_1F_1[1; 1+v; \mu(y-1)],$$

where $v = v/d$, $\mu = u/d$ and K is a normalization constant. Taking $y = 1$ in Eq. (C3), we can determine the normalization constant K as

$$K = \psi(1) = b\pi_{0,0,0},$$

where $b = ks/(k+f)v$. However, it follows from Eq. (C4) and the power series expansion of the confluent hypergeometric function that

$$\phi'(y) = \frac{bv\pi_{0,0,0}}{y-1}\{_1F_1[1; 1+v; \mu(y-1)] - 1\} = \frac{bv\pi_{0,0,0}}{y-1}\sum_{i=1}^{\infty}\frac{[\mu(y-1)]^i}{(1+v)_i}$$

$$= \frac{bv\mu\pi_{0,0,0}}{1+v}\sum_{i=0}^{\infty}\frac{[\mu(y-1)]^i}{(2+v)_i} = \frac{bv\mu\pi_{0,0,0}}{1+v}\,_1F_1[1; 2+v; \mu(y-1)], \tag{C5}$$

where $(x)_i = x(x+1)\dots(x+i-1)$ is the Pochhammer symbol. Thus, we obtain

$$\phi(y) = \frac{bv\mu\pi_{0,0,0}}{1+v}\int_1^y {}_1F_1[1; 2+v; \mu(z-1)]dz + C,$$

where C is an undetermined constant. It then follows from Eqs. (6) and (C2) that the generating function of the original model is given by

$$F(x_\star, x, y) = e^{\frac{a\pi_{0,0,0}x_\star + \phi(y) + x\psi(y) - 1}{\pi_{0,0,0}}} = e^{ax_\star + \frac{bv\mu}{1+v}\int_1^y {}_1F_1(1; 2+v; \mu(z-1))dz + bx\,_1F_1(1; 1+v; \mu(y-1)) + (C-1)/\pi_{0,0,0}}.$$

By using the fact that $F(1, 1, 1) = 1$, we can determined the constant C and thus the generating function can be rewritten as

$$F(x_\star, x, y) = e^{a(x_\star - 1) + \frac{bv\mu}{1+v}\int_1^y {}_1F_1[1; 2+v; \mu(z-1)]dz + b\{x\,_1F_1[1; 1+v; \mu(y-1)] - 1\}}. \tag{C6}$$

To proceed, recall that the confluent hypergeometric function satisfies the following recurrence relation [60, Eq. (13.3.3)]:

$$_1F_1[2; 2+v; \mu(z-1)] + v\,_1F_1[1; 2+v; \mu(z-1)] - (1+v)\,_1F_1[1; 1+v; \mu(z-1)] = 0. \tag{C7}$$

Moreover, it follows from the differentiation formula of confluent hypergeometric functions [60, Eq. (13.3.15)] that

$$\frac{d}{dz} \, _1F_1[1; 1+v; \mu(z-1)] = \frac{\mu}{1+v} \, _1F_1[2; 2+v; \mu(z-1)].$$

Integrating both sides of Eq. (C7) from 1 to y, we obtain

$$\frac{\mu v}{1+v} \int_1^y {}_1F_1[1; 2+v; \mu(z-1)]dz = \mu \int_1^y {}_1F_1[1; 1+v; \mu(z-1)]dz - {}_1F_1[1; 1+v; \mu(y-1)] + 1. \tag{C8}$$

Finally, inserting the above equation into Eq. (C6), we obtain Eq. (15) in the main text.

APPENDIX D: JOINT DISTRIBUTION FOR THE GENE REGULATORY MODEL WITH TRANSLATIONAL BURSTING

Let π_{n_1,n_2} denote the steady-state probability of observing microstate (n_1, n_2) for the modified model. From the transition diagram in Fig. 4(b), these steady-state probabilities satisfy the following equations:

$$
\begin{aligned}
& d_1\pi_{1,0} + d_2\pi_{0,1} - u_1\pi_{0,0} = 0, \\
& u_1\pi_{0,0} + d_2\pi_{1,1} + 2d_1\pi_{2,0} - (d_1 + u_2p)\pi_{1,0} = 0, \\
& d_1\pi_{1,n_2} + (n_2+1)d_2\pi_{0,n_2+1} - n_2d_2\pi_{0,n_2} = 0, \quad n_2 \geqslant 1, \\
& (n_2+1)d_2\pi_{1,n_2+1} + \sum_{i=0}^{n_2-1} u_2 p^{n_2-i} q \pi_{1,i} - (u_2p + n_2d_2 + d_1)\pi_{1,n_2} = 0, \quad n_2 \geqslant 1.
\end{aligned}
\tag{D1}
$$

To proceed, we define the following two generating functions:

$$\phi(y_2) = \sum_{n_2=0}^{\infty} \pi_{0,n_2} y_2^{n_2}, \quad \psi(y_2) = \sum_{n_2=0}^{\infty} \pi_{1,n_2} y_2^{n_2}.$$

Then the generating function of the modified model can be written as

$$H(y_1, y_2) = \phi(y_2) + y_1\psi(y_2). \tag{D2}$$

Note that Eq. (D1) can be converted into the following system of ODEs:

$$u_1\pi_{0,0} + \left[\frac{u_2p(y_2-1)}{1-py_2} - d_1\right]\psi(y_2) + d_2(1-y_2)\psi'(y_2) = 0, \tag{D3}$$

$$-u_1\pi_{0,0} + d_2(1-y_2)\phi'(y_2) + d_1\psi(y_2) = 0. \tag{D4}$$

Taking the derivative on both sides of Eq. (D3) yields

$$a(y_2)\psi''(y_2) + b(y_2)\psi'(y_2) + c(y_2)\psi(y_2) = 0,$$

where

$$a(y_2) = (py_2-1)^2(y_2-1), \quad b(y_2) = (py_2-1)[(\mu_2+v+1)py_2 - (\mu_2p+v+1)], \quad c(y_2) = \mu_2p(p-1).$$

This is a hypergeometric differential equation and its solution is given by

$$\psi(y_2) = K \, _2F_1[-\mu_2, 1; 1+v; \omega(y_2)],$$

where $\omega(y_2) = p(y_2-1)/(py_2-1)$ and K is a normalization constant. Taking $y_2 = 1$ in Eq. (D3), the normalization constant can be determined as

$$K = \psi(1) = \mu_1\pi_{0,0}.$$

Next, we compute $\phi(y_2)$ by using Eq. (D4). However, it follows from Eq. (D4) and the power series expansion of the hypergeometric function that

$$
\begin{aligned}
\phi'(y_2) &= \frac{\mu_1 v \pi_{0,0}}{y_2-1}\{_2F_1[-\mu_2, 1; 1+v; w(y_2)] - 1\} = \frac{\mu_1 v p \pi_{0,0}}{(py_2-1)w(y_2)} \sum_{i=1}^{\infty} \frac{(-\mu_2)_i w(y_2)^i}{(1+v)_i} \\
&= \frac{\mu_1 v p \pi_{0,0}}{(py_2-1)} \sum_{i=0}^{\infty} \frac{(-\mu_2)_{i+1} w(y_2)^i}{(1+v)_{i+1}} = \frac{-\mu_1\mu_2 v p \pi_{0,0}}{1+v} \frac{_2F_1[1-\mu_2, 1; 2+v; w(y_2)]}{py_2-1}.
\end{aligned}
$$

Thus, we obtain

$$\phi(y_2) = \frac{-\mu_1\mu_2 v p \pi_{0,0}}{1+v} \int_1^{y_2} \frac{_2F_1[1-\mu_2, 1; 2+v; w(z)]}{pz-1}dz + C, \tag{D5}$$

where C is an undetermined constant. It then follows from Eqs. (6) and (D2) that the generating function of the original model is given by

$$F(y_1, y_2) = e^{\frac{\phi(y_2)+y_1\psi(y_2)-1}{\pi_{0,0}}} = e^{\mu_1 y_1 {}_2F_1(-\mu_2,1;1+v;\omega(y_2))-\frac{\mu_1\mu_2 vp}{1+v}\int_1^{y_2}\frac{{}_2F_1(1-\mu_2,1;2+v;\omega(z))}{pz-1}dz+(C-1)/\pi_{0,0}}. \quad (D6)$$

By using the fact that $F(1,1) = 1$, we can determined the constant C and thus the generating function can be rewritten as

$$F(y_1, y_2) = e^{\mu_1[y_1 {}_2F_1(-\mu_2,1;1+v;\omega(y_2))-1]-\frac{\mu_2 vp}{1+v}\int_1^{y_2}\frac{{}_2F_1(1-\mu_2,1;2+v;\omega(z))}{pz-1}dz}. \quad (D7)$$

To proceed, recall that the hypergeometric function satisfies the following recurrence relation [60, Eqs. (15.5.13) and (15.5.15)]:

$$v\,{}_2F_1[1-\mu_2, 1; 2+v; \omega(z)] - (1+v)\,{}_2F_1[-\mu_2, 1; 1+v; \omega(z)] + [1-\omega(z)]\,{}_2F_1[1-\mu_2, 2; 2+v; \omega(z)] = 0. \quad (D8)$$

Since

$$\omega'(z) = \frac{p[1-\omega(z)]}{pz-1},$$

multiplying $\mu_1 p/(pz-1)$ on both sides of Eq. (D8) yields

$$\mu_1 vp\frac{{}_2F_1[1-\mu_2, 1; 2+v; \omega(z)]}{pz-1} - (1+v)\mu_1 p\frac{{}_2F_1[-\mu_2, 1; 1+v; \omega(z)]}{pz-1} + \mu_1\omega'(z)\,{}_2F_1[1-\mu_2, 2; 2+v; \omega(z)] = 0. \quad (D9)$$

Moreover, it follows from the differentiation formula of Gaussian hypergeometric functions [60, Eq. (15.5.1)] that

$$\frac{d}{dz}\,{}_2F_1(-\mu_2, 1; 1+v; \omega(z)) = \frac{-\mu_2}{1+v}\omega'(z)\,{}_2F_1[1-\mu_2, 2; 2+v; \omega(z)].$$

Integrating both sides of Eq. (D9) from 1 to y_2, we obtain

$$\frac{\mu_1\mu_2 vp}{1+v}\int_1^{y_2}\frac{{}_2F_1[1-\mu_2, 1; 2+v; \omega(z)]}{pz-1}dz$$
$$= \mu_1\mu_2 p\int_1^{y_2}\frac{{}_2F_1[-\mu_2, 1; 1+v; \omega(z)]}{pz-1}dz + \mu_1\,{}_2F_1[-\mu_2, 1; 1+v; \omega(y_2)] - \mu_1.$$

Inserting the above equation into Eq. (D7), we obtain

$$F(y_1, y_2) = e^{\mu_1 {}_2F_1[-\mu_2, 1; 1+v; \omega(y_2)](y_1-1)-\mu_1\mu_2 p\int_1^{y_2}\frac{{}_2F_1[-\mu_2, 1; 1+v; \omega(z)]}{pz-1}dz}.$$

Finally, using the Kummer's transformation [60, Eq. (15.5.1)], we obtain Eq. (19) in the main text.

APPENDIX E: JOINT DISTRIBUTION FOR THE GENE EXPRESSION MODEL WITH ALTERNATIVE SPLICING

Let π_{m,m_1,n_1,m_2,n_2} denote the steady-state probability of observing microstate (m, m_1, n_1, m_2, n_2) for the modified model. Given that there are n copies of the regulator, these steady-state probabilities satisfy the following equations:

$$f\pi_{1,0,0,0,0} + v_1\pi_{0,1,0,0,0} + v_2\pi_{0,0,0,1,0} + d_1\pi_{0,0,1,0,0} + d_2\pi_{0,0,0,0,1} - s\pi_{\mathbf{0}} = 0,$$
$$s\pi_{0,0,0,0,0} - (k_1(n) + k_2(n) + f)\pi_{1,0,0,0,0} = 0,$$
$$k_1(n)\pi_{1,0,0,0,0} + d_1\pi_{0,1,1,0,0} - (u_1 + v_1)\pi_{0,1,0,0,0} = 0,$$
$$k_2(n)\pi_{1,0,0,0,0} + d_2\pi_{0,0,0,1,1} - (u_2 + v_2)\pi_{0,0,0,1,0} = 0,$$
$$v_1\pi_{0,1,n_1,0,0} + (n_1+1)d_1\pi_{0,0,n_1+1,0,0} - n_1 d_1\pi_{0,0,n_1,0,0} = 0, \quad n_1 \geqslant 1,$$
$$u_1\pi_{0,1,n_1-1,0,0} + (n_1+1)d_1\pi_{0,0,n_1+1,0,0} - (n_1 d_1 + u_1 + v_1)\pi_{0,1,n_1,0,0} = 0, \quad n_1 \geqslant 1,$$
$$v_2\pi_{0,0,0,1,n_2} + (n_2+1)d_2\pi_{0,0,0,0,n_2+1} - n_2 d_2\pi_{0,0,0,0,n_2} = 0, \quad n_2 \geqslant 1,$$
$$u_2\pi_{0,0,0,1,n_2-1} + (n_2+1)d_2\pi_{0,0,0,0,n_2+1} - (n_2 d_2 + u_2 + v_2)\pi_{0,0,0,1,n_2} = 0, \quad n_2 \geqslant 1. \quad (E1)$$

To proceed, we define the following generating functions:

$$\phi_1(y_1) = \sum_{n_1=0}^{\infty}\pi_{0,0,n_1,0,0}y_1^{n_1}, \quad \psi_1(y_1) = \sum_{n_1=0}^{\infty}\pi_{0,1,n_1,0,0}y_1^{n_1},$$

$$\phi_2(y_2) = \sum_{n_2=1}^{\infty}\pi_{0,0,0,0,n_2}y_2^{n_2}, \quad \psi_2(y_2) = \sum_{n_2=0}^{\infty}\pi_{0,0,0,1,n_2}y_2^{n_2}.$$

Then, given that there are n copies of the regulator, the generating function of the modified model is given by

$$H(x, x_1, y_1, x_2, y_2 | n) = \pi_{1,0,0,0,0} x + \phi_1(y_1) + \psi_1(y_1) x_1 + \phi_2(y_2) + \psi_2(y_2) x_2. \tag{E2}$$

Note that Eq. (E1) can be converted into the following system of ODEs:

$$\pi_{1,0,0,0,0} k_i(n) + (u_i y_i - u_i - v_i) \psi_i(y_i) + d_i(1 - y_i) \psi_i'(y_i) = 0, \tag{E3}$$

$$-\pi_{1,0,0,0,0} k_i(n) + v_i \psi_i(y_i) + d_i(1 - y_i) \phi_i'(y_i) = 0, \tag{E4}$$

for $i = 1, 2$. By the second equation in Eq. (E1) we obtain

$$\pi_{1,0,0,0,0} = a(n)\pi_0,$$

where $a(n) = s/(k_1(n) + k_2(n) + f)$. Note that Eqs. (E3) and (E4) have a similar form as Eqs. (C3) and (C4). By using the same procedure used for solving Eqs. (C3) and (C4), we obtain

$$\psi_i(y_i) = K_i(n) b_i \pi_0 \, {}_1F_1[1; 1 + v_i; \mu(y_i - 1)], \quad \phi_i(y_i) = \frac{K_i(n) b_i v_i \mu_i \pi_0}{1 + v_i} \int_1^{y_i} {}_1F_1[1; 2 + v_i; \mu_i(z - 1)] dz + C_i,$$

where C_i are two undetermined constants and

$$K_1(n) = \frac{k_1(n)}{k_1(n) + k_2(n) + f}, \quad K_2(n) = \frac{k_2(n)}{k_1(n) + k_2(n) + f},$$

$$b_1 = \frac{s}{v_1}, \quad b_2 = \frac{s}{v_2}, \quad \mu_1 = \frac{u_1}{d_1}, \quad \mu_2 = \frac{u_2}{d_2}, \quad v_1 = \frac{v_1}{d_1}, \quad v_2 = \frac{v_2}{d_2}.$$

It thus follows from Eqs. (6) and (E2) that the generating function of the original model, given that there are n copies of the regulator, is given by

$$F(x, x_1, y_1, x_2, y_2 | n) = e^{a(n)(x-1) + \sum_{i=1}^2 K_i(n) b_i \left\{ x_i \, {}_1F_1[1; 1+v_i; \mu_i(y_i-1)] + \frac{\mu_i v_i}{1+v_i} \int_1^{y_i} {}_1F_1[1; 2+v_i; \mu_i(z-1)] dz \right\}}.$$

Replacing μ, v, and y in Eq. (C8) by μ_i, v_i, and y_i for $i = 1, 2$ and inserting the resulting two equations into the above equation give Eq. (26) in the main text.

Next we compute the correlation coefficients between the copy numbers of the two mRNA-protein isoforms under the assumption that the copy number of the regulator has a Poisson distribution with parameter λ. In this case, the generating function of the original model is given by

$$F(x, x_1, y_1, x_2, y_2) = \sum_{n=0}^\infty \frac{\lambda^n e^{-\lambda}}{n!} e^{a(n)(x-1) + \sum_{i=1}^2 K_i(n) b_i \left\{ (x_i - 1) \, {}_1F_1[1; 1+v_i; \mu_i(y_i-1)] + \mu_i \int_1^{y_i} {}_1F_1[1; 1+v_i; \mu_i(z-1)] dz \right\}}.$$

We first focus on the correlation between the two mRNA isoforms. Using the power series expansion and the Kummer transformation [60, Eq. (13.2.39)] of confluent hypergeometric functions, the derivative of F with respect to x_i is given by

$$\frac{\partial F}{\partial x_i}(1, 1, 1, 1, 1) = \sum_{n=0}^\infty K_i(n) b_i \frac{\lambda^n e^{-\lambda}}{n!} = \frac{b_i}{\xi_1 + \xi_2} \sum_{n=0}^\infty \left(\xi_i + \frac{\alpha_i \gamma}{n + \gamma} \right) \frac{\lambda^n e^{-\lambda}}{n!}$$

$$= \frac{b_i}{\xi_1 + \xi_2} \sum_{n=0}^\infty \left(\xi_i + \frac{\alpha_i(\gamma)_n}{(\gamma + 1)_n} \right) \frac{\lambda^n e^{-\lambda}}{n!} = \frac{b_i(\xi_i + \alpha_i h_1)}{\xi_1 + \xi_2}, \tag{E5}$$

and the second derivative of F with respect to x_i and x_j is given by

$$\frac{\partial^2 F}{\partial x_i \partial x_j}(1, 1, 1, 1, 1) = \sum_{n=0}^\infty K_i(n) K_j(n) b_i b_j \frac{\lambda^n e^{-\lambda}}{n!}$$

$$= \frac{b_i b_j}{(\xi_1 + \xi_2)^2} \sum_{n=0}^\infty \left[\xi_i \xi_j + \frac{(\alpha_i \xi_j + \alpha_j \xi_i)(\gamma)_n}{(\gamma + 1)_n} + \frac{\alpha_i \alpha_j (\gamma)_n (\gamma)_n}{(\gamma + 1)_n (\gamma + 1)_n} \right] \frac{\lambda^n e^{-\lambda}}{n!}$$

$$= \frac{b_i b_j [\xi_i \xi_j + (\alpha_i \xi_j + \alpha_j \xi_i) h_1 + \alpha_i \alpha_j h_2]}{(\xi_1 + \xi_2)^2}, \tag{E6}$$

where

$$\alpha_1 = \frac{\xi_2 \eta_1 - \xi_1 \eta_2 - \xi_1 f}{\eta_1 + \eta_2 + f}, \quad \alpha_2 = \frac{\xi_1 \eta_2 - \xi_2 \eta_1 - \xi_2 f}{\eta_1 + \eta_2 + f}, \quad \gamma = \frac{\eta_1 + \eta_2 + f}{\xi_1 + \xi_2},$$

$$h_1 = {}_1F_1(1; \gamma + 1; -\lambda), \quad h_2 = {}_2F_2(\gamma, \gamma; \gamma + 1, \gamma + 1; \lambda) e^{-\lambda}.$$

Inserting the above two equations into Eq. (10), we obtain Eq. (27) in the main text. We next focus on the correlation between the two protein isoforms. Using the power series expansion and the Kummer transformation [60, Eq. (13.2.39)] of confluent hypergeometric functions, it is not hard to prove that

$$\frac{\partial F}{\partial y_i}(1, 1, 1, 1, 1) = \sum_{n=0}^{\infty} K_i(n) \mu_i b_i \frac{\lambda^n e^{-\lambda}}{n!} = \frac{\mu_i b_i (\xi_i + \alpha_i h_1)}{\xi_1 + \xi_2}.$$

Similarly, the second derivative of F with respect to y_1 and y_2 is given by

$$\frac{\partial^2 F}{\partial y_1 \partial y_2}(1, 1, 1, 1, 1) = \sum_{n=0}^{\infty} K_1(n) K_2(n) \mu_1 \mu_2 b_1 b_2 \frac{\lambda^n e^{-\lambda}}{n!} = \frac{\mu_1 \mu_2 b_1 b_2 [\xi_1 \xi_2 + (\alpha_1 \xi_2 + \alpha_2 \xi_1) h_1 + \alpha_1 \alpha_2 h_2]}{(\xi_1 + \xi_2)^2},$$

and the second derivative of F with respect to y_i is given by

$$\frac{\partial^2 F}{\partial y_i^2}(1, 1, 1, 1, 1) = \sum_{n=0}^{\infty} \left\{ [K_i(n) \mu_i b_i]^2 + K_i(n) \mu_i b_i \frac{\mu_i}{1 + v_i} \right\} \frac{\lambda^n e^{-\lambda}}{n!} = \frac{\mu_i^2 b_i^2 (\xi_i^2 + 2\alpha_i \xi_i h_1 + \alpha_i^2 h_2)}{(\xi_1 + \xi_2)^2} + \frac{\mu_i^2 b_i (\xi_i + \alpha_i h_1)}{(\xi_1 + \xi_2)(1 + v_i)}.$$

Inserting the above three equations into Eq. (10) gives Eq. (28) in the main text.

Finally, we prove that $\xi_i + \alpha_i h_1$ and $h_2 - h_1^2$ are positive for any choice of rate constants. First, we note that

$$\frac{\alpha_i}{\xi_i} = \frac{\eta_i(\xi_1 + \xi_2)}{\xi_i(\eta_1 + \eta_2 + f)} - 1 > -1, \quad h_1 = e^{-\lambda} {}_1F_1(\gamma; \gamma + 1; \lambda) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\gamma)_n}{(\gamma + 1)_n} \frac{\lambda^n}{n!} < 1.$$

Combining the above inequalities shows that $\xi_i + \alpha_i h_1 > 0$. Second, it follows from the Cauchy product formula of two infinite series that

$$h_2 - h_1^2 = e^{-2\lambda} \{ e^{\lambda} {}_2F_2(\gamma, \gamma; \gamma + 1, \gamma + 1; \lambda) - [{}_1F_1(\gamma; \gamma + 1; \lambda)]^2 \}$$

$$= e^{-2\lambda} \left[\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \sum_{n=0}^{\infty} \left(\frac{\gamma}{\gamma + n} \right)^2 \frac{\lambda^n}{n!} - \left(\sum_{n=0}^{\infty} \frac{\gamma}{\gamma + n} \frac{\lambda^n}{n!} \right)^2 \right]$$

$$= e^{-2\lambda} \gamma^2 \left\{ \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \sum_{i=0}^{n} \binom{i}{n} \left[\frac{1}{(\gamma + i)^2} - \frac{1}{(\gamma + i)(\gamma + n - i)} \right] \right\}. \tag{E7}$$

We next prove that

$$\sum_{i=0}^{n} \binom{i}{n} \left[\frac{1}{(\gamma + i)^2} - \frac{1}{(\gamma + i)(\gamma + n - i)} \right] > 0, \tag{E8}$$

for any $\gamma > 0$ and $n > 0$. Putting the first term and the last term in the left-hand size of Eq. (E8) together yields

$$\left[\frac{1}{\gamma^2} - \frac{1}{\gamma(\gamma + n)} \right] + \left[\frac{1}{(\gamma + n)^2} - \frac{1}{\gamma(\gamma + n)} \right] = \frac{n}{\gamma^2(\gamma + n)} - \frac{n}{(\gamma + n)^2 \gamma} = \frac{n^2}{\gamma^2(\gamma + n)^2} > 0.$$

Similarly, putting the second term and the last but one term together gives

$$n \left[\frac{1}{(\gamma + 1)^2} - \frac{1}{(\gamma + 1)(\gamma + n - 1)} \right] + n \left[\frac{1}{(\gamma + n - 1)^2} - \frac{1}{(\gamma + 1)(\gamma + n - 1)} \right]$$

$$= n \left[\frac{n - 2}{(\gamma + 1)^2(\gamma + n - 1)} - \frac{n - 2}{(\gamma + n - 1)^2(\gamma + 1)} \right] = \frac{n(n - 2)^2}{(\gamma + 1)^2(\gamma + n - 1)^2} > 0.$$

If n is an odd number, then repeating the above procedure shows that the left-hand size of Eq. (E8) is positive. If n is an even number, then the $(n/2 + 1)$th term in the left-hand size of Eq. (E8) cannot be paired in the above manner. However, in this case it is easy to check the $(n/2 + 1)$th term must equal zero. Thus, we have proved Eq. (E8). Combining Eqs. (E7) and (E8) finally shows that $h_2 - h_1^2 > 0$.

[1] D. F. Anderson and T. G. Kurtz, *Stochastic Analysis of Biochemical Systems* (Springer, Berlin, 2015).

[2] H. Qian, and E. L. Elson, Single-molecule enzymology: Stochastic Michaelis–Menten kinetics, Biophys. Chem. **101**, 565 (2002).

[3] C. Jia, X.-F. Liu, M.-P. Qian, D.-Q. Jiang, and Y.-P. Zhang, Kinetic behavior of the general modifier mechanism of Botts and Morales with nonequilibrium binding, J. Theor. Biol. **296**, 13 (2012).

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

[5] D. Schnoerr, G. Sanguinetti, and R. Grima, The complex chemical Langevin equation, J. Chem. Phys. **141**, 024103 (2014).

[6] J. Holehouse, A. Sukys, and R. Grima, Stochastic time-dependent enzyme kinetics: Closed-form solution and transient bimodality, J. Chem. Phys. **153**, 164113 (2020).

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

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

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

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

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

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

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

[14] P. Bokes and A. Singh, Protein copy number distributions for a self-regulating gene in the presence of decoy binding sites, PloS ONE **10**, e0120555 (2015).

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

[16] C. Jia and R. Grima, Dynamical phase diagram of an autoregulating gene in fast switching conditions, J. Chem. Phys. **152**, 174110 (2020).

[17] B. Mélykúti, J. P. Hespanha, and M. Khammash, Equilibrium distributions of simple biochemical reaction systems for time-scale separation in stochastic reaction networks, J. R. Soc. Interface **11**, 20140054 (2014).

[18] E. Lakatos, A. Ale, P. D. Kirk, and M. P. Stumpf, Multivariate moment closure techniques for stochastic kinetic models, J. Chem. Phys. **143**, 094107 (2015).

[19] J. Zhang, Q. Nie, and T. Zhou, A moment-convergence method for stochastic analysis of biochemical reaction networks, J. Chem. Phys. **144**, 194109 (2016).

[20] P. Thomas, N. Popovic, and R. Grima, Phenotypic switching in gene regulatory networks, Proc. Natl. Acad. Sci. USA **111**, 6994 (2014).

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

[22] I. M. Krieger, and P. J. Gans, First-order stochastic processes, J. Chem. Phys. **32**, 247 (1960).

[23] I. Darvey and P. Staff, Stochastic approach to first-order chemical reaction kinetics, J. Chem. Phys. **44**, 990 (1966).

[24] N. G. Van Kampen, The equilibrium distribution of a chemical mixture, Phys. Lett. A **59**, 333 (1976).

[25] P. J. Gans, Open first-order stochastic processes, J. Chem. Phys. **33**, 691 (1960).

[26] C. Gadgil, C. H. Lee, and H. G. Othmer, A stochastic analysis of first-order reaction networks, Bull. Math. Biol. **67**, 901 (2005).

[27] W. J. Heuett and H. Qian, Grand canonical Markov model: A stochastic theory for open nonequilibrium biochemical networks, J. Chem. Phys. **124**, 044110 (2006).

[28] T. Jahnke and W. Huisinga, Solving the chemical master equation for monomolecular reaction systems analytically, J. Math. Biol. **54**, 1 (2007).

[29] F. J. M. Horn and R. Jackson, General mass action kinetics, Arch. Ration. Mech. An. **47**, 81 (1972).

[30] D. F. Anderson, G. Craciun, and T. G. Kurtz, Product-form stationary distributions for deficiency zero chemical reaction networks, Bull. Math. Biol. **72**, 1947 (2010).

[31] D. Cappelletti and C. Wiuf, Product-form poisson-like distributions and complex balanced reaction systems, SIAM J. Appl. Math. **76**, 411 (2016).

[32] M. Reis, J. A. Kromer, and E. Klipp, General solution of the chemical master equation and modality of marginal distributions for hierarchic first-order reaction networks, J. Math. Biol. **77**, 377 (2018).

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

[34] H. Pendar, T. Platini, and R. V. Kulkarni, Exact protein distributions for stochastic models of gene expression using partitioning of Poisson processes, Phys. Rev. E **87**, 042720 (2013).

[35] Q. Wang, and T. Zhou, Alternative-splicing-mediated gene expression, Phys. Rev. E **89**, 012713 (2014).

[36] J. R. Norris, J. R. Norris, and J. R. Norris, *Markov Chains*, No. 2 (Cambridge University Press, Cambridge, UK, 1998).

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

[38] W. P. Johnson, The curious history of Faà di Bruno's formula, Am. Math. Monthly **109**, 217 (2002).

[39] E. T. Bell, Partition polynomials, Ann. Math. **29**, 38 (1927).

[40] M. D. Ryan, A. M. Q. King, and G. P. Thomas, Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence, J. Gen. Virol. **72**, 2727 (1991).

[41] A. L. Szymczak and D. A. A. Vignali, Development of 2A peptide-based strategies in the design of multicistronic vectors, Expert Opin. Biol. Th. **5**, 627 (2005).

[42] Z. Liu *et al.*, Systematic comparison of 2A peptides for cloning multigenes in a polycistronic vector, Sci. Rep. **7**, 2193 (2017).

[43] M. D. Ryan *et al.*, Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal "skip," J. Gen. Virol. **82**, 1013 (2001).

[44] M. L. L. Donnelly, L. E. Hughes, G. Luke, H. Mendoza, and M. D. Ryan, The "cleavage" activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring "2A-like" sequences, J. Gen. Virol. **82**, 1027 (2001).

[45] S. Loukas and C. Kemp, The index of dispersion test for the bivariate Poisson distribution, Biometrics **42**, 941 (1986).

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

[47] N. Saitou, *Introduction to Evolutionary Genomics*, Computational Biology Book Series (COBO), Vol. 17 (Springer, Cham, 2018), 2nd ed.

[48] G. La Manno *et al.*, RNA velocity of single cells, Nature (London) **560**, 494 (2018).

[49] T. Li, J. Shi, Y. Wu, and P. Zhou, On the mathematics of RNA velocity I: Theoretical analysis, CSIAM Trans. Appl. Math. **2**, 1 (2021).

[50] C. Jia and R. Grima, 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).

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

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

[53] S. Ajith *et al.*, Position-dependent activity of CELF2 in the regulation of splicing and implications for signal-responsive regulation in T cells, RNA Biol. **13**, 569 (2016).

[54] X. D. Fu, and M. Ares, Context-dependent control of alternative splicing by RNA-binding proteins, Nat. Rev. Genet. **15**, 689 (2014).

[55] F. E. Baralle and J. Giudice, Alternative splicing as a regulator of development and tissue identity, Nat. Rev. Mol. Cell Biol. **18**, 437 (2017).

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

[57] C. Jia, P. Xie, M. Chen, and M. Q. Zhang, Stochastic fluctuations can reveal the feedback signs of gene regulatory networks at the single-molecule level, Sci. Rep. **7**, 1 (2017).

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

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

[60] F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, NIST Digital Library of Mathematical Functions, Release 1.1.2 of 2021-06-15 (2017), http://dlmf.nist.gov/.