Editors' Suggestion

Mean-field theory accurately captures the variation of copy number distributions across the mRNA life cycle

Juraj Szavits-Nossan* and Ramon Grima†
School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, United Kingdom

(Received 24 August 2021; accepted 15 November 2021; published 14 January 2022)

We consider a stochastic model where a gene switches between two states, an mRNA transcript is released in the active state, and subsequently it undergoes an arbitrary number of sequential unimolecular steps before being degraded. The reactions effectively describe various stages of the mRNA life cycle such as initiation, elongation, termination, splicing, export, and degradation. We construct a mean-field approach that leads to closed-form steady-state distributions for the number of transcript molecules at each stage of the mRNA life cycle. By comparison with stochastic simulations, we show that the approximation is highly accurate over all the parameter space, independent of the type of expression (constitutive or bursty) and of the shape of the distribution (unimodal, bimodal, and nearly bimodal). The theory predicts that in a population of identical cells, any bimodality is gradually washed away as the mRNA progresses through its life cycle.

DOI: 10.1103/PhysRevE.105.014410

I. INTRODUCTION

In recent years, the study of stochastic models of gene expression has received wide attention [1]. In the absence of transcriptional feedback, these models are often composed of first-order reactions, in which case exact expressions can be derived from the master equation for all the moments of the distribution of mRNA numbers. This is because for linear propensities, the moment equations are closed and can be solved straightforwardly [2].

However, the exact closed-form solution of the master equation for the probability distribution of mRNA numbers is a much harder problem. Popular methods such as the Poisson representation of Gardiner [3] and those devised more recently by Jahnke and Huisanga [4] are not applicable to models of gene expression composed of only first-order reactions because of the presence of a reaction of the type $G \to G + M$ which models transcription of mRNA (M) from the active state of a promoter (G). Hence, the general solution of these models is still an open research question. However, progress has been achieved on the solution of specific models.

A two-state model (commonly known as the telegraph model) whereby a gene switches between two states (an active and an inactive state) and mRNA is transcribed in the active state has been solved exactly in steady state [5] and in time [6] for the marginal distribution of mRNA numbers. Various other extensions of this model to include more biological realism have also been solved exactly or approximately for the marginal distribution of mRNA numbers. These include models that account for more than two gene states [7–9], polymerase dynamics [10], leaky expression from the inactive state [11], cell-to-cell variability (static and dynamic) [11,12], replication and binomial partitioning due to cell division [13], cell cycle duration variability [14], and modulation under environmental changes [15].

We here consider a different type of extension of the telegraph model that has recently received attention in three different biological contexts: (1) RNA polymerase (RNAP) movement along a gene during transcription [16,17], (2) multistep splicing [18], and (3) nuclear retention of mRNA [19]. In this stochastic model, a gene switches between two states (an active state G and an inactive one G^*), produces a transcript in the active state, and subsequently undergoes an arbitrary number L of sequential unimolecular steps, leading to L forms of the transcript (denoted by P_i, where $i \in [1, L]$). Each transcript either is removed after the L steps or else can also be degraded prematurely at an earlier step. A reaction scheme for this model is as follows:

$$\begin{aligned} &G \rightleftharpoons G^*, \quad G \to G + P_1, \\ &P_1 \to P_2 \to \cdots \to P_L \to \emptyset, \\ &P_1 \to \emptyset, \quad P_2 \to \emptyset, \ldots, P_L \to \emptyset. \end{aligned} \tag{1}$$

The L downstream processing steps can be interpreted in various ways. At the coarsest scale, the model with $L = 2$ steps can capture nuclear mRNA and cytoplasmic mRNA dynamics. At a less coarse scale, the model can capture the dynamics of the whole mRNA life cycle from birth (initiation) to death (in the cytoplasm). For example, one can associate steps $i = 1, \ldots, L_1 - 1$ with elongation of the nascent transcript, $i = L_1$ with termination, $i = L_1 + 1, \ldots, L_2 - 1$ with splicing, $i = L_2$ with export from the nucleus to the cytoplasm, and $i = L_2 + 1, \ldots, L$ with several reaction steps leading to mRNA degradation. For the last process, for example, the sequence of steps can model the fact that the polyadenylate tails of eukaryotic transcripts are sequentially chewed up before the protein-coding part of the message is degraded [20]. At a

*Juraj.Szavits.Nossan@ed.ac.uk
†Ramon.Grima@ed.ac.uk

fine scale, the model could capture the dynamics of nascent mRNA only, where the L steps represent the unidirectional movement of RNAP (with a nascent mRNA tail attached to it) along the gene. Hence, the reaction scheme (1) has myriad molecular interpretations according to the desired application.

In the past, several models have been proposed that couple gene promoter switching to downstream processing of RNA. The telegraph model coupled to deterministic nascent RNA elongation has been studied in Ref. [21]. For the model with stochastic RNA processing described by (1), the mean and variance of the number of molecules of P_i have been computed numerically or analytically [16,17]. Recently, an analytical expression for the probability distribution to find n_1 copies of P_1, n_2 copies of P_2, etc. has been found in a similar model in which transcription initiation is allowed from both on and off states [18]. The probability distribution was expressed in a power series in the difference between the transcription initiation rates from on and off states. We also note that a model similar to (1), but with excluded volume interactions restricting copy numbers of P_i to 0 or 1, has been recently studied in Refs. [22,23]. However, only numerical results have been provided.

In addition, several important results have been obtained in the limiting case in which the gene is always on and transcription initiation occurs in bursts. In that case the first two steps in (1)—the promoter switching ($G \rightleftharpoons G^*$) and transcription initiation ($G \to G + P_1$)—are combined into a single step, $G \to G + n \times P_1$, where n is usually taken from a geometric distribution. The case $L = 2$ describing nuclear (P_1) and cytoplasmic (P_2) RNA has been solved in full in the steady state [24], and a formal steady-state solution has also been found recently for arbitrary L [25]. In the $L = 2$ case, the mean and variance of the cytoplasmic RNA have also been found for an arbitrary distribution of the nuclear RNA retention time [26]. This model can be seen as a coarse-grained version of (1) in which multiple RNA processing steps are implicit in the probability distribution of the nuclear RNA retention time.

Despite these successes, an exact, steady-state, and closed-form solution for the marginal distribution of the number of molecules of P_i in the full model described by (1) is unknown. In this paper, we report an approximate solution to this open problem. Because the method we devise is nonperturbative and does not make assumptions on the strength of correlations between the forms of the product, we find by comparison with stochastic simulations that the approximation is highly accurate across all the parameter space.

The paper is divided as follows. In Sec. II we describe the model in detail, formulate its master equation, and discuss some known exactly solvable cases. In Sec. III we derive exact expressions for the moments of the distribution of P_i using a lattice path method. In Sec. IV we use the first and second moments and mean-field approximation to construct marginal distributions for the number of molecules of P_i. We show that while the use of the standard mean-field approximation is valid in only certain regions of parameter space, the use of an improved mean-field approximation leads to a uniformly accurate approximation over all parameter space. In particular, we show that the latter is practically indistinguishable from distributions calculated using the stochastic simulation algorithm. We conclude in Sec. V with a discussion of our method vs those already in the literature.

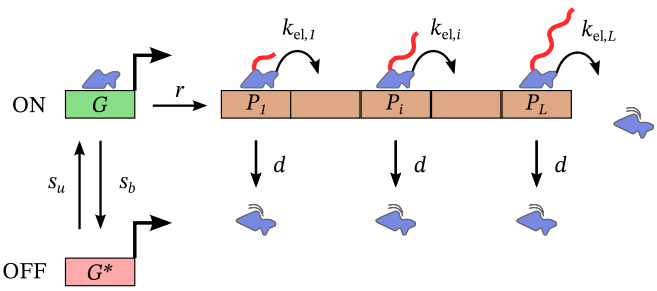

FIG. 1. Schematic of the stochastic transcription model. The gene body is divided into L segments. The promoter can be in two states: active (on) and inactive (off). Promoter on and off switching rates are s_u and s_b, respectively. Transcription initiation occurs when the promoter is in the on state at the rate r by which a new RNAP is added to the first segment. No transcription initiation is allowed when the promoter is in the off state. Following initiation, RNAPs move along the gene with segment-dependent elongation rates $k_{el,i}$. They can detach from DNA at any segment with rate d (premature termination). Once they reach the stop site, they terminate transcription at rate $k_{el,L}$. We note that the model ignores excluded volume interactions between neighboring RNAPs, i.e., the number of RNAPs in each segment is unbounded.

II. STOCHASTIC MODEL OF TRANSCRIPTION KINETICS

In this section we define the model and introduce the master equation that describes its stochastic evolution in time. While, as mentioned in the Introduction, the model has many interpretations, in what follows we will describe it in terms of RNAP dynamics during transcription.

A. Definition of the model

Transcription is modeled as a series of steps involving promoter activation, initiation at the transcription start site, RNAP movement along the gene, and termination at the stop site that frees the newly synthesized mRNA.

The promoter is modeled having two states, activated (on state G) and deactivated (off state G^*). The gene body—the part of the gene between the start and stop sites—is divided into L segments. The state of the system is described by the state of the promoter (on or off), and the number of RNA polymerases in each of the L segments, n_i for $i = 1, \ldots, L$.

The model is summarized in Fig. 1; note that this is a cartoon version of the reaction scheme (1), which is specific to the context of transcription. In particular, the RNAP on gene segment i is what was labeled as P_i in scheme (1). Note that the number of RNAPs on a gene is equal to the number of nascent mRNA; this is because each RNAP has attached to it a nascent mRNA tail whose length increases as elongation proceeds. The rates of gene activation and deactivation are s_u and s_b, respectively. Once the promoter is active (in the on state), transcription initiation occurs at rate r, whereby a new RNAP enters the first segment. We do not consider leaky transcription, i.e., transcription initiation is not allowed when the promoter is in the off state. After initiation, the

RNA polymerase moves along the gene body from segment to segment with, in general, a segment-dependent elongation rate $k_{\mathrm{el},i}$. During elongation, the RNAP can prematurely detach from the gene at rate d. Transcription termination occurs at the last segment with rate $k_{\mathrm{el},L}$, whereby the RNAP is removed from the gene and the newly synthesized mRNA is released.

We note that the model does not explicitly take into account excluded volume interactions between RNAPs. RNAP has a footprint of $\ell_{RNAP} = 35$ base pairs [23]. The size of each segment is $\ell_{\mathrm{segment}} = L_{\mathrm{gene}}/L$, where L_{gene} is the number of base pairs in the gene body, typically measured in thousands of base pairs. Thus, the maximum number of RNAPs in each segment is $c = \ell_{\mathrm{segment}}/\ell_{\mathrm{RNAP}}$, whereas in our model the number of RNAPs in each segment is unbounded.

As we show later, this simplification allows us to obtain analytical expressions for the distribution of RNAPs along the gene. In order to compensate for this simplification, one can choose the size of the segment ℓ_{segment} such that the average number of RNAPs in any segment is less than the maximum capacity c,

$$\langle n_i \rangle < c, \quad i = 1, \ldots, L. \tag{2}$$

Alternatively, one can choose ℓ_{segment} such that the probability of observing more than c RNAPs in each segment is small,

$$P(n_i > c) \ll 1, \quad i = 1, \ldots, L. \tag{3}$$

We provide analytical expression for both $\langle n_i \rangle$ and $P(n_i)$ so that either of these two conditions can be easily checked if the model parameters are known. Transcription is typically rate limited by initiation, which means there are only a few active RNAPs on the gene at a given time, and hence conditions (2) and (3) are automatically satisfied.

B. Master equations for the RNAP distribution

The central information that we are interested in is the probability $P(n_1, \ldots, n_L; t)$ to find n_1 RNAPs in segment 1, n_2 RNAPs in segment 2, and so on, at a given time t. This probability is in fact a sum of two probabilities $P_{\mathrm{on}}(n_1, \ldots, n_L; t)$ and $P_{\mathrm{off}}(n_1, \ldots, n_L; t)$, which are conditioned on the state of the promoter. The probabilities P_{on} and P_{off} satisfy the following master equations:

$$\frac{\partial}{\partial t} P_{\mathrm{on}}(n_1, \ldots, n_L; t) = \sum_{i=1}^{L-1} k_{\mathrm{el},i}(n_i + 1)\mathbb{E}_i \mathbb{E}_{i+1}^{-1} P_{\mathrm{on}} + k_{\mathrm{el},L}(n_L + 1)\mathbb{E}_L P_{\mathrm{on}} + \sum_{i=1}^{L} d(n_i + 1)\mathbb{E}_i P_{\mathrm{on}}$$
$$+ r\mathbb{E}_1^{-1} P_{\mathrm{on}} + s_u P_{\mathrm{off}} - \sum_{i=1}^{L} (k_{\mathrm{el},i} + d) n_i P_{\mathrm{on}} - (s_b + r) P_{\mathrm{on}}, \tag{4a}$$

$$\frac{\partial}{\partial t} P_{\mathrm{off}}(n_1, \ldots, n_L; t) = \sum_{i=1}^{L-1} k_{\mathrm{el},i}(n_i + 1)\mathbb{E}_i \mathbb{E}_{i+1}^{-1} P_{\mathrm{off}} + k_{\mathrm{el},L}(n_L + 1)\mathbb{E}_L P_{\mathrm{off}}$$
$$+ \sum_{i=1}^{L} d(n_i + 1)\mathbb{E}_i P_{\mathrm{off}} + s_b P_{\mathrm{on}} - \sum_{i=1}^{L} (k_{\mathrm{el},i} + d) n_i P_{\mathrm{off}} - s_u P_{\mathrm{off}}. \tag{4b}$$

Here we have shortened the notation by introducing step operators $\mathbb{E}_i$ and $\mathbb{E}_i^{-1}$ [27] that increase and decrease the number of particles n_i in segment i, respectively,

$$\mathbb{E}_i P(\ldots, n_i, \ldots) = P(\ldots, n_i + 1, \ldots), \tag{5a}$$
$$\mathbb{E}_i^{-1} P(\ldots, n_i, \ldots) = P(\ldots, n_i - 1, \ldots), \quad n_i \geqslant 1, \tag{5b}$$
$$\mathbb{E}_i^{-1} P(\ldots, n_i = 0, \ldots) = 0. \tag{5c}$$

Hereafter, we are interested only in the steady state, so that $\partial P_{\mathrm{on}}/\partial t = 0$ and $\partial P_{\mathrm{off}}/\partial t = 0$, and so we drop the time dependence from P_{on} and P_{off}.

Rather than working with the master equation directly, we introduce the probability-generating functions G_{on}, G_{off} for each of the promoter states and $G = G_{\mathrm{on}} + G_{\mathrm{off}}$, whereby

$$G_{\mathrm{on}}(z_1, \ldots, z_L) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_L=0}^{\infty} z_1^{n_1} \ldots z_L^{n_L} \times P_{\mathrm{on}}(n_1, \ldots, n_L), \tag{6a}$$

$$G_{\mathrm{off}}(z_1, \ldots, z_L) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_L=0}^{\infty} z_1^{n_1} \ldots z_L^{n_L} \times P_{\mathrm{off}}(n_1, \ldots, n_L). \tag{6b}$$

From the master equation we get the following steady-state partial differential equations for G_{on} and G_{off}:

$$\sum_{i=1}^{L-1} [(k_{\mathrm{el},i} + d) z_i - k_{\mathrm{el},i} z_{i+1} - d] \frac{\partial G_{\mathrm{on}}}{\partial z_i} - (k_{\mathrm{el},L} + d)(1 - z_L) \frac{\partial G_{\mathrm{on}}}{\partial z_L} = s_u G_{\mathrm{off}} - s_b G_{\mathrm{on}} - r(1 - z_1) G_{\mathrm{on}}, \tag{7a}$$

$$\sum_{i=1}^{L-1} [(k_{\mathrm{el},i} + d) z_i - k_{\mathrm{el},i} z_{i+1} - d] \frac{\partial G_{\mathrm{off}}}{\partial z_i} - (k_{\mathrm{el},L} + d)(1 - z_L) \frac{\partial G_{\mathrm{off}}}{\partial z_L} = s_b G_{\mathrm{on}} - s_u G_{\mathrm{off}}. \tag{7b}$$

To the best of our knowledge, this system of equations has not been solved in closed-form before, except in three special cases: the telegraph model ($L = 1$), constitutive gene expression ($s_b = 0$), and deterministic elongation with mature RNA production. For completeness, we summarize these three cases below.

C. Exactly solvable cases

1. The telegraph model

The case of $L = 1$ is known as the telegraph model [5], whose solution for the generating function reads

$$G(z_1) = e^{-\alpha(1-z_1)} M(\sigma_b; \sigma_u + \sigma_b; \alpha(1 - z_1)), \tag{8}$$

where $k_{\text{el},1} = k_{\text{el}}$ and

$$\alpha = \frac{r}{k_{\text{el}} + d}, \quad \sigma_u = \frac{s_u}{k_{\text{el}} + d}, \quad \sigma_b = \frac{s_b}{k_{\text{el}} + d}. \tag{9}$$

In the expression for $G(z_1)$, M is Kummer's (confluent hypergeometric) function [28]. The probability distribution of n_1 is given by

$$P(n_1) = \frac{\alpha^{n_1} e^{-\alpha}}{n_1!} \frac{(\sigma_u)_{n_1}}{(\sigma_u + \sigma_b)_{n_1}} M(\sigma_b, \sigma_u + \sigma_b + n_1, \alpha). \tag{10}$$

We note this expression also holds for the marginal distribution $P(n_1)$ of the model with $L \geqslant 1$; this is because the fluctuations in the RNAP numbers on segment 1 are unaffected by the fluctuations on the other segments due to the irreversible nature of the reactions between segments.

2. Constitutive gene expression

A constitutive promoter is always in the on state, which means that $s_b = 0$. In that case there is only one equation to solve, that for $G_{\text{on}} \equiv G$, since $G_{\text{off}} = 0$. We can easily check that the solution is given by

$$G(z_1, \ldots, z_L) = \prod_{i=1}^{L} e^{-\mu_i(1-z_i)}, \tag{11}$$

where μ_i is the average number of RNAPs in segment i and is given by Eq. (23) in which $\eta = s_u/(s_u + s_b) = 1$. This generating function leads to a probability distribution that is a product of Poisson distributions,

$$P(n_1, \ldots, n_L) = \prod_{j=1}^{L} \frac{\mu_j^{n_j} e^{-\mu_j}}{n_j!}. \tag{12}$$

Consequently there are no correlations between segments. Models of this type have been used to describe mRNA senescence [29] and splicing [30].

3. Deterministic elongation with mature RNA production

We can extend this model to include mature RNA production by another stage after transcription termination. The model now has $L + 1$ segments, whereby the first L segments describe RNAP dynamics and the last segment counts the number of mature RNA m. We further assume uniform elongation rates in segments $i = 1, \ldots, L$ so that $k_{el,i} = k_{\text{el}}$. At the last segment, the mature RNA is degraded at rate d_m, which in our notation is equivalent to saying that $d_m \equiv k_{\text{el},L+1} + d$.

For this model the distribution of mature RNA, $P(m)$, has been found analytically [17] in the limit in which the elongation is deterministic, i.e., when $d_m/k_{\text{el}} \to 0$ and $L \to \infty$ but keeping the transcription time $T = L/k_{\text{el}}$ fixed. In this limit, the distribution of the mature RNA is the same as in the telegraph model; hence, this limit is a more formal way of deriving the telegraph model from a detailed model of RNAP dynamics.

III. EXACT MOMENTS OF THE RNAP DISTRIBUTION $P(n_i)$

In this section we consider the first three moments of the RNAP distribution $P(n_i)$ in the steady state. We will need these moments later in Sec. IV, where we derive an analytical expression for the RNAP distribution using mean-field theory.

We begin by deriving recurrence relations for the first three moments, which we solve analytically in the uniform case (all $k_{\text{el},i}$ are equal to k_{el}). In the nonuniform case (arbitrary values of $k_{\text{el},i}$ in each segment) we obtain the first moments analytically and the second and third moments numerically.

In the former case, the first two moments have been previously derived in Ref. [17]. Here for the uniform case we present an alternative derivation using lattice paths which has the advantage that it can be extended to higher moments, as we demonstrate by computing the third moment.

We first make the following change of variables:

$$u_i = \lambda_i(1 - z_i), \tag{13}$$

where λ_i is given by

$$\lambda_1 = 1, \quad \lambda_i = \prod_{j=1}^{i-1} \frac{k_{\text{el},j}}{k_{\text{el},j} + d}, \quad i = 2, \ldots, L. \tag{14}$$

The equations for G_{on} and G_{off} now take a simpler form,

$$\sum_{i=1}^{L-1} \omega_i(u_i - u_{i+1}) \frac{\partial G_{\text{on}}}{\partial u_i} + \omega_L u_L \frac{\partial G_{\text{on}}}{\partial u_L} = s_u G_{\text{off}} - (s_b + r u_1) G_{\text{on}}, \tag{15a}$$

$$\sum_{i=1}^{L-1} \omega_i(u_i - u_{i+1}) \frac{\partial G_{\text{off}}}{\partial u_i} + \omega_L u_L \frac{\partial G_{\text{off}}}{\partial u_L} = s_b G_{\text{on}} - s_u G_{\text{off}}, \tag{15b}$$

where we have introduced ω_i defined as

$$\omega_i = k_{\text{el},i} + d. \tag{16}$$

The recurrence relations for the moments of $P(n_i)$ can be found by taking partial derivatives of the equations above and setting $u_1 = \cdots = u_L = 0$.

A. First moment of $P(n_i)$

We introduce the following notation:

$$g^{(i)} = g^{(i)}_{\text{on}} + g^{(i)}_{\text{off}}, \tag{17}$$

where

$$g^{(i)}_{\text{on}} = \left.\frac{\partial G_{\text{on}}}{\partial u_i}\right|_{\{u_j\}=0}, \quad g^{(i)}_{\text{off}} = \left.\frac{\partial G_{\text{off}}}{\partial u_i}\right|_{\{u_j\}=0}, \tag{18}$$

and $\{u_j\} = 0$ means that $u_1 = \cdots = u_L = 0$. Using $g^{(i)}$, we can express the first moment of $P(n_i)$ as

$$\mu_i = \langle n_i \rangle = -\lambda_i g^{(i)}. \tag{19}$$

In order to find $g^{(i)}$ we add Eqs. (15a) and (15b) together, then take a partial derivative with respect to u_i, and set all

$u_1 = \cdots = u_L = 0$. The resulting recurrence relation for $g^{(i)}$ reads

$$\omega_i g^{(i)} = \omega_{i-1} g^{(i-1)}, \quad \omega_1 g^{(1)} = -r\eta, \tag{20}$$

where η is the fraction of the time that the promoter spends in the on state,

$$\eta = \frac{s_u}{s_u + s_b}. \tag{21}$$

From this we get

$$g^{(i)} = -\frac{r\eta}{\omega_i}. \tag{22}$$

The first moment of P in segment i is thus given by

$$\mu_i = \langle n_i \rangle = \frac{r\eta\lambda_i}{\omega_i} = \frac{r\eta\lambda_{i+1}}{k_{\mathrm{el},i}} = \frac{r\eta}{k_{\mathrm{el},i}} \prod_{j=1}^{i} \frac{k_{\mathrm{el},j}}{k_{\mathrm{el},j} + d}. \tag{23}$$

For uniform rates with no detachment, the mean is the same across the gene. Otherwise, the mean varies across the gene, for example, if the elongation rates are nonuniform with no detachment or if the elongation rates are uniform with nonzero detachment.

We also write an analytical expression for $g^{(i)}_{\mathrm{on}}$ because we will need it later for computing the second moment of $P(n_i)$,

$$g^{(i)}_{\mathrm{on}} = -\frac{r\eta}{\omega_i} \sum_{n=1}^{i} \left(\delta_{n,1} + \frac{s_u}{\omega_n} \right) \times \prod_{k=n}^{i} \frac{\omega_k}{s_u + s_b + \omega_k}, \tag{24}$$

where $\delta_{i,j}$ is the Kronecker delta function. In the case of uniform elongation rates, the expression for $g^{(i)}_{\mathrm{on}}$ takes a simpler form:

$$g^{(i)}_{\mathrm{on}} = -\alpha\eta \left[\eta + \frac{1-\eta}{(1 + \sigma_u + \sigma_b)^i} \right], \tag{25}$$

where the rescaled parameters α, σ_u, and σ_b are defined in Eq. (9).

B. Second moment of $P(n_i)$

Next, we define

$$g^{(ij)} = g^{(ij)}_{\mathrm{on}} + g^{(ij)}_{\mathrm{off}}, \tag{26}$$

where

$$g^{(ij)}_{\mathrm{on}} = \left. \frac{\partial^2 G_{\mathrm{on}}}{\partial u_i \partial u_j} \right|_{\{u_j\}=0}, \quad g^{(ij)}_{\mathrm{off}} = \left. \frac{\partial^2 G_{\mathrm{off}}}{\partial u_i \partial u_j} \right|_{\{u_j\}=0}. \tag{27}$$

The second central moment of $P(n_i)$ (the covariance) is related to $g^{(ij)}$ as follows:

$$\begin{aligned} \mathrm{cov}(n_i, n_j) &= \langle (n_i - \mu_i)(n_j - \mu_j) \rangle \\ &= \delta_{ij}\mu_i + \lambda_i \lambda_j (g^{(ij)} - g^{(i)} g^{(j)}). \end{aligned} \tag{28}$$

In order to find $g^{(ij)}$, we add Eqs. (15a) and (15b) together, then take partial derivatives with respect to u_i and u_j, and set $u_1 = \cdots = u_L = 0$, yielding the following recurrence relations for $g^{(ij)}$:

$$g^{(ij)} = \frac{\omega_{i-1}}{\omega_i + \omega_j} g^{(i-1\,j)} + \frac{\omega_{j-1}}{\omega_i + \omega_j} g^{(ij-1)}, \quad i, j = 2, \ldots, L, \tag{29a}$$

$$g^{(1j)} = \frac{\omega_{j-1}}{\omega_1 + \omega_j} g^{(1j-1)} - \frac{r g^{(j)}_{\mathrm{on}}}{\omega_1 + \omega_j}, \quad j = 2, \ldots, L, \tag{29b}$$

$$g^{(i1)} = \frac{\omega_{i-1}}{\omega_i + \omega_1} g^{(i-11)} - \frac{r g^{(i)}_{\mathrm{on}}}{\omega_i + \omega_1}, \quad i = 2, \ldots, L, \tag{29c}$$

$$g^{(11)} = -\frac{r}{\omega_1} g^{(1)}_{\mathrm{on}}. \tag{29d}$$

A closed-form expression for $g^{(ij)}$ can be found for uniform elongation rates, in which case the coefficients in Eqs. (29) are all equal to $1/2$. We can think of Eqs. (29) in terms of two-dimensional lattice paths with unit steps $(m, n) \to (m-1, n)$ and $(m, n) \to (m, n-1)$. We are interested in the paths that start at (i, j) and end at one of the boundary points, $(1, n)$ or $(m, 1)$, for $1 \leqslant m \leqslant i$ and $1 \leqslant n \leqslant j$. Let us say we start at (i, j) and end at $(m, 1)$ for $1 \leqslant m \leqslant i$. Each step we make is weighted by factor $1/2$ until we reach the end point at $(m, 1)$, which is weighted by $(-\alpha g^{(m)}_{\mathrm{on}}/2)$, where $\alpha = r/(k_{\mathrm{el}} + d)$. Similarly, the end point at $(1, n)$ for $1 \leqslant n \leqslant j$ is weighted by $(-\alpha g^{(n)}_{\mathrm{on}})/2$, whereas the end point at (1,1) is weighted by $-\alpha g^{(1)}_{\mathrm{on}}$.

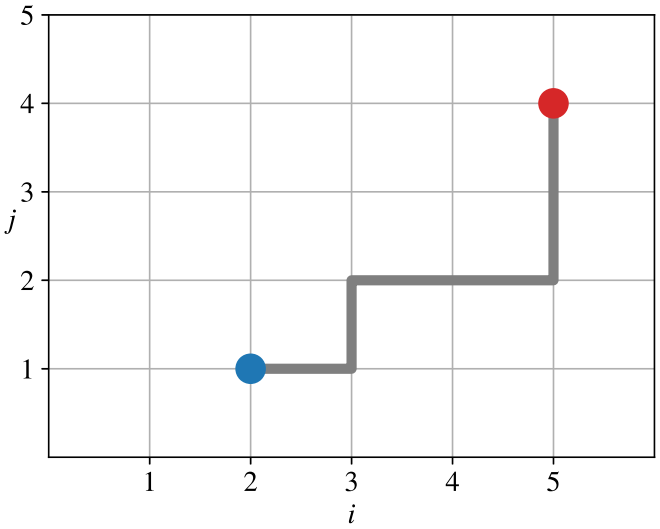

FIG. 2. A lattice path from point $(i, j) = (5, 4)$ (red) to point (2,1) (blue). The contribution from this path to $g^{(54)}$ is $(1/2)^6(-\alpha g^{(2)}_{\mathrm{on}}/2)$.

One such lattice path is presented in Fig. 2 with the start point at (5,4) and the end point at (2,1). The total weight of this path is $1/2^6$, and the length of the path is 6, multiplied by the weight at the end point, $(-\alpha g^{(2)}_{\mathrm{on}}/2)$, which gives

$$\text{path weight in Fig. 2} = \left(\frac{1}{2}\right)^6 \left(-\frac{\alpha g^{(2)}_{\mathrm{on}}}{2}\right). \tag{30}$$

Because all paths between two given points on the lattice carry the same weight, we can simply count them, multiply their weight by the weight at the boundary, and then sum over

all boundary points. This simplification is not possible in the case of nonuniform elongation rates because paths between two given points on the lattice will in general all have different weights.

In order to count the paths between two fixed points, we denote by S moving south and by W moving west. We can then describe a path of length n as an n-letter word consisting of the letters S and W. For example, the path in Fig. 2 can be written as SSWWSWW.

Next, we count all paths from a given point (i, j) to the boundary point $(1, n)$ for each $n = 2, \ldots, j$. To get to this point, we need to make $i - 1$ steps to the south and $j - n$ steps to the west. The total length of this path is $i - 1 + j - n$, and there are $\binom{i-1+j-n}{i-1}$ such paths. We repeat this procedure for all paths from (i, j) to the boundary points $(m, 1)$ for each $m = 2, \ldots, i$. Finally, we count all paths from (i, j) to (1,1). After rearranging terms, the total contribution from all paths can be written as

$$g^{(ij)} = \sum_{m=1}^{i} \frac{1}{2^{i-m+j-1}} \binom{i-m+j-1}{j-1} \left(-\frac{\alpha g_{\text{on}}^{(m)}}{2}\right) + \sum_{n=1}^{j} \frac{1}{2^{i-1+j-n}} \binom{i-1+j-n}{i-1} \left(-\frac{\alpha g_{\text{on}}^{(n)}}{2}\right). \quad (31)$$

Inserting the expression for $g_{\text{on}}^{(m)}$ from Eq. (25) into (31) we get a closed-form expression for the second moment of $P(n_i)$ (the covariance),

$$\text{cov}(n_i, n_j) = \delta_{i,j}\mu_i + \mu_i\mu_j \frac{\sigma_b/\sigma_u}{1 + \sigma_u + \sigma_b} \gamma_{ij}(\sigma_u, \sigma_b), \quad (32)$$

where $\gamma_{i,j}(\sigma_u, \sigma_b)$ [31] is given by

$$\gamma_{ij}(\sigma_u, \sigma_b) = \sum_{q=0}^{i-1} \binom{j-1+q}{q} \frac{1}{2^{j+q}(1 + \sigma_u + \sigma_b)^{i-q-1}} + \sum_{q=0}^{j-1} \binom{i-1+q}{q} \frac{1}{2^{i+q}(1 + \sigma_u + \sigma_b)^{j-q-1}}. \quad (33)$$

The derivation of the second moments conditioned on the promoter being in the on state can be done is a similar fashion. We omit here the details and state only the final result:

$$g_{\text{on}}^{(ij)} = \frac{1}{2 + \sigma_u + \sigma_b} \sum_{m=1}^{i} \sum_{n=1}^{j} \binom{i-m+j-n}{i-m} \times \frac{(\sigma_u g^{(mn)} - \delta_{m1}\alpha g_{\text{on}}^{(n)} - \delta_{1n}\alpha g_{\text{on}}^{(m)})}{(2 + \sigma_u + \sigma_b)^{i-m+j-n}}. \quad (34)$$

It is interesting that while RNAPs in this model are noninteracting (dynamics of a single RNAP does not depend on the presence of other RNAPs), yet the model exhibits long-range correlations in RNAP numbers between segments. These correlations are entirely due to promoter switching, because the model with constitutive gene expression ($s_b = 0$) has covariance $\text{cov}(n_i, n_j) = 0$ for $i \neq j$; see Sec. II C.

In the case of nonuniform elongation rates, the lattice path method cannot be used because paths that share the same end points in general have different weights. Instead, the recurrence relation (29) for $g^{(ij)}$ must be solved directly. leading to rather cumbersome expressions.

First, we solve the recurrence relation for $g^{(1j)}$, which reads

$$g^{(1j)} = \frac{1}{\omega_1\omega_j} \sum_{m=1}^{j} \frac{\omega_1 + \delta_{1m}\omega_m}{\omega_1 + \omega_m} \frac{Q_{1j}}{Q_{1m}} \left(-r\omega_m g_{\text{on}}^{(m)}\right), \quad (35)$$

where δ_{ij} is the Kronecker delta, $g_{\text{on}}^{(i)}$ is given by Eq. (24), and Q_{ij} is defined as

$$Q_{ij} = \prod_{k=1}^{j} \frac{\omega_k}{\omega_i + \omega_k}. \quad (36)$$

Next, we can express $g^{(ij)}$ in terms of $g^{(i-1m)}$ for $m = 1, \ldots, j$,

$$g^{(ij)} = \frac{1}{\omega_i\omega_j} \sum_{m=1}^{j} \frac{\omega_i + \delta_{1m}\omega_m}{\omega_i + \omega_m} \frac{Q_{ij}}{Q_{im}} (\omega_{i-1}\omega_m g^{(i-1m)}). \quad (37)$$

The final expression for $g^{(ij)}$ can be obtained by iterating (37) until $g^{(ij)}$ is expressed in terms of $g^{(1k)}$ for $k = 1, \ldots, j$ and then inserting expression (35) for $g^{(1k)}$.

C. Third moment of $P(n_i)$

In order to compute the third moment of $P(n_i)$, we consider third-order partial derivatives of G_{on} and of G_{off},

$$g^{(ijk)} = g_{\text{on}}^{(ijk)} + g_{\text{off}}^{(ijk)}, \quad (38)$$

where

$$g_{\text{on}}^{(ijk)} = \left.\frac{\partial^3 G_{\text{on}}}{\partial u_i \partial u_j \partial u_k}\right|_{\{u_j\}=0}, \quad (39a)$$

$$g_{\text{off}}^{(ijk)} = \left.\frac{\partial^3 G_{\text{off}}}{\partial u_i \partial u_j \partial u_k}\right|_{\{u_j\}=0}. \quad (39b)$$

Of particular importance is the third standardized central moment of $P(n_i)$, which is given by

$$\kappa_i = \frac{\langle (n_i - \langle n_i \rangle)^3 \rangle}{\sigma_i^3} = \frac{1}{\sigma_i^3}\{-\lambda_i^3(g^{(iii)} - 3g^{(ii)}g^{(i)} + 2g^{(i)}) + 3\lambda_i^2[g^{(ii)} - (g^{(i)})^2] - \lambda_i g^{(i)}\}, \quad (40)$$

where $g^{(i)}$, $g^{(ij)}$, and $g^{(ijk)}$ are given by Eqs. (22), (31), and (43), respectively, and σ_i is the standard deviation of n_i,

$$\sigma_i = \sqrt{\langle (n_i - \langle n_i \rangle)^2 \rangle} = \sqrt{\text{cov}(n_i, n_i)}. \quad (41)$$

The third standardized central moment is useful for understanding the shape of the distribution, in particular its (a)symmetry. We will use this expression later in Sec. IV.

The recipe to obtain the recurrence relation for $g^{(ijk)}$ is the same as before. We first add Eqs. (15a) and (15b) together, then take the partial derivatives with respect to u_i, u_j, and u_k, and finally set $u_1 = \cdots = u_L = 0$. The resulting recurrence

equations for $g^{(ijk)}$ are given by

$$g^{(ijk)} = \frac{\omega_{i-1}}{\omega_i+\omega_j+\omega_k} g^{i-1jk} + \frac{\omega_{j-1}}{\omega_i+\omega_j+\omega_k} g^{ij-1k} + \frac{\omega_{k-1}}{\omega_i+\omega_j+\omega_k} g^{ijk-1}, \quad i,j,k = 2,\ldots,L, \tag{42a}$$

$$g^{(1jk)} = \frac{\omega_{j-1}}{\omega_1+\omega_j+\omega_k} g^{1j-1k} + \frac{\omega_{k-1}}{\omega_1+\omega_j+\omega_k} g^{1jk-1} - \frac{r g_{\rm on}^{(jk)}}{\omega_1+\omega_j+\omega_k} \quad j,k = 2,\ldots,L, \tag{42b}$$

$$g^{(11k)} = \frac{\omega_{k-1}}{2\omega_1+\omega_k} - \frac{2r g_{on}^{(1k)}}{2\omega_1+\omega_k}, \quad k = 2,\ldots,L, \tag{42c}$$

$$g^{(111)} = -r g_{\rm on}^{(11)}. \tag{42d}$$

Other equations, for example, for $g^{(i1k)}$, can be obtained from these equations using the fact $g^{(ijk)}$ is invariant under the permutation of indices i, j, k.

Equations (42a) and (42b) can be solved analytically in the case of uniform elongation rates using lattice paths, or numerically in the nonuniform case. In the former case, we consider three-dimensional lattice paths that start at (i, j, k) and end at one of the boundary points $(1, n, o)$, $(m, 1, o)$, or $(m, n, 1)$, where $1 \leqslant m \leqslant i$, $1 \leqslant n \leqslant j$, and $1 \leqslant o \leqslant k$. Instead of $1/2$, each step is now weighted by $1/3$. If the end point is, say, $(1, n, o)$, where $n, o \neq 1$, then the weight of that point is $-\alpha g_{\rm on}^{(no)}/3$, where $g_{\rm on}^{(no)}$ is the second moment conditioned on the promoter being in the on state. If the end point is $(1, 1, o)$ for $o \neq 1$, then the weight is $-2\alpha g_{\rm on}^{(1o)}/3$. If $o = 1$ then the weight is $-\alpha g_{\rm on}^{(no)}$. The final result for $g^{(ijk)}$ is

$$\begin{aligned} g^{(ijk)} = &\sum_{m=1}^{i}\sum_{n=1}^{j} \binom{i-m+j-n+k-1}{i-m, j-n, k-1} \times \frac{1}{3^{i-m+j-n+k-1}} \left(-\frac{\alpha g_{\rm on}^{(mn)}}{3}\right) \\ &+ \sum_{m=1}^{i}\sum_{o=1}^{k} \binom{i-m+j-1+k-o}{i-m, j-1, k-o} \times \frac{1}{3^{i-m+j-1+k-o}} \left(-\frac{\alpha g_{\rm on}^{(mo)}}{3}\right) \\ &+ \sum_{n=1}^{j}\sum_{o=1}^{k} \binom{i-1+j-n+k-o}{i-1, j-n, k-o} \times \frac{1}{3^{i-1+j-n+k-o}} \left(-\frac{\alpha g_{\rm on}^{(no)}}{3}\right), \end{aligned} \tag{43}$$

where $g_{\rm on}^{(ij)}$ is given by Eq. (34).

IV. THE RNAP DISTRIBUTION $P(n_i)$ IN THE MEAN-FIELD APPROXIMATION

As we mentioned earlier, solving the system of partial differential equations for $G_{\rm on}$ and $G_{\rm off}$ for arbitrary number of segments L is a difficult problem. In order to make progress, we focus on the marginal distribution of n_i in the steady state defined as

$$P(n_i) = \sum_{\{n_i\}\setminus n_i} P(n_1,\ldots,n_L), \tag{44}$$

where the summation goes over all values of variables $n_1, \ldots, n_L$ except for the variable n_i which is kept fixed. In this section we provide an analytical expression for $P(n_i)$ in the mean-field approximation.

A. Naive mean-field approximation

We first consider an approximation of the mean-field type that we call the naive mean-field (NMF) approximation for reasons that will be clear in a moment. This type of approximation is well known in statistical physics and often gives satisfactory results in the absence of long-range correlations (for example, away from phase transitions). However, we will show that for this model the NMF approximation does not always work, and later we find an improved mean-field approximation using exact results from Sec. III, which shows excellent agreement with the exact results from stochastic simulations. Since both of these approximations are of the mean-field type, we use terms "naive" and "improved" to distinguish between them.

The main idea of the NMF approximation is to find closed master equations for the marginal distributions $P_{\rm on}(n_i)$ and $P_{\rm off}(n_i)$, which is done by decoupling fluctuations in segments $i-1$ and i. To find the master equation for $P_{\rm on}(n_i)$ and $P_{\rm off}(n_i)$, we sum Eqs. (4a) and (4b) over all n_j for all $j \neq i$. In each of the resulting master equations, there will be exactly one term that couples segments i and $i-1$, and that is the term that describes elongation from $i-1$ to i. For example, in the master equation for $P_{\rm on}$, this term is given by

$$\begin{aligned} &\sum_{n_{i-1}=0}^{\infty} k_{{\rm el},i-1}(n_{i-1}+1) P_{\rm on}(n_{i-1}+1, n_i-1) \\ &= \left(k_{{\rm el},i-1} \sum_{n_{i-1}=0}^{\infty} n_{i-1} P_{\rm on}(n_{i-1}|n_i-1)\right) \times P_{\rm on}(n_i-1). \end{aligned} \tag{45}$$

Here $P_{\rm on}(n_{i-1}+1, n_i-1)$ describes the coupling between segments $i-1$ and i. This type of hierarchy in which a n-segment probability distribution depends on a $(n+1)$-segment probability distribution is typical of an interacting many-body system and cannot be solved. However, we can rewrite this term using the conditional distribution $P_{\rm on}(n_{i-1}|n_i-1)$, so that the expression in the parentheses is an effective rate at which new RNAPs are added to segment i from segment $i-1$. The NMF approximation amounts to ignoring the dependence of $P(n_{i-1}|n_i-1)$ on n_i, i.e., replacing

$$P_{\rm on}(n_{i-1}|n_i-1) \approx \frac{P_{\rm on}(n_{i-1})}{\sum_{n_{i-1}} P_{\rm on}(n_{i-1})}, \tag{46a}$$

$$P_{\rm off}(n_{i-1}|n_i-1) \approx \frac{P_{\rm off}(n_{i-1})}{\sum_{n_{i-1}} P_{\rm off}(n_{i-1})}. \tag{46b}$$

In other words we ignore correlations between segments $i-1$ and i, except for the fact that both segments have the same promoter state. By ignoring these correlations, we obtain

master equations that contain only terms with $P_{\rm on}(n_i)$ and $P_{\rm off}(n_i)$. The effective rates p_i and q_i at which new RNAPs are added to the segment i in the on and off state are thus given by

$$p_i = \frac{k_{{\rm el},i-1}\sum_{n_{i-1}} n_{i-1}P_{\rm on}(n_{i-1})}{\sum_{n_{i-1}} P_{\rm on}(n_{i-1})} = \frac{-k_{{\rm el},i-1}\ \lambda_{i-1}g_{\rm on}^{(i-1)}}{\eta}, \quad i=2,\ldots,L, \tag{47a}$$

$$q_i = \frac{k_{{\rm el},i-1}\sum_{n_{i-1}} n_{i-1}P_{\rm off}(n_{i-1})}{\sum_{n_{i-1}} P_{\rm off}(n_{i-1})} = \frac{-k_{{\rm el},i-1}\ g_{\rm off}^{(i-1)}}{1-\eta}, \quad i=2,\ldots,L. \tag{47b}$$

The effective model that governs the stochastic dynamics of n_i in the NMF approximation is equivalent to the telegraph model for RNA production [5] in which transcription rates in the on and off states are p_i and q_i, respectively, and the mRNA degradation rate is $k_{{\rm el},i}+d$. The reactions for this model with their corresponding rates are given by

$$G \underset{s_u}{\overset{s_b}{\rightleftharpoons}} G^*, \tag{48a}$$

$$G \xrightarrow{p_i} G + P_i, \tag{48b}$$

$$G^* \xrightarrow{q_i} G^* + P_i, \tag{48c}$$

$$P_i \xrightarrow{k_{{\rm el},i}+d} \emptyset, \tag{48d}$$

where G and G^* represent the on and off states of the gene and P_i is the species associated with the ith segment [see scheme (1)]. We note that the telegraph model in which transcription is allowed from both on and off states is also known as the leaky telegraph model [11].

The telegraph model described by reactions (48a)–(48d) can be solved exactly, and the probability-generating function reads

$$G_0(z_i) = \sum_{n_i=0}^{\infty} z_i^{n_i}P_0(n_i) = e^{-\alpha_i(1-z_i)} \times M(\sigma_b;\sigma_u+\sigma_b;(\alpha_i-\beta_i)(1-z_i)), \tag{49}$$

where α_i, β_i, $\sigma_{u,i}$, and $\sigma_{b,i}$ are defined as

$$\alpha_i = \frac{p_i}{k_{el,i}+d}, \quad \beta_i = \frac{q_i}{k_{el,i}+d}, \quad \sigma_{u,i} = \frac{s_u}{k_{el,i}+d}, \quad \sigma_{b,i} = \frac{s_b}{k_{el,i}+d}. \tag{50}$$

The probability distribution $P(n_i)$ in the NMF approximation reads

$$P(n_i) = \frac{e^{-\alpha_i}}{n_i!}\sum_{k=0}^{n_i}\binom{n_i}{k}\beta_i^{n_i-k}(\alpha_i-\beta_i)^k \times \frac{(\sigma_{u,i})_k}{(\sigma_{u,i}+\sigma_{b,i})_k}M(\sigma_{b,i};\sigma_{u,i}+\sigma_{b,i}+k;(\alpha_i-\beta_i)), \tag{51}$$

and the mean and the variance are given by, respectively,

$$\mu_{0,i} = \alpha_i\eta + \beta_i(1-\eta), \tag{52}$$

$$\sigma_{0,i}^2 = \mu_{0,i} + (\alpha_i-\beta_i)^2\frac{\eta(1-\eta)}{1+\sigma_{u,i}+\sigma_{b,i}}. \tag{53}$$

We can check that the choice of p_i and q_i in Eqs. (47a) and (47b) ensures that $\mu_{0,i}=\mu_i$, i.e., the mean of n_i computed in Eq. (23) is the same as the mean in the NMF approximation. However, the value of the variance is generally different from the exact value.

In the case of uniform elongation rate the expression for the variance simplifies to

$$\sigma_{0,i}^2 = \mu_i + \mu_i^2\frac{\sigma_b/\sigma_u}{1+\sigma_u+\sigma_b}\gamma_{0,ii}(\sigma_u,\sigma_b), \tag{54}$$

where $\gamma_{0,ii}(\sigma_u,\sigma_b)$ is given by

$$\gamma_{0,ii}(\sigma_u,\sigma_b) = \frac{1}{(1+\sigma_u+\sigma_b)^{2i-2}}. \tag{55}$$

The difference between the exact variance in Eq. (32) and Eq. (54) is in the factor γ_{ii} for $i\geqslant 2$ (the NMF approximation is exact in segment 1). In general, we find that $\gamma_{0,ii}<\gamma_{ii}$ for any $i=2,\ldots,L$ for $\sigma_u\neq 0$ and $\sigma_b\neq 0$, whereas $\gamma_{0,ii}=\gamma_{ii}=1$ for $\sigma_u=\sigma_b=0$. The NMF approximation underestimates the exact variance, which is expected given that we ignored correlations. When $\sigma_b=0$ (constitutive gene expression), there are no correlations between segments, the NMF approximation becomes exact and the resulting distribution is Poisson; see Eq. (12).

In general, the NMF approximation is valid if the correlations between n_{i-1} and n_i are small. Fortunately, for this model we can compute these correlations exactly from the covariance $\mathrm{cov}(n_{i-1},n_i)$ in Eq. (28). In particular, we can compute the correlation coefficient defined as

$$R_i = \frac{\mathrm{cov}(n_{i-1},n_i)}{\sigma_{i-1}\sigma_i}, \tag{56}$$

where σ_i is the standard deviation of n_i; see Eq. (41). The naive mean field is valid provided $R_i\approx 0$, i.e., when n_{i-1} and n_i are not correlated. Computing R_i thus gives us a direct method to check the validity of the NMF theory.

In Fig. 3 we compare RNAP number distribution in the NMF approximation with the exact distribution obtained from stochastic simulations for two segments of the gene, $i=2$ and $i=L=30$. In the case of constitutive gene expression [Fig. 3(a)] we find an excellent agreement between the distributions, in accordance with the fact $R_2=1.5\times10^{-3}$ and $R_{30}=10^{-4}$. In the case of bursty expression, the agreement is not so good for $i=2$, but improves for $i=L$ [Fig. 3(b)]. Again, this can be explained by stronger correlations at the start of the gene than at the end, $R_2=0.35$ and $R_{30}=0.18$. The disagreement between the distributions is strongest in the bimodal and nearly bimodal cases [Figs. 3(c) and 3(d)]. In those cases the correlations are the strongest: we find $R_2=0.74$ and $R_{30}=0.58$ for the bimodal and $R_2=0.76$ and $R_{30}=0.59$ for the nearly bimodal distributions.

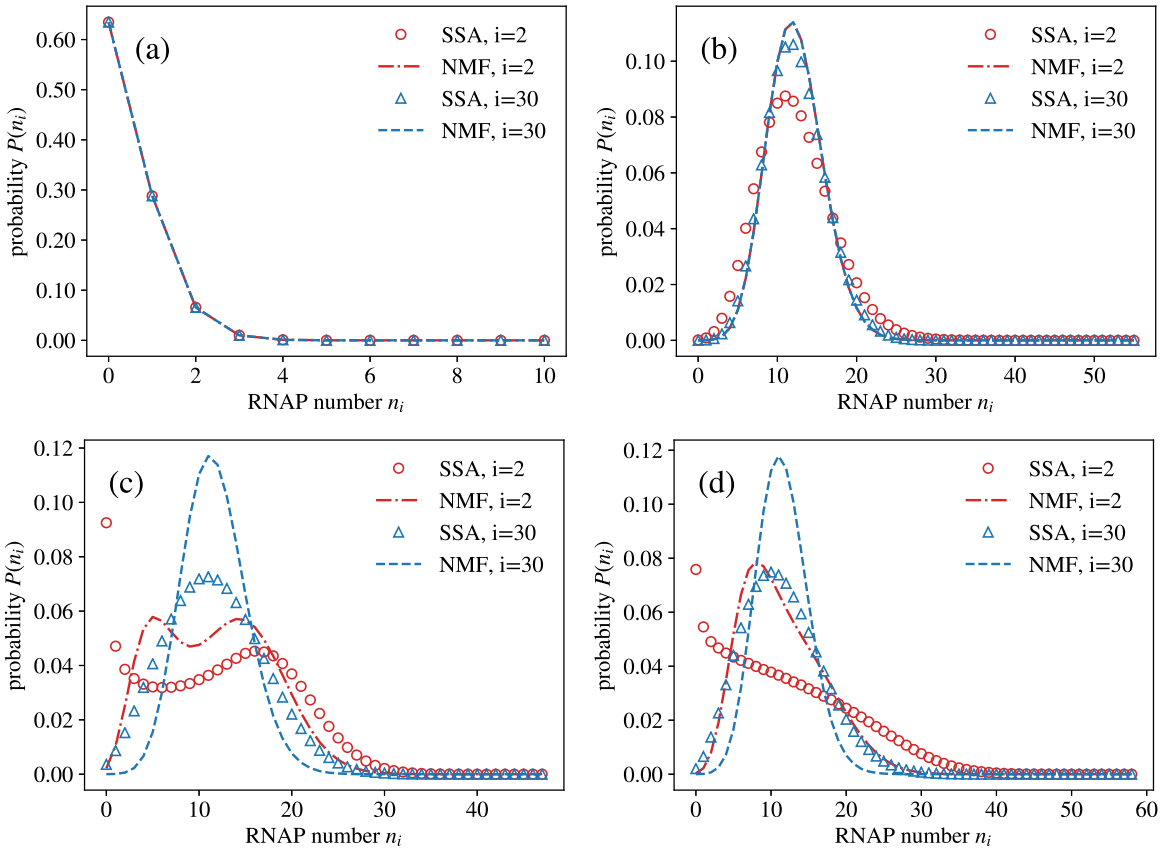

FIG. 3. RNAP number distribution in segments $i = 2$ and $i = L = 30$ in the naive mean-field approximation [NMF; the distribution is given by Eq. (51)], compared to the exact distribution obtained by using the stochastic simulation algorithm (SSA). Elongation rates are assumed to be uniform ($k_{\mathrm{el},i} = k_{\mathrm{el}}$). (a) Constitutive expression: $r = 0.5$, $k_{\mathrm{el}} = 1$, $d = 0$, $s_u = 10$, $s_b = 1$. (b) Bursty expression: $r = 100$, $k_{\mathrm{el}} = 1$, $d = 0$, $s_u = 7$, $s_b = 50$. (c) Bimodal distribution: $r = 20$, $k_{\mathrm{el}} = 1$, $d = 0$, $s_u = 0.35$, $s_b = 0.25$. (d) Nearly bimodal distribution: $r = 30$, $k_{\mathrm{el}} = 1$, $d = 0$, $s_u = 0.5$, $s_b = 0.8$.

B. Improved mean-field approximation

We saw in the previous section that the NMF approximation is equivalent to a telegraph model that produces exact mean but wrong variance. In this section we consider a telegraph model with general rates,

$$G \underset{s'_{u,i}}{\overset{s'_{b,i}}{\rightleftharpoons}} G^*, \tag{57a}$$

$$G \xrightarrow{p'_i} G + P_i, \tag{57b}$$

$$G^* \xrightarrow{q'_i} G^* + P_i, \tag{57c}$$

$$P_i \xrightarrow{k'_{\mathrm{el},i}+d'_i} \emptyset. \tag{57d}$$

We want to find rates p'_i, q'_i, $s'_{u,i}$, $s'_{b,i}$, $k'_{\mathrm{el},i}$, and d'_i such that the mean μ'_i and variance σ'_i of the telegraph model described by Eqs. (57a)–(57d) are both matched to their exact values μ_i in Eq. (23) and $\sigma_i^2 = \mathrm{cov}(n_i, n_i)$ in Eq. (28), respectively. In particular, we want to solve

$$\mu'_i = \alpha'_i \eta' + \beta'_i(1-\eta') = \mu_i, \tag{58a}$$

$$\sigma'_i = \mu'_i + (\alpha'_i - \beta'_i)^2 \frac{\eta'(1-\eta')}{1+\sigma'_{u,i}+\sigma'_{b,i}} = \sigma_i, \tag{58b}$$

where α'_i, β'_i, $\sigma'_{u,i}$, and $\sigma'_{b,i}$ are obtained as before by rescaling p'_i, q'_i, $s'_{u,i}$, $s'_{b,i}$ by $k'_{\mathrm{el},i} + d'_i$, and $\eta'_i = \sigma'_{u,i}/(\sigma'_{u,i} + \sigma'_{b,i})$. In other words, we decouple the segments $i-1$ and i as before resulting in an effective telegraph model, but now we reinstate some of these correlations implicitly throughout by matching the variance.

Obviously we cannot find unique rates that solve Eqs. (58a) and (58b) because the problem is overdetermined. We therefore keep the elongation rate $k'_{\mathrm{el},i}$ and the detachment rate d'_i the same as in the original model,

$$k'_{\mathrm{el},i} = k_{\mathrm{el},i}, \quad d'_i = d. \tag{59}$$

That leaves us with four parameters to adjust by solving two equations, Eqs. (58a) and (58b). We further require that the fraction of time the gene spends in the on state, η'_i, is preserved,

$$\eta'_i = \eta. \tag{60}$$

We consider three options for the remaining parameters:

Option 1: We set $s'_{u,i} = s_u$ and $s'_{b,i} = s_b$ so that Eq. (60) is automatically satisfied. We then adjust the effective transcription rates p'_i and q'_i to match the mean and variance. This is a leaky telegraph model with the same gene switching rates as in the original model.

Option 2: We set $q'_i = 0$ and adjust p'_i, $s'_{u,i}$, and $s'_{b,i}$ by solving Eqs. (58a), (58b), and (60). This is a nonleaky telegraph

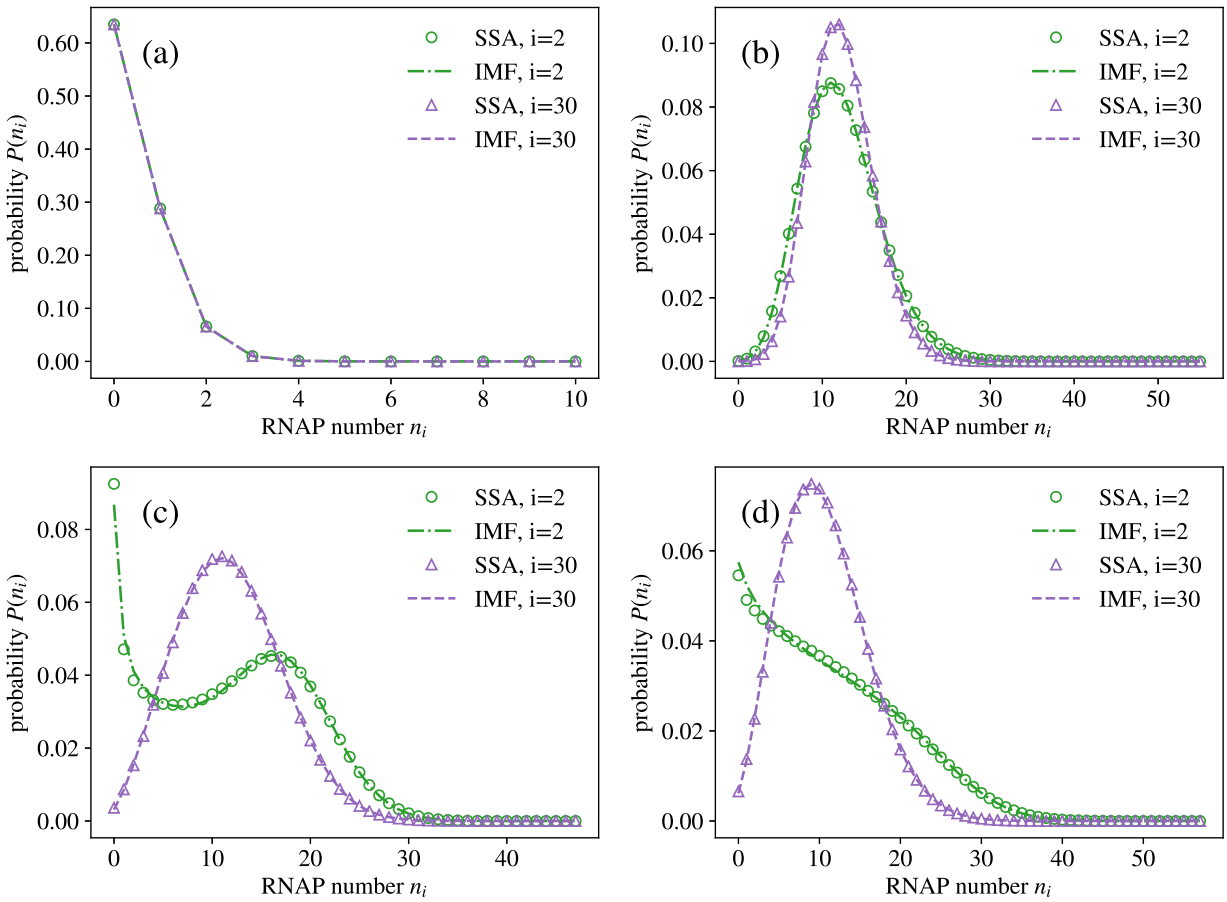

FIG. 4. RNAP number distribution in segments $i = 2$ and $i = L = 30$ in the improved mean-field approximation [IMF; the distribution is given by Eq. (62)], compared to the exact distribution obtained by the stochastic simulation algorithm (SSA). Model parameters are the same as in Fig. 3. The IMF approximation performs significantly better than the NMF approximation and is accurate even when the number correlations between segments are large. (a) Constitutive expression; (b) bursty expression; (c) bimodal distribution; (d) nearly bimodal distribution.

model with effective, segment-dependent gene switching rates.

Option 3: We adjust all four rates p_i', q_i', $s_{u,i}'$, $s_{b,i}'$ by solving Eqs. (58a), (58b), and (60). Additionally, we require that the skewness predicted by the telegraph model, Eq. (63), is matched to the exact skewness computed in Eq. (40). This is a leaky telegraph model with effective, segment-dependent gene switching rates.

Surprisingly, option 1 does not lead to a noticeable improvement over the NMF approximation, and we discard it. In contrast, option 2 significantly improves the NMF approximation in all four cases: the constitutive, bursty, bimodal, and nearly bimodal, as we demonstrate below. Interestingly, option 3 does not provide noticeable improvement over option 2. We will show later that that is because option 2 predicts skewness that is very close to the exact one. Since option 2 has a simpler expression for the probability density, we from now on consider only option 2, to which we refer to as the improved mean-field (IMF) approximation.

For option 2 the solution to Eqs. (58a), (58b), and (60) with $\beta_i' = 0$ is unique and reads

$$\alpha_i' = \frac{p_i'}{k_{\mathrm{el},i} + d} = \frac{\mu_i}{\eta}, \tag{61a}$$

$$\sigma_{u,i}' = \frac{s_{u,i}'}{k_{\mathrm{el},i} + d} = \eta\left(\frac{\mu_i^2}{\sigma_i^2 - \mu_i}\frac{1-\eta}{\eta} - 1\right), \tag{61b}$$

$$\sigma_{b,i}' = \frac{s_{b,i}'}{k_{\mathrm{el},i} + d} = (1-\eta)\left(\frac{\mu_i^2}{\sigma_i^2 - \mu_i}\frac{1-\eta}{\eta} - 1\right). \tag{61c}$$

One can show that the effective parameters $\sigma_{u,i}'$ and $\sigma_{b,i}'$ are always positive, and thus the effective rates can be found for any choice of the model parameters.

These parameter values can then be used to compute the probability distribution in the IMF approximation from the telegraph model,

$$P(n_i) = \frac{\alpha_i'^{n_i} e^{-\alpha_i'}}{n_i!}\frac{(\sigma_{u,i}')_{n_i}}{(\sigma_{u,i}' + \sigma_{b,i}')_{n_i}} \times M(\sigma_{b,i}'; \sigma_{u,i}' + \sigma_{b,i}' + n_i; \alpha_i'). \tag{62}$$

In Fig. 4 we compare the approximate distribution in Eq. (62) to the exact distribution obtained by stochastic simulations using the same model parameters α, σ_u, and σ_b as in Fig. 3. We find an excellent agreement in all four cases, the constitutive, bursty, bimodal, and nearly bimodal. We argue that option 2 performs significantly better than option 1 because the correlations between segments are solely due to promoter

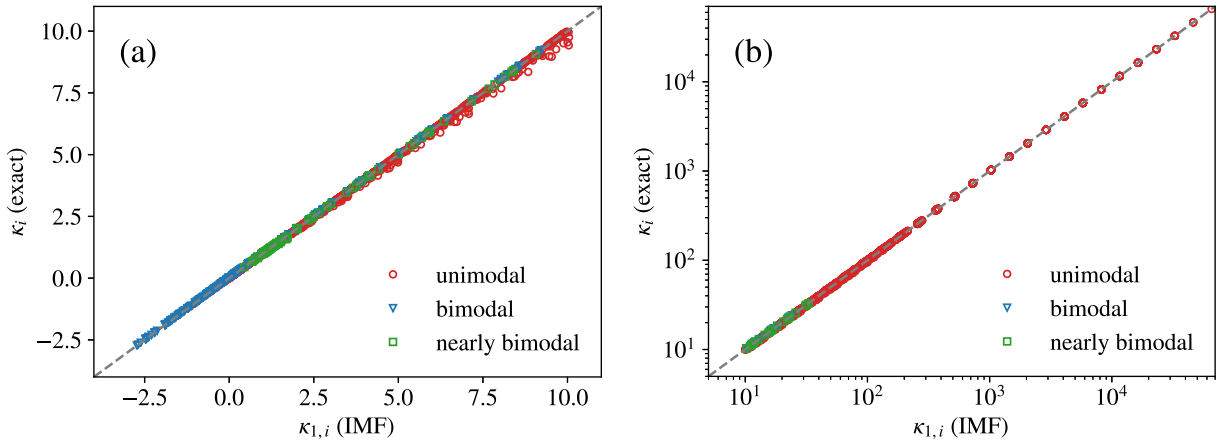

FIG. 5. Skewness $\kappa_{1,i}$ predicted by the improved mean-field approximation [IMF; see Eq. (63)] compared to the exact value κ_i [given by Eq. (40)] for 8000 combinations of α, σ_u, and σ_b. Each point $(\kappa_{1,i}, \kappa_i)$ corresponds to one combination of parameters α, σ_u, and σ_b. For each combination, the segment i with the largest relative error between $\kappa_{1,i}$ and κ_i was selected. The results were divided in two sets: (a) is for $\kappa_i < 10$ (linear scale) and (b) is for $\kappa_i > 10$ (log scale). The dashed line is for reference only and given by $\kappa_{1,i} = \kappa_i$.

switching. These correlations can be better taken into account by adjusting the effective on and off rates, rather than keeping them the same.

Because the IMF approximation matches only the first two moments, we wanted to check its accuracy in predicting the third standardized central moment or skewness. To this end we generated 8000 unique values of the parameters α, σ_u, and σ_b. The values for each parameter were generated uniformly on a logarithmic scale according to the formula 2^k for integer k between -13 and 6. That gave us 20 values for each parameter spanning over five orders of magnitude in the range from 1.2×10^{-4} to 64. For each combination we computed the exact skewness κ_i according to Eq. (40), and compared it to the one predicted by the IMF approximation [computed from Eq. (62)],

$$\kappa_{1,i} = \frac{1}{\sigma_i^3}\left[-\frac{2\alpha_i'^3\eta(1-\eta)(2\eta-1)}{(1+\sigma_{u,i}'+\sigma_{b,i}')(2+\sigma_{u,i}'+\sigma_{b,i}')} + 3\sigma_i^2 - 2\mu_i\right]. \tag{63}$$

We did this for all segments of the gene, after which we selected the segment with the largest relative error $\epsilon = 100\% \times |\kappa_{1,i} - \kappa_i|/\kappa_i$. For this segment we further computed the RNAP number distribution in the IMF approximation according to Eq. (62). Depending on the shape, we determined to which of the three categories the distribution corresponds to unimodal (without an inflection point), bimodal, and nearly bimodal (unimodal with an inflection point). Out of 8000 distributions, 7190 were unimodal, 487 bimodal, and 323 nearly bimodal.

The results comparing $\kappa_{1,i}$ vs exact κ_i are presented in Fig. 5. Due to the large range of κ_i, we separated the data according to $\kappa_i < 10$ [Fig. 5(a), linear scale] and $\kappa_i > 10$ [Fig. 5(b), log scale]. Next, we inspected in more detail how the relative error ϵ is distributed among the four aforementioned categories. We focused on distributions for which the relative error $\epsilon > 10\%$. We found 74 such distributions in total; 64 were unimodal and 10 were bimodal. A closer inspection of these distributions revealed that they all had in common skewness $\kappa_i \approx 0$, which caused the relative error to be large. We inspected the distributions with the two largest ϵ values and used the Hellinger distance [32] to quantify their similarity compared to the exact distributions obtained by stochastic simulations. Indeed, we found very small Hellinger distances for both distributions, 3.7×10^{-3} and 3.3×10^{-3}, confirming the high accuracy of our IMF approximation.

C. Other applications: Nuclear retention and export of mRNA

So far we have analyzed the performance of the IMF theory in the context of transcription elongation, for which the assumption of uniform elongation seems reasonable. However, in other multistep downstream processes such as splicing or transport of nuclear mRNA to the cytoplasm, we cannot expect the rates of different downstream processing steps to be equal because each step may represent a physically different process.

Here we demonstrate the performance of the IMF theory for the case of nonuniform rates. We consider the model for nuclear mRNA retention developed in Ref. [19]. A reaction scheme for this model with the corresponding rates is given by

$$G \underset{s_u}{\overset{s_b}{\rightleftharpoons}} G^*, \quad G \xrightarrow{r} G + P_1, \quad P_1 \xrightarrow{k_1} P_2 \xrightarrow{k_2} \emptyset. \tag{64}$$

In this model, nuclear mRNA P_1 is produced at rate r from a gene in the on state and exported at rate k_1 from the nucleus to the cytoplasm [this is a $L = 2$ version of the reaction scheme (1)]. The cytoplasmic mRNA P_2 is degraded at rate k_2. Parameters for this model were quantified experimentally for a set of genes in mouse liver cells in Ref. [19]. For our purposes, we will use the following parameters obtained for the Mlxipl gene:

$$r = 77\,\mathrm{hr}^{-1}, \quad k_1 = 0.9\,\mathrm{hr}^{-1}, \quad k_2 = 3.8\,\mathrm{hr}^{-1}. \tag{65}$$

In Ref. [19] only the fraction of active Mlxipl genes $\eta = 0.425$ was measured but not the absolute on and off rates s_u and s_b. In addition, the coefficient of variance (CV) for cytoplasmic mRNA was measured to be CV $= 0.46$. By matching η and CV to those predicted by our theory (using the results for the

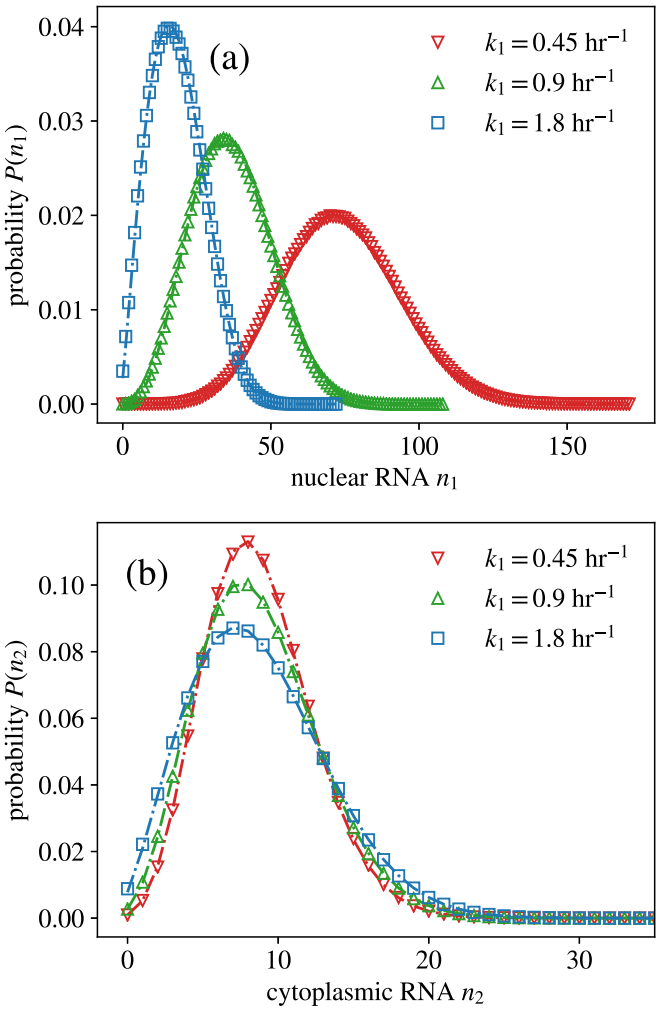

FIG. 6. Probability distributions of nuclear and cytoplasmic mRNA numbers in a two-state model of transcription with nuclear retention. Distributions were obtained by stochastic simulations (symbols) and compared to the prediction of the IMF theory (lines). (a) Distribution of nuclear mRNA number $P(n_1)$. (b) Distribution of cytoplasmic mRNA number $P(n_2)$. The reaction scheme is given by (64), and the parameters are given in the main text.

mean and variance), we estimated the switching parameters: $s_u = 4.3$ hr^{-1} and $s_b = 5.7$ hr^{-1}. In addition to these values, we considered two additional values of export rate k_1 that were two times smaller and larger than the experimental value.

Using these parameters, we performed stochastic simulations of the model and compared the distributions of P_1 and P_2 with predictions of the IMF theory. The results for nuclear mRNA are presented in Fig. 6(a) and for cytoplasmic mRNA in Fig. 6(b). In all cases we find excellent agreement with the predictions of the IMF theory, showing that it is equally accurate for nonequal rates of downstream processing steps.

D. Variation of the shape of the distribution with increasing number of downstream processing steps

In Fig. 4(c), we saw how a distribution that is bimodal for small i can become unimodal for large enough i. The question we want to address here is whether this observation is a special case or if it generally holds.

Where the model is interpreted as one for RNAP dynamics during transcription, assuming uniform elongation rates across the gene ($k_{\mathrm{el},i} = k_{\mathrm{el}}$) and no premature detachment ($d = 0$), and using the exact Eqs. (23) and (32), one can easily deduce from Eqs. (32) and (33) that the factor $\frac{\mu_i^2}{\sigma_i^2 - \mu_i}$ increases monotonically with i. Hence, from Eqs. (61b) and (61c), it follows that the effective switching rates in the IMF theory increase with distance i from the start site. This implies that any bimodality in the RNAP distribution is washed out as i increases because bimodality typically manifests due to slow switching between inactive and active states [33,34].

Remarkably, this is true also in the general case of nonuniform elongation rates (including premature detachment), as we prove in the Appendix. Hence, we can state that for the general model defined in (1), any bimodality in the distribution of P_i will decrease as i increases. Interpreting reaction scheme (1) as a model of the whole mRNA life cycle, this implies that in a population of identical cells (since our model does not consider cell-to-cell variability), it is easier to observe bimodality in nascent or nuclear mRNA distributions than in cytoplasmic distributions.

The increase of the effective switching rates with i also has implications for the inference of the extent of transcriptional bursting [35]. This phenomenon is characterized by bursts of mRNA expression separated by long intervals of no expression, i.e., large mean burst size and low burst frequency. These two parameters are often estimated by fitting the telegraph model to measured distributions of the mRNA copy number [36].

Now, it is well known that within the mathematical framework of the (nonleaky) telegraph model of gene expression, the mean burst size is the initiation rate divided by the off switching rate, while the burst frequency is the on switching rate. We have shown above that the effective on rate increases with i, which means that fitting the telegraph model to mRNA data primarily described by the ith stage of the life cycle will necessarily overestimate the true burst frequency. The effective mean burst size is obtained by dividing Eq. (61a) by Eq. (61c). While the latter equation increases with i, the former equation can be independent or increase or decrease with i. Hence, fitting the telegraph model to mRNA data primarily described by the ith stage of the life cycle can under or overestimate the true burst size.

These results show that noise due to downstream processing steps can have a significant effect on the estimation of transcriptional parameters; this makes the case that for the accurate estimation of these parameters, the use of nascent mRNA data (which would correspond to small i in our model) is preferable to the use of cytoplasmic or whole cell mRNA data [37].

V. SUMMARY AND DISCUSSION

In this paper, we have devised a simple but very effective approximation to the steady-state distribution of the reaction scheme (1). The main idea behind our work was to approximate the dynamics of P_i in this model by that of an effective

(nonleaky) telegraph model:

$$G \rightleftharpoons G^{\star}, \quad G \to G + P_i, \quad P_i \to \emptyset. \tag{66}$$

In the improved mean-field approximation, the effective rates of the latter are deduced by matching the first two moments of $P(n_i)$ in the full and reduced models such that the ratio of the switching rates is the same in both models (but the absolute rates are not the same). This procedure is possible because the moments of both models are known exactly due to the linearity of their propensities. We have shown that this automatically means that the third moments of both models are very close to each other, which suggests that the distributions of both models are also very close. By an extensive search over parameter space, we verified that the steady-state distribution of the reduced model distribution (which is known in closed form) was very close to that obtained from stochastic simulations of the full model: the maximum Hellinger distance between the two distributions is of the order of 10^{-3}, and in fact in all cases one could not easily distinguish by eye any difference between the two.

The study by Vastola *et al.* [18] is the only other study of the steady-state distribution of (1) of which we are aware. In this paper, the same model is considered, but there is no premature degradation and there is a nonzero transcription rate from the off state G^*. The model is studied in the context of transcription coupled to a switching gene with multiple downstream splicing steps. In this case, P_i can be interpreted as the number of mRNA molecules after $i-1$ splicing steps. The paper uses tools inspired by quantum mechanics, including ladder operators and Feynman-like diagrams, to write a formula for the joint distribution in terms of an infinite sum. This cannot be generally written in closed form. In particular the solution is in powers of the difference between the transcription rates of the active and inactive states, and since this is not typically small, one must compute many series terms in order to closely approximate the correct answer. This is a disadvantage compared to our method, which leads to a simple closed-form solution that is easy to evaluate and numerically stable. The disadvantage of our method relative to that of Vastola *et al.* is that it is not based on a systematic derivation; however, as we have shown, our method leads to an impressively accurate approximation of the full model over all parameter space. Another work related to our study is that by Bokes and Singh [24], which studied a one-state gene system where mRNA molecules are produced in bursts (with a size that is distributed according to the geometric distribution) and then they are exported to the nucleus. This is likely [38] a special case of our model, i.e., the bursty limit where r and $s_b \to \infty$ at constant r/s_b for two segments $L = 2$. The authors find an exact solution to this special case; the marginal distribution of mRNA is found to be always unimodal. The advantage of our approximation is that it is valid not just in the bursty limit but all across parameter space, thus being able to capture the variation of the modality of the distribution through the mRNA life cycle (all three types of distributions—unimodal, bimodal, and nearly bimodal—have been measured [39–41]).

Finally, we note that our present model could be improved to include further biological realism, e.g., an arbitrary number of gene states and time-dependent switching rates. The latter is important since the presence of additional states is sometimes needed to explain the nonexponential duration of the off state for mammalian genes [42]; indeed, a recent paper [43] found that a model equivalent to our reaction scheme (1) but with at most three gene states was sufficient to explain the general features of transcription and pervasive stochastic splice site selection. The extension of our model to consider time-dependent switching rates is also important to describe, for example, the activation of a promoter by a time-dependent transcriptional factor signal; e.g., the identities and intensities of different stresses in budding yeast are transmitted by modulation of the amplitude, duration, or frequency of nuclear translocation of the general stress response transcription factor Msn2 [44]. The modification of our improved mean-field approach to describe these more complex scenarios will be reported in a separate paper. Ultimately, we would also like to use the model for interpreting live cell imaging using fluorescence microscopy, for example, as was done in Ref. [45] but taking into account stochastic rather than deterministic transcriptional elongation.

ACKNOWLEDGMENT

We are grateful to Tatiana Filatova for critical reading of an earlier version of this manuscript. This work was supported by a Leverhulme Trust research award (Grant No. RPG-2020-327).

APPENDIX: MATHEMATICAL PROOF THAT THE EFFECTIVE SWITCHING RATES INCREASE ALONG THE GENE

We prove here that $\sigma_{u,i}$ and $\sigma_{b,i}$, given by Eqs. (61b) and (61c), respectively, are monotonically increasing along the gene, i.e., that

$$\sigma_{u,i-1} < \sigma_{u,i}, \quad \sigma_{b,i-1} < \sigma_{b,i}, \quad i = 2, \ldots, L. \tag{A1}$$

We prove that this statement is true for *any* choice of the elongation rates $k_{\mathrm{el},i}$.

According to Eqs. (61b) and (61c), the dependence on i in both $\sigma_{u,i}$ and $\sigma_{b,i}$ comes from the expression

$$\frac{\mu_i^2}{\sigma_i^2 - \mu_i} = \frac{1}{g^{(ii)}/(g^{(i)})^2 - 1}. \tag{A2}$$

If we define the ratio $w^{(ij)}$ as

$$w^{(ij)} = \frac{g^{(ij)}}{g^{(i)} g^{(j)}}, \tag{A3}$$

then the claim in Eq. (A1) is equivalent to the claim that $w^{(ii)}$ is decreasing, i.e., that

$$w^{(ii)} < w^{(i-1i-1)}, \quad i = 2, \ldots, L. \tag{A4}$$

In order to prove (A4), we use Eqs. (29) to derive recurrence relations for $w^{(ij)}$, which read

$$w^{(ij)} = \frac{\omega_i}{\omega_i + \omega_j} w^{(i-1j)} + \frac{\omega_j}{\omega_i + \omega_j} w^{(ij-1)}, \quad i, j = 2, \ldots, L, \tag{A5a}$$

$$w^{(1j)} = \frac{\omega_j}{\omega_1 + \omega_j} w^{(1j-1)} - \frac{\omega_1}{\omega_1 + \omega_j} \left(\frac{r g_{\mathrm{on}}^{(j)}}{g^{(j)}} \right), \quad j = 2, \ldots, L, \tag{A5b}$$

$$w^{(i1)} = \frac{\omega_i}{\omega_i+\omega_1} w^{(i-11)} - \frac{\omega_1}{\omega_i+\omega_1}\left(\frac{r g_{\rm on}^{(i)}}{g^{(i)}}\right), \quad i = 2,\dots,L, \tag{A5c}$$

$$w^{(11)} = \frac{r g_{\rm on}^{(1)}}{g^{(1)}}. \tag{A5d}$$

Note that $w^{(ij)} = w^{(ji)}$, and thus it is sufficient to consider only $j \geqslant i$. For $j = i$, Eq. (A5a) gives

$$w^{(ii)} = w^{(i-1i)}, \quad i = 2,\dots,L. \tag{A6}$$

Next, if we can show that the following inequalities hold for $j > i$:

$$w^{(i-1j)} < w^{(ij)} < w^{(ij-1)}, \quad j > i, \tag{A7}$$

then it is easy to show that (A4) is true since

$$\begin{aligned} w^{(ii)} &= \frac{1}{2}w^{(i-1i)} + \frac{1}{2}w^{(ii-1)} = w^{(i-1i)} \\ &= \frac{\omega_{i-1}}{\omega_{i-1}+\omega_i} w^{(i-2i)} + \frac{\omega_i}{\omega_{i-1}+\omega_i} w^{(i-1i-1)} \\ &< \frac{\omega_{i-1}}{\omega_{i-1}+\omega_i} w^{(i-2i-1)} + \frac{\omega_i}{\omega_{i-1}+\omega_i} w^{(i-1i-1)} \\ &= \frac{\omega_{i-1}}{\omega_{i-1}+\omega_i} w^{(i-1i-1)} + \frac{\omega_i}{\omega_{i-1}+\omega_i} w^{(i-1i-1)} \\ &= w^{(i-1i-1)}. \end{aligned} \tag{A8}$$

We first prove that $w^{(1j)} < w^{(1j-1)}$. According to Eqs. (22) and (24), $g^{(j)}$ and $g_{\rm on}^{(j)}$ are both negative for all $j = 1,\dots,L$, and thus their ratio is always positive. Since $w^{(11)}$ is positive, and $\omega_j/(\omega_1+\omega_j) < 1$ and $\omega_i/(\omega_i+\omega_1) < 1$, we conclude that

$$w^{(1j)} < w^{(1j-1)}, \quad j > 1. \tag{A9}$$

Next, we prove (A7) by induction for any $2 \leqslant i \leqslant L$. First, we show that (A7) is true for $i = 2$ and all $j > 2$. We then assume that (A7) holds for an arbitrary $i = k$ and all $j > k$. Then we show that (A7) also holds for $i = k+1$ and $j > k+1$.

We first consider $i = 2$ and $j = 3$. Using (A5a) we obtain

$$w^{(23)} = \frac{\omega_2}{\omega_2+\omega_3} w^{(13)} + \frac{\omega_2}{\omega_2+\omega_3} w^{(22)}, \tag{A10}$$

and since $w^{(22)} = w^{(12)} > w^{(13)}$, it follows that $w^{(23)} < w^{(22)}$ and $w^{(23)} > w^{(13)}$. Next, we assume that (A7) holds for $i = 2$ and some arbitrary $j = k > 3$. Using (A5a) we get that

$$w^{(2k+1)} = \frac{\omega_2}{\omega_2+\omega_{k+1}} w^{(1k+1)} + \frac{\omega_{k+1}}{\omega_2+\omega_{k+1}} w^{(2k)}. \tag{A11}$$

Since $w^{(1k+1)} < w^{(1k)} < w^{(2k)}$, we obtain that $w^{(2k+1)} < w^{(2k)}$ and $w^{(2k+1)} > w^{(1k+1)}$. This proves (A7) for $i = 2$ and $j > 2$.

Next, we assume that (A7) holds for some arbitrary $i = k > 2$ and all $j > k$. We then consider the case of $i = k+1$. For $i = k+1$ and $j = k+2$ we get that

$$w^{(k+1k+2)} = \frac{\omega_{k+1}}{\omega_{k+1}+\omega_{k+2}} w^{(kk+2)} + \frac{\omega_{k+2}}{\omega_{k+1}+\omega_{k+2}} w^{(k+1k+1)}. \tag{A12}$$

Since $w^{(kk+2)} < w^{(kk+1)} = w^{(k+1k+1)}$, we conclude that $w^{(k+1k+2)} < w^{(k+1k+1)}$. Now, we assume that (A7) holds for $i = k+1$ and some arbitrary $j = m > k+2$. Then using (A5a) we get that

$$w^{(k+1m+1)} = \frac{\omega_{k+1}}{\omega_{k+1}+\omega_{m+1}} w^{(km+1)} + \frac{\omega_{m+1}}{\omega_{k+1}+\omega_{m+1}} w^{(k+1m)}. \tag{A13}$$

Since $w^{(km+1)} < w^{(km)} < w^{(k+1m)}$, it follows that $w^{(k+1m+1)} < w^{(k+1m)}$ and $w^{(k+1m+1)} > w^{(km+1)}$, and thus (A7) also holds for $i = k+1$ and $j = m+1$. This completes our proof of (A7) and thus the proof of (A4).

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

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

[3] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer, Berlin, 2009).

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

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

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

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

[8] L. Ham, D. Schnoerr, R. D. Brackston, and M. P. H. Stumpf, Exactly solvable models of stochastic gene expression, J. Chem. Phys. **152**, 144106 (2020).

[9] C. Shi, Y. Jiang, and T. Zhou, Queuing models of gene expression: Analytical distributions and beyond, Biophys. J. **119**, 1606 (2020).

[10] Z. Cao, T. Filatova, D. A. Oyarzún, and R. Grima, A stochastic model of gene expression with polymerase recruitment and pause release, Biophys. J. **119**, 1002 (2020).

[11] L. Ham, R. D. Brackston, and M. P. H. Stumpf, Extrinsic Noise and Heavy-Tailed Laws in Gene Expression, Phys. Rev. Lett. **124**, 108101 (2020).

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

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

[14] R. Perez-Carrasco, C. Beentjes, and R. Grima, Effects of cell cycle variability on lineage and population measurements of messenger RNA abundance, J. R. Soc. Interface **17**, 20200360 (2020).

[15] C. Zhu, G. Han, and F. Jiao, Dynamical regulation of mRNA distribution by cross-talking signaling pathways, Complexity **2020**, 6402703 (2020).

[16] S. Choubey, J. Kondev, and A. Sanchez, Deciphering transcriptional dynamics in vivo by counting nascent RNA molecules, PLoS Comput. Biol. **11**, e1004345 (2015).

[17] T. Filatova, N. Popovic, and R. Grima, Statistics of nascent and mature RNA fluctuations in a stochastic model of transcriptional initiation, elongation, pausing, and termination, Bull. Math. Biol. **83**, 3 (2020).

[18] J. J. Vastola, G. Gorin, L. Pachter, and W. R. Holmes, Analytic solution of chemical master equations involving gene switching. I: Representation theory and diagrammatic approach to exact solution, arXiv:2103.10992.

[19] K. Bahar Halpern, I. Caspi, D. Lemze, M. Levy, S. Landen, E. Elinav, I. Ulitsky, and S. Itzkovitz, Nuclear retention of mRNA in mammalian tissues, Cell Rep. **13**, 2653 (2015).

[20] D. Cao and R. Parker, Computational modeling of eukaryotic mRNA turnover, RNA **7**, 1192 (2001).

[21] B. Zoller, S. C. Little, and T. Gregor, Diverse spatial expression patterns emerge from unified kinetics of transcriptional bursting, Cell **175**, 835 (2018).

[22] J. Rodriguez, G. Ren, C. R. Day, K. Zhao, C. C. Chow, and D. R. Larson, Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity, Cell **176**, 213 (2019).

[23] M. Z. Ali, S. Choubey, D. Das, and R. C. Brewster, Probing mechanisms of transcription elongation through cell-to-cell variability of RNA polymerase, Biophys. J. **118**, 1769 (2020).

[24] A. Singh and P. Bokes, Consequences of mRNA transport on stochastic variability in protein levels, Biophys. J. **103**, 1087 (2012).

[25] G. Gorin and L. Pachter, Analytical solutions of the chemical master equation with bursty production and isomerization reactions, bioRxiv 2021.03.24.436847 (2021).

[26] M. Smith, M. Soltani, R. Kulkarni, and A. Singh, Amplification and attenuation of noisy expression by export processes, bioRxiv 2021.10.06.463423 (2021).

[27] N. Van Kampen, in *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, Amsterdam, 2007), p. 139.

[28] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1974).

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

[30] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan *et al.*, RNA velocity of single cells, Nature (London) **560**, 494 (2018).

[31] We note that in the derivation of $\mathrm{cov}(n_i, n_j)$ we have used the fact that $\gamma_{i,j}(0,0) = 1$, which is not at all obvious that is true. One can prove this relation by showing that $\gamma_{i,j}(0,0) = \gamma_{i,j-1}(0,0)/2 + \gamma_{i-1,j}(0,0)/2$ for any $i, j \geqslant 2$, and also that $\gamma_{i,1}(0,0) = \gamma_{1,j}(0,0) = 1$ for any $i, j \geqslant 1$. This recurrence relation has a simple solution $\gamma_{ij}(0,0) = 1$ for any $i, j \geqslant 1$.

[32] The Hellinger distance takes values between 0 and 1, where 0 is achieved if two distributions are equal, and 1 is achieved if one distribution assigns probability 0 to every set to which the other assigns a positive probability and vice versa.

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

[34] T. Plesa, R. Erban, and H. G. Othmer, Noise-induced mixing and multimodality in reaction networks, Eur. J. Appl. Math. **30**, 887 (2019).

[35] R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger, Transcriptional burst frequency and burst size are equally modulated across the human genome, Proc. Natl. Acad. Sci. USA **109**, 17454 (2012).

[36] A. J. M. Larsson, P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, Å. Segerstolpe, C. M. Rivera, B. Ren, and R. Sandberg, Genomic encoding of transcriptional burst kinetics, Nature (London) **565**, 251 (2019).

[37] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Stochastic Kinetics of Nascent RNA, Phys. Rev. Lett. **117**, 128101 (2016).

[38] It is yet to be formally proven using singular perturbation theory that this is indeed the case. A proof can be constructed similar to that reported for other systems [13].

[39] Z. S. Singer, J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz, Dynamic heterogeneity and DNA methylation in embryonic stem cells, Mol. Cell **55**, 319 (2014).

[40] H. Xu, L. A. Sepúlveda, L. Figard, A. M. Sokac, and I. Golding, Combining protein and mRNA quantification to decipher transcriptional regulation, Nat. Methods **12**, 739 (2015).

[41] S. Hocine, M. Vera, D. Zenklusen, and R. H. Singer, Promoter-autonomous functioning in a controlled environment using single molecule fish, Sci. Rep. **5**, 9934 (2015).

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

[43] Y. Wan, D. G. Anastasakis, J. Rodriguez, M. Palangat, P. Gudla, G. Zaki, M. Tandon, G. Pegoraro, C. C. Chow, M. Hafner, and D. R. Larson, Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection, Cell **184**, 2878 (2021).

[44] N. Hao and E. K. O'Shea, Signal-dependent dynamics of transcription factor translocation controls gene expression, Nat, Struct. Mol. Biol. **19**, 31 (2012).

[45] J. Liu, D. Hansen, E. Eck, Y. J. Kim, M. Turner, S. Alamos, and H. G. Garcia, Real-time single-cell characterization of the eukaryotic transcription cycle reveals correlations between RNA initiation, elongation, and cleavage, PLoS Comput. Biol. **17**, e1008999 (2021).