Bulletin of Mathematical Biology (2021) 83:3
https://doi.org/10.1007/s11538-020-00827-7

Society for Mathematical Biology

ORIGINAL ARTICLE

Statistics of Nascent and Mature RNA Fluctuations in a Stochastic Model of Transcriptional Initiation, Elongation, Pausing, and Termination

Tatiana Filatova[1,2] · Nikola Popovic[2] · Ramon Grima[1]

Received: 13 May 2020 / Accepted: 26 October 2020 / Published online: 22 December 2020

Abstract
Recent advances in fluorescence microscopy have made it possible to measure the fluctuations of nascent (actively transcribed) RNA. These closely reflect transcription kinetics, as opposed to conventional measurements of mature (cellular) RNA, whose kinetics is affected by additional processes downstream of transcription. Here, we formulate a stochastic model which describes promoter switching, initiation, elongation, premature detachment, pausing, and termination while being analytically tractable. We derive exact closed-form expressions for the mean and variance of nascent RNA fluctuations on gene segments, as well as of total nascent RNA on a gene. We also obtain exact expressions for the first two moments of mature RNA fluctuations and approximate distributions for total numbers of nascent and mature RNA. Our results, which are verified by stochastic simulation, uncover the explicit dependence of the statistics of both types of RNA on transcriptional parameters and potentially provide a means to estimate parameter values from experimental data.

Keywords Stochastic gene expression · Master equation · RNA fluctuations · Singular perturbation theory · Distributions of RNA molecules · Stochastic simulations

1 Introduction

Transcription, the production of RNA from a gene, is an inherently stochastic process. Specifically, the interval of time between two successive transcription events is a random variable whose statistics depend on multiple single-molecule events behind

✉ Ramon Grima
Ramon.Grima@ed.ac.uk

1 School of Biological Sciences, The University of Edinburgh, Edinburgh, UK

2 School of Mathematics and Maxwell Institute for Mathematical Sciences, The University of Edinburgh, Edinburgh, UK

transcription (Sanchez and Golding 2013). When the distribution of this random variable is exponential, we say that expression is constitutive; in that case, the number of transcripts produced in a certain interval of time follows a Poisson distribution. On the other hand, when the distribution of times between two successive transcripts is non-exponential, then the number of transcripts is non-Poissonian. A special case of such non-constitutive behaviour is bursty expression, whereby transcripts are produced in short bursts that are separated by long silent intervals (Suter et al. 2011; Halpern et al. 2015). In yeast, genes whose expression is constitutive include MDN1, KAP104, and DOA1, whereas PDR5 is an example of a gene whose expression is bursty (Zenklusen et al. 2008).

For two decades, mathematical models of gene expression have been developed to predict the distribution of RNA abundance. By matching the theoretical distribution with experimental measurements from microscopy-based methods (Raj et al. 2008), one hopes to obtain insight into the underlying kinetics of transcription and to estimate transcriptional parameters. The standard model of gene expression which has been used for these analyses is the telegraph model (Peccoud and Ycart 1995), whereby a gene can be in two states. Transcription occurs in one of the states, whereupon RNA degrades; first-order kinetics is assumed for all processes. While the distribution obtained from the telegraph model can typically fit cellular RNA abundance data, there are innate difficulties with the interpretation of that fit: fluctuations in cellular RNA numbers and, hence, the shape of the experimental RNA distribution do not only reflect transcription, but also many processes downstream thereof, such as splicing, RNA degradation, and partitioning during cell division.

To counteract these difficulties, in the past few years, mathematical models (Choubey et al. 2015; Choubey 2018; Heng et al. 2016; Cao and Grima 2020) have been developed to predict the statistics of nascent RNA, i.e. of RNA in the process of being synthesised by the RNA polymerase molecule (RNAP), which can be visualised and quantified due to recent advances in fluorescence microscopy (Lenstra et al. 2016; Skinner et al. 2016; Larson et al. 2011; Antoine et al. 2014; Brouwer and Lenstra 2019). In contrast to cellular RNA, the statistics of nascent RNA is a direct reflection of the transcription process; hence, these models can potentially give more insight than the simpler, but cruder telegraph model. Choubey and collaborators (Choubey et al. 2015; Choubey 2018) have developed a stochastic model with the following properties: (i) a gene can be in two states (active or inactive); (ii) from the active state, transcription initiation occurs in two sequential steps: the pre-initiation complex is formed, after which the RNA polymerase escapes the promoter; (iii) once on the gene, the polymerase moves from one base pair to the next (with some probability) until the end of the gene is reached, when transcription is terminated and polymerase detaches. Queuing theory is used to derive analytical expressions for the transient and steady-state means and variances of numbers of RNAP that are attached to the gene in the long-gene limit when the elongation time is practically deterministic. Heng et al. (2016) have considered a coarse-grained version of that model, whereby the movement of RNAP from one base pair to the next is not explicitly modelled, obtaining an analytical expression for the total RNAP distribution in steady-state conditions. More recently, Cao and Grima (2020) have studied a model of eukaryotic gene expression that yields approximate time-dependent distributions of both nascent and cellular RNA

abundance as a function of the parameters controlling gene switching, DNA duplication, partitioning at cell division, gene dosage compensation, and RNA degradation; in their coarse-grained model, the movement of RNAP is not explicitly modelled, while the elongation time is assumed to be exponentially distributed, which simplifies the requisite analysis.

The complexity of nascent RNA models has thus far not allowed the same detailed level of analysis as has been possible with the much simpler telegraph model. A few shortcomings of current models can be summarised as follows: (i) distributions of nascent RNA have been derived from models that do not explicitly model the movement of RNAP along a gene (Heng et al. 2016; Cao and Grima 2020), resulting in a disconnect between theoretical description and the microscopic processes underlying transcription; (ii) while the analysis of single-cell sequencing data and electron micrograph data yields the positions of individual polymerases along the gene, allowing for the calculation of statistics (means and variances) of the numbers of RNAP on gene segments that are obtained after binning, detailed models of RNAP elongation (Choubey et al. 2015; Choubey 2018) provide analytical results only for total RNAP on a gene and hence cannot be used to understand gene segment data; (iii) analytical calculations of the statistics of nascent RNA ignore important details of the transcription process such as pausing, traffic jams, backtracking, and premature termination, some of which have to date been explored via stochastic simulation (Klumpp and Hwa 2008; Rajala et al. 2010; Choubey et al. 2015; Rodriguez et al. 2019; Md Zulfikar et al. 2020).

In this paper, we overcome some of the aforementioned shortcomings of analytically tractable models for the transcription process. In Sect. 2, we study a stochastic model for promoter switching and the stochastic movement of RNAP along a gene, allowing for premature termination. We derive exact closed-form expressions for the first and second moments (means and variances) of local RNAP fluctuations on gene segments of arbitrary length, which allows us to study how these statistics vary along a gene as a function of transcriptional parameters; we also obtain expressions for the mean and variance of the total RNAP on the gene which generalise previous work by Choubey et al. (2015). In Sect. 3, we investigate approximations for the distributions of total RNAP and mature RNA, showing in particular that Negative Binomial distributions can provide an accurate approximation in certain biologically meaningful limits. In Sect. 4, we illustrate the difference between the statistics of local and total RNAP fluctuations and those of light fluorescence due to tagged nascent RNA. In Sect. 5, we extend our model to include pausing by deriving approximate expressions for the mean, variance, and distribution of observables. We conclude with a discussion of our results in Sect. 6.

2 Detailed Stochastic Model of Transcription: Set-up and Analysis

In this section, we specify the stochastic model studied here; then, we derive closed-form expressions for the moments of mature RNA and of local and total RNAP fluctuations in various parameter regimes.

2.1 Set-up of Model

We consider a stochastic model of transcription that includes the processes of initiation, elongation, and termination, as illustrated in Fig. 1. For simplicity, we divide the gene into L segments; the RNAP on gene segment i is then denoted by P_i. The promoter can be either in the inactive state (G_{off}) or the active state (G_{on}), switching from the inactive state to the active one with rate s_u and from the active state to the inactive one with rate s_b. When the promoter is active, initiation commences via the binding of an RNAP with rate r, denoted by P_1. Subsequently, the RNAP either moves from a gene segment to the neighbouring segment with rate k, or it prematurely detaches with rate d. Note that here we have made two assumptions: (i) the movement of RNAP is unidirectional, away from the promoter site and hence left to right, with no pausing or backtracking allowed; (ii) the detachment and elongation rates are independent of the position of RNAP on the gene. Each RNAP has associated with it a nascent RNA tail that grows longer as the RNAP transcribes more of the gene. When the RNAP reaches the last gene segment, termination occurs, i.e. the RNAP–nascent RNA complex gets dissociated from the gene leading to a mature RNA (M) which degrades with rate d_m. Note that for simplicity, we have not considered excluded-volume interaction between adjacent RNAPs here; hence, we make the implicit assumption of low 'traffic', which is plausible when the initiation rate is sufficiently low. (We test the validity of this assumption through simulations below.)

Note that, while the choice of L is arbitrary, it should be kept in mind that L needs to be sufficiently large for the dynamics to be described at a fine spatial resolution. However, L also has to be small enough for the length of each gene segment to be much

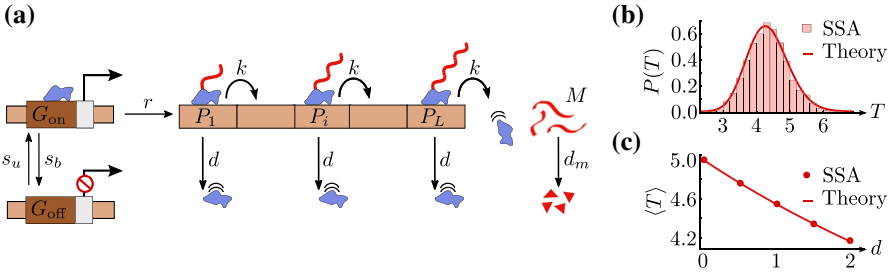


Fig. 1 (Color Figure Online) Model of transcription. **a** The gene is arbitrarily divided into L segments, with RNAP (blue) on gene segment i denoted by P_i. The promoter switches from the active state G_{on} to the inactive state G_{off} with rate s_b, while the reverse switching occurs with rate s_u. When the promoter is active, initiation of RNAP occurs with rate r. Initiation is followed by elongation, which is modelled as RNAP 'hopping' from gene segment i to the neighbouring segment $i+1$ with rate k, i.e. as the transformation of species P_i to P_{i+1}. RNAP prematurely detaches from the gene with rate d. A nascent RNA tail (red), attached to the RNAP, grows as elongation proceeds. Termination is modelled by the change of P_L with rate k to mature RNA (M), which subsequently degrades with rate d_m. In **b**, we show the probability distribution $P(T)$ of the total elongation time T—the time between initiation and termination—as predicted by the stochastic simulation algorithm (SSA; histogram) and our theory (Erlang distribution with shape parameter L and rate $k+d$; solid line). The parameter values used are $L = 50$, $k = 10$/min, and $d = 1.5$/min. In **c**, we show the dependence of the mean of the distribution $P(T)$ on the RNAP detachment rate (d), as predicted by SSA (dots) and our theory ($\langle T \rangle = L/(k+d)$; solid line). The relevant parameter values are $L = 50$ and $k = 10$/min (Color figure online)

larger than the footprint of an RNAP; the latter is needed to ensure the validity of the low-traffic assumption. The elongation time which is the total time T from initiation to termination, that is, conditioning on those realisations for which the RNAP does not prematurely detach, is Erlang distributed with mean $L/(k+d)$ and coefficient of variation $1/\sqrt{L}$; see 'Appendix A' for a derivation and Fig. 1b, c for verification through stochastic simulation (SSA).

Note that the total number of RNAPs transcribing the gene is equal to the number of nascent RNA molecules present, *irrespective of their lengths*; to shed light on the fluctuations of nascent RNA, in this section we therefore focus on the calculation of statistics of local and total RNAP fluctuations. We define the vector of molecule numbers $\vec{m} = (n_0, n_1, \ldots, n_L, n)$, and we write $\langle n_0 \rangle$, $\langle n_i \rangle$ $(i = 1, 2, \ldots, L)$, and $\langle n \rangle$ for the average numbers of molecules of active gene, RNAP, and mature RNA, respectively. The above model can then be conveniently described by $L+2$ species interacting via a set of $2L+4$ reactions with the following rate functions:

Species	Molecule numbers	Position (in $\vec{m}$)
G_{on}	n_0	1
$P_i, \quad i \in \{1, \ldots, L\}$	n_i	$i+1$
M	n	$L+2$

Reaction	Rate function f_j
$G_{\text{on}} \xrightarrow{s_b} G_{\text{off}}$	$f_1 = s_b \langle n_0 \rangle$
$G_{\text{off}} \xrightarrow{s_u} G_{\text{on}}$	$f_2 = s_u (1 - \langle n_0 \rangle)$
$G_{\text{on}} \xrightarrow{r} G_{\text{on}} + P_1$	$f_3 = r \langle n_0 \rangle$
$P_i \xrightarrow{k} P_{i+1}, \quad i \in \{1, \ldots, L-1\}$	$f_{i+3} = k \langle n_i \rangle$
$P_L \xrightarrow{k} M$	$f_{L+3} = k \langle n_L \rangle$
$P_i \xrightarrow{d} \varnothing, \quad i \in \{1, \ldots, L\}$	$f_{i+L+3} = d \langle n_i \rangle$
$M \xrightarrow{d_m} \varnothing$	$f_{2L+4} = d_m \langle n \rangle$

Note that G_{off} is not an independent species; the reason is that the binary state of the gene implies a conservation law, with the sum of the numbers of G_{on} and G_{off} equalling 1. Hence, the number of independent species in the model is $L+2$. The rate functions f_j are the averaged propensities from the underlying chemical master equation (CME); note that, because our reaction network is composed of first-order reactions, these rate functions also equal the reaction rates in the corresponding deterministic rate equations. The description of our model is completed by the $(L+2) \times (2L+4)$-dimensional stoichiometric matrix $\mathbf{S}$; the element $\mathbf{S}_{ij}$ of $\mathbf{S}$ gives the net change in the number of molecules of the ith species when the jth reaction occurs. Given the ordering of species and reactions as described in the tables above, it follows that the matrix $\mathbf{S}$ has the simple form

$$\begin{aligned} \mathbf{S}_{11} &= -1, & \mathbf{S}_{12} &= 1, & & \\ \mathbf{S}_{i,i+1} &= 1, & \mathbf{S}_{i,i+2} &= -1, & \mathbf{S}_{i,i+L+2} &= -1, \\ \mathbf{S}_{L+2,L+3} &= 1, & \mathbf{S}_{L+2,2L+4} &= -1, & & \end{aligned} \tag{1}$$

where $i = 2, \ldots, L+1$.

2.2 Closed-Form Expressions for Moments of Mature RNA and Local RNAP

In this subsection, we outline the derivation of the steady-state means and variances of local RNAP fluctuations (on each gene segment), as well as of mature RNA. Our results are summarised in the following two propositions.

Proposition 1 *Let $\eta = s_u/(s_u + s_b)$ be the fraction of time the gene spends in the active state, let $\rho_k = r/k$ be the mean number of RNAPs binding to the promoter site in the time it takes for a single RNAP to move from one gene segment to the next, let $\rho = r/d_m$ be the mean number of RNAPs binding to the promoter site in the time it takes for a mature RNA to decay, and let $\mu = k/(k+d)$ be the probability that an RNAP molecule moves to the next gene segment rather than detaching prematurely. Then, the steady-state mean numbers of molecules of active gene, RNAP, and mature RNA are given by*

$$\langle n_0 \rangle = \eta, \tag{2a}$$

$$\langle n_i \rangle = \eta \rho_k \mu^i \quad \text{for } i = 1, \ldots, L, \tag{2b}$$

$$\langle n \rangle = \eta \rho \mu^L, \tag{2c}$$

respectively.

Proposition 1 can be proved in a straightforward fashion, as follows. Using the underlying CME, one can show from the corresponding moment equations (Warren et al. 2006) that the time evolution of the vector $\langle \vec{m} \rangle$ of mean molecule numbers in a system of zeroth-order or first-order reactions, i.e. with propensities that are linear in the number of molecules, is given by the time derivative $\mathrm{d}\langle \vec{m} \rangle / \mathrm{d}t = \mathbf{S} \cdot \vec{f}$. Given the form of the stoichiometric matrix $\mathbf{S}$ and of the rate functions f_j, as described in Sect. 2.1, it follows that the mean numbers of all species in steady state can be obtained by solving the following system of $L+2$ algebraic equations:

$$\begin{aligned} 0 &= s_u(1 - \langle n_0 \rangle) - s_b \langle n_0 \rangle, \\ 0 &= r \langle n_0 \rangle - (k+d) \langle n_1 \rangle, \\ 0 &= k \langle n_{i-1} \rangle - (k+d) \langle n_i \rangle \quad \text{for } i = 2, \ldots, L, \\ 0 &= k \langle n_L \rangle - d_m \langle n \rangle. \end{aligned} \tag{3}$$

These equations can easily be solved simultaneously to yield the steady-state value of $\langle \vec{m} \rangle$, as given in Eq. (2).

Proposition 2 *Let $\tau_p = 1/(d+k)$, $\tau_g = 1/(s_u+s_b)$, and $\tau_m = 1/d_m$ be the timescales of fluctuations of RNAP, gene, and mature RNA, respectively, and define the three new parameters*

$$\alpha = \frac{1}{1+\tau_p/\tau_g}, \quad \gamma = \frac{1}{1+\tau_p/\tau_m}, \quad \textit{and} \quad \theta = \frac{1}{1+\tau_m/\tau_g}.$$

Furthermore, let $\beta = s_b/s_u$ denote the ratio of gene inactivation and activation rates. Then, the variances and covariances of molecule number fluctuations of active gene, RNAP, and mature RNA are given by

$$\mathrm{Var}(n_0) = \langle n_0\rangle^2\beta, \tag{4a}$$

$$\mathrm{Cov}(n_0, n_i) = \langle n_0\rangle\langle n_i\rangle\alpha\beta \cdot f_{1i}, \qquad \textit{where } f_{1i} = \alpha^{i-1}; \tag{4b}$$

$$\mathrm{Cov}(n_0, n) = \langle n_0\rangle\langle n\rangle\alpha\beta \cdot f_{1M}, \qquad \textit{where } f_{1M} = \theta\alpha^{L-1}, \tag{4c}$$

$$\mathrm{Cov}(n_i, n_j) = \delta_{ij}\langle n_i\rangle + \langle n_i\rangle\langle n_j\rangle\alpha\beta \cdot f_{ij}, \qquad \textit{where } f_{ij} = f(i,j) + f(j,i), \tag{4d}$$

$$\mathrm{Cov}(n_i, n) = \langle n_i\rangle\langle n\rangle\alpha\beta \cdot f_{iM}, \qquad \textit{where } f_{iM} = \gamma^i\theta\alpha^{L-1} + (1-\gamma)\sum_{q=1}^{i}\gamma^{i-q} f_{qL}, \tag{4e}$$

$$\mathrm{Var}(n, n) = \langle n\rangle + \langle n\rangle^2\alpha\beta \cdot f_{MM}, \qquad \textit{where } f_{MM} = f_{LM}, \tag{4f}$$

and where $i, j = 1, \ldots, L$. Here, δ_{ij} is the Kronecker delta; moreover,

$$f(i,j) = \frac{\alpha^{i+j-1}}{(2\alpha-1)^i} + \frac{1}{2^{i+j-1}}\binom{i+j-1}{i}\Big[1 - \frac{2\alpha-1}{2\alpha}\,{}_2F_1\big(1, i+j; j; \tfrac{1}{2\alpha}\big)\Big],$$

where ${}_2F_1$ denotes the generalised hypergeometric function of the second kind (Digital Library of Mathematical Functions 2020a), which is defined as

$${}_2F_1(a_1, a_2; b_1; z) = \sum_{s=0}^{\infty}\frac{(a_1)_s(a_2)_s}{(b_1)_s}\frac{z^s}{s!},$$

with $(a)_s = \Gamma(a+s)/\Gamma(a)$ the Pochhammer symbol.

Here, we note that an alternative representation of the functions f_{ij} in Eq. (4d), in terms of finite sums, is given in Eq. (B.33) of 'Appendix B'.

As above, since the underlying propensities are linear in the number of molecules, the CME implies (Warren et al. 2006) that the corresponding second moments in steady state are exactly given by a Lyapunov equation. That equation, which is precisely the same as the one that is obtained from the linear-noise approximation (LNA) (Elf and Ehrenberg 2003), takes the form

$$\mathbf{J}\cdot\mathbf{C} + \mathbf{C}\cdot\mathbf{J}^T + \mathbf{D} = \mathbf{0}. \tag{5}$$

Here, $\mathbf{C}$, $\mathbf{J}$, and $\mathbf{D}$ are $(L+2)\times(L+2)$-dimensional matrices; $\mathbf{C}$ is a variance–covariance matrix that is symmetric ($\mathbf{C}_{ij} = \mathbf{C}_{ji}$), $\mathbf{J}$ is the Jacobian matrix with

elements $\mathbf{J}_{ij} = \partial(\mathbf{S}\cdot\vec{f})_i/\partial\langle n_j\rangle$, and $\mathbf{D} = \mathbf{S}\cdot\mathbf{Diag}(\vec{f})\cdot\mathbf{S}^T$ is a diffusion matrix, where $\mathbf{Diag}(\vec{f})$ is a diagonal matrix whose elements are the entries in the rate function vector $\vec{f}$. The nonzero elements of $\mathbf{J}$ are given by

$$\begin{aligned}
\mathbf{J}_{11} &= -(s_u + s_b), & & \\
\mathbf{J}_{21} &= r, & \mathbf{J}_{22} &= -(k+d), & \\
\mathbf{J}_{i,i-1} &= k, & \mathbf{J}_{ii} &= -(k+d) & \text{for } i = 3,\ldots,L+1,\\
\mathbf{J}_{L+2,L+1} &= k, & \mathbf{J}_{L+2,L+2} &= -d_m, &
\end{aligned} \tag{6}$$

while the nonzero elements $\mathbf{D}_i$ read

$$\begin{aligned}
\mathbf{D}_{11} &= s_b\langle n_0\rangle + s_u(1-\langle n_0\rangle), & & & \\
\mathbf{D}_{22} &= r\langle n_0\rangle + (k+d)\langle n_1\rangle, & \mathbf{D}_{23} &= -k\langle n_1\rangle, & \\
\mathbf{D}_{i,i-1} &= -k\langle n_{i-2}\rangle, & \mathbf{D}_{ii} &= k\langle n_{i-1}\rangle + (k+d)\langle n_i\rangle & \text{for } i = 3,\ldots,L+1,\\
\mathbf{D}_{i,i+1} &= -k\langle n_{i-1}\rangle & & & \text{for } i = 3,\ldots,L,\\
\mathbf{D}_{L+2,L+1} &= -k\langle n_L\rangle, & \mathbf{D}_{L+2,L+2} &= k\langle n_L\rangle + d_m\langle n\rangle. &
\end{aligned} \tag{7}$$

Given the structure of the matrices $\mathbf{J}$ and $\mathbf{D}$ above, the Lyapunov Eq. (5) can be solved explicitly for the covariance matrix $\mathbf{C}$ whose elements are given by Eq. (4). The solution by induction is involved and can be found in 'Appendix B', which proves Proposition 2.

2.2.1 Simplification in Bursty and Constitutive Limits

Bursty limit: We now consider a particular parameter regime—the limit of large initiation rate r and large gene inactivation rate s_b such that $b = r/s_b$ is constant. Since the fraction of time spent in the active state is η, it follows that the gene is mostly in the inactive state in that limit. During the short periods of time when it transitions to the active state, a burst of initiation events occur; in particular, a mean number b of RNAPs bind to the promoter during activation. Hence, such genes are often termed bursty, since transcription proceeds via sporadic bursts of activity and b is called the mean transcriptional burst size. For r and s_b large with b constant, the expressions for the first two moments of RNAP at every gene segment and of mature RNA from Eqs. (2) and (4), respectively, simplify to

$$\langle n_i\rangle_b = b\upsilon_k\mu^i, \tag{8a}$$

$$\langle n\rangle_b = b\upsilon_m\mu^L, \tag{8b}$$

$$\mathrm{Cov}(n_i,n_j)_b = \delta_{ij}\langle n_i\rangle_b + \langle n_i\rangle_b\langle n_j\rangle_b(\upsilon_k\mu)^{-1}\cdot h_{ij}, \quad \text{where } h_{ij} = \frac{1}{2^{i+j-2}}\frac{\Gamma(i+j-1)}{\Gamma(i)\Gamma(j)}, \tag{8c}$$

$$\mathrm{Cov}(n_i,n)_b = \langle n_i\rangle_b\langle n\rangle_b(\upsilon_k\mu)^{-1}\cdot h_{iM}, \quad \text{where } h_{iM} = (1-\gamma)\sum_{q=1}^{i}\gamma^{i-q}\cdot h_{qL} \tag{8d}$$

$$\mathrm{Var}(n)_b = \langle n\rangle_b + \langle n\rangle_b^2(\upsilon_k\mu)^{-1}\cdot h_{MM}, \qquad \text{where } h_{MM} = h_{LM}; \tag{8e}$$

here, the subscript b denotes the moments in the bursty limit. Moreover, $\upsilon_k = s_u/k$, $\upsilon_m = s_u/d_m$, and $h_{ij} = f_{ij}|_{\alpha\to 0}$ denotes the simplified function f_{ij} in the limit of $\alpha\longrightarrow 0$, which is achieved when $s_b \to \infty$. We note that the above expressions for the functions h_{ij} are derived from the expressions for f_{ij} that are given in Eq. (B.33), rather than from those in Eq. (4d). The reason is that, in the bursty limit, we have that $\frac{1}{2\alpha}\to\infty$, in which case the identity in Eq. (B.36) does not hold. The bursty limit in Eq. (B.33) is simply taken by collecting terms that are not dependent on α, since $\alpha\longrightarrow 0$ in that limit.

To test the accuracy of our theory, in Fig. 2 we compare our analytical expressions for the mean of local RNAP numbers, as well as for various measures of local RNAP fluctuations—the coefficient of variation CV, the Fano factor FF, and the Pearson correlation coefficient CC—with those calculated from stochastic simulation using Gillespie's algorithm (SSA) (Gillespie 1977). Simulations are performed for two different scenarios: (i) without volume exclusion, where the footprint of RNAPs is not taken into account; and (ii) with volume exclusion, where RNAPs are treated as solid objects with a footprint of 35 bp, which is the value reported in Md Zulfikar et al. (2020). For our simulations in Fig. 2, we use parameter values characteristic for the gene PDR5 of length 3070 bp, as reported in Zenklusen et al. (2008). Our choice of $L = 30$ implies that the length of each gene segment is about 100 bp and, hence, that at most 3 RNAPs can fit in each segment when volume exclusion is taken into account. In this case, Gillespie's algorithm is modified such that the initiation and RNAP 'hopping' rates are proportional to the available volume in the gene segment which the RNAP is moving to. That is achieved by rescaling the transcription initiation rate as $r\mapsto r(1-n_1/3)$ and the RNAP hopping rate from the ith to the $(i+1)$th gene segment as $k\mapsto k(1-n_{i+1}/3)$. Since we use parameters measured for a gene that demonstrates bursty expression (PDR5) (Zenklusen et al. 2008), we test the accuracy of both the exact theory from Eqs. (2) and (4) and the approximate expressions given in Eq. (8).

The perfect agreement between our exact theory (solid lines) and simulation without volume exclusion (dots) provides a numerical validation of that theory. Our approximate theory (dashed lines) also yields a reasonably good approximation; the mismatch can be decreased if the degree of burstiness is increased, i.e. by increasing the parameters r and s_b relative to the other rates in the model. We also note that the theory is in good agreement with simulation with volume exclusion (open circles), which shows that the 'low traffic' assumption upon which our theory is based is valid.

The following interesting observations can be made from these figures: (i) if the rate of premature detachment is greater than zero, then the mean of local RNAP decreases monotonically with the distance i from the promoter according to a power law, whereas that mean is constant along the gene if there is no premature detachment, as expected; (ii) the size of RNAP fluctuations, as measured by CV, decreases with i for small premature detachment rates, but increases with i for sufficiently large values of the detachment rate; (iii) the Fano factor approaches 1—the value of FF for a Poissonian distribution—as i increases, which is due to the dispersal of the burst as stochastic elongation proceeds; (iv) the correlation coefficient between the local RNAP on two

neighbouring gene segments decreases monotonically with i, which is exacerbated by premature detachment and is a direct result of the stochasticity inherent in the elongation process.

The observation in (iii) can be explained in detail as follows. When the detachment rate is zero, a burst of RNAPs rapidly bind to the promoter, leading to large fluctuations near that site; however, thereafter each RNAP moves distinctly from all others due to stochastic elongation. Hence, the burst is gradually dispersed as elongation proceeds, which implies a decrease in the variance of fluctuations with increasing i. When the detachment rate is nonzero, then the same effect is at play; however, the increase in the variance of fluctuations along the gene is now counteracted by the decrease in mean RNAP numbers, which leads to two types of behaviour: for small i, CV decreases with i, since the variance dominates over the mean, while for large i, the opposite occurs and CV increases with i.

Constitutive limit: The other common parameter regime is that of constitutive gene expression, where the gene spends most of its time in the active state and transcription is continuous, which corresponds to the limit of very small s_b. In that limit, the expressions from Eqs. (2) and (4) simplify to

$$\langle n_i \rangle_c = \mathrm{Var}(n_i)_c = \rho_k \mu^i \qquad \text{and} \qquad \langle n \rangle_c = \mathrm{Var}(n)_c = b\rho\mu^L, \tag{9}$$

while the covariances $\mathrm{Cov}(n_i, n_j)_c$ and $\mathrm{Cov}(n_i, n)_c$ between the species are zero; here, the subscript c denotes the constitutive limit. This drastic simplification reflects the fact that, in the constitutive limit, the distributions of mature RNA and local RNAP are Poissonian: as the regulatory network is effectively given by $\emptyset \to P_1 \to P_2 \to \ldots \to P_L \to M \to \emptyset$ then, the result follows directly from the exact solution provided in Jahnke and Huisinga (2007).

To further test the accuracy of our theory, in Fig. 3 we compare our analytical expressions for the mean of local RNAP numbers, as well as for various measures of local RNAP fluctuations, with those calculated from stochastic simulation using Gillespie's algorithm, where we use parameters measured for a gene that demonstrates constitutive expression (DOA1) (Zenklusen et al. 2008). As before, we test the accuracy of both the exact theory given by Eqs. (2) and (4) and the approximate expressions from Eq. (9). Unsurprisingly, we observe agreement between exact theory (solid lines) and simulation (dots); the mismatch between our approximate theory and simulation is due to the fact that the gene does not spend 100% of its time in the active state—the true constitutive limit—but, rather, $s_u/(s_u + s_b) \approx 85\%$. The local mean RNAP number decreases with distance from the promoter, as was the case for bursty expression in the previous subsubsection, which is to be expected. The various measures which depend on the second moments are, however, considerably different: CV increases monotonically with i, independently of the rate of premature detachment, while FF and CC are very close to 1 and zero, respectively; moreover, the latter two measures practically show very little variation along the gene. The lack of transcriptional bursting explains all these effects in a straightforward fashion.

Finally, we remark that the accuracy of our expressions for the mean and variance of mature RNA, as given in Eq. (2) and (4), is verified by simulation (SSA) in Fig. 4a,

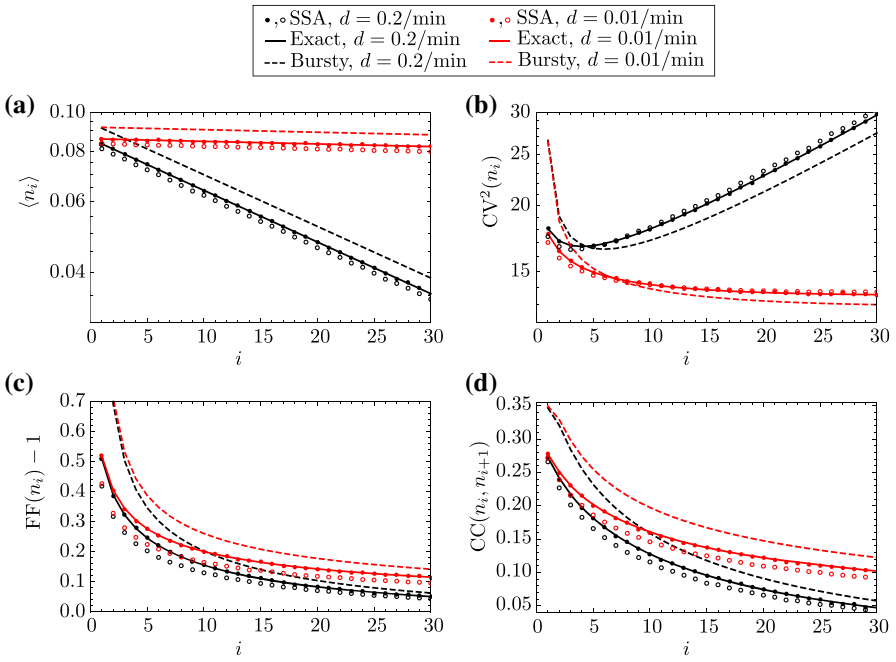


Fig. 2 (Color Figure Online) First and second moments of the distribution of local RNAP for the PDR5 gene in yeast, which demonstrates bursty expression. In **a–d**, we show the dependence of the mean, coefficient of variation squared, Fano factor, and Pearson correlation coefficient, respectively, of local RNAP fluctuations on gene segment i, as predicted by our exact theory (Eqs. (2), (4); solid lines), the approximate theory in the bursty limit (Eq. (8); dashed lines), and simulation via Gillespie's stochastic simulation algorithm (SSA), respectively. We performed simulations for two different cases: without volume exclusion (dots) and with volume exclusion (open circles). The parameters are fixed to s_u=0.44/min, s_b=4.7/min, and r=6.7/min, which are characteristic of the PDR5 gene in yeast, as reported in Supplemental Table 2 of Zenklusen et al. (2008). The number of gene segments is arbitrarily chosen to be $L = 30$. The total elongation time $\langle T \rangle = 4.5$ min is also reported for PDR5, described as the synthesis time and denoted by τ in Zenklusen et al. (2008). The elongation rate by definition takes the value of the ratio $k = L/\langle T \rangle - d \approx L/\langle T \rangle$, since $d \ll k$. The detachment rate d is arbitrarily chosen to be $d = 0.01$/min (red lines and dots) or $d = 0.2$/min (black lines and dots). Note that, for the SSA, moments are calculated from one long trajectory with a few million time points, sampled at unit intervals (Color figure online)

b for parameters typical of the bursty PDR5 gene. The meaning of the dependence of descriptive statistics on L is discussed in the next section.

2.3 Closed-Form Expressions for Moments of Total RNAP

While local RNAP fluctuations are measurable in experiment, as discussed in the Introduction, measurements of total RNAP on a gene are typically reported. Hence, in this section, we briefly discuss descriptive statistics of total RNAP fluctuations.

Recalling that n_i is the number of RNAP molecules on the ith gene segment, the total number of RNAPs on the gene—arbitrarily divided into L segments—is given by $n_{\mathrm{tot}} = \sum_{i=1}^{L} n_i$. Given Eq. (2) and (4), the steady-state mean $\langle n_{\mathrm{tot}} \rangle = \sum_{i=1}^{L} \langle n_i \rangle$

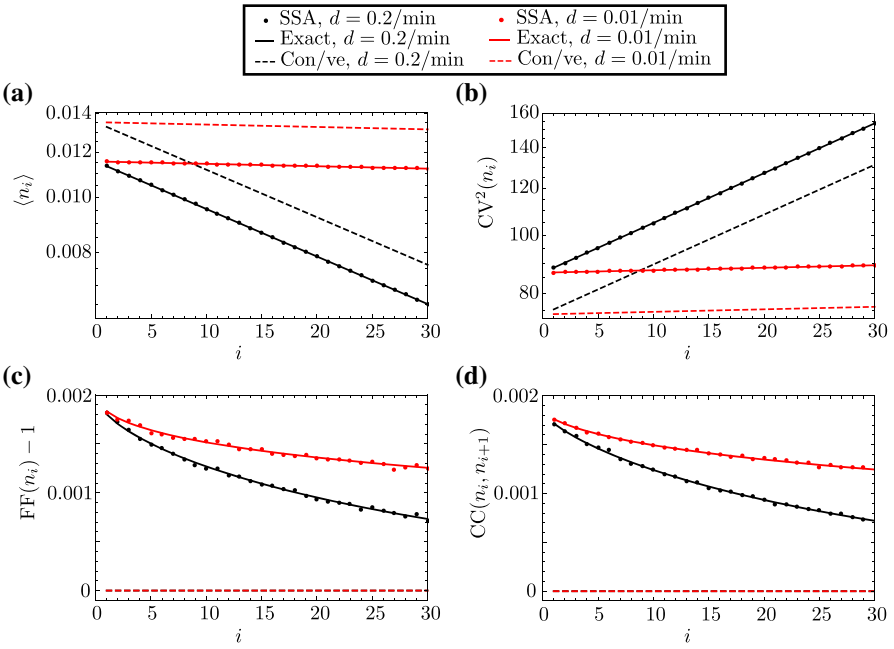


Fig. 3 (Color Figure Online) First and second moments of the distribution of local RNAP for the DOA1 gene in yeast, which demonstrates constitutive expression. In **a**–**d**, we show the dependence of the mean, coefficient of variation squared, Fano factor, and Pearson correlation coefficient, respectively, of local RNAP fluctuations on gene segment i, as predicted by our exact theory (Eqs. (2) and (4); solid lines), the approximate theory in the constitutive limit (Eq. (9); dashed lines), and simulation via Gillespie's stochastic simulation algorithm (SSA; dots), respectively. The parameters are fixed to $s_u = 0.7$/min, $s_b = 0.12$/min and $r = 0.14$/min, which are characteristic of the DOA1 gene in yeast, as reported in Supplemental Table 2 of Zenklusen et al. (2008). The number of gene segments is arbitrarily chosen to be $L = 30$. The total elongation time $\langle T \rangle = 2.9$ min is also reported for DOA1, described as the synthesis time and denoted by τ in Zenklusen et al. (2008). The elongation rate by definition takes the value of the ratio $k = L/\langle T \rangle - d \approx L/\langle T \rangle$, since $d \ll k$. The detachment rate d is arbitrarily chosen to be $d = 0.01$/min (red lines and dots) or $d = 0.2$/min (black lines and dots). Note that, for the SSA, moments are calculated from one long trajectory with a few billion time points, sampled at unit intervals (Color figure online)

and the steady-state variance $\mathrm{Var}(n_{\mathrm{tot}}) = \sum_{i,j=1}^{L} \mathrm{Cov}(n_i, n_j)$ of the total RNAP distribution are given by

$$\langle n_{\mathrm{tot}} \rangle = \eta \rho_k \mu \frac{\mu^L - 1}{\mu - 1} \quad \text{and} \quad \mathrm{Var}(n_{\mathrm{tot}}) = \langle n_{\mathrm{tot}} \rangle + \alpha\beta(\eta\rho_k)^2 \sum_{i,j=1}^{L} \mu^{i+j} \cdot f_{ij}. \tag{10}$$

For a detailed derivation of the variance in Eq. (10), we refer to 'Appendix C'. These expressions for the mean and variance of the total RNAP distribution simplify in the bursty and constitutive limits, as can be seen in 'Appendix D'. The accuracy of Eq. (10) is tested by comparing against stochastic simulation with SSA in Fig. 4c, d. Both mean and variance are seen to increase monotonically with the number of

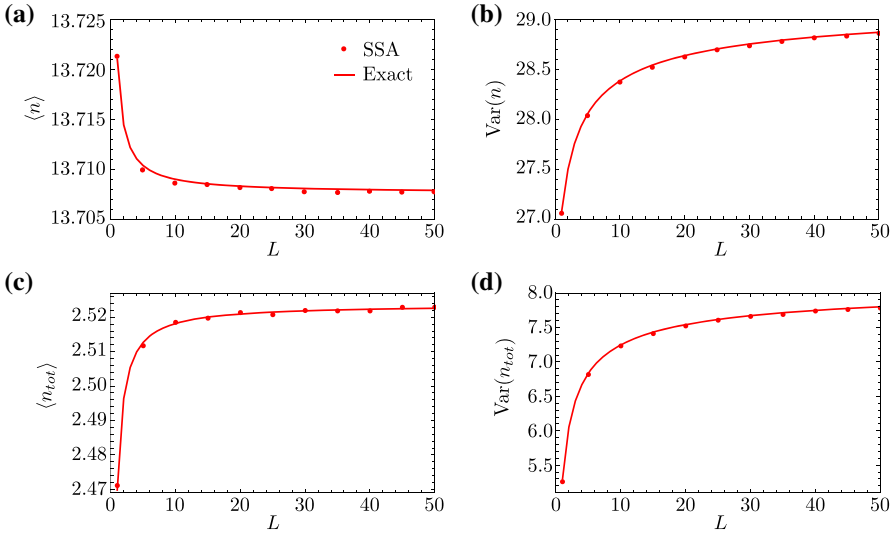


Fig. 4 Mean and variance of the distributions of mature RNA and total RNAP for the PDR5 gene in yeast. In **a**, **b**, we show the dependence of the moments of mature RNA fluctuations on the number of gene segments L, as predicted by our theory (Eqs. (2) and (4); solid lines) and SSA (dots). In **c**, **d**, we show the dependence of the moments of total RNAP on L, as predicted by our exact theory (Eq. (10); solid lines) and SSA (dots). The parameters s_u, s_b, r, and $\langle T \rangle$ are characteristic of the PDR5 gene and are the same as in Fig. 2. The premature detachment rate is chosen to be $d = 0.01$/min; the elongation rate is then given by $k \approx L/\langle T \rangle$. The degradation rate of mature RNA is $d_m = 0.04$/min, which is chosen such that the mean mature RNA is roughly consistent with that reported in Fig. 6(b) of Zenklusen et al. (2008). Note that, for the SSA, moments are calculated from one long trajectory with a few billion time points, sampled at unit intervals

gene segments L, as we keep the mean elongation time constant; the mean shows very little dependence on L, while the dependence of the variance is more pronounced. We recall that, while the parameter L is arbitrary in principle, it actually determines the size of fluctuations in the elongation time. Since that time is the sum of L independent exponential variables with mean $1/(k+d)$ each, it follows that the distribution of the elongation time T is Erlang with mean $\langle T \rangle = L/(k+d)$ and coefficient of variation squared equal to $1/L$. Hence, the larger L is, the narrower is the distribution of T and the more deterministic is elongation itself. Thus, Fig. 4c, d predicts that the mean and variance of total RNAP increase rapidly with decreasing fluctuations in the elongation time T. It hence follows that models in which the elongation rate is assumed to be exponentially distributed (Cao and Grima 2020), which correspond to the case where $L = 1$ in our model, underestimate the size of nascent RNA fluctuations.

2.4 Special Case of Deterministic Elongation

Next, we derive expressions for the descriptive statistics of total RNAP and mature RNA in the limit of large L taken at constant mean elongation time, which corresponds to deterministic elongation. As is shown in Fig. 4, these statistics converge quickly to

the ones obtained in the large-L limit; hence, the resulting limiting expressions are likely to be useful across a variety of genes.

Moments of total RNAP distribution: We define the non-dimensional parameters $\delta_g = \tau_g/\tau_d$, $T_g = \langle T\rangle/\tau_g$, and $T_d = \langle T\rangle/\tau_d$, which correspond to the ratio of the gene timescale and the polymerase detachment timescale, the ratio of the mean elongation time and the gene timescale, and the ratio of the mean elongation time and the polymerase detachment timescale, respectively; here, $\tau_d = 1/d$, as before. Substituting $k \mapsto L/\langle T\rangle - d$ into Eq. (10) and taking the limit of deterministic elongation, i.e. letting $L \to \infty$ at constant $\langle T\rangle$, we obtain the following expressions for the mean, variance, and CV^2 of total RNAP:

$$\begin{aligned}
\langle n_{\mathrm{tot}}\rangle_\infty &= \eta\frac{r}{d}(1-\mathrm{e}^{-T_d}),\\
\mathrm{Var}(n_{\mathrm{tot}})_\infty &= \langle n_{\mathrm{tot}}\rangle_\infty + \langle n_{\mathrm{tot}}\rangle_\infty^2\cdot\beta\delta_g\frac{(\delta_g-1)+(\delta_g+1)\mathrm{e}^{-2T_d}-2\delta_g\mathrm{e}^{-T_g}\mathrm{e}^{-T_d}}{(\delta_g-1)(\delta_g+1)(1-\mathrm{e}^{-T_d})^2},\\
\mathrm{CV}^2(n_{\mathrm{tot}})_\infty &= \langle n_{\mathrm{tot}}\rangle_\infty^{-1} + \beta\delta_g\frac{(\delta_g-1)+(\delta_g+1)\mathrm{e}^{-2T_d}-2\delta_g\mathrm{e}^{-T_g}\mathrm{e}^{-T_d}}{(\delta_g-1)(\delta_g+1)(1-\mathrm{e}^{-T_d})^2}.
\end{aligned} \tag{11}$$

Here, the subscript ∞ denotes the limit of $L \to \infty$. A detailed derivation of the variance in Eq. (11) can be found in Lemma C.1 of 'Appendix C'.

In the special case when RNAP does not prematurely detach from the gene, i.e. for $d = 0$, the expressions in Eq. (11) simplify to

$$\begin{aligned}
\langle n_{\mathrm{tot}}\rangle_{(\infty;0)} &= \eta r\langle T\rangle,\\
\mathrm{Var}(n_{\mathrm{tot}})_{(\infty;0)} &= \langle n_{\mathrm{tot}}\rangle_{(\infty;0)} + \langle n_{\mathrm{tot}}\rangle_{(\infty;0)}^2\cdot 2\beta T_g^{-1}\big(1-T_g^{-1}+T_g^{-1}\mathrm{e}^{-T_g}\big),\\
\mathrm{CV}^2_{(\infty;0)} &= \langle n_{\mathrm{tot}}\rangle_{(\infty;0)}^{-1} + 2\beta T_g^{-1}\big(1-T_g^{-1}+T_g^{-1}\mathrm{e}^{-T_g}\big),
\end{aligned} \tag{12}$$

where the subscript $(\infty; 0)$ denotes the limit of $(L, d) \to (\infty, 0)$. The expressions in Eq. (12) have been previously reported in Choubey et al. (2015), where they were derived using queuing theory. Hence, our expressions in Eq. (11) constitute a generalisation of known results, by further taking into account premature detachment of RNAP from the gene.

Equation (12) shows that the coefficient of variation squared of total RNAP, denoted by $\mathrm{CV}^2_{(\infty;0)}$, can be written as the sum of two terms: (i) the inverse of the mean which is expected if the distribution of total RNAP is Poissonian, and (ii) a term that increases with increasing β and decreasing T_g. Hence, the latter term provides a measure for the deviation of the total RNAP distribution from a Poissonian. In particular, it shows that the deviation is significant in genes for which (i) the fraction of time spent in the inactive state is large (large β), and (ii) the elongation time is much shorter than the switching time between the active and inactive states (small T_g).

Moments of mature RNA distribution: Similarly, in the limit of deterministic elongation, it is straightforward to show that the expressions for the mean and variance of

the distribution of mature RNA given by Eqs. (2) and (4) reduce to

$$\langle n\rangle_\infty = \eta\rho \mathrm{e}^{-T_d} \quad \text{and} \quad \mathrm{Var}(n)_\infty = \langle n\rangle_\infty + \langle n\rangle_\infty^2 \cdot \beta\theta. \tag{13}$$

These expressions can be further simplified in the special case of no premature detachment to read

$$\langle n\rangle_{(\infty;0)} = \eta\rho \quad \text{and} \quad \mathrm{Var}(n)_{(\infty;0)} = \langle n\rangle_{(\infty;0)} + \langle n\rangle_{(\infty;0)}^2 \cdot \beta\theta. \tag{14}$$

Note that the mean and variance are precisely the same as would be obtained from the telegraph model, for which the corresponding Fano factor in the bursty limit is given by Eq. (16) below. Hence, we anticipate that, in the limit of no premature detachment and deterministic elongation, the distribution of mature RNA from our transcription model is the same as the distribution obtained from the coarser telegraph model. A formal proof of that claim will be given in Sect. 3.

Relationship between Fano factors of total RNAP and mature RNA: Specifying to the case of no premature detachment, it is interesting to note that in the bursty limit, i.e. for $r, s_b \to \infty$ at constant mean burst size $b = r/s_b$ in Eq. (12), the Fano factor of total RNAP is given by

$$\mathrm{FF}_{n(b;\infty;0)} = 1 + 2b; \tag{15}$$

see also Eq. (D.3) in 'Appendix D'. Here, the subscript n denotes nascent RNA (total RNAP). Eq. (15) is in contrast to the Fano factor of mature RNA in the same bursty limit:

$$\mathrm{FF}_{m(b;\infty;0)} = 1 + b, \tag{16}$$

see Eq. (D.8) in 'Appendix D', where the subscript m denotes mature RNA. (Note that $\mathrm{FF}_{m(b;\infty;0)}$ also equals the Fano factor of the telegraph model in the same bursty limit (Raj et al. 2006).) Hence, by comparing Eqs. (15) and (16), we can deduce the following for bursty expression: (i) if the telegraph model is used to estimate the mean transcriptional burst size from total RNAP data where the elongation time is deterministic, then the mean burst size will be overestimated by a factor of two—in other words, the implicit assumption that the elongation time is exponentially distributed is inadequate; (ii) fluctuations in total RNAP (nascent RNA) deviate more from Poisson statistics, for which the Fano factor equals one, than fluctuations in mature RNA.

More generally, if we do not enforce the bursty limit, then we find the following relationship between the Fano factors of total RNAP and mature RNA, which are calculated from Eqs. (12) and (14), respectively:

$$\frac{\mathrm{FF}_{n(\infty;0)}}{\mathrm{FF}_{m(\infty;0)}} = 1 + \frac{\mathrm{e}^{-T_g} T_r T_{s_b} \Xi}{T_g^2\left[T_r T_{s_b} + T_g(T_g + T_m)\right]}. \tag{17}$$

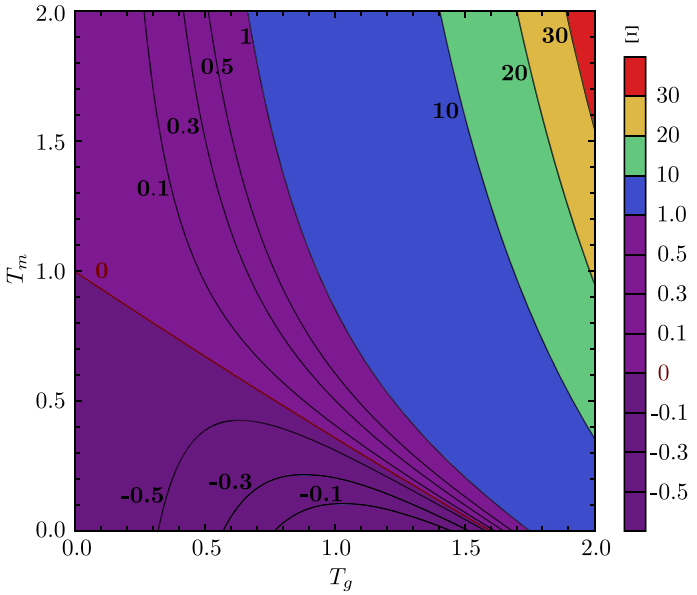


Fig. 5 Comparison between the Fano factors of nascent and mature RNA. Contour plot showing the variation of Ξ—a measure of the difference between the two Fano factors which is defined in Eq. (18)—with the non-dimensional parameters T_g and T_m which denote the ratio of the mean elongation time to the timescales of promoter switching and decay of mature RNA, respectively. As can be appreciated from Eq. (17), Ξ is positive if the Fano factor of nascent RNA is larger than that of mature RNA and negative if the reverse is true. The line $T_m \approx 1 - \frac{5}{8}T_g$, where $\Xi = 0$, shows where the two Fano factors are identical

Here,

$$\Xi = 2(T_g + T_m) + \mathrm{e}^{T_g}[2(T_g - 1)T_m + (T_g - 2)T_g], \tag{18}$$

while $T_g = (s_u + s_b)\langle T\rangle$, $T_r = r\langle T\rangle$, $T_m = d_m\langle T\rangle$, and $T_{s_b} = s_b\langle T\rangle$ are non-dimensional parameters representing the ratio of the mean elongation time to the timescales of promoter switching, initiation, decay of mature RNA, and gene deactivation, respectively. From Eq. (17), we deduce that $\mathrm{FF}_{\mathrm{n}(\infty;0)} > \mathrm{FF}_{\mathrm{m}(\infty;0)}$ if and only if $\Xi > 0$. From the contour plot of Ξ in Fig. 5, one can deduce that

$$\Xi > 0 \quad \text{if and only if} \quad T_m \gtrapprox 1 - \frac{5}{8}T_g. \tag{19}$$

Hence, the Fano factor of nascent RNA is larger than that of mature RNA if and only if the above (approximate) condition is satisfied. In the bursty limit, $T_g \to \infty$ due to $s_b \to \infty$ which, together with $T_m > 0$, implies that Eq. (19) holds; the condition is also satisfied if promoter switching is very fast compared to elongation. By contrast, if $T_m < 1$ and $T_g < 1$, then it is possible to have the opposite scenario where the Fano factor of mature RNA is larger than that of nascent RNA, which occurs, for example, if promoter switching and mature RNA decay are very slow compared to elongation.

Sensitivity of coefficient of variation of total RNAP and mature RNA: Since we have found explicit expressions for the first two moments of the distributions of total RNAP and of mature RNA, we can now estimate the sensitivity of the noise in each of those to small perturbations in the transcriptional parameters. Specifically, we calculate the logarithmic sensitivity (LS), which is also known as the relativity sensitivity, of the coefficient of variation (CV) to a parameter s, which is defined as $\Lambda_s = (s/\mathrm{CV})(\partial \mathrm{CV}/\partial s)$. (That definition implies that a 1% change in the value of the parameter s results in a change of Λ_s% in CV.)

In Table 1b, we report the logarithmic sensitivity of the coefficient of variation of total RNAP fluctuations, which is obtained from Eq. (12), to perturbations in the parameters s_u, s_b, r, and $\langle T \rangle$. Similarly, in Table 1c, we report the logarithmic sensitivity of the coefficient of variation of mature RNA fluctuations from Eq. (14) to perturbations in the parameters s_u, s_b, r, and d_m. In both cases, these sensitivities are calculated for parameter values estimated for five genes in yeast, as reported in Zenklusen et al. (2008); see Table 1a.

The following observations can be made regarding the sensitivity of the noise in total RNAP fluctuations: (i) for the two genes PDR5 and POL1 which spend most of their time in the inactive state due to $s_b \gg s_u$, CV is most sensitive to changes in the parameters s_u and $\langle T \rangle$; (ii) for the genes DOA1, MDN1, and KAP104 which spend most of their time in the active state due to $s_u \gg s_b$, CV is most sensitive to changes in the parameters r and $\langle T \rangle$; (iii) the size of mature RNA fluctuations is found to be most sensitive to perturbations in s_u and d_m for PDR5 and POL1, and to perturbations in r and d_m for the other three genes. We furthermore note that for both total RNAP and mature RNA, r is the least sensitive parameter for the genes which are mostly inactive, whereas it is among the most sensitive parameters for genes that are mostly active.

3 Approximate Distributions of Total RNAP and Mature RNA

Thus far, we have derived expressions for the first two moments of the distributions of total RNAP and mature RNA. Naturally, it would also be useful to derive closed-form expressions for the distributions themselves; such a derivation is, however, analytically intractable in general (Jahnke and Huisinga 2007) due to the presence of the catalytic reaction $G_{\mathrm{on}} \rightarrow G_{\mathrm{on}} + P_1$, which models initiation of the transcription process. Still, there are two special cases where analytical distributions are known: (i) when the elongation time is considered to be fixed, which corresponds to our model with $L \rightarrow \infty$ at constant $\langle T \rangle$ (Heng et al. 2016); (ii) when the elongation time is exponentially distributed, corresponding to our model with $L = 1$, in which case the distribution of total RNAP is identical to the one which is derived from the telegraph model (Peccoud and Ycart 1995; Raj et al. 2006). While one may argue that the analytical distribution of RNAP for deterministic elongation times may well approximate the stochastic (finite-L) case, the issue remains that the exact solution is not given in terms of simple functions unless promoter switching is slow compared to initiation, elongation, and termination, in which case the solution reduces to a weighted sum of two Poisson distributions (Heng et al. 2016). Hence, it is generally very difficult to apply in practice,

Table 1 Logarithmic sensitivity (LS) of the coefficient of variation CV of total RNAP and mature RNA fluctuations for five genes in yeast; see Sect. 2.4 for a discussion. (a) Parameter values from Supplemental Tables 2 and 4 in Zenklusen et al. (2008). The degradation rate d_m of mature mRNA is estimated from the reported mean number of mature RNA, the parameters s_u, s_b, r, and Eq. (14) for the mean. (b) Logarithmic sensitivity of CV of total RNAP fluctuations. (c) Logarithmic sensitivity of CV of mature mRNA fluctuations. The most sensitive parameter and the next most sensitive one are marked in dark bold and italic, respectively

	PDR5	POL1	DOA1	MDN1	KAP104
(a)					
Mean mature RNA #	13.40	3.13	2.59	6.12	4.93
$\langle T \rangle$(min)	4.50	3.75	2.90	16.75	3.50
$s_u(\text{min}^{-1})$	0.44	0.07	0.70	0.70	0.70
$s_b(\text{min}^{-1})$	4.70	0.68	0.12	0.12	0.12
$r(\text{min}^{-1})$	6.70	2.00	0.14	0.19	0.27
$d_m(\text{min}^{-1})$	0.04	0.06	0.05	0.03	0.05
LS	PDR5	POL1	DOA1	MDN1	KAP104
(b)					
Λ_{s_u}	**− 0.52**	**− 0.51**	− 0.09	− 0.12	− 0.11
Λ_{s_b}	0.18	0.29	0.09	0.09	0.10
Λ_r	− 0.15	− 0.12	**− 0.49**	*− 0.47*	*− 0.47*
$\Lambda_{\langle T \rangle}$	*− 0.48*	*− 0.34*	**− 0.49**	**− 0.50**	**− 0.49**
LS	PDR5	POL1	DOA1	MDN1	KAP104
(c)					
Λ_{s_u}	**− 0.50**	**− 0.52**	− 0.09	− 0.10	− 0.11
Λ_{s_b}	0.23	0.20	0.08	0.08	0.09
Λ_r	− 0.23	− 0.15	*− 0.49*	*− 0.48*	*− 0.48*
Λ_{d_m}	**0.50**	*0.47*	**0.50**	**0.50**	**0.50**

such as to infer parameters from data using a Bayesian approach. Moreover, to our knowledge, no exact solutions are known for the distribution of mature RNA in our model. In this section, we aim to devise a simple approximation for the distribution of total RNAP numbers in terms of the Negative Binomial (NB) distribution; these simple distributions have shown great flexibility in describing complex gene expression models with a large number of parameters (Cao and Grima 2020). Finally, by means of singular perturbation theory, we will obtain the distribution of mature RNA under the assumption that RNA polymerase elongation is faster than degradation of mature RNA.

3.1 Approximation of Total RNAP Distribution

We approximate the distribution of total RNAP transcribing the gene via a Negative Binomial distribution, as follows. The mean and variance of the Negative Binomial

distribution NB(q, p) are given by $pq/(1-p)$ and $pq/(1-p)^2$, respectively. By assuming that these are equal to the exact mean and variance, respectively, of the total RNAP distribution, see Eq. (10), we obtain effective values for the parameters p and q:

$$n_{\text{tot}} \sim \text{NB}(q, p) \equiv \text{NB}\left(\frac{\langle n_{\text{tot}}\rangle^2}{\text{Var}(n_{\text{tot}}) - \langle n_{\text{tot}}\rangle}, \frac{\text{Var}(n_{\text{tot}}) - \langle n_{\text{tot}}\rangle}{\text{Var}(n_{\text{tot}})}\right). \tag{20}$$

In Fig. 6, we show a comparison between the distributions of total RNAP obtained from SSA (dots) and the Negative Binomial approximation in Eq. (20) (solid lines). Our results are presented for two different values of the number of gene segments: $L = 1$ (exponentially distributed elongation time; left column) and $L = 50$ (quasi-deterministic elongation time; right column). Additionally, we rescale our gene inactivation rate as $s_b \mapsto s_b\epsilon$, and we present results for three different values of the parameter ϵ: 10^{-3}, the constitutive limit of the gene being mostly in the active state (top row); 10^{-1}, where the gene spends almost equal amounts of time in the active and inactive states, with $s_b \approx s_u$ (middle row); and 1, the bursty limit, where the gene spends most of its time in the inactive state (bottom row).

We can make several observations, as follows. For both $L = 1$ and $L = 50$, the Negative Binomial approximation performs well for bursting and constitutive expression (top and bottom rows), whereas it is appreciably poor when expression is in between those two limits (middle row). Intuitively, this observation can be explained via the following reasoning. In the limits of the gene being mostly in the active state (constitutive expression) or the inactive state (bursty expression), the distribution of total RNAP is necessarily unimodal. However, when the gene spends a considerable amount of time in each state, the distribution is the sum of two conditional distributions which can manifest either as bimodality or as a wide unimodal distribution, neither of which can be captured by a Negative Binomial distribution. Assuming bursty expression, the Negative Binomial distribution is a more accurate approximation to the distribution obtained from SSA for $L = 1$ than it is for $L = 50$; the reason is that $L = 1$ corresponds to the telegraph model (Raj et al. 2006), in which case it can be proven analytically that the distribution reduces to a Negative Binomial in the limit of bursty expression. For constitutive expression, the Negative Binomial approximation is equally good for $L = 1$ and $L = 50$, as the distribution is necessarily Poissonian then and as it is well known that a Negative Binomial distribution can approximate a Poissonian to a high degree of accuracy. In summary, our results hence indicate that Eq. (20) yields a good approximation for the total RNAP distribution of bursty and constitutively expressed genes.

We also note from Fig. 6 that the comparison between the SSA distributions for $L = 1$ and $L = 50$, with equal mean elongation times, highlights the importance of modelling elongation with the correct distribution of elongation times for genes that are non-constitutive, i.e. for $\epsilon = 10^{-1}$ or $\epsilon = 1$. In particular, if the elongation time is quasi-deterministic ($L = 50$), there appears to be a significant increase in the probability of observing zero total RNAP transcribing the gene compared to models with an exponentially distributed elongation time ($L = 1$).

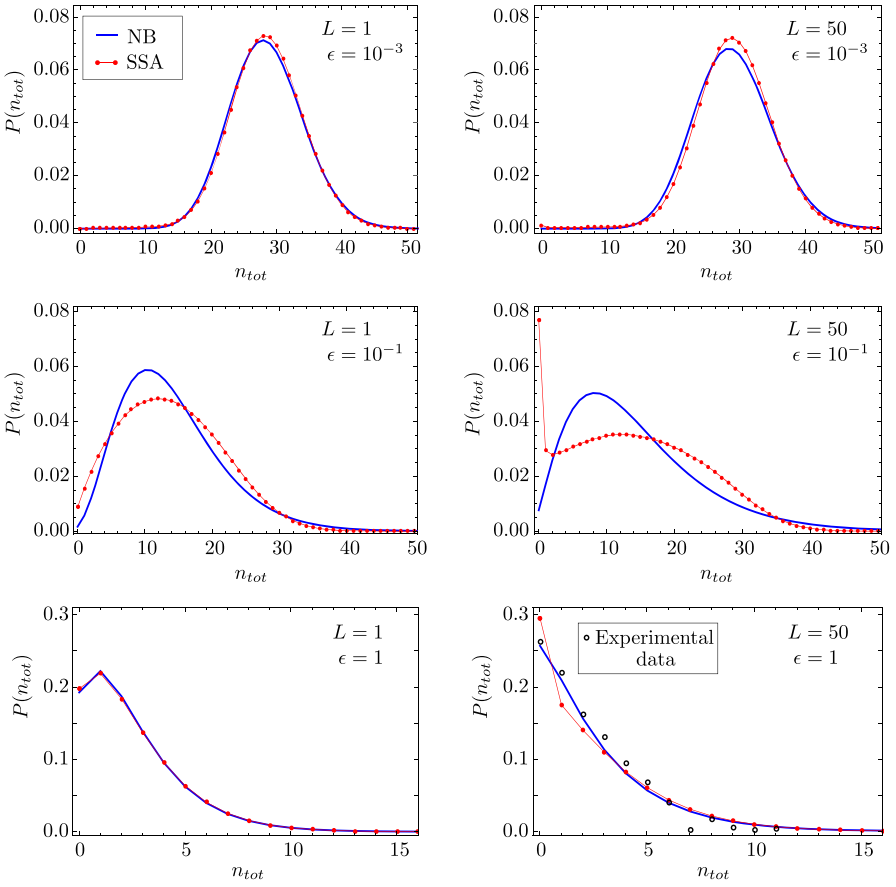


Fig. 6 (Color Figure Online) Steady-state distribution of total RNAP and its approximation by a Negative Binomial distribution. We compare the approximation from Eq. (20) (blue lines) with the distribution of total RNAP obtained from stochastic simulation (SSA; red dots). With the exception of s_b, the parameters are for the PDR5 gene in yeast and are hence the same as in Fig. 2, with $d = 0.01/\text{min}$. Results are presented for two different values of L, corresponding to an exponentially distributed elongation time ($L = 1$) and a quasi-deterministic elongation time ($L = 50$); k is rescaled such that the two have the same mean elongation time. Additionally, we rescale the gene inactivation rate via $s_b \mapsto s_b\epsilon$, where $\epsilon = 10^{-3}, 10^{-1}, 1$, corresponding to constitutive, general, and bursty expression, respectively. (Here, general expression is neither clearly constitutive nor bursty, since the gene spends roughly equal amounts of time in the inactive and active states.) Note that $\epsilon = 1$ results in a distribution of nascent RNA that is consistent with that measured for PDR5; the experimental data from Fig. 6(b) of Zenklusen et al. (2008) are plotted for comparison. The Negative Binomial approximation is found to be accurate in the limits of constitutive and bursty expression (top and bottom rows), independently of L

3.2 Approximation of Mature RNA Distribution

Next, we apply singular perturbation theory to formally derive the distribution of mature RNA when the elongation rate is much larger than the degradation rate of mature RNA.

We start by defining $P_j(\vec{n};t)$ $(j = 0, 1)$ as the probability of the state $\vec{n} = (n_1, \ldots, n_L, n)$ at time t while the gene is either active (0) or inactive (1). Note that n_i is the number of RNAPs on gene segment i for $i = 1, \ldots, L$, while n is the number of mature RNAs. The time evolution of the probabilities $P_j(\vec{n};t)$ can be described by a system of coupled CMEs:

$$\begin{aligned}
\partial_t P_0 &= s_u P_1 - s_b P_0 + r(\mathbb{E}_{n_1}^{-1} - 1)P_0 + k\sum_{i=1}^{L-1}\big(\mathbb{E}_{n_i}\mathbb{E}_{n_{i+1}}^{-1} - 1\big)n_i P_0 \\
&\quad + k\big(\mathbb{E}_{n_L}\mathbb{E}_{n}^{-1} - 1\big)n_L P_0 + d\sum_{i=1}^{L}(\mathbb{E}_{n_i} - 1)n_i P_0 + d_m(\mathbb{E}_n - 1)nP_0, \\
\partial_t P_1 &= s_b P_0 - s_u P_1 + k\sum_{i=1}^{L-1}\big(\mathbb{E}_{n_i}\mathbb{E}_{n_{i+1}}^{-1} - 1\big)n_i P_1 + k\big(\mathbb{E}_{n_L}\mathbb{E}_{n}^{-1} - 1\big)n_L P_1 \\
&\quad + d\sum_{i=1}^{L}(\mathbb{E}_{n_i} - 1)n_i P_1 + d_m(\mathbb{E}_n - 1)nP_1,
\end{aligned} \tag{21}$$

where $\mathbb{E}_{n_i}^{c}[f(\vec{n})] = f(n_1, n_2, \ldots, n_i + c, \ldots, n_L, n)$, with $c \in \mathbb{Z}$, denotes the standard step operator. We assume that the elongation rate k is faster than the degradation rate d_m of mature RNA, i.e. that $k/d_m \gg 1$. Since $k = L/\langle T\rangle - d$, it follows that in the limit of deterministic elongation ($k \to \infty$), i.e. for $L \to \infty$ at constant mean elongation time $\langle T\rangle$, the condition $k/d_m \gg 1$ is naturally satisfied.

In order to find an analytical expression for the propagator probabilities $P(\vec{n};t)$ which satisfies the system of CMEs in Eq. (21), we define the probability-generating function as $F = \sum_j F_j$, with $F_j(\vec{z};t) = \sum_{\vec{n}=\vec{0}}^{\infty} P_j(\vec{n};t)\vec{z}^{\vec{n}}$; here, $\vec{z} = (z_1, \ldots, z_L, z)$ is a vector of variables corresponding to the state $\vec{n}$. Given the equations for $P_j(\vec{n};t)$ from Eq. (21), we obtain the following systems of PDEs for the corresponding generating functions $F_j(\vec{z};t)$:

$$\begin{aligned}
\mathbb{L}[F_0] &= s_u F_1 - s_b F_0 + r(z_1 - 1)F_0, \\
\mathbb{L}[F_1] &= s_b F_0 - s_u F_1,
\end{aligned} \tag{22}$$

where

$$\mathbb{L} = \partial_t + k\sum_{i=1}^{L-1}(z_i - z_{i+1})\partial_{z_i} + k(z_L - z)\partial_{z_L} + d\sum_{i=1}^{L}(z_i - 1)\partial_{z_i} + d_m(z-1)\partial_z \tag{23}$$

is a differential operator acting on the generating functions F_0 and F_1. Eq. (22) represents a system of coupled, linear, first-order partial differential equations (PDEs). Now, we introduce the new variables $u_i = z_i - 1$ $(i = 1, \ldots, L)$ and $u = z - 1$ to

rewrite Eq. (22) as

$$\begin{aligned}\mathbb{L}[F_0] &= s_u F_1 - s_b F_0 + r u_1 F_0,\\ \mathbb{L}[F_1] &= s_b F_0 - s_u F_1;\end{aligned} \tag{24}$$

here, the operator in Eq. (23) now takes the form

$$\mathbb{L} = \partial_t + k\sum_{i=1}^{L-1}(u_i - u_{i+1})\partial_{u_i} + k(u_L - u)\partial_{u_L} + d\sum_{i=1}^{L} u_i\partial_{u_i} + d_m u\partial_u. \tag{25}$$

In order to find an analytical solution to Eq. (24), we rescale all rates and the time variable by the decay rate of mature RNA; then, we apply the method of characteristics, with s being the characteristic variable. The first characteristic equation gives $d_m(\mathrm{d}t/\mathrm{d}s) = 1$, with solution $s \equiv t' = d_m t$; hence, we can use the variable t' as the independent variable and thus convert the system of PDEs in Eq. (24) into a characteristic system of ordinary differential equations (ODEs),

$$\dot{u}_i = (k/d_m)[u_i - u_{i+1} + (d/k)u_i] \qquad \text{for } i = 1,\ldots,L-1, \tag{26a}$$
$$\dot{u}_L = (k/d_m)[u_L - u + (d/k)u_L], \tag{26b}$$
$$\dot{u} = u, \tag{26c}$$
$$\dot{F}_0 = (s_u/d_m)F_1 - (s_b/d_m)F_0 + (r/d_m)u_1 F_0, \tag{26d}$$
$$\dot{F}_1 = (s_b/d_m)F_0 - (s_u/d_m)F_1, \tag{26e}$$

where the overdot denotes differentiation with respect to t'. The existence of an integral-form solution to Eq. (26) follows from the fact that the reaction scheme in Fig. 1 contains first-order reactions only. Under the assumption that $k \gg d_m$, we define $\varepsilon = d_m/k$; then, we apply Geometric Singular Perturbation Theory (GSPT) (Fenichel 1979; Jones 1995), with $0 < \varepsilon \ll 1$ as the (small) singular perturbation parameter. We hence separate the system in Eq. (26) into fast and slow dynamics, which will allow us to find an asymptotic approximation for F_0 and F_1 in steady state. A brief introduction to GSPT can be found in 'Appendix E'. Given the above definition of ε, Eqs. (26a) and (26b), the governing equations for u_i in the 'slow system', become

$$\begin{aligned}\varepsilon\dot{u}_i &= u_i - u_{i+1} + (d/k)u_i \qquad \text{for } i = 1,\ldots,L-1,\\ \varepsilon\dot{u}_L &= u_L - u + (d/k)u_L,\end{aligned} \tag{27}$$

where u_i $(i,\ldots,L)$ are the fast variables and u, F_0, and F_1 are the slow ones. Setting $\varepsilon = 0$ in Eq. (27), we can express the variables u_i as $u_i = \mu\cdot u_{i+1}$, with $\mu = k/(k+d)$ for $i = 1,\ldots,L$. Finally, we write the variable u_1 as $u_1 = \mu^L\cdot u$. Next, given Eq. (26c), we apply the chain rule, with $\mathrm{d}t' \equiv \mathrm{d}u\cdot u$, to rewrite Eqs. (26d) and (26e) as

$$F_0' d_m u = s_u F_1 - s_b F_0 + r\mu^L u F_0, \tag{28a}$$
$$F_1' d_m u = s_b F_0 - s_u F_1, \tag{28b}$$

where the prime now denotes differentiation with respect to u. Solving Eq. (28a) for F_1 and substituting the result into Eq. (28b), we obtain the second-order ODE

$$d_m^2 u F_0'' + d_m(d_m + s_b + s_u - r\mu^L u)F_0' - r\mu^L(d_m + s_u)F_0 = 0 \tag{29}$$

for $F_0(u)$. Eq. (29) is a confluent hypergeometric differential equation (Kummer's equation) (Digital Library of Mathematical Functions 2020b) which admits the solution

$$F_0(u) = C \cdot {}_1F_1\Big(\frac{d_m + s_u}{d_m}; \frac{d_m + s_b + s_u}{d_m}; \frac{r}{d_m}\mu^L u\Big), \tag{30}$$

where ${}_1F_1$ denotes the confluent hypergeometric function; here, we consider only one of two independent fundamental solutions of Kummer's differential equation, as we are seeking a solution in steady state where the variable u is bounded. The constant C in Eq. (30) is a constant of integration that is determined from the normalisation condition on the full generating function: $F = F_0 + F_1$. From Eq. (28), one finds that F satisfies

$$F' = \frac{r}{d_m}\mu^L F_0. \tag{31}$$

Making use of Eq. (31) and applying the normalisation condition $F|_{u=0} = 1$, we find that the generating function in steady state reads

$$F(z) = {}_1F_1\Big(\frac{s_u}{d_m}; \frac{s_b + s_u}{d_m}; \frac{r}{d_m}\mu^L(z-1)\Big). \tag{32}$$

The probability distribution $P(n)$ of mature RNA can be found from the formula

$$P(n) = \frac{1}{n!}\frac{d^n}{dz^n}F(z)|_{z=0},$$

which yields the analytical expression

$$P(n) = \frac{1}{n!}\frac{(s_u)_n}{(s_b + s_u)_n}\Big(\frac{r}{d_m}\Big)^n(\mu^L)^n\,{}_1F_1\big(\tfrac{s_u}{d_m} + n; \tfrac{s_b+s_u}{d_m} + n; -\tfrac{r}{d_m}\mu^L\big), \tag{33}$$

where $(\cdot)_n$ is the Pochhammer symbol, as before. Note that the mean and variance of mature mRNA, as calculated from the distribution in Eq. (33), agree exactly with Eqs. (2c) and (4f) in the limit of fast elongation ($k \to \infty$). Note also that the solution in Eq. (33) depends on the parameter μ^L, which represents the survival probability of an RNAP molecule, i.e. the probability that RNAP will not prematurely detach from the gene. Finally, we take the limit of deterministic elongation, letting $L \to \infty$ at constant $\langle T\rangle$, which leads to

$$P(n) = \frac{1}{n!}\frac{(s_u)_n}{(s_b + s_u)_n}\Big(\frac{r}{d_m}\Big)^n \mathrm{e}^{-nd\langle T\rangle}\,{}_1F_1\big(\tfrac{s_u}{d_m} + n; \tfrac{s_b+s_u}{d_m} + n; -\tfrac{r}{d_m}\mathrm{e}^{-d\langle T\rangle}\big). \tag{34}$$

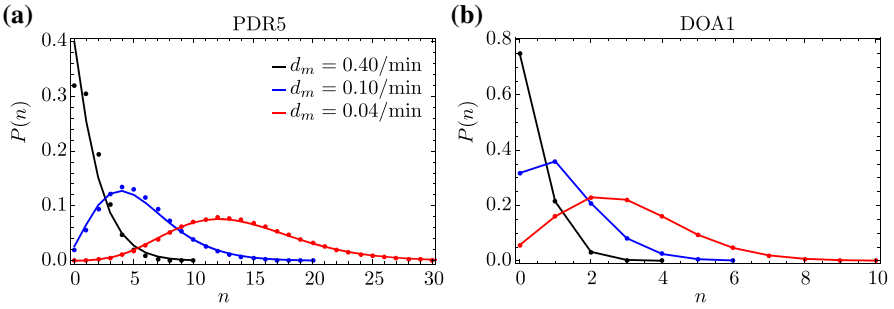


Fig. 7 Steady-state distribution of mature RNA for two different genes in yeast. We compare the distribution obtained from SSA (dots) to the perturbative approximation in Eq. (33) (solid lines) for two different genes. In **a**, we consider the PDR5 gene, fixing the parameters as in Fig. 2: $s_u = 0.44$/min, $s_b = 4.7$/min, $r = 6.7$/min, $d = 0.01$/min, and $\langle T \rangle = 4.5$ min. The degradation rate of mature RNA takes the values $d_m = 0.04, 0.10, 0.40$/min; note that the experimental value is $d_m = 0.04$/min. In **b**, we consider the DOA1 gene, fixing the parameters as in Fig. 3: $s_u = 0.7$/min, $s_b = 0.12$/min, $r = 0.14$/min, $d = 0.01$/min, and $\langle T \rangle = 2.9$ min. The degradation rate of mature RNA again takes the values $d_m = 0.04, 0.10, 0.40$/min; the experimental value is $d_m = 0.05$/min. For both genes, the agreement between SSA and our perturbative approximation increases with k/d_m, as expected, since Eq. (33) is derived under the assumption that $k \gg d_m$. Note that the distribution is practically independent of L, since Eq. (33) depends on L only through μ^L, which for small premature detachment rates d implies $\mu^L \approx 1$ for any L (Color figure online)

Note that in the limit of no premature detachment ($d = 0$), Eq. (34) is precisely equal to the distribution of mature RNA predicted by the telegraph model, which is in wide use in the literature (Raj et al. 2006). Hence, our perturbative approach can be seen as a means to formally derive the conventional telegraph model of gene expression starting from a more fundamental and microscopic model. In Fig. 7, we verify our analytical solution with stochastic simulation for two different genes in yeast. We also note that, for nonzero premature detachment rates ($d \neq 0$), Eq. (34) is the steady-state solution predicted by the telegraph model, with parameter r renormalised to $r\mathrm{e}^{-d\langle T \rangle}$; that is to be expected, as the latter is the rate at which RNAPs undergo termination, leading to mature RNAs.

4 Statistics of Fluorescent Nascent RNA Signal

Thus far, we have determined the statistics of the total number of RNAP transcribing the given gene; these are also the statistics of the number of nascent RNA molecules. However, in experiments using single-molecule fluorescence in situ hybridisation [smFISH (Heng et al. 2016)], molecule numbers of nascent RNA cannot be directly determined. Rather, the experimentally measured RNA 'abundance' is the fluorescent signal emitted by oligonucleotide probes bound to the RNA. Since the length of the nascent RNA grows as RNAP moves away from the promoter, it follows that we must account for the increase in the fluorescent signal as elongation proceeds.

In this section, we take into account these experimental details to obtain closed-form expressions for the mean and variance of the fluorescent signal of local and total nascent RNA. We assume that the signal from nascent RNA on the ith gene segment

is given by $r_i = (\nu/L) i n_i$ for $i = 1, \ldots, L$, where ν is some experimental constant; the value of the parameter $(\nu/L)i$ is increasing with i, which models the fact that the fluorescent signal becomes stronger as RNAP moves along the gene. The formula for the mean fluorescent signal at gene segment i is then given by $\langle r_i \rangle = (\nu/L) i \langle n_i \rangle$, where $\langle n_i \rangle$ follows from Eq. (2b); the covariance of two fluorescent signals along the gene, r_i and r_j $(i, j = 1, \ldots, L)$, is given by $\mathrm{Cov}(r_i, r_j) = (\nu/L)^2 ij \mathrm{Cov}(n_i, n_j)$, where $\mathrm{Cov}(n_i, n_j)$ is obtained from Eq. (4d). In Fig. 8a, b, we plot the mean and Fano factor of the local signal as a function of the gene segment i; note the contrast between the statistics of the fluorescent signal and the corresponding statistics of local RNAP—which is the statistics of nascent RNA—shown in Fig. 2a, c.

Similarly, denoting by $r_{\mathrm{tot}} = \sum_{i=1}^{L} r_i$ the total fluorescent signal across the gene, we find the following expressions for the steady-state mean $\langle r_{\mathrm{tot}} \rangle = \sum_{i=1}^{L} \langle r_i \rangle$ and the steady-state variance $\mathrm{Var}(r_{\mathrm{tot}}) = \sum_{i,j=1}^{L} \mathrm{Cov}(r_i, r_j)$:

$$\begin{aligned} \langle r_{\mathrm{tot}} \rangle &= \nu \eta \rho_k \mu \frac{\mu^L [L\mu - (L+1)] + 1}{L(\mu - 1)^2}, \\ \mathrm{Var}(r_{\mathrm{tot}}) &= \left(\frac{\nu}{L}\right)^2 \eta \rho_k \sum_{i=1}^{L} i^2 \mu^i + \left(\frac{\nu}{L}\right)^2 \alpha \beta (\eta \rho_k)^2 \sum_{i,j=1}^{L} ij \cdot \mu^{i+j} \cdot f_{ij}. \end{aligned} \tag{35}$$

For a detailed derivation of the variance in Eq. (35), see Eq. (F.1) in 'Appendix F'; see also 'Appendix G' for the corresponding expressions in the bursty, constitutive, and deterministic elongation limits. In Fig. 8c, d, we show the mean and Fano factor of the total signal as a function of the number of gene segments (L); as above, we note the contrasting difference between the statistics of the fluorescent signal and the corresponding statistics of total RNAP—which is the statistics of total nascent RNA—shown in Fig. 4c, d.

Hence, the calculation of the statistics of the number of nascent RNAs from the raw signal intensity presents a challenge and has to be approached carefully. The expressions presented above allow for the inference of transcriptional parameters from the first two moments of the fluorescent signal by means of moment-based inference techniques (Zechner et al. 2012). Quantitative information about nascent RNA can also be obtained from electron micrograph images (El Hage et al. 2010), which avoids the challenges presented by smFISH.

5 Model Extension with Pausing of RNAP

Thus far, we have studied a model where RNAPs do not pause as they move along the gene. A natural extension is provided by a modified model in which RNAPs pause along the gene at random sites and elongation is characterised by three processes: forward hopping, pausing, and unpausing of RNAP. The motivation for studying this extended model, which has recently been considered via stochastic simulation in Md Zulfikar et al. (2020), is that experiments have revealed that RNAP exhibits

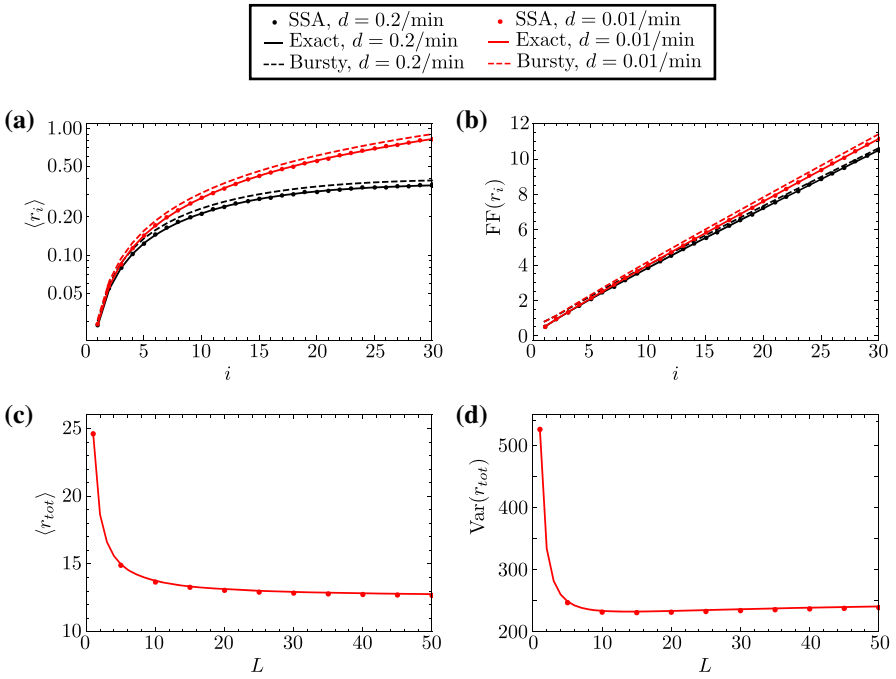


Fig. 8 First and second moments of the local and total fluorescent signal for the bursty gene PDR5 in yeast. In **a**, **b**, we show the dependence of the mean and the Fano factor of local fluorescent signal fluctuations on the gene segment i, as predicted by our exact theory (solid lines) and SSA (dots), respectively. The plots for $CV^2(r_i)$ and $CC(r_i, r_{i+1})$ are identical to those of $CV^2(n_i)$ and $CC(n_i, n_{i+1})$ in Fig. 2. The number of gene segments is arbitrarily chosen to be $L = 30$. In **c**, **d**, we show the dependence of the mean and variance of total fluorescent signal fluctuations on the number of gene segments L, as predicted by our exact theory (Eq. (35); solid lines) and SSA (dots). The parameters s_u, s_b, r, and $\langle T \rangle$ are characteristic of the PDR5 gene and take the same values as in Fig. 2, as do the rates of elongation and RNAP detachment. The value of the parameter ν is arbitrarily chosen to be $\nu = 10$

pauses of varying duration, typically on the timescale of few seconds (Forde et al. 2002; Adelman et al. 2002).

5.1 Closed-Form Expressions for Moments of Local RNAP Fluctuations

We extend the model described in Fig. 1 by assuming that the RNAP on gene segment i can switch between a non-paused (actively moving) state P_i and a paused state $\bar{P}_i$. The actively moving state P_i switches to $\bar{P}_i$ with rate r_p, while the reverse reaction occurs with rate r_a. Premature detachment from the actively moving RNAP occurs with rate d_a, whereas it occurs with rate d_p from the paused RNAP. The resulting extended model is illustrated in Fig. 9a. In 'Appendix A', we derive the mean and variance of the corresponding elongation time, which is not Erlang distributed now, as was the case for the model without pausing. Furthermore we find two interesting properties of the coefficient of variation CV_T^2 of the elongation time: (i) in the limit of large L at constant mean elongation time, CV_T^2 does not tend to zero, which implies

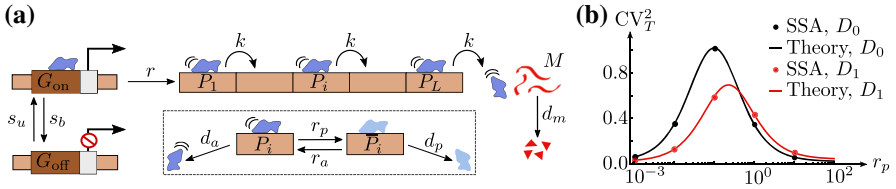


Fig. 9 Model of transcription that includes RNAP pausing. In **a**, we extend the model in Fig. 1 so that it takes into account pausing of RNAP at random segments on the gene. Pausing on gene segment i is modelled by the transition from the active state P_i to the paused state $\bar{P}_i$ with rate r_p, while the reverse ('unpausing') transition occurs with rate r_a. Premature termination of RNAP occurs with rate d_a from the actively moving state, and with rate d_p from the paused state. In **b**, we show the dependence of the coefficient of variation squared (CV_T^2) of the elongation time distribution on the pausing rate (r_p), as predicted from SSA (dots) and theory (Eq. (A.7); solid lines). Results are shown for two different parameter regimes: $D_0 \equiv \{d_a = 0/\text{min} = d_p\}$ (no premature polymerase detachment) and $D_1 \equiv \{d_a = 0.05/\text{min} = d_p\}$ (premature polymerase detachment). The remaining parameters are fixed to $L = 50$, $k = 10/\text{min}$, and $r_a = 0.1/\text{min}$

that elongation is not deterministic; (ii) for small rates of premature detachment, CV_T^2 is at its maximum when $r_p \approx r_a$, i.e. when RNAP spends roughly half of its time in the paused state. See 'Appendix A' for details and Fig. 9b for a confirmation through stochastic simulation.

Proposition 3 *Let the number of RNAP molecules in the active state P_i be denoted by n_i^a, let the number of molecules in the paused state $\bar{P}_i$ be n_i^p, and let the number of molecules of mature RNA be denoted by n. Let $\sigma = r_p/r_a$ be the ratio of the pausing and activation rates, let $\pi_{r_a} = r_a/(r_a + d_p)$ be the probability of RNAP switching to the actively moving state from the paused state, and let $\pi_{d_p} = d_p/(r_a + d_p)$ be the probability of premature RNAP detachment from the paused state. Furthermore, define the new parameters $\tilde{\mu} = k/(k + d_a + r_p\pi_{d_p})$ and $\lambda = \sigma\pi_{r_a}$.*

Then, it follows that the steady-state mean number of RNAP molecules in the active and paused states on gene segment i ($i = 1, \ldots L$) is given by

$$\langle n_i^a \rangle = \eta\rho_k\tilde{\mu}^i \quad \text{and} \quad \langle n_i^p \rangle = \langle n_i^a \rangle\lambda. \tag{36}$$

Hence, the total mean number of RNAP molecules on each gene segment i reads

$$\langle n_i \rangle = \langle n_i^a \rangle + \langle n_i^p \rangle = \langle n_i^a \rangle(1 + \lambda). \tag{37}$$

The proof of Proposition 3 can be found in 'Appendix H'. Note that in the limit of no pausing, i.e. for $r_p = 0$, Eq. (37) reduces to the expression for the mean of RNAP reported in Eq. (2b).

Proposition 4 *Let $\tau_{r_a} = 1/r_a$ be the timescale of RNAP activation from the paused state, let $\tau_{d_p} = 1/d_p$ be the timescale of premature termination of paused RNAP, let $\tau_p = 1/(k + d_a)$ be the typical time that an actively moving RNAP spends on a gene segment, and let $\tau_{pp} = 1/(r_a + d_p)$ be the typical time spent in the paused state. Furthermore, define the new parameters $\lambda_{r_p} = \pi_{r_p}/(1 - \pi_{r_p})$, where $\pi_{r_p} =$*

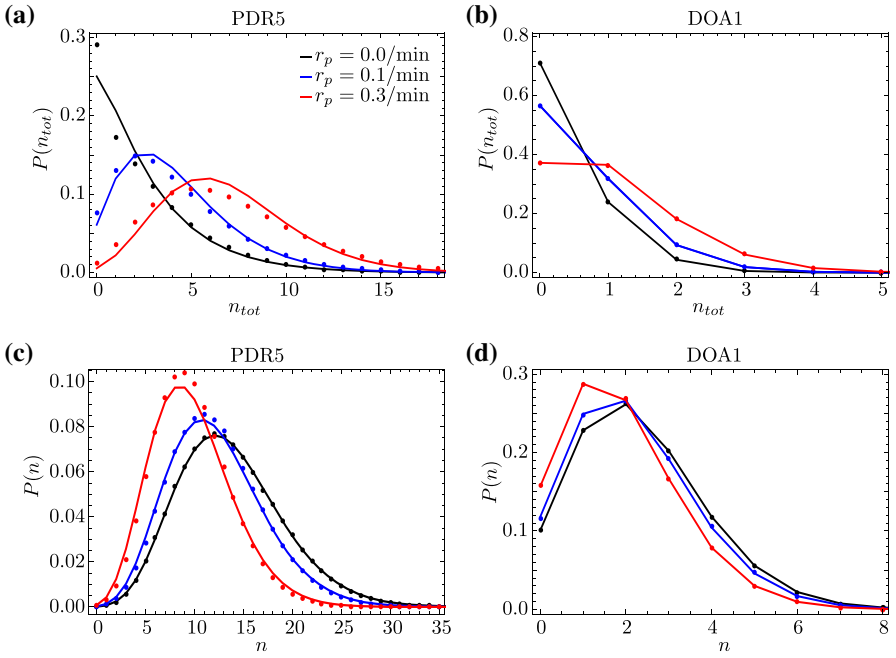


Fig. 10 Dependence of the steady-state probability distributions of total RNAP and mature RNA on the RNAP pausing rate r_p for two different genes in yeast. In **a**, **b**, we compare the distribution $P(n_{\text{tot}})$ of the total number of RNAP molecules, as predicted by our model (solid lines), with that obtained from SSA (dots) for yeast genes PDR5 and DOA1, respectively. The model prediction involves fitting a Negative Binomial distribution with a mean and variance given by the closed-form expressions in Eqs. (41) and (42). In **c**, **d**, we compare the distribution $P(n)$ of mature RNA, as obtained from singular perturbation theory (Eq. (43); solid lines) with the SSA (dots) for yeast genes PDR5 and DOA1, respectively. Note that for both genes, we keep all parameters fixed (including the elongation rate k) while varying the pausing rate r_p to simulate an experiment where the pausing rate can be perturbed directly. The parameters for each gene can be found in Table 1a; we furthermore used $L = 50$ and fixed k to $L/\langle T\rangle$, where $\langle T\rangle$ is the mean elongation time measured experimentally and reported in Table 1a. Note that the actual mean elongation time is not fixed, as it depends on the pausing rate (r_p) via Eq. (40). The remaining parameters are fixed to $r_a = 0.1$/min, $d_a = 0.01$/min, and $d_p = 0.03$/min. The value of d_a is taken from Table 1 in Rajala et al. (2010),where it is reported as the premature termination rate of polymerase in *E. coli*; the value of d_p was chosen to be larger than that of d_a to simulate a scenario where premature detachment is enhanced in the paused state. Note that our theory is less accurate for PDR5 than it is for DOA1, as all parameters are very small compared to the elongation rate in the latter case, hence satisfying better the assumptions behind the theory (Color figure online)

$r_p/(r_p + k + d_a)$ is the probability of the actively moving RNAP switching to the paused state, as well as

$$\omega_{r_a} = \frac{\pi_{r_a}\tau_g}{\pi_{r_a}\tau_{r_a} + \tau_g}, \quad \tilde{\alpha} = \frac{\tau_g + \lambda_{r_p}\pi_{d_p}\tau_g}{\tau_g + \tau_p + \lambda_{r_p}\tau_g(1-\omega_{r_a})}, \quad \textit{and} \quad \omega = \frac{\tau_g}{\tau_{pp} + \tau_g}. \tag{38}$$

Assume that the elongation rate is faster than the rates of RNAP pausing, activation, and premature termination, i.e. that $k \gg r_a, r_p, d_a, d_p$. Then, it follows that to leading

order in $1/k$*, asymptotic expressions for the variances and covariances of molecule number fluctuations of active and paused RNAP are given by*

$$\begin{aligned}
&\mathrm{Cov}(n_i^a, n_j^a) = \delta_{ij}\langle n_i^a\rangle + \langle n_i^a\rangle\langle n_j^a\rangle\tilde{\alpha}\beta \cdot g_{ij}^{aa}, && \text{where } g_{ij}^{aa} = g^{aa}(i,j) + g^{aa}(j,i),\\
&\mathrm{Cov}(n_i^a, n_j^p) = \langle n_i^a\rangle\langle n_j^p\rangle\tilde{\alpha}\beta \cdot g_{ij}^{ap}, && \text{where } g_{ij}^{ap} = \omega\tilde{\alpha}^{j-1},\\
&\mathrm{Cov}(n_i^p, n_j^a) = \langle n_i^p\rangle\langle n_j^a\rangle\tilde{\alpha}\beta \cdot g_{ij}^{pa}, && \text{where } g_{ij}^{pa} = \omega\tilde{\alpha}^{i-1},\\
&\mathrm{Cov}(n_i^p, n_j^p) = \delta_{ij}\langle n_i^p\rangle + \langle n_i^p\rangle\langle n_j^p\rangle\tilde{\alpha}\beta \cdot g_{ij}^{pp}, && \text{where } g_{ij}^{pp} = (g_{ij}^{ap} + g_{ij}^{pa})/2;
\end{aligned} \tag{39}$$

here, $i, j = 1, 2, \ldots, L$ *and*

$$g^{aa}(i,j) = \frac{\tilde{\alpha}^{i+j-1}}{(2\tilde{\alpha}-1)^i} + \frac{1}{2^{i+j-1}}\binom{i+j-1}{i}\Big[1 - \frac{2\tilde{\alpha}-1}{2\tilde{\alpha}}\,{}_2F_1\big(1, i+j; j; \tfrac{1}{2\tilde{\alpha}}\big)\Big].$$

These results are proved in full in 'Appendix H'. From 'Appendix A', we also have that the mean elongation time in the pausing model is given by

$$\langle T\rangle = L\frac{(r_a + d_p)^2 + r_a r_p}{(r_a + d_p)[(k + d_a)(r_a + d_p) + d_p r_p]}. \tag{40}$$

Solving Eq. (40) for the elongation rate k, we find that in the limit of $L \to \infty$ taken at constant mean elongation time, k tends to infinity and hence is much larger than r_a, r_p, d_a, and d_p, which implies that the results of Proposition 4 hold naturally in that limit.

5.2 Approximate Distributions of Total RNAP and Mature RNA

Negative Binomial approximation of total RNAP distribution: We define the total number of RNAP molecules as $n_{\mathrm{tot}} = \sum_{i=1}^{L} n_i$. It then immediately follows from Eq. (37) that the mean of the total RNAP distribution in the pausing model is given by

$$\langle n_{\mathrm{tot}}\rangle = \eta\rho_k(1+\lambda)\tilde{\mu}\frac{\tilde{\mu}^L - 1}{\tilde{\mu} - 1}. \tag{41}$$

It can also be shown that the variance of total RNAP fluctuations reads

$$\mathrm{Var}(n_{\mathrm{tot}}) = \langle n_{\mathrm{tot}}\rangle + (\eta\rho_k)^2\tilde{\alpha}\beta\Big[2\sum_{i,j=1}^{L} g^{aa}(i,j) + \lambda(2+\lambda)\omega L\frac{\tilde{\alpha}^L - 1}{\tilde{\alpha} - 1}\Big]; \tag{42}$$

see 'Appendix H'. Next, we approximate the distribution of total RNAP by a Negative Binomial distribution whose mean and variance match those just derived, i.e. we consider Eq. (20) with the mean and variance of the total RNAP distribution given by

Eqs. (41) and (42) now, respectively. The resulting approximate Negative Binomial distribution is compared with the distribution obtained from SSA in Fig. 10a, b for two different yeast genes, PDR5 and DOA1. The results verify that our approximation is accurate provided the elongation rate k is significantly larger than the other parameters, as assumed in Proposition 4.

Perturbative approximation of mature RNA distribution: We can apply singular perturbation theory to formally derive the distribution of mature RNA, assuming that $k/d_m \gg 1$ and $r_a/d_m \gg 1$. Following the derivation in Sect. 3.2, we find the following analytical expression for the steady-state probability distribution of mature RNA:

$$P(n) = \frac{1}{n!}\frac{(s_u)_n}{(s_b+s_u)_n}\left(\frac{r}{d_m}\right)^n \left(\tilde{\mu}^L\right)^n {}_1F_1\left(\frac{s_u}{d_m}+n; \frac{s_b+s_u}{d_m}+n; -\frac{r}{d_m}\tilde{\mu}^L\right); \quad (43)$$

see 'Appendix I' for details. Note that the solution in Eq. (43) is dependent on the parameter $\tilde{\mu}^L$, which gives the probability that an RNAP molecule does not prematurely detach before termination; see 'Appendix A'. Also, note that in the limit of zero premature termination, i.e. for $d_a = 0 = d_p$, Eq. (43) is identical to the distribution of mature RNA predicted by the telegraph model. Finally, by solving Eq. (40) for k, then substituting the resulting expression into Eq. (43) and taking the long-gene limit of $L \to \infty$ at constant $\langle T \rangle$, we obtain that the probability distribution of mature RNA has the same functional form as in Eq. (43), albeit with

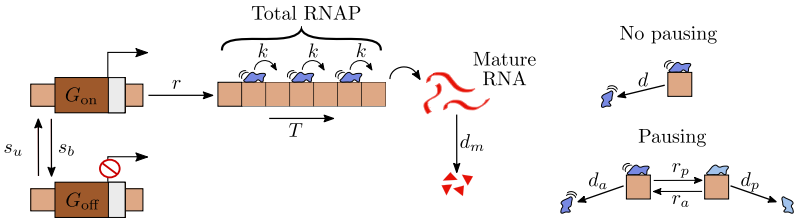
$$\lim_{L\to\infty} \tilde{\mu}^L = \mathrm{e}^{-\psi\langle T\rangle}, \quad \text{where } \psi = \frac{d_a + r_p \pi_{d_p}}{1+\sigma\pi_{r_a}}. \quad (44)$$

Note that Eqs. (43) and (44) equal the steady-state solution predicted by the telegraph model, with the initiation rate r renormalised to $r\tilde{\mu}^L$ or $r\mathrm{e}^{-\psi\langle T\rangle}$, respectively. In Fig. 10c, d, we verify the accuracy of our analytical solution using stochastic simulation for two different genes in yeast. Note that a change in the pausing rate r_p has relatively little effect on the distribution of mature RNA, as compared to the effect on the distribution of total RNAP; cf. panels (a) and (b) of Fig. 10 in comparison with panels (c) and (d), respectively.

6 Summary and Conclusion

In this paper, we have analysed a detailed stochastic model of transcription. Our model extends previous analytical work (Choubey et al. 2015; Heng et al. 2016) by (i) taking into account salient processes, such as premature detachment and pausing of RNAP, that were previously not considered analytically; (ii) deriving explicit expressions for the mean and variance of RNAP numbers (nascent RNA) on gene segments as well as on the entire gene; (iii) deriving explicit expressions for the mean and variance of the fluorescent nascent RNA signal obtained from smFISH and identifying differences between the statistics thereof and those of direct measurements of nascent RNA; and (iv) finding approximate distributions of total nascent RNA fluctuations on a gene,

Table 2 Summary of main results



Approximate total RNAP distribution for stochastic elongation

$$n_{tot} \sim \mathrm{NB}\left(\frac{\langle n_{tot}\rangle^2}{\mathrm{Var}(n_{tot})-\langle n_{tot}\rangle}, \frac{\mathrm{Var}(n_{tot})-\langle n_{tot}\rangle}{\mathrm{Var}(n_{tot})}\right)$$

No pausing

$$\langle n_{tot}\rangle = \eta\rho_k\mu\frac{\mu^L-1}{\mu-1}$$

$$\mathrm{Var}(n_{tot}) = \langle n_{tot}\rangle + \alpha\beta(\eta\rho_k)^2\sum_{i,j=1}^{L}\mu^{i+j}\cdot[f(i,j)+f(j,i)]$$

Pausing

$$\langle n_{tot}\rangle = \eta\rho_k(1+\lambda)\tilde{\mu}\frac{\tilde{\mu}^L-1}{\tilde{\mu}-1}$$

$$\mathrm{Var}(n_{tot}) = \langle n_{tot}\rangle + \tilde{\alpha}\beta(\eta\rho_k)^2\left[2\sum_{i,j=1}^{L}g^{aa}(i,j)+\lambda(2+\lambda)\omega L\frac{\tilde{\alpha}^L-1}{\tilde{\alpha}-1}\right]$$

Approximate mature RNA distribution ($k/d_m \gg 1$) as the solution of the telegraph model with renormalised transcription rate rP_μ and RNAP survival probability P_μ

$$n \sim P(n) = \frac{1}{n!}\frac{(s_u)_n}{(s_b+s_u)_n}\left(\frac{r}{d_m}\right)^n (P_\mu)^n\, {}_1F_1\left(\frac{s_u}{d_m}+n; \frac{s_b+s_u}{d_m}+n; -\frac{r}{d_m}P_\mu\right)$$

	Stochastic elongation		Deterministic elongation	
	Detachment	No detachment	Detachment	No detachment
No pausing Pausing	$d \neq 0$ $d_a, d_p \neq 0$	$d = 0$ $d_a = d_p = 0$	$d \neq 0$ $d_a, d_p \neq 0$	$d = 0$ $d_a = d_p = 0$
No pausing Pausing	$P_\mu = \mu^L$ $P_\mu = \tilde{\mu}^L$	$P_\mu = 1$ $P_\mu = 1$	$P_\mu = e^{-d\langle T\rangle}$ $P_\mu = e^{-\psi\langle T\rangle}$	$P_\mu = 1$ $P_\mu = 1$

Ratio of Fano factors (FF) of total RNAP and mature RNA for deterministic elongation without detachment ($d = 0$)

$$R_{FF} = \mathrm{FF_n}/\mathrm{FF_m}$$

	General	Bursty ($s_b, r \to \infty$)
No pausing	$R_{FF} = 1 + \frac{e^{-T_g}T_rT_{s_b}\Xi}{T_g^2[T_rT_{s_b}+T_g(T_g+T_m)]}$	$R_{FF} = 1 + \frac{b}{1+b}$

The cartoon represents our model in various limits: no pausing ($r_p = 0$), pausing ($r_p \neq 0$), stochastic elongation (T Erlang distributed), deterministic elongation (T fixed), bursty limit ($r, s_b \to \infty$), and premature RNAP detachment ($d, d_a, d_p \neq 0$). We summarise our analytical expressions for the approximate distributions and moments of total RNAP and mature RNA

Table 3 Definition of parameters and functions

$f(i,j) = \frac{\alpha^{i+j-1}}{(2\alpha-1)^i} + \frac{1}{2^{i+j-1}}\binom{i+j-1}{i}\Big[1 - \frac{2\alpha-1}{2\alpha}\,{}_2F_1\big(1, i+j; j; \frac{1}{2\alpha}\big)\Big]$,	
$g^{aa}(i,j) = \frac{\tilde{\alpha}^{i+j-1}}{(2\tilde{\alpha}-1)^i} + \frac{1}{2^{i+j-1}}\binom{i+j-1}{i}\Big[1 - \frac{2\tilde{\alpha}-1}{2\tilde{\alpha}}\,{}_2F_1\big(1, i+j; j; \frac{1}{2\tilde{\alpha}}\big)\Big]$,	
$\Xi = 2(T_g + T_m) + \mathrm{e}^{T_g}[2(T_g-1)T_m + (T_g-2)T_g]$	
$\eta = s_u/(s_u + s_b)$	Fraction of time the gene spends in the active state
$\rho_k = r/k$	Mean number of bound RNAPs in the time $1/k$
$\rho = r/d_m$	Mean number of bound RNAPs in the time $1/d_m$
$\mu = k/(k+d)$	Local RNAP survival probability (no-pausing case)
$\tau_p = 1/(d+k)$	Timescale of fluctuations of RNAP
$\tau_g = 1/(s_u + s_b)$	Timescale of fluctuations of gene
$\tau_d = 1/d$	Timescale of RNAP detachment
$\tau_m = 1/d_m$	Timescales of fluctuations of mature RNA
$\alpha = 1/(1+\tau_p/\tau_g)$	Non-dimensional parameter
$\gamma = 1/(1+\tau_p/\tau_m)$	Non-dimensional parameter
$\theta = 1/(1+\tau_m/\tau_g)$	Non-dimensional parameter
$\beta = s_b/s_u$	Ratio of gene inactivation and activation rates
$b = r/s_b$	Mean burst size
$\upsilon_k = s_u/k$	Ratio of gene activation and RNAP elongation rates
$\upsilon_m = s_u/d_m$	Ratio of gene activation and mature RNA degradation rates
$\delta_g = \tau_g/\tau_d$	Ratio of gene timescale and RNAP detachment timescale
$T_g = \langle T\rangle/\tau_g$	Ratio of elongation timescale and gene timescale
$T_d = \langle T\rangle/\tau_d$	Ratio of elongation timescale and RNAP detachment timescale
$T_r = r\langle T\rangle$	Ratio of the mean elongation time to the timescale of initiation
$T_m = d_m\langle T\rangle$	Ratio of the mean elongation time to the timescale of decay of mature RNA
$T_{s_b} = s_b\langle T\rangle$	Ratio of the mean elongation time to the timescale of gene deactivation
$\sigma = r_p/r_a$	Ratio of the pausing and activation rates
$\pi_{r_a} = r_a/(r_a + d_p)$	Probability of RNAP activation

Table 3 continued

$\pi_{d_p} = d_p/(r_a + d_p)$	Probability of premature RNAP detachment from the paused state
$\lambda = \sigma\pi_{r_a}$	Probability of RNAP pausing from active state
$\tilde{\mu} = k/(k + d_a + r_p\pi_{d_p})$	Local RNAP survival probability (in pausing case)
$\tau_{r_a} = 1/r_a$	Timescale of RNAP activation from the paused state
$\tau_{d_p} = 1/d_p$	Timescale of premature termination of paused RNAP
$\tau_p = 1/(k + d_a)$	Typical time that an actively moving RNAP spends on a gene segment
$\tau_{pp} = 1/(r_a + d_p)$	Typical time spent in the paused state
$\lambda_{r_p} = \pi_{r_p}/(1 - \pi_{r_p})$	Ratio of active RNAP timescale over RNAP pausing timescale
$\pi_{r_p} = r_p/(r_p + k + d_a)$	Probability of the actively moving RNAP switching to the paused state
$\omega_{r_a} = \pi_{r_a}\tau_g/(\pi_{r_a}\tau_{r_a} + \tau_g)$	Non-dimensional parameter
$\tilde{\alpha} = (\tau_g + \lambda_{r_p}\pi_{d_p}\tau_g)/(\tau_g + \tau_p + \lambda_{r_p}\tau_g(1 - \omega_{r_a}))$	Non-dimensional parameter.

without assuming slow promoter switching. A number of interesting observations from our work include the following:

(i) When the premature detachment rate of RNAP is nonzero and gene expression is bursty, the coefficient of variation of local RNAP fluctuations can either decrease or increase with distance from the promoter. By contrast, when expression is constitutive, the coefficient of variation increases monotonically with distance from the promoter. Other statistical measures such as the mean, Fano factor, and correlation coefficient of local RNAP numbers decrease monotonically with distance from the promoter.

(ii) In the limits of bursty expression, deterministic elongation, and no premature detachment or pausing, the Fano factor of total nascent RNA equals $1+2b$, whereas that of mature RNA is $1 + b$, where b denotes the mean burst size. An implication is that the telegraph model will result in an overestimate of the mean burst size from nascent RNA data by a factor of 2. Another implication is that deviations from Poisson fluctuations are more apparent in data for nascent RNA than they are for mature RNA. One can further state the following relationship: the Fano factor of nascent RNA equals twice the Fano factor of mature RNA, minus 1. If expression is non-bursty, then the Fano factor of nascent RNA can be larger or smaller than that of mature RNA, as determined by the condition in Eq. (19).

(iii) For genes characterised by bursty expression, the sensitivity of the noise in total RNAP fluctuations is highest to perturbations in the gene activation rate and the mean elongation time; for constitutive genes, the most sensitive parameters are the initiation rate and the mean elongation time.

(iv) A Negative Binomial distribution, parameterised with the expressions for the mean and variance of total nascent RNA derived here, provides a good approximation to the true distribution of total nascent RNA fluctuations on a gene when expression is either bursty or constitutive; the approximation is not accurate when the gene spends roughly equal amounts of time in the active and inactive states. We show that the distribution of nascent RNA is highly sensitive to the distribution of elongation times. In particular, if the elongation time is assumed to be exponentially distributed, as is implicitly assumed by telegraph models of nascent RNA, then the probability of observing zero RNA is much lower than if the elongation time is assumed to be fixed.

(v) Using geometric singular perturbation theory (GSPT), we have rigorously proven that, in the limit of deterministic elongation (or fast elongation), no pausing and premature detachment, the steady-state distribution of mature RNA in our model is identical to that in the telegraph model (Raj et al. 2006). Consideration of pausing and premature detachment leads to a distribution that can also be obtained from a telegraph model with appropriately renormalised parameters.

A summary of the main theoretical results can be found in Table 2, with all requisite parameters and functions defined in Table 3. The main limiting assumption of our theoretical approach is that the initiation rate is slow enough such that RNAP molecules do not frequently collide with each other while moving along the gene. Hence, the expressions we have derived are reasonable for all but the strongest promoters which are characterised by very fast initiation rates. We anticipate that approximate closed-form expressions for the corresponding moments can also be derived when volume exclusion between RNAPs is taken into account by a modification of methods previously devised to understand molecular movement and kinetics in crowded conditions (Cianci et al. 2016; Smith et al. 2017). It is also possible to extend our model by including translation of mature RNA to protein; one can then again apply GSPT to derive distributions for protein numbers in the limit of RNA decaying much faster than protein; however, given item (v) above, we anticipate that the resulting protein distribution will be very similar to those derived from models that do not explicitly take into account nascent RNA (Shahrezaei and Swain 2008; Popović et al. 2016). Further research is required to develop simple approximations of the nascent RNA distribution that are accurate independently of the ratio of gene switching rates. Finally, given the strong recent interest in the development of statistical inference techniques in molecular biology (Gorin et al. 2020; Zechner et al. 2012; Kaan Öcal et al. 2019), we expect that our closed-form expressions for the moments and distributions of nascent and mature RNA will be useful for developing computationally efficient and accurate methods for estimating transcriptional parameters.

Acknowledgements R.G. acknowledges useful discussions with Zhixing Cao and Tineke Lenkstra. This work was supported by a departmental PhD studentship to T.F.

Appendix

A. Distribution of Elongation Time

In this section, we answer the following question: what is the distribution of the elongation time, i.e. the time between initiation and termination? In other words, with reference to Fig. 9—which includes the non-pausing model in Fig. 1 as a special case-we want to find the distribution of the time at which RNAP leaves gene segment L (termination) if it was in the active state on gene segment 1 at time $t = 0$ (initiation).

Let $z_i(t)$ be the probability of an RNAP to be on gene segment i in the active state at time t, let $\tilde{z}_i(t)$ be the probability of the RNAP to be on gene segment i in the paused state at time t, and let $z_i^*(t)$ be the probability of the RNAP moving to gene segment $i+1$ at time t; note that $z_L^*(t)$ is the probability of the RNAP falling off the gene and forming a mature RNA, since for $i = L$, gene segment $L+1$ does not exist. Then, it follows from the reaction scheme illustrated in Fig. 9 that the master equations describing the Markovian dynamics on gene segment i are given by

$$\partial_t z_i(t) = -(r_p + k + d_a)z_i(t) + r_a\tilde{z}_i(t), \tag{A.1a}$$
$$\partial_t \tilde{z}_i(t) = -(d_p + r_a)\tilde{z}_i(t) + r_p z_i(t), \tag{A.1b}$$
$$\partial_t z_i^*(t) = k z_i(t). \tag{A.1c}$$

Now, we use these equations to find the distribution of the time when RNAP jumps to gene segment $i+1$, given that it is on gene segment i in the active state at $t = 0$, i.e. that $z_i(0) = 1$ and $\tilde{z}_i(0) = 0$. Taking the Laplace transform of Eqs. (A.1a) and (A.1b), we find

$$s\hat{z}_i(s) - 1 = -(r_p + k + d_a)\hat{z}_i(s) + r_a\hat{\tilde{z}}_i(s), \tag{A.2a}$$
$$s\hat{\tilde{z}}_i(s) = -(d_p + r_a)\hat{\tilde{z}}_i(s) + r_p\hat{z}_i(s), \tag{A.2b}$$

where $\hat{f}(s) = \int_0^\infty \mathrm{e}^{-st} f(t)\,\mathrm{d}t$. Solving these equations simultaneously, we obtain

$$\hat{z}_i(s) = \frac{s + d_p + r_a}{(s + k + d_a)(s + d_p + r_a) + r_p(s + d_p)} \tag{A.3}$$

Let $w(t)\mathrm{d}t$ be the probability that the RNAP moves from segment i to $i+1$ in the time interval $(t, t + \mathrm{d}t)$. Then, it follows from Eq. (A.1c) that $w(t) = \partial_t z_i^*(t) = k z_i(t)$. Integrating $w(t)$ over all times gives us the probability that the RNAP ultimately moves

to the next segment $i+1$,

$$\int_0^\infty w(t)\,\mathrm{d}t = \hat{w}(0) = k\hat{z}_i(0) = \frac{k\left(r_a+d_p\right)}{\left(d_a+k\right)\left(r_a+d_p\right)+d_p r_p}. \tag{A.4}$$

Note that $\hat{w}(0)$ is identical to the parameter $\tilde{\mu}$, as defined in Proposition 3. Let $y(t)\mathrm{d}t$ be the probability that the RNAP moves from gene segment i to segment $i+1$ in the time interval $(t, t+\mathrm{d}t)$, *conditioned* on those realisations that lead to an RNAP moving to the next gene segment $i+1$. (In other words, we exclude those realisations that lead to premature detachment.) Then, it follows by the definition of conditional probabilities that $y(t) = w(t)/\hat{w}(0)$, which implies

$$\hat{y}(s) = \frac{\hat{w}(s)}{\hat{w}(0)} = \frac{[(d_a+k)(r_a+d_p)+d_p r_p](r_a+d_p+s)}{(r_a+d_p)[(d_a+k+s)(r_a+d_p+s)+r_p(d_p+s)]}. \tag{A.5}$$

It follows that the mean $\langle t\rangle$ and variance $\mathrm{Var}(t)$ of the time t it takes RNAP to move to the next gene segment are given by

$$\langle t\rangle = -\frac{d\hat{y}(s)}{\mathrm{d}s}\bigg|_{s=0} = \frac{(r_a+d_p)^2+r_a r_p}{(r_a+d_p)[(d_a+k)(r_a+d_p)+d_p r_p]}, \tag{A.6a}$$

$$\begin{aligned}\mathrm{Var}(t) &= \frac{d^2\hat{y}(s)}{\mathrm{d}s^2}\bigg|_{s=0} - \left(\frac{d\hat{y}(s)}{\mathrm{d}s}\bigg|_{s=0}\right)^2 \\ &= \frac{2r_a r_p(r_a+d_p)(d_a+r_a+d_p+k)+(r_a+d_p)^4+r_a r_p^2(r_a+2d_p)}{(r_a+d_p)^2[(d_a+k)(r_a+d_p)+d_p r_p]^2},\end{aligned} \tag{A.6b}$$

respectively. Since RNAP can only move forwards in our model (irreversible motion), it follows that the time it takes an RNAP to move from the ith to the $(i+1)$th gene segment is independent of the time taken to move from another, jth segment to the $(j+1)$th segment. Hence, the time required for an RNAP to move across the entire gene from the first to the Lth segment, i.e. the 'elongation' time T from initiation to termination, is a sum of L independent and identical random variables. Thus, we can immediately state that the mean elongation time is $\langle T\rangle = L\langle t\rangle$, whereas the variance of the elongation time is $\mathrm{Var}(T) = L\mathrm{Var}(t)$. The coefficient of variation squared takes the form

$$\mathrm{CV}_T^2 = \frac{\mathrm{Var}(T)}{\langle T\rangle^2} = 1 + \frac{2r_a r_p[(d_a+k)(r_a+d_p)+d_p r_p]}{[(r_a+d_p)^2+r_a r_p]^2}. \tag{A.7}$$

From Eq. (A.7), it can be shown that for small premature detachment rates, the coefficient of variation of the elongation time is maximised when $r_p \approx r_a$. Taking the limit of infinitely many gene segments at constant mean elongation time, i.e. solving for k from the expression for the mean elongation time in Eq. (A.6), substituting into

Eq. (A.7), and taking the limit of $L \to \infty$, we obtain

$$\lim_{L\to\infty} \mathrm{CV}_T^2 = \frac{2r_a r_p}{\langle T\rangle (r_a + d_p)[(r_a + d_p)^2 + r_a r_p]}. \tag{A.8}$$

For the non-pausing model shown in Fig. 1, the above results simplify considerably due to $r_p = 0 = d_p$ and $d_a = d$; in that case, the inverse Laplace transform of Eq. (A.5) implies that $y(t)$ is an exponential distribution with parameter $k + d$. Hence, the total time it takes an RNAP to move across the entire gene is the sum of L independent and identically distributed exponential random variables, i.e. an Erlang distribution with shape parameter L and rate $k + d$, which implies that the mean elongation time is $L/(k + d)$, with coefficient of variation $1/\sqrt{L}$. It can be seen from Eq. (A.8) that deterministic elongation can only be observed when there is no pausing, i.e. when $r_p = 0$.

B. Solution of Lyapunov Equation

Proof of Proposition 2 We start by defining the symmetric functions $f_{ij} = f_{ji}$ for $i, j = 1, \ldots, L$ as

$$\begin{aligned} &f_{00} = 1, && f_{0j} = \alpha^{j-1}, && f_{0M} = \theta\alpha^{L-1}, \\ &f_{ij} = (f_{i-1,j} + f_{i,j-1})/2, && f_{iM} = \gamma f_{i-1,M} + (1-\gamma) f_{iL}, && f_{MM} = f_{LM}, \end{aligned} \tag{B.1}$$

where the non-dimensional parameters α, γ, and θ are defined in Proposition 2. The elements of the Lyapunov equation given by Eq. (5) can be written explicitly as a set of simultaneous equations:

$$\mathbf{C}_{11} \cdot 2\mathbf{J}_{11} = -\mathbf{D}_{11}, \tag{B.2a}$$
$$\mathbf{C}_{12} \cdot (\mathbf{J}_{11} + \mathbf{J}_{22}) = -\mathbf{J}_{21}\mathbf{C}_{11}, \tag{B.2b}$$
$$\mathbf{C}_{1j} \cdot (\mathbf{J}_{11} + \mathbf{J}_{jj}) = -\mathbf{J}_{j,j-1}\mathbf{C}_{1,j-1} \quad \text{for } j = 3, \ldots, L+1, \tag{B.2c}$$
$$\mathbf{C}_{1,L+2} \cdot (\mathbf{J}_{11} + \mathbf{J}_{L+2,L+2}) = -\mathbf{J}_{L+2,L+1}\mathbf{C}_{1,L+1}, \tag{B.2d}$$
$$\mathbf{C}_{22} \cdot 2\mathbf{J}_{22} = -2\mathbf{J}_{21}\mathbf{C}_{12} - \mathbf{D}_{22}, \tag{B.2e}$$
$$\mathbf{C}_{23} \cdot (\mathbf{J}_{22} + \mathbf{J}_{33}) = -\mathbf{J}_{21}\mathbf{C}_{13} - \mathbf{J}_{32}\mathbf{C}_{22} - \mathbf{D}_{23}, \tag{B.2f}$$
$$\mathbf{C}_{2j} \cdot (\mathbf{J}_{22} + \mathbf{J}_{jj}) = -\mathbf{J}_{21}\mathbf{C}_{1j} - \mathbf{J}_{j,j-1}\mathbf{C}_{2,j-1} \quad \text{for } j = 4, \ldots, L+1, \tag{B.2g}$$
$$\mathbf{C}_{2,L+2} \cdot (\mathbf{J}_{22} + \mathbf{J}_{L+2,L+2}) = -\mathbf{J}_{21}\mathbf{C}_{1,L+2} - \mathbf{J}_{L+2,L+1}\mathbf{C}_{2,L+1}, \tag{B.2h}$$
$$\mathbf{C}_{ii} \cdot 2\mathbf{J}_{ii} = -2\mathbf{J}_{i,i-1}\mathbf{C}_{i-1,i} - \mathbf{D}_{ii} \quad \text{for } i = 3, \ldots, L+1, \tag{B.2i}$$
$$\mathbf{C}_{i,i+1} \cdot (\mathbf{J}_{ii} + \mathbf{J}_{i+1,i+1}) = -\mathbf{J}_{i,i-1}\mathbf{C}_{i-1,i+1} - \mathbf{J}_{i+1,i}\mathbf{C}_{ii} - \mathbf{D}_{i,i+1} \quad \text{for } i = 3, \ldots, L, \tag{B.2j}$$
$$\mathbf{C}_{ij} \cdot (\mathbf{J}_{ii} + \mathbf{J}_{jj}) = -\mathbf{J}_{i,i-1}\mathbf{C}_{i-1,j} - \mathbf{J}_{j,j-1}\mathbf{C}_{i,j-1} \quad \text{for } i = 3, \ldots, L+1 \text{ and } j = i+2, \ldots, L+1, \tag{B.2k}$$
$$\mathbf{C}_{i,L+2} \cdot (\mathbf{J}_{ii} + \mathbf{J}_{L+2,L+2}) = -\mathbf{J}_{i,i-1}\mathbf{C}_{i-1,L+2} - \mathbf{J}_{L+2,L+1}\mathbf{C}_{i,L+1} \quad \text{for } i = 3, \ldots, L+1, \tag{B.2l}$$
$$\mathbf{C}_{L+2,L+2} \cdot 2\mathbf{J}_{L+2,L+2} = -2\mathbf{J}_{L+2,L+1}\mathbf{C}_{L+1,L+2} - \mathbf{D}_{L+2,L+2}. \tag{B.2m}$$

Now, we substitute the elements of the Jacobian matrix $\mathbf{J}$ and the diffusion matrix $\mathbf{D}$ from Eqs. (6) and (7), respectively, into the above system of algebraic equations,

which we then solve to find the elements of the covariance matrix $\mathbf{C}$. Note that, for the following mathematical derivation, we take into account the expressions for the steady-state mean numbers of species given in Eq. (2), as well as the definition of the functions f_{ij} in Eq. (B.1).

From Eq. (B.2a), one easily obtains $\mathbf{C}_{11} = \eta^2\beta$. Then, it follows from Eq. (B.2b) that

$$\mathbf{C}_{12} = \frac{r}{s_u + s_b + k + d}\mathbf{C}_{11} = \rho_k\mu\alpha(\eta^2\beta) = \eta(\eta\rho_k\mu)\alpha\beta = \eta\langle n_1\rangle\alpha\beta \cdot f_{01}. \quad \text{(B.3)}$$

Eq. (B.2c) implies that, for $j = 3, \ldots, L+1$:

$$\begin{aligned}\mathbf{C}_{1j} &= \frac{k}{s_u + s_b + k + d}\mathbf{C}_{1,j-1} = \mu\alpha \cdot \mathbf{C}_{1,j-1} = (\mu\alpha)^{j-2}\mathbf{C}_{12} \\ &= (\mu\alpha)^{j-2}(\eta\langle n_1\rangle\alpha\beta) = \eta\langle n_{j-1}\rangle\alpha\beta \cdot f_{0,j-1}.\end{aligned} \quad \text{(B.4)}$$

From Eq. (B.2d), we have that

$$\begin{aligned}\mathbf{C}_{1,L+2} &= \frac{k}{s_u + s_b + d_m}\mathbf{C}_{1,L+1} = \frac{k}{d_m}\theta(\langle n_L\rangle\alpha\beta \cdot f_{0L}) \\ &= \eta(\frac{k}{d_m}\langle n_L\rangle)(\alpha\beta)(\theta \cdot f_{0L}) = \eta\langle n\rangle \cdot f_{0M};\end{aligned} \quad \text{(B.5)}$$

from Eq. (B.2e), we find

$$\begin{aligned}\mathbf{C}_{22} &= \frac{r\langle n_0\rangle + (k+d)\langle n_1\rangle}{2(k+d)} + \frac{r}{k+d}\mathbf{C}_{12} = \frac{\rho_k\mu\eta + \langle n_1\rangle}{2} \\ &+ (\rho_k\mu)(\eta\langle n_1\rangle\alpha\beta \cdot f_{01}) = \langle n_1\rangle + \langle n_1\rangle^2\alpha\beta \cdot f_{11},\end{aligned} \quad \text{(B.6)}$$

since $f_{11} = (f_{01} + f_{10})/2 = f_{01}$ from the definition in Eq. (B.1).
From Eq. (B.2f), we obtain

$$\begin{aligned}\mathbf{C}_{23} &= -\frac{k}{2(d+k)}\langle n_1\rangle + \frac{r}{2(k+d)}\mathbf{C}_{13} + \frac{k}{2(d+k)}\mathbf{C}_{22} \\ &= -\frac{\langle n_2\rangle}{2} + \frac{1}{2}(\rho_k\mu\eta)\langle n_2\rangle\alpha\beta \cdot f_{02} + \frac{1}{2}\mu[\langle n_1\rangle + \langle n_1\rangle^2\alpha\beta \cdot f_{11}] \\ &= -\frac{\langle n_2\rangle}{2} + \frac{1}{2}\langle n_1\rangle\langle n_2\rangle\alpha\beta \cdot f_{02} + \frac{\langle n_2\rangle}{2} + \frac{1}{2}(\mu\langle n_1\rangle)\langle n_1\rangle\alpha\beta \cdot f_{11} \\ &= \frac{1}{2}\langle n_1\rangle\langle n_2\rangle\alpha\beta \cdot f_{02} + \frac{1}{2}\langle n_2\rangle\langle n_1\rangle\alpha\beta \cdot f_{11} = \langle n_1\rangle\langle n_2\rangle\alpha\beta\frac{1}{2}(f_{02} + f_{11}) \\ &= \langle n_1\rangle\langle n_2\rangle\alpha\beta \cdot f_{12},\end{aligned} \quad \text{(B.7)}$$

since $f_{12} = (f_{02} + f_{11})/2$ from the definition in Eq. (B.1).

From Eq. (B.2g), we have that, for $j = 4, \ldots, L+1$,

$$\begin{aligned}\mathbf{C}_{2j} &= \frac{r}{2(k+d)}\mathbf{C}_{1j} + \frac{k}{2(k+d)}\mathbf{C}_{2,j-1} = \frac{\rho_k\mu}{2}\mathbf{C}_{1j} + \frac{\mu}{2}\mathbf{C}_{2,j-1} \\ &= \frac{\rho_k\mu}{2}\sum_{q=0}^{j-4}\left(\frac{\mu}{2}\right)^q \mathbf{C}_{1,j-q} + \left(\frac{\mu}{2}\right)^{j-3}\mathbf{C}_{23}.\end{aligned} \tag{B.8}$$

The proof of Eq. (B.8) is given in Lemma B.1. The above expression for $\mathbf{C}_{2j}$ can be further simplified to

$$\begin{aligned}\mathbf{C}_{2j} &= \frac{\rho_k\mu}{2}\sum_{q=0}^{j-4}\left(\frac{\mu}{2}\right)^q \eta\langle n_{j-q-1}\rangle\alpha\beta\cdot f_{0,j-q-1} + \left(\frac{\mu}{2}\right)^{j-3}\langle n_1\rangle\langle n_2\rangle\alpha\beta\cdot f_{12} \\ &= \sum_{q=0}^{j-4}\left(\frac{1}{2}\right)^{q+1}(\rho_k\mu\eta)(\mu^q\langle n_{j-q-1}\rangle)\alpha\beta\cdot f_{0,j-q-1} + \left(\frac{1}{2}\right)^{j-3}\langle n_1\rangle(\mu^{j-3}\langle n_2\rangle)\alpha\beta\cdot f_{12} \\ &= \sum_{q=0}^{j-4}\left(\frac{1}{2}\right)^{q+1}\langle n_1\rangle\langle n_{j-1}\rangle\alpha\beta\cdot f_{1,j-q-1} + \left(\frac{1}{2}\right)^{j-3}\langle n_1\rangle\langle n_{j-1}\rangle\alpha\beta\cdot f_{12} \\ &= \langle n_1\rangle\langle n_{j-1}\rangle\alpha\beta\left[\sum_{q=0}^{j-4}\left(\frac{1}{2}\right)^{q+1} f_{1,j-q-1} + \left(\frac{1}{2}\right)^{j-3} f_{12}\right] = \langle n_1\rangle\langle n_{j-1}\rangle\alpha\beta\cdot f_{1,j-1}.\end{aligned} \tag{B.9}$$

For the proof of the last equality in Eq. (B.9), see Lemma B.2.
From Eq. (B.2h), we have that

$$\begin{aligned}\mathbf{C}_{2,L+2} &= \frac{r}{k+d+d_m}\mathbf{C}_{1,L+2} + \frac{k}{k+d+d_m}\mathbf{C}_{2,L+1} = \rho_k\mu\gamma\mathbf{C}_{1,L+2} + \mu\gamma\mathbf{C}_{2,L+1} \\ &= (\rho_k\mu\gamma)(\eta\langle n\rangle\alpha\beta\cdot f_{0M}) + (\mu\gamma)(\langle n_1\rangle\langle n_L\rangle\alpha\beta\cdot f_{1L}) \\ &= (\rho_k\eta\mu)\langle n\rangle\alpha\beta\cdot\gamma f_{0M} + \mu\frac{d_m}{k}\langle n_1\rangle\frac{k}{d_m}\langle n_L\rangle\alpha\beta\cdot\gamma f_{1L} \\ &= \langle n_1\rangle\langle n\rangle\alpha\beta\cdot\left[\gamma f_{0M} + \mu\frac{d_m}{k}\gamma f_{1L}\right] \\ &= \langle n_1\rangle\langle n\rangle\alpha\beta\cdot[\gamma f_{0M} + (1-\gamma)\cdot f_{1L}] = \langle n_1\rangle\langle n\rangle\alpha\beta\cdot f_{1M},\end{aligned} \tag{B.10}$$

where f_{1M} is defined in Eq. (B.1).
Eqs. (B.2i) through (B.2k) yield the system

$$\begin{aligned}\mathbf{C}_{ii} &= \frac{k\langle n_{i-2}\rangle + (k+d)\langle n_{i-1}\rangle}{2(k+d)} + \frac{k}{k+d}\mathbf{C}_{i-1,i} = \langle n_{i-1}\rangle + \mu\mathbf{C}_{i-1,i}, \\ \mathbf{C}_{i,i+1} &= \frac{\mu}{2}\mathbf{C}_{i-1,i+1} + \frac{\mu}{2}\mathbf{C}_{ii} - \frac{\mu}{2}\langle n_{i-1}\rangle = \frac{\mu}{2}(\mathbf{C}_{i-1,i+1} + \mu\mathbf{C}_{i-1,i}), \\ \mathbf{C}_{ij} &= \frac{\mu}{2}(\mathbf{C}_{i-1,j} + \mathbf{C}_{i,j-1}),\end{aligned} \tag{B.11}$$

which can be rewritten more compactly as

$$\mathbf{C}_{ij} = \delta_{ij}\langle n_{i-1}\rangle + \langle n_{i-1}\rangle\langle n_{j-1}\rangle\alpha\beta \cdot f_{i-1,j-1} \qquad \text{for } i, j = 3, \ldots, L+1, \tag{B.12}$$

where δ_{ij} is the Kronecker delta. A detailed derivation is given in Lemma B.3.
From Eq. (B.2l), we have that for $i = 3, \ldots, L+1$,

$$\begin{aligned}
\mathbf{C}_{i,L+2} &= \frac{k}{k+d+d_m}\mathbf{C}_{i-1,L+2} + \frac{k}{k+d+d_m}\mathbf{C}_{i,L+2} \\
&= \mu\gamma\mathbf{C}_{i-1,L+2} + (k/d_m)(1-\gamma)\mathbf{C}_{i,L+1} \\
&= \gamma(\mu\langle n_{i-2}\rangle)\langle n\rangle\alpha\beta \cdot f_{i-2,M} + (1-\gamma)\langle n_{i-1}\rangle(k/d_m\langle n_L\rangle)\alpha\beta \cdot f_{i-1,L} \\
&= \langle n_{i-1}\rangle\langle n\rangle\alpha\beta \cdot [\gamma f_{i-2,M} + (1-\gamma) f_{i-1,L}] = \langle n_{i-1}\rangle\langle n\rangle\alpha\beta \cdot f_{i-1,M},
\end{aligned} \tag{B.13}$$

where f_{iM} is defined in Eq. (B.1).
Finally, Eq. (B.2m) yields

$$\begin{aligned}
\mathbf{C}_{L+2,L+2} &= \frac{k\langle n_L\rangle + d_m\langle n\rangle}{2d_m} + \frac{k}{d_m}\mathbf{C}_{L+1,L+2} \\
&= \langle n\rangle + (k/d_m)\langle n_L\rangle\langle n\rangle\alpha\beta \cdot f_{LM} = \langle n\rangle + \langle n\rangle^2\alpha\beta \cdot f_{MM},
\end{aligned} \tag{B.14}$$

where $f_{MM} = f_{LM}$ is defined in Eq. (B.1).
Summarising the above results, we conclude that the solution for the symmetric covariance matrix $\mathbf{C}$ is given by the system in Eq. (4), where we have that $\mathrm{Cov}(n_i, n_j) = \mathbf{C}_{i+1,j+1}$, $\mathrm{Cov}(n_i, n) = \mathbf{C}_{i+1,L+2}$ for $i, j = 0, \ldots, L$, and $\mathrm{Var}(n, n) = \mathbf{C}_{L+2,L+2}$. Here, the functions f_{ij} are defined as in Eq. (B.1). Now, the recurrence relation $f_{ij} = (f_{i-1,j} + f_{i,j-1})/2$ in Eq. (B.1) can be solved for $i, j = 1, 2, \ldots, L$ via the method of generating functions, which gives the following analytical expression:

$$f_{ij} = f(i, j) + f(j, i), \tag{B.15}$$

where

$$f(i, j) = \frac{\alpha^{i+j-1}}{(2\alpha-1)^i} + \frac{1}{2^{i+j-1}}\binom{i+j-1}{i}\Big[1 - \frac{2\alpha-1}{2\alpha}\,{}_2F_1\big(1, i+j; j; \tfrac{1}{2\alpha}\big)\Big];$$

see Lemma B.5 for a detailed derivation. Additionally, we can easily prove that the function f_{iM} in Eq. (B.1) can be rewritten as

$$f_{iM} = \gamma^i f_{0M} + (1-\gamma)\sum_{q=1}^{i}\gamma^{i-q} f_{qL}, \tag{B.16}$$

as shown in Lemma B.4. □

Lemma B.1 *For $j = 4, \ldots, L+1$, we have the identity*

$$\mathbf{C}_{2j} = \frac{\rho_k \mu}{2}\mathbf{C}_{1j} + \frac{\mu}{2}\mathbf{C}_{2,j-1} = \frac{\rho_k \mu}{2}\sum_{q=0}^{j-4}\left(\frac{\mu}{2}\right)^q \mathbf{C}_{1,j-q} + \left(\frac{\mu}{2}\right)^{j-3}\mathbf{C}_{23}, \quad \text{(B.17)}$$

as stated in Eq. (B.8).

Proof The identity in Eq. (B.17) will be proved by induction: one can easily show that it holds for $j = 4$. Now, we assume that Eq. (B.17) is true for some $j \geq 5$; hence, for $j+1$, we have

$$\begin{aligned}
\mathbf{C}_{2,j+1} &= \sum_{q=0}^{j-3}\left(\frac{\mu}{2}\right)^q \frac{\rho_k \mu}{2}\mathbf{C}_{1,j+1-q} + \left(\frac{\mu}{2}\right)^{j-2}\mathbf{C}_{23} \\
&= \frac{\rho_k \mu}{2}\mathbf{C}_{1,j+1} + \sum_{q=1}^{j-3}\left(\frac{\mu}{2}\right)^q \frac{\rho_k \mu}{2}\mathbf{C}_{1,j+1-q} + \left(\frac{\mu}{2}\right)^{j-2}\mathbf{C}_{23} \\
&= \frac{\rho_k \mu}{2}\mathbf{C}_{1,j+1} + \frac{\mu}{2}\left[\sum_{q=1}^{j-3}\left(\frac{\mu}{2}\right)^{q-1} \frac{\rho_k \mu}{2}\mathbf{C}_{1,j+1-q} + \left(\frac{\mu}{2}\right)^{j-3}\mathbf{C}_{23}\right] \\
&= \frac{\rho_k \mu}{2}\mathbf{C}_{1,j+1} + \frac{\mu}{2}\left[\sum_{q=0}^{j-4}\left(\frac{\mu}{2}\right)^q \frac{\rho_k \mu}{2}\mathbf{C}_{1,j-q} + \left(\frac{\mu}{2}\right)^{j-3}\mathbf{C}_{23}\right] \\
&= \frac{\rho_k \mu}{2}\mathbf{C}_{1,j+1} + \frac{\mu}{2}\mathbf{C}_{2j}, \qquad \text{(B.18)}
\end{aligned}$$

as claimed, which implies that the identity in Eq. (B.17) holds for all $j = 4, \ldots, L+1$. □

Lemma B.2 *The function f_{1j}, which is defined by the recurrence relation $f_{1j} = (f_{0j} + f_{1,j-1})/2$ in Eq. (B.1), satisfies the identity*

$$f_{1j} = \sum_{q=0}^{j-3}\left(\frac{1}{2}\right)^{q+1} f_{0,j-q} + \left(\frac{1}{2}\right)^{j-2} f_{12} \quad \text{for } j = 3, \ldots, L, \qquad \text{(B.19)}$$

as stated in Eq. (B.9).

Proof We will again prove Eq. (B.19) by induction. For $j = 3$, we have from Eq. (B.19) that $f_{13} = (f_{03} + f_{12})/2$, which is true by the definition of f_{13}. We assume that the identity in Eq. (B.19) is correct for some $j \geq 4$; then, for $j+1$, the definition of

$f_{1,j+1}$, in combination with our assumption, implies

$$\begin{aligned} f_{1,j+1} &= \frac{1}{2} f_{0,j+1} + \frac{1}{2} f_{1j} = \frac{1}{2} f_{0,j+1} + \frac{1}{2}\Big[\sum_{q=0}^{j-3} \Big(\frac{1}{2}\Big)^{q+1} f_{0,j-q} + \Big(\frac{1}{2}\Big)^{j-2} f_{12}\Big] \\ &= \frac{1}{2} f_{0,j+1} + \frac{1}{2}\Big[\sum_{q=1}^{j-2} \Big(\frac{1}{2}\Big)^{q} f_{0,j+1-q} + \Big(\frac{1}{2}\Big)^{j-2} f_{12}\Big] \\ &= \sum_{q=0}^{j-2} \Big(\frac{1}{2}\Big)^{q+1} f_{1,j+1-q} + \Big(\frac{1}{2}\Big)^{j-1} f_{12}, \end{aligned}$$

as claimed. Hence, the equality in Eq. (B.19) is true for all $j = 3, \ldots, L$. □

Lemma B.3 *The system in Eq. (B.11), which is given by*

$$\begin{aligned} \mathbf{C}_{ii} &= \langle n_{i-1}\rangle + \mu \mathbf{C}_{i-1,i} \qquad \text{for } i = 3, \ldots, L, \\ \mathbf{C}_{i,i+1} &= \frac{\mu}{2}(\mathbf{C}_{i-1,i+1} + \mu \mathbf{C}_{i-1,i}) \quad \text{for } i = 3, \ldots, L, \\ \mathbf{C}_{ij} &= \frac{\mu}{2}(\mathbf{C}_{i-1,j} + \mathbf{C}_{i,j-1}) \quad \text{for } i = 3, \ldots, L+1 \text{ and } j = i+1, \ldots, L+1, \end{aligned} \tag{B.20}$$

is equivalent to the system

$$\mathbf{C}_{ij} = \delta_{ij}\langle n_{i-1}\rangle + \langle n_{i-1}\rangle\langle n_{j-1}\rangle\alpha\beta \cdot f_{i-1,j-1} \qquad \text{for } i, j = 3, \ldots, L+1, \tag{B.21}$$

as stated in Eq. (B.12). Here, the functions f_{ij} are defined as in Eq. (B.1).

Proof We again use the method of induction. For $i = 3$, we have

$$\begin{aligned} \mathbf{C}_{33} &= \langle n_2\rangle + \langle n_2\rangle^2\alpha\beta \cdot f_{22} = \langle n_2\rangle + \mu\langle n_1\rangle\langle n_2\rangle\alpha\beta \cdot f_{12} = \langle n_2\rangle + \mu\mathbf{C}_{23}, \\ \mathbf{C}_{34} &= \langle n_2\rangle\langle n_3\rangle(\beta\alpha) \cdot f_{34} = \langle n_2\rangle\langle n_3\rangle\alpha\beta(f_{22} + f_{13})/2 \\ &= [\langle n_2\rangle\langle n_3\rangle\alpha\beta \cdot f_{22} + \langle n_2\rangle\langle n_3\rangle\alpha\beta \cdot f_{13}]/2 \\ &= [\mu\langle n_1\rangle\langle n_3\rangle\alpha\beta \cdot f_{13} + \mu^2\langle n_1\rangle\langle n_2\rangle\alpha\beta \cdot f_{12}]/2 \\ &= \frac{\mu}{2}[\langle n_1\rangle\langle n_3\rangle\alpha\beta \cdot f_{13} + \mu\langle n_1\rangle\langle n_2\rangle\alpha\beta \cdot f_{12}] = \frac{\mu}{2}(\mathbf{C}_{24} + \mu\mathbf{C}_{23}), \\ \mathbf{C}_{3j} &= \langle n_2\rangle\langle n_{j-1}\rangle\alpha\beta \cdot f_{2,j-1} = \langle n_2\rangle\langle n_{j-1}\rangle\alpha\beta(f_{2,j-2} + f_{1,j-1})/2 \\ &= [\langle n_2\rangle\langle n_{j-1}\rangle\alpha\beta \cdot f_{2,j-2} + \langle n_2\rangle\langle n_{j-1}\rangle\alpha\beta \cdot f_{1,j-1}]/2 \\ &= [\langle n_2\rangle\mu\langle n_{j-2}\rangle\alpha\beta \cdot f_{2,j-2} + \mu\langle n_1\rangle\langle n_{j-1}\rangle\alpha\beta \cdot f_{1,j-1}]/2 \\ &= \frac{\mu}{2}[\langle n_2\rangle\langle n_{j-2}\rangle\alpha\beta \cdot f_{2,j-2} + \langle n_1\rangle\langle n_{j-1}\rangle\alpha\beta \cdot f_{1,j-1}] \\ &= \frac{\mu}{2}(\mathbf{C}_{3,j-1} + \mathbf{C}_{2j}) \end{aligned} \tag{B.22}$$

Now, we assume that the statement is true for some $i \geq 4$; then, for $i+1$, we have

$$\begin{aligned}
\mathbf{C}_{i+1,i+1} &= \langle n_i\rangle + \langle n_i\rangle^2\alpha\beta\cdot f_{ii} = \langle n_i\rangle + \mu\langle n_{i-1}\rangle\langle n_i\rangle\alpha\beta\cdot f_{i-1,i} = \langle n_i\rangle + \mu\mathbf{C}_{i,i+1},\\
\mathbf{C}_{i+1,i+2} &= \langle n_i\rangle\langle n_{i+1}\rangle\alpha\beta\cdot f_{i,i+1} = \langle n_i\rangle\langle n_{i+1}\rangle\alpha\beta(f_{i-1,i+1}+f_{ii})/2\\
&= [\langle n_i\rangle\langle n_{i+1}\rangle\alpha\beta\cdot f_{i-1,i+1} + \langle n_i\rangle\langle n_{i+1}\rangle\alpha\beta\cdot f_{ii}]/2\\
&= [\mu\langle n_{i-1}\rangle\langle n_{i+1}\rangle\alpha\beta\cdot f_{i-1,i+1} + \mu^2\langle n_{i-1}\rangle\langle n_i\rangle\alpha\beta\cdot f_{i-1,i}]/2\\
&= \frac{\mu}{2}[\langle n_{i-1}\rangle\langle n_{i+1}\rangle\alpha\beta\cdot f_{i-1,i+1} + \mu\langle n_{i-1}\rangle\langle n_i\rangle\alpha\beta\cdot f_{i-1,i}]\\
&= \frac{\mu}{2}(\mathbf{C}_{i,i+2} + \mu\mathbf{C}_{i,i+1}),\\
\mathbf{C}_{i+1,j} &= \langle n_i\rangle\langle n_{j-1}\rangle\alpha\beta\cdot f_{i,j-1} = \langle n_i\rangle\langle n_{j-1}\rangle\alpha\beta(f_{i-1,j-1}+f_{i,j-2})/2\\
&= [\langle n_i\rangle\langle n_{j-1}\rangle\alpha\beta\cdot f_{i-1,j-1} + \langle n_i\rangle\langle n_{j-1}\rangle\alpha\beta\cdot f_{i,j-2}]/2\\
&= [\mu\langle n_{i-1}\rangle\langle n_{j-1}\rangle\alpha\beta\cdot f_{i-1,j-1} + \mu\langle n_i\rangle\langle n_{j-2}\rangle\alpha\beta\cdot f_{i,j-2}]/2\\
&= \frac{\mu}{2}[\langle n_{i-1}\rangle\langle n_{j-1}\rangle\alpha\beta\cdot f_{i-1,j-1} + \langle n_i\rangle\langle n_{j-2}\rangle\alpha\beta\cdot f_{i,j-2}]\\
&= \frac{\mu}{2}(\mathbf{C}_{ij} + \mathbf{C}_{i+1,j-1}),
\end{aligned} \tag{B.23}$$

which is also correct. Hence, the statement of the lemma is true for all i and j, as stated. □

Lemma B.4 *For $i = 1,\ldots,L$, the function f_{iM} defined in Eq. (B.1) can be simplified as in Eq. (B.16); specifically, we have the identity*

$$f_{iM} = \gamma f_{i-1,M} + (1-\gamma)f_{i,L} = \gamma^i\cdot f_{0M} + (1-\gamma)\sum_{q=1}^{i}\gamma^{i-q}\cdot f_{qL}. \tag{B.24}$$

Proof The proof is by induction: for $i = 1$, the identity is obvious. We now suppose that Eq. (B.24) is true for some $i \geq 2$; hence, for $i+1$, we have

$$\begin{aligned}
f_{i+1,M} &= \gamma^{i+1}\cdot f_{0M} + (1-\gamma)\sum_{q=1}^{i+1}\gamma^{i+1-q}\cdot f_{qL}\\
&= \gamma\left[\gamma^i\cdot f_{0M} + (1-\gamma)\sum_{q=1}^{i}\gamma^{i-q}\cdot f_{qL}\right] + (1-\gamma)f_{i+1,L}\\
&= \gamma f_{iM} + (1-\gamma)f_{i+1,L},
\end{aligned} \tag{B.25}$$

which is correct. Hence, Eq. (B.24) is true for all i, as stated. □

Lemma B.5 *For $i, j = 1,\ldots,L$, the solution of the recurrence relation $f_{ij} = (f_{i,j-1} + f_{i-1,j})/2$ in Eq. (B.1) is given by $f_{ij} = f(i,j) + f(j,i)$, where*

$$f(i,j) = \frac{\alpha^{i+j-1}}{(2\alpha-1)^i} + \frac{1}{2^{i+j-1}}\binom{i+j-1}{i}\Big[1 - \frac{2\alpha-1}{2\alpha}\,{}_2F_1\big(1, i+j; j; \tfrac{1}{2\alpha}\big)\Big]. \tag{B.26}$$

Proof In order to solve the recurrence relation for the function f_{ij}, we take into account the initial conditions $f_{00} = 1$ and $f_{0j} = f_{j0} = \alpha^{j-1}$. Then, we define a generating function $g(x, y)$ via

$$g(x,y) = \sum_{i,j\geq 0} f_{ij}x^i y^j = f_{00} + \sum_{j\geq 1} f_{0j}y^j + \sum_{i\geq 1} f_{i0}x^i + \sum_{i,j\geq 1} f_{ij}x^i y^j, \tag{B.27}$$

where the last term can be rewritten as

$$\begin{aligned}
\sum_{i,j\geq 1} f_{ij}x^i y^j &= \sum_{i,j\geq 1} \frac{1}{2}(f_{i-1,j} + f_{i,j-1})x^i y^j \\
&= \frac{1}{2}x \sum_{i,j\geq 1} f_{i-1,j}x^{i-1}y^j + \frac{1}{2}y \sum_{i,j\geq 1} f_{i,j-1}x^i y^{j-1} \\
&= \frac{1}{2}x \sum_{i\geq 0}\sum_{j\geq 1} f_{ij}x^i y^j + \frac{1}{2}y \sum_{i\geq 1}\sum_{j\geq 0} f_{ij}x^i y^j \\
&= \frac{1}{2}x\Big(\sum_{i,j\geq 0} f_{ij}x^i y^j - \sum_{i\geq 0} f_{i0}x^i\Big) + \frac{1}{2}y\Big(\sum_{i,j\geq 0} f_{ij}x^i y^j - \sum_{j\geq 0} f_{0j}y^j\Big) \\
&= \frac{1}{2}x\Big(g(x,y) - \sum_{i\geq 0} f_{i0}x^i\Big) + \frac{1}{2}y\Big(g(x,y) - \sum_{j\geq 0} f_{0j}y^j\Big).
\end{aligned} \tag{B.28}$$

Hence, Eq. (B.27) becomes

$$\begin{aligned}
g(x,y) \;&= f_{00} + \sum_{j\geq 1} f_{0j}y^j + \sum_{i\geq 1} f_{i0}x^i + \frac{1}{2}x\Big(g(x,y) - \sum_{i\geq 0} f_{i0}x^i\Big) \\
&\quad + \frac{1}{2}y\Big(g(x,y) - \sum_{j\geq 0} f_{0j}y^j\Big),
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
g(x,y)\Big(1 - \frac{1}{2}x - \frac{1}{2}y\Big) \;&= f_{00}\Big(1 - \frac{1}{2}x - \frac{1}{2}y\Big) + \Big(1 - \frac{1}{2}y\Big)\sum_{j\geq 1} f_{0j}y^j \\
&\quad + \Big(1 - \frac{1}{2}x\Big)\sum_{i\geq 1} f_{i0}x^i
\end{aligned}$$

or

$$g(x,y) = f_{00} + \left(1-\frac{1}{2}y\right)\frac{1}{1-\frac{1}{2}x-\frac{1}{2}y}\sum_{j\geq 1} f_{0j}y^j + \left(1-\frac{1}{2}x\right)\frac{1}{1-\frac{1}{2}x-\frac{1}{2}y}\sum_{i\geq 1} f_{i0}x^i. \tag{B.29}$$

Taking into account the initial conditions, we find that

$$\sum_{j\geq 1} f_{0j}y^j = \sum_{j\geq 1}\alpha^{j-1}y^j = \frac{1}{\alpha}\sum_{j\geq 1}(\alpha y)^j \quad \text{and} \quad \sum_{i\geq 1} f_{i0}x^i = \frac{1}{\alpha}\sum_{i\geq 1}(\alpha x)^i, \tag{B.30}$$

which we substitute into Eq. (B.29) to obtain

$$g(x,y) = 1 + \left(1-\frac{1}{2}y\right)\frac{1}{1-\frac{1}{2}x-\frac{1}{2}y}\frac{1}{\alpha}\sum_{j\geq 1}(\alpha y)^j + \left(1-\frac{1}{2}x\right)\frac{1}{1-\frac{1}{2}x-\frac{1}{2}y}\frac{1}{\alpha}\sum_{i\geq 1}(\alpha x)^i. \tag{B.31}$$

Making use of the well-known symmetric, bivariate generating function of the binomial coefficients

$$\frac{1}{1-s-t} = \sum_{i,j\geq 0}\binom{i+j}{i}s^i t^j, \tag{B.32}$$

we can rewrite Eq. (B.31) as

$$\begin{aligned} g(x,y) &= 1 + \left(1-\frac{1}{2}y\right)\frac{1}{\alpha}\sum_{j\geq 1}(\alpha y)^j \sum_{i,j\geq 0}\binom{i+j}{i}\frac{x^i y^j}{2^{i+j}} \\ &\quad + \left(1-\frac{1}{2}x\right)\frac{1}{\alpha}\sum_{i\geq 1}(\alpha x)^i \sum_{i,j\geq 0}\binom{i+j}{i}\frac{x^i y^j}{2^{i+j}} \\ &= \left(1-\frac{1}{2}y\right)\sum_{i,j\geq 0}\sum_{q=0}^{j-1}\binom{i+q}{i}\frac{\alpha^{j-q-1}}{2^{i+q}}x^i y^j \\ &\quad + \left(1-\frac{1}{2}x\right)\sum_{i,j\geq 0}\sum_{q=0}^{i-1}\binom{j+q}{q}\frac{\alpha^{i-q-1}}{2^{j+q}}x^i y^j. \end{aligned}$$

Rearranging sums in the above expression, we find

$$g(x,y) = \sum_{i,j\geq 0}\Bigg[\sum_{q=0}^{j-1}\binom{i+q}{i}\frac{\alpha^{j-q-1}}{2^{i+q}} - \sum_{q=0}^{j-2}\binom{i+q}{i}\frac{\alpha^{j-q-2}}{2^{i+q+1}} + \sum_{q=0}^{i-1}\binom{j+q}{q}\frac{\alpha^{i-q-1}}{2^{j+q}} - \sum_{q=0}^{i-2}\binom{j+q}{q}\frac{\alpha^{i-q-2}}{2^{j+q+1}}\Bigg]x^i y^j.$$

Hence, we obtain the following exact expression for the function f_{ij},

$$f_{ij} = \sum_{q=0}^{j-1}\binom{i+q}{q}\frac{\alpha^{j-q-1}}{2^{i+q}} - \sum_{q=0}^{j-2}\binom{i+q}{q}\frac{\alpha^{j-q-2}}{2^{i+q+1}} + \sum_{q=0}^{i-1}\binom{j+q}{q}\frac{\alpha^{i-q-1}}{2^{j+q}} - \sum_{q=0}^{i-2}\binom{j+q}{q}\frac{\alpha^{i-q-2}}{2^{j+q+1}}. \tag{B.33}$$

The expression in Eq. (B.33) can be simplified further due to its symmetry with respect to the indices i and j: we write $f_{ij} = f(i,j) + f(j,i)$, where $f(i,j)$ is defined as

$$f(i,j) = \sum_{q=0}^{j-1}\binom{i+q}{q}\frac{\alpha^{j-q-1}}{2^{i+q}} - \sum_{q=0}^{j-2}\binom{i+q}{q}\frac{\alpha^{j-q-2}}{2^{i+q+1}}. \tag{B.34}$$

The function $f(i,j)$ can be further simplified as

$$\begin{aligned} f(i,j) &= \binom{i+j-1}{j-1}\frac{1}{2^{i+j-1}} + 2\alpha\sum_{q=0}^{j-2}\binom{i+q}{q}\frac{\alpha^{j-q-2}}{2^{i+q+1}} - \sum_{q=0}^{j-2}\binom{i+q}{q}\frac{\alpha^{j-q-2}}{2^{i+q+1}} \\ &= \binom{i+j-1}{j-1}\frac{1}{2^{i+j-1}} + (2\alpha-1)\frac{\alpha^{j-2}}{2^{i+1}}\sum_{q=0}^{j-2}\binom{i+q}{q}\Big(\frac{1}{2\alpha}\Big)^q; \end{aligned} \tag{B.35}$$

next, we use the identity

$$\sum_{q=0}^{j}\binom{i+q}{i}x^q = \frac{1}{(1-x)^{i+1}} - x^{j+1}\binom{j+1+i}{j+1}{}_2F_1(1, j+i+2; j+2; x), \tag{B.36}$$

where ${}_2F_1$ is again the generalised hypergeometric function of the second kind (Digital Library of Mathematical Functions 2020a). Note that the above identity can be used only when $|x| < 1$, as the hypergeometric function ${}_2F_1$ is not defined otherwise.

Hence, Eq. (B.35) becomes

$$
\begin{aligned}
f(i,j) &= \binom{i+j-1}{j-1}\frac{1}{2^{i+j-1}} + (2\alpha-1)\frac{\alpha^{j-2}}{2^{i+1}} \\
&\quad \times\left[\left(\frac{2\alpha}{2\alpha-1}\right)^{i+1} - \frac{1}{(2\alpha)^{j-1}}\binom{j+i-1}{j-1}{}_2F_1\left(1, j+i; j; \tfrac{1}{2\alpha}\right)\right] \\
&= \frac{\alpha^{i+j-1}}{(2\alpha-1)^i} + \frac{1}{2^{i+j-1}}\binom{i+j-1}{i}\left[1 - \frac{2\alpha-1}{2\alpha}{}_2F_1\left(1, j+i; j; \tfrac{1}{2\alpha}\right)\right]
\end{aligned}
\tag{B.37}
$$

Given the expression for $f(i,j)$ in Eq. (B.37), one can find the corresponding expression for $f(j,i)$ by exchanging the indexes $i \leftrightarrow j$. □

C. Variance of Total RNAP Distribution

In this section, we derive the exact expression for the variance of the total RNAP distribution, as stated in Eq. (10), which is given by the sum over the covariances $\mathrm{Cov}(x_i, x_j)$ $(i, j = 1, \ldots, L)$, as defined in Eq. (4d). Hence, we have

$$
\begin{aligned}
\mathrm{Var}(n_{\mathrm{tot}}) &= \sum_{i,j=1}^{L}\mathrm{Cov}(n_i, n_j) = \sum_{i=1}^{L}\mathrm{Var}(n_i) + \sum_{i\neq j}\mathrm{Cov}(n_i, n_j) \\
&= \sum_{i=1}^{L}\left[\langle n_i\rangle + \langle n_i\rangle^2\alpha\beta\cdot f_{ii}\right] + \sum_{i\neq j}\langle n_i\rangle\langle n_j\rangle\alpha\beta\cdot f_{ij} \\
&= \sum_{i=1}^{L}\langle n_i\rangle + \alpha\beta\left(\sum_{i=1}^{L}\langle n_i\rangle^2\cdot f_{ii} + \sum_{i\neq j}\langle n_i\rangle\langle n_j\rangle\cdot f_{ij}\right) \\
&= \sum_{i=1}^{L}\langle n_i\rangle + \alpha\beta\sum_{i,j=1}^{L}\langle n_i\rangle\langle n_j\rangle\cdot f_{ij},
\end{aligned}
\tag{C.1}
$$

where the function f_{ij} is given in Eq. (10). The first term in Eq. (C.1) equals $\langle n_{\mathrm{tot}}\rangle$, the mean of the total RNAP distribution, as stated in Eq. (10); substituting in the expressions for the means $\langle n_i\rangle$ from Eq. (2b), as well, we obtain

$$
\mathrm{Var}(n_{\mathrm{tot}}) = \langle n_{\mathrm{tot}}\rangle + \alpha\beta(\eta\rho_k)^2\sum_{i,j=1}^{L}\mu^{i+j}\cdot f_{ij}. \tag{C.2}
$$

Lemma C.1 *In the limit of deterministic elongation, i.e. for* $L \to \infty$*, the expression for* $\mathrm{Var}(n_{\mathrm{tot}})$ *in Eq. (10) simplifies to*

$$\begin{aligned}\mathrm{Var}(n_{\mathrm{tot}})_\infty &= \langle n_{\mathrm{tot}}\rangle_\infty \\ &\quad +\beta(\eta r)^2\frac{(s_b+s_u-d)-(s_b+s_u+d)e^{-2d\langle T\rangle}+2de^{-(s_b+s_u+d)\langle T\rangle}}{d(s_b+s_u+d)(s_b+s_u-d)},\end{aligned} \tag{C.3}$$

which can be further simplified to the expression in Eq. (11).

Proof In order to find the limit of $L \to \infty$ in Eq. (10) (or Eq. (C.2)), we have to evaluate the term $\sum_{i,j=1}^{L}\mu^{i+j}\cdot f_{ij}$ in that limit. For the following derivation, we consider the function $f_{ij} = f(i,j)+f(j,i)$, where $f(i,j)$ is defined in terms of sums in Eq. (B.34). Hence, we have

$$\begin{aligned}\sum_{i,j=1}^{L}\mu^{i+j}\cdot f_{ij} &= \sum_{i,j=1}^{L}\mu^{i+j}f(i,j)+\sum_{i,j=1}^{L}\mu^{i+j}f(j,i) = 2\sum_{i,j=1}^{L}\mu^{i+j}f(i,j)\\ &= 2\Bigg[\sum_{i,j=1}^{L}\sum_{q=0}^{j-1}\mu^{i+j}\binom{i+q}{q}\frac{\alpha^{j-q-1}}{2^{i+q}}-\sum_{i,j=1}^{L}\sum_{q=0}^{j-2}\mu^{i+j}\binom{i+q}{q}\frac{\alpha^{j-q-2}}{2^{i+q+1}}\Bigg]\\ &= \underbrace{2\sum_{i,j=1}^{L}\mu^{i+j}\binom{i+j-1}{i}\frac{1}{2^{i+j-1}}}_{G_1}+\underbrace{2(2\alpha-1)\sum_{i,j=1}^{L}\sum_{q=0}^{j-2}\mu^{i+j}\binom{i+q}{q}\frac{\alpha^{j-q-2}}{2^{i+q+1}}}_{G_2}.\end{aligned} \tag{C.4}$$

Substituting $k \to L/\langle T\rangle - d$ in Eq. (C.4) and taking the limit of $L \to \infty$, we have that $G_1 \underset{L\to\infty}{\longrightarrow} 0$; hence, $\mathrm{Var}(n_{\mathrm{tot}})$ evaluates to

$$\mathrm{Var}(n_{\mathrm{tot}})_\infty = \langle n_{\mathrm{tot}}\rangle_\infty + \lim_{L\to\infty}[\alpha\beta(\eta\rho_k)^2 G_2] \tag{C.5}$$

in that limit, which yields the expression in Eq. (C.3), as can easily be verified with the computer algebra package Mathematica. Hence, in the limit of deterministic elongation, the expression for the variance of the RNAP distribution in Eq. (10) reduces to the one in Eq. (11), as claimed. □

D. Moments of Total RNAP and Mature RNA in Bursty and Constitutive Limits

Moments of total RNAP in the bursty limit: In the bursty limit, the expressions for the mean and variance of the total RNAP distribution given in Eq. (10) simplify to

$$\langle n_{\rm tot}\rangle_b = b\frac{s_u}{k}\mu\frac{\mu^L-1}{\mu-1}, \quad \text{and} \quad \text{Var}(n_{\rm tot})_b = \langle n_{\rm tot}\rangle_b + b^2\frac{s_u}{k}\sum_{i,j=1}^{L}\mu^{i+j}\cdot h_{ij}. \tag{D.1}$$

If, furthermore, we take the limit of deterministic elongation, with $L\to\infty$ at constant $\langle T\rangle$, Eq. (D.1) simplifies to

$$\begin{aligned}\langle n_{\rm tot}\rangle_{(b;\infty)} &= b\frac{s_u}{d}(1-\mathrm{e}^{-T_d}) \quad \text{and} \quad \text{Var}(n_{\rm tot})_{(b;\infty)}\\ &= \langle n_{\rm tot}\rangle_{(b;\infty)} + \langle n_{\rm tot}\rangle^2_{(b;\infty)}\frac{d}{s_u}\frac{1+\mathrm{e}^{-T_d}}{1-\mathrm{e}^{-T_d}},\end{aligned} \tag{D.2}$$

where the subscript $(b;\infty)$ denotes the bursty limit with infinite L. In the limit of zero RNAP detachment, Eq. (D.2) further simplifies to

$$\langle n_{\rm tot}\rangle_{(b;\infty;0)} = bs_u\langle T\rangle \quad \text{and} \quad \text{Var}(n_{\rm tot})_{(b;\infty;0)} = \langle n_{\rm tot}\rangle_{(b;\infty;0)}(1+2b), \tag{D.3}$$

where the subscript $(b;\infty;0)$ denotes the bursty limit, with $L\to\infty$ and $d\to 0$.

Moments of total RNAP in the constitutive limit: In the constitutive limit, Eq. (10) simplifies to

$$\langle n_{\rm tot}\rangle_c = \frac{r}{k}\mu\frac{\mu^L-1}{\mu-1} = \text{Var}(n_{\rm tot})_c. \tag{D.4}$$

If, furthermore, we take the limit of deterministic elongation, i.e. $L\to\infty$ at constant $\langle T\rangle$, Eq. (D.4) simplifies to

$$\langle n_{\rm tot}\rangle_{(c;\infty)} = \frac{r}{d}(1-\mathrm{e}^{-T_d}) = \text{Var}(n_{\rm tot})_{(c;\infty)}; \tag{D.5}$$

finally, in the limit of zero RNAP detachment, Eq. (D.5) further simplifies to

$$\langle n_{\rm tot}\rangle_{(c;\infty;0)} = r\langle T\rangle = \text{Var}(n_{\rm tot})_{(c;\infty;0)}. \tag{D.6}$$

Moments of mature RNA distribution in the bursty limit: In that limit, the closed-form expressions in Eq. (8) are given by

$$\langle n\rangle_b = b\upsilon_m\mu^L \quad \text{and} \quad \text{Var}(n)_b = \langle n\rangle_b + \langle n\rangle_b^2(\upsilon_k\mu)^{-1}\cdot h_{MM}, \tag{D.7}$$

which in the limit of deterministic elongation simplify to

$$\langle n\rangle_{(b;\infty)} = b\upsilon_m\mathrm{e}^{-T_d} \quad \text{and} \quad \text{Var}(n)_{(b;\infty)} = \langle n\rangle_{(b;\infty)} + \langle n\rangle^2_{(b;\infty)}\upsilon_m^{-1}. \tag{D.8}$$

In the limit of zero RNAP detachment, these expressions further simplify to

$$\langle n\rangle_{(b;\infty;0)} = b\upsilon_m \quad \text{and} \quad \text{Var}(n)_{(b;\infty;0)} = \langle n\rangle_{(b;\infty;0)} + \langle n\rangle^2_{(b;\infty;0)}\upsilon_m^{-1}. \tag{D.9}$$

E. Introduction to Geometric Singular Perturbation Theory (GSPT)

We consider a system of first-order autonomous ordinary differential equations in the general ('standard') form

$$\varepsilon \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{y}, \varepsilon), \tag{E.1}$$
$$\dot{\mathbf{y}} = \mathbf{g}(\mathbf{x}, \mathbf{y}, \varepsilon), \tag{E.2}$$

where $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^m \times \mathbb{R}^l$, with $m, l \in \mathbb{N}$. Here, $0 < \varepsilon \ll 1$ is a (real) singular perturbation parameter, and the overdot denotes differentiation with respect to the 'slow' time t. (Correspondingly, Eq. (E.1) is referred to as the 'slow' system.) The variable $\mathbf{x}$ is referred to as the 'fast variable', while $\mathbf{y}$ is the 'slow variable'. For simplicity, the functions $\mathbf{f} : \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R}^+ \to \mathbb{R}^m$ and $\mathbf{g} : \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R}^+ \to \mathbb{R}^l$ are assumed to be $\mathcal{C}^\infty$-smooth in all their arguments. In the context of our analysis of the characteristic system in Eq. (26), we have the 'slow system'

$$\varepsilon \dot{u}_i = u_i - u_{i+1} + (d/k)u_i \qquad \text{for } i = 1, \ldots, L-1, \tag{E.3a}$$
$$\varepsilon \dot{u}_L = u_L - u + (d/k)u_L, \tag{E.3b}$$
$$\dot{u} = u, \tag{E.3c}$$
$$\dot{F}_0 = (s_u/d_m)F_1 - (s_b/d_m)F_0 + (r/d_m)u_1 F_0, \tag{E.3d}$$
$$\dot{F}_1 = (s_b/d_m)F_0 - (s_u/d_m)F_1. \tag{E.3e}$$

By comparing the system of equations in Eq. (E.3) with the general form in Eq. (E.1), we see that u_i $(i = 1, \ldots, L)$ are the fast variables, while u, F_0, and F_1 are slow. Correspondingly, we have $m = L$ and $l = 3$ in the above notation, which implies $\mathbf{f} = (f_1, f_2, \ldots, f_L)$, with $f_i = f_i(u_i, u_{i+1}) = u_i - u_{i+1} + (d/k)u_i$ for $i = 1, \ldots, L-1$, $f_L = f_L(u_L, u) = u_L - u + (d/k)u_L$, and $\mathbf{g} = (g_1, g_2, g_3)(u_1, u, F_0, F_1) = (u, (s_u/d_m)F_1 - (s_b/d_m)F_0 + (r/d_m)u_1 F_0, (s_b/d_m)F_0 - (s_u/d_m)F_1)$.

Now, we introduce a new 'fast' time $\tau = t/\varepsilon$, which we substitute into Eq. (E.1) to find the 'fast system'

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}, \mathbf{y}, \varepsilon), \tag{E.4a}$$
$$\mathbf{y}' = \varepsilon \mathbf{g}(\mathbf{x}, \mathbf{y}, \varepsilon) \tag{E.4b}$$

corresponding to Eq. (E.1); here, the prime denotes the derivative with respect to τ. Hence, rewriting Eq. (E.3) in the fast formulation, we find

$$u_i' = u_i - u_{i+1} + (d/k)u_i \qquad \text{for } i = 1, \ldots, L-1, \tag{E.5a}$$
$$u_L' = u_L - u + (d/k)u_L, \tag{E.5b}$$
$$u' = \varepsilon u, \tag{E.5c}$$
$$F_0' = \varepsilon[(s_u/d_m)F_1 - (s_b/d_m)F_0 + (r/d_m)u_1 F_0], \tag{E.5d}$$
$$F_1' = \varepsilon[(s_b/d_m)F_0 - (s_u/d_m)F_1]. \tag{E.5e}$$

For positive ε, the systems in Eqs. (E.1) and (E.4)—and, correspondingly, the systems in Eqs. (E.3) and (E.5)—are equivalent; however, in the singular limit of $\varepsilon \to 0$, we obtain two different systems: setting $\varepsilon = 0$ in Eq. (E.1), we have the 'reduced problem'

$$\mathbf{0} = \mathbf{f}(\mathbf{x}, \mathbf{y}, 0), \tag{E.6a}$$
$$\dot{\mathbf{y}} = \mathbf{g}(\mathbf{x}, \mathbf{y}, 0), \tag{E.6b}$$

while we obtain the 'layer problem'

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}, \mathbf{y}, 0), \tag{E.7a}$$
$$\mathbf{y}' = \mathbf{0} \tag{E.7b}$$

for $\varepsilon = 0$ in Eq. (E.4). The 'reduced problem' for the system in Eq. (E.3) implies that the flow of (u, F_0, F_1) is constrained to lie on the $(l = 3)$-dimensional 'critical manifold' $\mathcal{S}_0$ that is defined by $\mathbf{f} = \mathbf{0}$:

$$u_i = \mu \cdot u_{i+1} = \mu^{i+L-1} \cdot u \qquad \text{for } i = 1, \ldots, L, \tag{E.8}$$

where $u_{L+1} \equiv u$ and (F_0, F_1) are assumed to vary in an appropriately chosen subset of $\mathbb{R}^2$.

From the 'layer problem' of the system in Eq. (E.3), we conclude that $\mathbf{y} = (u, F_0, F_1)$ is a parameter which parameterises the $(m = L)$-dimensional flow of $u_i' = f_i$ $(i = 1, \ldots, L)$, the equilibria of which are located on $\mathcal{S}_0$.

The Jacobian matrix $\mathbf{D_x f}(\mathbf{x}, \mathbf{y}, 0)$ of the 'layer problem' corresponding to Eq. (E.5) about $\mathcal{S}_0$ has the eigenvalues

$$\lambda_i = k(1 + (d/k) - u_{i+1}) = (k + d)(1 - \mu^{i+L+1} u) \qquad \text{for } i = 1, \ldots, L. \tag{E.9}$$

Since our definition of the generating function $F(z, \tau)$ in Sect. 3.2 assumed $z \in [-1, 1]$, we may restrict to $u \in [-2, 0]$ which, by Eq. (E.9), implies that $\lambda_i > 0$. Hence, the critical manifold $\mathcal{S}_0$ is 'normally hyperbolic'—and, in fact, normally repelling—with an $(m + l = L + 3)$-dimensional unstable manifold $\mathcal{W}^{\mathrm{u}}(\mathcal{S}_0)$.

The geometric singular perturbation theory due to Fenichel (1979) thus implies that $\mathcal{S}_0$ will persist, for ε positive and sufficiently small, as a slow manifold' $\mathcal{S}_\varepsilon$ that is (locally) invariant, smooth, and $\mathcal{O}(\varepsilon)$-close to $\mathcal{S}_0$. (As the unstable manifold $\mathcal{W}^{\mathrm{u}}(\mathcal{S}_0)$ equals the entire phase space of Eq. (E.3), it trivially persists as the unstable manifold $\mathcal{W}^{\mathrm{u}}(\mathcal{S}_\varepsilon)$ for $\mathcal{S}_\varepsilon$.) In particular, as $\mathcal{S}_0$ is repelling in forward time, it follows that the inverse characteristic transformation corresponding to Eq. (26) is well defined in backward time; details can be found in Veerman et al. (2018), Popović et al. (2016).

F. Variance of Fluctuating Total Fluorescent Signal

By definition, the variance of the total fluorescent signal is given by the sum over all elements $\mathrm{Cov}(r_i, r_j)$ for $i, j = 1, \ldots, L$, where $r_i = (\nu/L) i n_i$; the corresponding definitions can be found in Sect. 4 of the main text. Hence, we have that

$$\begin{aligned}
\mathrm{Var}(r_{\mathrm{tot}}) &= \sum_{i,j=1}^{L} \mathrm{Cov}(r_i, r_j) = \sum_{i,j=1}^{L} \mathrm{Cov}\Big(\frac{\nu}{L} i n_i, \frac{\nu}{L} j n_j\Big) = \Big(\frac{\nu}{L}\Big)^2 \sum_{i,j=1}^{L} ij \cdot \mathrm{Cov}(n_i, n_j) \\
&= \Big(\frac{\nu}{L}\Big)^2 \Big(\sum_{i=1}^{L} i^2 \mathrm{Var}(n_i) + \sum_{i \neq j} ij \cdot \mathrm{Cov}(n_i, n_j) \Big) \\
&= \Big(\frac{\nu}{L}\Big)^2 \Big(\sum_{i=1}^{L} i^2 [\langle n_i \rangle + \langle n_i \rangle^2 \alpha\beta \cdot f_{ii}] + \sum_{i \neq j} ij \langle n_i \rangle \langle n_j \rangle \alpha\beta \cdot f_{ij} \Big) \qquad \text{(F.1)} \\
&= \Big(\frac{\nu}{L}\Big)^2 \sum_{i=1}^{L} i^2 \langle n_i \rangle + \Big(\frac{\nu}{L}\Big)^2 \alpha\beta \Big(\sum_{i=1}^{L} i^2 \langle n_i \rangle^2 \cdot f_{ii} + \sum_{i \neq j} ij \langle n_i \rangle \langle n_j \rangle \cdot f_{ij} \Big) \\
&= \Big(\frac{\nu}{L}\Big)^2 \sum_{i=1}^{L} i^2 \langle n_i \rangle + \Big(\frac{\nu}{L}\Big)^2 \alpha\beta \sum_{i,j=1}^{L} ij \langle n_i \rangle \langle n_j \rangle \cdot f_{ij}.
\end{aligned}$$

Substituting the expressions for the means $\langle n_i \rangle$ from Eq. (2b) into Eq. (F.1), we obtain

$$\mathrm{Var}(r_{\mathrm{tot}}) = \Big(\frac{\nu}{L}\Big)^2 \eta\rho_k \sum_{i=1}^{L} i^2 \mu^i + \Big(\frac{\nu}{L}\Big)^2 \alpha\beta (\eta\rho_k)^2 \sum_{i,j=1}^{L} ij \cdot \mu^{i+j} \cdot f_{ij}, \qquad \text{(F.2)}$$

which is the expression stated in Eq. (35).

G. Moments of Fluctuations in Total Fluorescent Signal in Various Limits

Deterministic elongation Substituting $k \mapsto L/\langle T \rangle - d$ and taking the long-gene limit of $L \to \infty$ in Eq. (35), we obtain the simplified expressions

$$\begin{aligned}
\langle r_{\mathrm{tot}} \rangle_\infty &= \frac{\nu\eta r}{\mathrm{d} T_d} [1 - (1 + T_d)\mathrm{e}^{-T_d}], \\
\mathrm{Var}(r_{\mathrm{tot}})_\infty &= \langle r_{\mathrm{tot}} \rangle_\infty \cdot \mathcal{F}_0 + \langle r_{\mathrm{tot}} \rangle_\infty^2 \cdot \beta\delta_g \frac{\mathcal{F}_1 + \mathcal{F}_2 + \mathcal{F}_3}{2(\delta_g - 1)^2 (\delta_g + 1)^2 [1 - (1 + T_d)\mathrm{e}^{-T_d}]^2},
\end{aligned} \qquad \text{(G.1)}$$

where

$$\begin{aligned}
\mathcal{F}_0 &= \nu\left[\frac{2}{T_d} - \frac{T_d \mathrm{e}^{-T_d}}{1-(1+T_d)\mathrm{e}^{-T_d}}\right],\\
\mathcal{F}_1 &= (\delta_g - 1)^2(2\delta_g + 1),\\
\mathcal{F}_2 &= (\delta_g + 1)^2[2\delta_g(1+T_d)(1+T_d-T_g) - 1]\mathrm{e}^{-2T_d},\\
\mathcal{F}_3 &= -4\delta_g^3(1+T_d+T_g)\mathrm{e}^{-T_g}\mathrm{e}^{-T_d};
\end{aligned} \tag{G.2}$$

the expression for the variance in Eq. (G.1) is found via the same method as is used in Lemma C.1 of 'Appendix C'. When there is no detachment of RNAP from the gene, i.e. when $d = 0$, Eq. (G.1) simplifies to

$$\begin{aligned}
\langle r_{\mathrm{tot}}\rangle_{(\infty;0)} &= \frac{1}{2}\nu\eta r\langle T\rangle,\\
\mathrm{Var}(r_{\mathrm{tot}})_{(\infty;0)} &= \langle r_{\mathrm{tot}}\rangle_{(\infty;0)}\frac{2\nu}{3}\\
&\quad + \langle r_{\mathrm{tot}}\rangle^2_{(\infty;0)}\cdot 8\beta T_g^{-1}\Big[\frac{1}{3} - \frac{1}{2}T_g^{-1} + T_g^{-3} - T_g^{-3}\big(1+T_g\big)\mathrm{e}^{-T_g}\Big].
\end{aligned} \tag{G.3}$$

Bursty limit: In the limit when the rates s_b and r are large, the expressions for the mean and variance of the total fluorescent signal given in Eq. (35) become

$$\begin{aligned}
\langle r_{\mathrm{tot}}\rangle_b &= \nu b\frac{s_u}{d}\Big(\frac{k}{d}\frac{(1-\mu^L)}{\mu L} - \mu^L\Big),\\
\mathrm{Var}(r_{\mathrm{tot}})_b &= \Big(\frac{\nu}{L}\Big)^2 b\frac{s_u}{k}\sum_{i=1}^{L} i^2\mu^i + \Big(\frac{\nu}{L}\Big)^2 b^2\frac{s_u}{k}\sum_{i,j=1}^{L} ij\cdot\mu^{i+j}\cdot f_{ij}.
\end{aligned} \tag{G.4}$$

Constitutive limit: When the gene spends most of its time in the active state, Eq. (35) simplifies to

$$\begin{aligned}
\langle r_{\mathrm{tot}}\rangle_c &= \frac{\nu}{L}\rho_k\mu\frac{1+\mu^L[L(\mu-1)-1]}{(\mu-1)^2},\\
\mathrm{Var}(r_{\mathrm{tot}})_c &= \Big(\frac{\nu}{L}\Big)^2\rho_k\mu\frac{1+\mu-\mu^L[L^2\mu^2+(1+L)^2\mu-(2L^2+2L-1)]}{(1-\mu)^3}.
\end{aligned} \tag{G.5}$$

Bursty expression with deterministic elongation. In this case, Eq. (G.4) simplifies to

$$\begin{aligned}
\langle r_{\mathrm{tot}}\rangle_{(b;\infty)} &= \frac{\nu b s_u}{\mathrm{d}T_d}\big[1-(1+T_d)\mathrm{e}^{-T_d}\big],\\
\mathrm{Var}(r_{\mathrm{tot}})_{(b;\infty)} &= \langle r_{\mathrm{tot}}\rangle_{(b;\infty)}\cdot\mathcal{F}_0 + \langle r_{\mathrm{tot}}\rangle^2_{(b;\infty)}\cdot\frac{d}{2s_u}\frac{1-(1+2T_d+2T_d^2)\mathrm{e}^{-2T_d}}{\big[1-(1+T_d)\mathrm{e}^{-T_d}\big]^2},
\end{aligned} \tag{G.6}$$

where $\mathcal{F}_0$ is given by Eq. (G.2). In the special case of no premature RNAP detachment from the gene ($d \to 0$), Eq. (G.6) can be further simplified to

$$\begin{aligned} \langle r_{\text{tot}} \rangle_{(b;\infty;0)} &= \frac{1}{2} \nu b s_u \langle T \rangle, \\ \text{Var}(r_{\text{tot}})_{(b;\infty;0)} &= \langle r_{\text{tot}} \rangle_{(b;\infty;0)} \cdot \frac{2\nu}{3} + \langle r_{\text{tot}} \rangle^2_{(b;\infty;0)} \cdot \frac{8}{3 s_u \langle T \rangle}. \end{aligned} \tag{G.7}$$

Constitutive expression with deterministic elongation: In this case, Eq. (G.5) simplifies to

$$\begin{aligned} \langle r_{\text{tot}} \rangle_{(c;\infty)} &= \frac{\nu}{T_d} \frac{r}{d} \left[1 - (1 + T_d) \mathrm{e}^{-T_d} \right] \quad \text{and} \\ \text{Var}(r_{\text{tot}})_{(c;\infty)} &= \frac{\nu^2}{T_d^2} \frac{r}{d} \left[2 - (2 + 2T_d + T_d^2) \mathrm{e}^{-T_d} \right], \end{aligned} \tag{G.8}$$

which reduces to

$$\langle r_{\text{tot}} \rangle_{(c;\infty;0)} = \frac{1}{2} \nu r \langle T \rangle \quad \text{and} \quad \text{Var}(r_{\text{tot}})_{(c;\infty;0)} = \frac{1}{3} \nu^2 r \langle T \rangle \tag{G.9}$$

for the special case of zero RNAP detachment from the gene.

H. Extended Model with RNAP Pausing

Proof of Proposition 3 The new pausing model presented in Fig. 9 can be conveniently described by $2L + 2$ species interacting via an effective set of $5L + 4$ reactions. The vector $\vec{m}$ of the number of molecules of the respective species is given by $\vec{m} = (n_0, n_1^a, \ldots, n_L^a, n_1^p, \ldots, n_L^p, n)$; in the table below, we summarise the respective positions of each entry in $\vec{m}$, as well as the definition of the rate functions f_j, for $j = 1, \ldots, 5L + 4$. Note that we do not consider G_{off} as an independent species,

Species	Molecule numbers	Position (in $\vec{m}$)
G_{on}	n_0	1
$P_i, \quad i \in \{1, \ldots, L\}$	n_i^a	$i + 1$
$\bar{P}_i, \quad i \in \{1, \ldots, L\}$	n_i^p	$i + L + 1$
M	n	$2L + 2$

as a conservation law implies $\langle G_{\text{off}} \rangle = 1 - \langle n_0 \rangle$. Given the ordering of species and reactions as described in the above tables, we can define the $(2L + 2) \times (5L + 4)$-dimensional stoichiometry matrix $\mathbf{S}$, with nonzero elements given by

Reaction	Rate function f_j
$G_{\text{on}} \xrightarrow{s_b} G_{\text{off}}$	$f_1 = s_b\langle n_0\rangle$
$G_{\text{off}} \xrightarrow{s_u} G_{\text{on}}$	$f_2 = s_u(1-\langle n_0\rangle)$
$G_{\text{on}} \xrightarrow{r} G_{\text{on}} + P_1$	$f_3 = r\langle n_0\rangle$
$P_i \xrightarrow{k} P_{i+1},\ \ i \in \{1,\ldots,L-1\}$	$f_{i+3} = k\langle n_i^a\rangle$
$P_L \xrightarrow{k} M$	$f_{L+3} = k\langle n_L^a\rangle$
$P_i \xrightarrow{d_a} \varnothing,\ \ i \in \{1,\ldots,L\}$	$f_{i+L+3} = d_a\langle n_i^a\rangle$
$P_i \xrightarrow{r_p} \bar{P}_i,\ \ i \in \{1,\ldots,L\}$	$f_{i+2L+3} = r_p\langle n_i^a\rangle$
$\bar{P}_i \xrightarrow{r_a} P_i,\ \ i \in \{1,\ldots,L\}$	$f_{i+3L+3} = r_a\langle n_i^p\rangle$
$\bar{P}_i \xrightarrow{d_p} \varnothing,\ \ i \in \{1,\ldots,L\}$	$f_{i+4L+3} = d_p\langle n_i^p\rangle$
$M \xrightarrow{d_m} \varnothing$	$f_{5L+4} = d_m\langle n\rangle$

$$
\begin{aligned}
\mathbf{S}_{11} &= -1, & \mathbf{S}_{12} &= 1, & & \\
\mathbf{S}_{i,i+1} &= 1, & \mathbf{S}_{i,i+2} &= -1, & \mathbf{S}_{i,i+L+2} &= -1, \\
\mathbf{S}_{i,i+2L+2} &= -1, & \mathbf{S}_{i,i+3L+2} &= 1, & & \\
\mathbf{S}_{i+L,i+2L+2} &= 1, & \mathbf{S}_{i+L,i+3L+2} &= -1, & \mathbf{S}_{i+L,i+4L+2} &= -1, \\
\mathbf{S}_{2L+2,L+3} &= 1, & \mathbf{S}_{2L+2,5L+4} &= -1, &
\end{aligned}
\tag{H.1}
$$

where $i = 2,\ldots,L+1$. From the associated CME, it can be shown via the moment equations that the time evolution of the vector $\langle\vec{m}\rangle$ of mean molecule numbers in a system of reactions with propensities that are linear in the number of molecules is determined by $d\langle\vec{m}\rangle/\mathrm{d}t = \mathbf{S}\cdot\vec{f}$. Given the form of the stoichiometric matrix $\mathbf{S}$ and of the rate functions f_j, it follows that the mean numbers of molecules of active gene, active and paused RNAP, and mature RNA in steady-state can be obtained by solving the following system of $2L+2$ algebraic equations:

$$
\begin{aligned}
0 &= s_u(1-\langle n_0\rangle) - s_b\langle n_0\rangle, \\
0 &= r\langle n_0\rangle - (k+d_a+r_p)\langle n_1^a\rangle + r_a\langle n_1^p\rangle, \\
0 &= k\langle n_{i-1}^a\rangle - (k+d_a+r_p)\langle n_i^a\rangle + r_a\langle n_i^p\rangle \qquad \text{for } i = 2,\ldots,L, \\
0 &= r_p\langle n_i^a\rangle - (r_a+d_p)\langle n_i^p\rangle \qquad \text{for } i = 1,\ldots,L, \\
0 &= k\langle n_L^a\rangle - d_m\langle n\rangle.
\end{aligned}
\tag{H.2}
$$

Here, we recall the definition of the following parameters from the main text: $\eta = s_u\tau_g$, where $\tau_g = 1/(s_u+s_b)$ is the gene switching timescale, $\rho_k = r/k$, and $\rho = r/d_m$. Also, we define several new parameters: $\sigma = r_p/r_a$ as the ratio of the pausing and activation rates; $\pi_{r_a} = r_a/(r_a+d_p)$, which is the probability of RNAP switching to the active state; $\pi_{d_p} = d_p/(r_a+d_p)$, which is the probability of premature termination from the paused RNAP state; $\tilde{\mu} = k/(k+d_a+r_p\pi_{d_p})$; and $\lambda = \sigma\pi_{r_a}$. It follows that the solution of Eq. (H.2) can be written as

$$
\langle n_0\rangle = \eta, \quad \langle n_i^a\rangle = \eta\rho_k\tilde{\mu}^i, \quad \langle n_i^p\rangle = \langle n_i^a\rangle\lambda, \quad \text{and} \quad \langle n\rangle = \eta\rho\tilde{\mu}^L. \tag{H.3}
$$

Proof of Proposition 4 In order to solve the Lyapunov equation $\mathbf{J}\cdot\mathbf{C}+\mathbf{C}\cdot\mathbf{J}^T+\mathbf{D}=\mathbf{0}$ for the symmetric elements $\mathbf{C}_{ij}=\mathbf{C}_{ji}$ of the $(2L+2)\times(2L+2)$-dimensional covariance matrix $\mathbf{C}$, we will follow the same approach as in 'Appendix B'. First, we define the $(2L+2)\times(2L+2)$-dimensional Jacobian and diffusion matrices for our system. The Jacobian matrix $\mathbf{J}$ has the following nonzero elements:

$$\begin{aligned}
&\mathbf{J}_{11}=-(s_u+s_b),\\
&\mathbf{J}_{21}=r, && \mathbf{J}_{22}=-(k+d_a+r_p), && \mathbf{J}_{2,2+L}=r_a,\\
&\mathbf{J}_{i,i-1}=k, && \mathbf{J}_{ii}=-(k+d_a+r_p), && \mathbf{J}_{i,i+L}=r_a && \text{for } i=3,\ldots,L+1,\\
&\mathbf{J}_{i+L,i}=r_p, && \mathbf{J}_{i+L,i+L}=-(r_a+d_p) && && \text{for } i=2,\ldots,L+1,\\
&\mathbf{J}_{2L+2,L+1}=k, && \mathbf{J}_{2L+2,2L+2}=-d_m,
\end{aligned} \tag{H.4}$$

while the nonzero elements of the symmetric diffusion matrix $\mathbf{D}$ are given by

$$\begin{aligned}
&\mathbf{D}_{11}=s_u(1-\langle n_0\rangle)+s_b\langle n_0\rangle,\\
&\mathbf{D}_{22}=r\langle n_0\rangle+(k+d_a+r_p)\langle n_1^a\rangle+r_a\langle n_1^p\rangle, && \mathbf{D}_{23}=-k\langle n_1^a\rangle, && \mathbf{D}_{2,2+L}=-r_p\langle n_1^a\rangle-r_a\langle n_1^p\rangle;\\
&\text{for } i=3,\ldots,L+1:\\
&\mathbf{D}_{ii}=k\langle n_{i-2}^a\rangle+(k+d_a+r_p)\langle n_{i-1}^a\rangle+r_a\langle n_{i-1}^p\rangle, && \mathbf{D}_{i,i+1[i\le L]}=-k\langle n_{i-1}^a\rangle, && \mathbf{D}_{L+1,2L+2}=-k\langle n_L^a\rangle,\\
&\mathbf{D}_{i,i+L}=-r_p\langle n_{i-1}^a\rangle-r_a\langle n_{i-1}^p\rangle;\\
&\text{for } i=2,\ldots,L+1:\\
&\mathbf{D}_{i+L,i+L}=r_p\langle n_{i-1}^a\rangle+(r_a+d_p)\langle n_{i-1}^p\rangle,\\
&\mathbf{D}_{2L+2,2L+2}=k\langle n_L^a\rangle+d_m\langle n\rangle.
\end{aligned} \tag{H.5}$$

Next, using the definition of $\mathbf{J}$ and $\mathbf{D}$ from Eqs. (H.4) and (H.5), respectively, we solve the Lyapunov equation. Here, we note that we are only interested in expressions for the covariances of fluctuations in active and paused RNAP, but not of mature RNA fluctuations; hence, we require closed-form expressions for the elements $\mathbf{C}_{ij}$ with $i,j\neq 2L+2$, which we derive by following the same procedure as in 'Appendix B'.

Now, we recall that $\beta=s_b/s_u$ is the ratio of gene deactivation and activation rates, while $\tau_p=1/(k+d_a)$ is the typical time that an actively moving RNAP spends on a gene segment. Additionally, let $\tau_{r_a}=1/r_a$ be the timescale of RNAP activation from the paused state, let $\tau_{d_p}=1/d_p$ be the timescale of premature termination of paused RNAP, and let $\tau_{pp}=1/(r_a+d_p)$ be the typical time spent in the paused state. Finally, we define the following new parameters: $\lambda_{r_p}=\pi_{r_p}/(1-\pi_{r_p})$, where $\pi_{r_p}=r_p/(r_p+k+d_a)$ is the probability of actively moving RNAP switching to the paused state, as well as

$$\omega_{r_a}=\frac{\pi_{r_a}\tau_g}{\pi_{r_a}\tau_{r_a}+\tau_g},\quad \tilde{\alpha}=\frac{\tau_g+\lambda_{r_p}\pi_{d_p}\tau_g}{\tau_g+\tau_p+\lambda_{r_p}\tau_g(1-\omega_{r_a})},\quad \text{and}\quad \omega=\frac{\tau_g}{\tau_{pp}+\tau_g}; \tag{H.6}$$

then, closed-form expressions for the covariances of the active gene with itself and the remaining species are given by

$$\begin{aligned}
\mathrm{Var}(n_0) &= \eta^2\beta \cdot g_{00}^{aa}, && \text{where } g_{00}^{aa} = 1,\\
\mathrm{Cov}(n_0, n_j^a) &= \eta\langle n_j^a\rangle\tilde{\alpha}\beta \cdot g_{0j}^{aa}, && \text{where } g_{0j}^{aa} = \tilde{\alpha}^{j-1},\\
\mathrm{Cov}(n_0, n_j^p) &= \eta\langle n_j^p\rangle\tilde{\alpha}\beta \cdot g_{0j}^{ap}, && \text{where } g_{0j}^{ap} = \omega\tilde{\alpha}^{j-1}.
\end{aligned} \tag{H.7}$$

Similarly, closed-form expressions for the covariances between all RNAP species read

$$\begin{aligned}
\mathrm{Cov}(n_i^a, n_j^a) &= \delta_{ij}\langle n_i^a\rangle + \langle n_i^a\rangle\langle n_j^a\rangle\tilde{\alpha}\beta \cdot g_{ij}^{aa},\\
\mathrm{Cov}(n_i^a, n_j^p) &= \langle n_i^a\rangle\langle n_j^p\rangle\tilde{\alpha}\beta \cdot g_{ij}^{ap},\\
\mathrm{Cov}(n_i^p, n_j^a) &= \langle n_i^p\rangle\langle n_j^a\rangle\tilde{\alpha}\beta \cdot g_{ij}^{pa},\\
\mathrm{Cov}(n_i^p, n_j^p) &= \delta_{ij}\langle n_i^p\rangle + \langle n_i^p\rangle\langle n_j^p\rangle\tilde{\alpha}\beta \cdot g_{ij}^{pp},
\end{aligned} \tag{H.8}$$

where the functions $g_{ij}^{aa} = g_{ji}^{aa}$, $g_{ij}^{ap} = g_{ji}^{pa}$, and $g_{ij}^{pp} = g_{ji}^{pp}$ satisfy the following recurrence relations:

$$\begin{aligned}
g_{ij}^{aa} &= \frac{[(k+d_a)(r_a+d_p)+r_pd_p](g_{i-1,j}^{aa}+g_{i,j-1}^{aa})+r_ar_p(g_{ij}^{ap}+g_{ij}^{pa})}{2(k+d_a+r_p)(r_a+d_p)},\\
g_{ij}^{ap} &= \frac{[(k+d_a)(r_a+d_p)+r_pd_p]g_{i-1,j}^{ap}+(r_a+d_p)^2g_{ij}^{aa}+r_ar_pg_{ij}^{pp}}{(k+d_a+r_a+r_p+d_p)(r_a+d_p)},\\
g_{ij}^{pp} &= \frac{g_{ij}^{ap}+g_{ij}^{pa}}{2}.
\end{aligned} \tag{H.9}$$

Now, we assume that the elongation rate is faster than the rates of RNAP pausing, activation, and premature termination, i.e. that $k \gg r_a, r_p, d_a, d_p$ in Eq. (H.9). Taking the limit of $k \to \infty$, we find that the expressions in Eqs. (H.7) and (H.8) remain unchanged, while Eq. (H.9) simplifies to

$$g_{ij}^{aa} = (g_{i-1,j}^{aa} + g_{i,j-1}^{aa})/2, \tag{H.10a}$$

$$g_{ij}^{ap} = g_{i-1,j}^{ap}, \tag{H.10b}$$

$$g_{ij}^{pp} = (g_{ij}^{ap} + g_{ij}^{pa})/2; \tag{H.10c}$$

in particular, to leading order in $1/k$, the functions g_{ij}^{aa}, g_{ij}^{ap}, g_{ij}^{pa}, and g_{ij}^{pp} hence do not depend on k. Eq. (H.10a) defines a recurrence relation for the symmetric function $g_{ij}^{aa} = g_{ji}^{aa}$ with initial conditions g_{00}^{aa} and g_{0j}^{aa} from Eq. (H.7). Using the same mathematical technique as in Lemma B.5, we find that the solution for the function g_{ij}^{aa} is given by $g_{ij}^{aa} = g^{aa}(i,j) + g^{aa}(j,i)$, where

$$g^{aa}(i,j) = \frac{\tilde{\alpha}^{i+j-1}}{(2\tilde{\alpha}-1)^i} + \frac{1}{2^{i+j-1}}\binom{i+j-1}{i}\Big[1 - \frac{2\tilde{\alpha}-1}{2\tilde{\alpha}}\,{}_2F_1\big(1, i+j; j; \tfrac{1}{2\tilde{\alpha}}\big)\Big]; \tag{H.11}$$

Eq. (H.10b) is a recurrence relation for the function g_{ij}^{ap} with initial conditions g_{0j}^{ap} from Eq. (H.7); the corresponding solution is then given by $g_{ij}^{ap} = \omega\tilde{\alpha}^{j-1}$. Finally, the solution of the recurrence relation in Eq. (H.10c) for g_{ij}^{pp} is given by $g_{ij}^{pp} = \omega(\tilde{\alpha}^{j-1} + \tilde{\alpha}^{i-1})/2$. In sum, the leading-order asymptotics (in $1/k$) of the covariances between the various RNAP species for k large is hence given by Eq. (H.8), with g_{ij}^{aa}, $g_{ij}^{ap} = g_{ij}^{pa}$, and g_{ij}^{pp} as stated above. □

Asymptotics of variance of total RNAP distribution: The variance of the total RNAP distribution for the pausing model is given by

$$\mathrm{Var}(n_{\mathrm{tot}}) = \sum_{i,j=1}^{L}\big(\mathrm{Cov}(n_i^a, n_j^a) + \mathrm{Cov}(n_i^a, n_j^p) + \mathrm{Cov}(n_i^p, n_j^a) + \mathrm{Cov}(n_i^p, n_j^p)\big), \tag{H.12}$$

where the expressions for the corresponding covariances are given in Eq. (39). In order to simplify the above expression, we consider each term on the right-hand side in Eq. (H.12) separately, as follows:

$$\begin{aligned}
\sum_{i,j=1}^{L}\mathrm{Cov}(n_i^a, n_j^a) &= \sum_{i,j=1}^{L}\delta_{ij}\langle n_i^a\rangle + (\eta\rho_k)^2\tilde{\alpha}\beta\sum_{i,j=1}^{L} g_{ij}^{aa},\\
\sum_{i,j=1}^{L}\mathrm{Cov}(n_i^a, n_j^p) &= (\eta\rho_k)^2\tilde{\alpha}\beta\lambda\sum_{i,j=1}^{L} g_{ij}^{ap},\\
\sum_{i,j=1}^{L}\mathrm{Cov}(n_i^p, n_j^a) &= (\eta\rho_k)^2\tilde{\alpha}\beta\lambda\sum_{i,j=1}^{L} g_{ij}^{pa},\\
\sum_{i,j=1}^{L}\mathrm{Cov}(n_i^p, n_j^p) &= \sum_{i,j=1}^{L}\delta_{ij}\langle n_i^p\rangle + (\eta\rho_k)^2\tilde{\alpha}\beta\lambda^2\sum_{i,j=1}^{L} g_{ij}^{pp}.
\end{aligned} \tag{H.13}$$

Since $\sum_{i,j=1}^{L}\big(\delta_{ij}\langle n_i^a\rangle + \delta_{ij}\langle n_i^p\rangle\big) = \sum_{i=1}^{L}\langle n_i\rangle = \langle n_{\mathrm{tot}}\rangle$, Eq. (H.12) becomes

$$\mathrm{Var}(n_{\mathrm{tot}}) = \langle n_{\mathrm{tot}}\rangle + (\eta\rho_k)^2\tilde{\alpha}\beta\sum_{i,j=1}^{L}\big(g_{ij}^{aa} + \lambda g_{ij}^{ap} + \lambda g_{ij}^{pa} + \lambda^2 g_{ij}^{pp}\big). \tag{H.14}$$

Using the expressions for the functions g_{ij}^{aa}, g_{ij}^{ap}, g_{ij}^{pa}, and g_{ij}^{pp} from Eq. (39), we conclude that Eq. (H.14) further simplifies to

$$\mathrm{Var}(n_{\mathrm{tot}}) = \langle n_{\mathrm{tot}} \rangle + (\eta\rho_k)^2 \tilde{\alpha}\beta \left[2 \sum_{i,j=1}^{L} g^{aa}(i,j) + \lambda(2+\lambda)\omega L \frac{\tilde{\alpha}^L - 1}{\tilde{\alpha} - 1} \right]. \tag{H.15}$$

I. Approximation of Mature RNA Distribution in Extended Model

Similarly to Sect. 3.2, we apply geometric singular perturbation theory (GSPT) to formally derive the distribution of mature RNA for the extended pausing model. As was done there, we define $P_j(\vec{n};t)$ $(j = 0, 1)$ as the probability of the state $\vec{n} = (n_1^a, \ldots, n_L^a, n_1^p, \ldots, n_L^p, n)$ at time t while the gene is either active (0) or inactive (1); then, the time evolution of these probabilities can be described by a system of coupled CMEs:

$$\begin{aligned}
\partial_t P_0 &= s_u P_1 - s_b P_0 + r(\mathbb{E}_{n_1}^{-1} - 1)P_0 + k \sum_{i=1}^{L-1} (\mathbb{E}_{n_i^a} \mathbb{E}_{n_{i+1}^a}^{-1} - 1) n_i^a P_0 + k(\mathbb{E}_{n_L^a} \mathbb{E}_n^{-1} - 1) n_L^a P_0 \\
&\quad + d_a \sum_{i=1}^{L} (\mathbb{E}_{n_i^a} - 1) n_i^a P_0 + r_p \sum_{i=1}^{L} (\mathbb{E}_{n_i^a} \mathbb{E}_{n_i^p}^{-1} - 1) n_i^a P_0 + r_a \sum_{i=1}^{L} (\mathbb{E}_{n_i^p} \mathbb{E}_{n_i^a}^{-1} - 1) n_i^p P_0 \\
&\quad + d_p \sum_{i=1}^{L} (\mathbb{E}_{n_i^p} - 1) n_i^p P_0 + d_m (\mathbb{E}_n - 1) n P_0, \\
\partial_t P_1 &= s_b P_0 - s_u P_1 + k \sum_{i=1}^{L-1} (\mathbb{E}_{n_i^a} \mathbb{E}_{n_{i+1}^a}^{-1} - 1) n_i^a P_1 + k(\mathbb{E}_{n_L^a} \mathbb{E}_n^{-1} - 1) n_L^a P_1 \\
&\quad + d_a \sum_{i=1}^{L} (\mathbb{E}_{n_i^a} - 1) n_i^a P_1 + r_p \sum_{i=1}^{L} (\mathbb{E}_{n_i^a} \mathbb{E}_{n_i^p}^{-1} - 1) n_i^a P_1 + r_a \sum_{i=1}^{L} (\mathbb{E}_{n_i^p} \mathbb{E}_{n_i^a}^{-1} - 1) n_i^p P_1 \\
&\quad + d_p \sum_{i=1}^{L} (\mathbb{E}_{n_i^p} - 1) n_i^p P_1 + d_m (\mathbb{E}_n - 1) n P_1.
\end{aligned} \tag{I.1}$$

In order to find analytical expressions for the propagator probabilities $P(\vec{n};t)$ which satisfy the system of CMEs in Eq. (I.1), we define the probability-generating functions $F_j(\vec{z};t)$, where $\vec{z} = (z_1^a, \ldots, z_L^a, z_1^p, \ldots, z_L^p, z)$ is a vector of variables corresponding to the state $\vec{n}$. Given the equations for $P_j(\vec{n};t)$ from Eq. (I.1), we obtain the following system of PDEs for the corresponding generating functions $F_j(\vec{z};t)$:

$$\begin{aligned}
\mathbb{L}[F_0] &= s_u F_1 - s_b F_0 + r(z_1^a - 1) F_0, \\
\mathbb{L}[F_1] &= s_b F_0 - s_u F_1;
\end{aligned} \tag{I.2}$$

here,

$$\begin{aligned}\mathbb{L} = \partial_t + d_m(z-1)\partial_z + k\sum_{i=1}^{L-1}(z_i^a - z_{i+1}^a)\partial_{z_i^a} + k(z_L^a - z)\partial_{z_L^a} + d_a\sum_{i=1}^{L}(z_i^a - 1)\partial_{z_i^a} \\ + r_p\sum_{i=1}^{L}(z_i^a - z_i^p)\partial_{z_i^a} + r_a\sum_{i=1}^{L}(z_i^p - z_i^a)\partial_{z_i^p} + d_p\sum_{i=1}^{L}(z_i^p - 1)\partial_{z_i^p}\end{aligned} \tag{I.3}$$

is a differential operator acting on the functions F_0 and F_1. Eq. (I.2) represents a system of coupled, linear, first-order PDEs. Now, we introduce new variables $u_i^a = z_i^a - 1$, $u_i^p = z_i^p - 1$, and $u = z - 1$; we also rescale all rates and the time variable with the degradation rate d_m of mature RNA. Next, we apply the method of characteristics, with s being the characteristic variable. The first characteristic equation will give us $d_m(\mathrm{d}t/\mathrm{d}s) = 1$, with solution $s \equiv d_m t$; hence, we can use the variable $t' = d_m t$ as the independent characteristic variable and thus convert the system of PDEs in Eq. (I.2) into a characteristic system of ODEs:

$$\dot{u}_i^a = (k/d_m)[(u_i^a - u_{i+1}^a) + (d_a/k)u_i^a + (r_p/k)(u_i^a - u_i^p)] \qquad \text{for } i = 1, \ldots, L-1, \tag{I.4a}$$

$$\dot{u}_L^a = (k/d_m)[(u_L^a - u) + (d_a/k)u_L^a + (r_p/k)(u_L^a - u_L^p)], \tag{I.4b}$$

$$\dot{u}_i^p = (r_a/d_m)[(u_i^p - u_i^a) + (d_p/r_a)u_i^p] \qquad \text{for } i = 1, \ldots, L, \tag{I.4c}$$

$$\dot{u} = u, \tag{I.4d}$$

$$\dot{F}_0 = (s_u/d_m)F_1 - (s_b/d_m)F_0 + (r/d_m)u_1^a F_0, \tag{I.4e}$$

$$\dot{F}_1 = (s_b/d_m)F_0 - (s_u/d_m)F_1, \tag{I.4f}$$

where the overdot denotes differentiation with respect to t. Here, we assume that $k/d_m \gg 1$ and $r_a/d_m \gg 1$; hence, we define $\varepsilon = d_m/k$ as the singular perturbation parameter, and we write $d_m/r_a = \varepsilon\delta$, where $\delta = k/r_a = \mathcal{O}(1)$ by assumption. Since $0 < \varepsilon \ll 1$ is small, we can apply GSPT in order to separate the system in Eq. (I.4) into fast and slow dynamics, which will allow us to find an asymptotic approximation for F_0 and F_1 in steady state. With the above definitions, the governing equations for u_i^a and u_i^p in the 'slow system' in Eqs. (I.4a) through (I.4c) become

$$\varepsilon\dot{u}_i^a = (u_i^a - u_{i+1}^a) + (d_a/k)u_i^a + (r_p/k)(u_i^a - u_i^p) \qquad \text{for } i = 1, \ldots, L-1, \tag{I.5a}$$

$$\varepsilon\dot{u}_L^a = (u_L^a - u) + (d_a/k)u_L^a + (r_p/k)(u_L^a - u_L^p), \tag{I.5b}$$

$$\varepsilon\dot{u}_i^p = [(u_i^p - u_i^a) + (d_p/r_a)u_i^p]/\delta \qquad \text{for } i = 1, \ldots, L. \tag{I.5c}$$

It follows that u_i^a and u_i^p $(i = 1, \ldots, L)$ are the fast variables in our system, while u, F_0, and F_1 are the slow ones; see 'Appendix E'. Setting $\varepsilon = 0$ and solving the system in Eq. (I.5), we find $u_1^a = \tilde{\mu}^L \cdot u$, where $\tilde{\mu} = k/(k + d_a + r_p\pi_{d_p})$ has previously been

defined in Proposition 3. Now, given Eq. (I.4d), we apply the chain rule, $\mathrm{d}t' \equiv \mathrm{d}u \cdot u$, to rewrite Eqs. (I.4e) and (I.4f) as:

$$F_0' d_m u = s_u F_1 - s_b F_0 + r\tilde{\mu}^L u F_0, \tag{I.6a}$$
$$F_1' d_m u = s_b F_0 - s_u F_1, \tag{I.6b}$$

where the prime now denotes differentiation with respect to u. The system in Eq. (I.6) is the same as that in Eq. (28), with the substitution $\mu \mapsto \tilde{\mu}$; hence, following the same derivation as in Sect. 3.2, we conclude that the steady-state analytical expression for the probability distribution of mature RNA is given by

$$P(n) = \frac{1}{n!}\frac{(s_u)_n}{(s_b+s_u)_n}\left(\frac{r}{d_m}\right)^n (\tilde{\mu}^L)^n {}_1F_1\left(\frac{s_u}{d_m}+n; \frac{s_b+s_u}{d_m}+n; -\frac{r}{d_m}\tilde{\mu}^L\right). \tag{I.7}$$

References

Adelman K, La Porta A, Santangelo TJ, Lis JT, Roberts JW, Wang MD (2002) Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. Proc Nat Acad Sci 99(21):13538–13543

Ali MZ, Choubey S, Das D, Brewster RC (2020) Probing mechanisms of transcription elongation through cell-to-cell variability of RNA polymerase. Biophys J 118(7):1769–1781

Brouwer I, Lenstra TL (2019) Visualizing transcription: key to understanding gene expression dynamics. Curr Opin Chem Biol 51:122–129

Cao Z, Grima R (2020) Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. Proc Nat Acad Sci 117(9):4682–4692

Choubey S (2018) Nascent RNA kinetics: transient and steady state behavior of models of transcription. Phys Rev E 97(2):022402

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

Cianci C, Smith S, Grima R (2016) Molecular finite-size effects in stochastic models of equilibrium chemical systems. J Chem Phys 144(8):084101

Coulon A, Ferguson ML, Va de Turris M, Palangat CCC, Larson DR (2014) Kinetic competition during the transcription cycle results in stochastic RNA processing. eLife 3:e03939

Digital Library of Mathematical Functions (2020a) Chapter 15: https://dlmf.nist.gov/15. Accessed 15 May 2020

Digital Library of Mathematical Functions (2020b) Chapter 15: https://dlmf.nist.gov/13. Accessed 15 May 2020

El Hage A, French SL, Beyer AL, Tollervey D (2010) Loss of topoisomerase i leads to r-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev 24(14):1546–1558

Elf J, Ehrenberg M (2003) Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. Genome Res 13(11):2475–2484

Fenichel N (1979) Geometric singular perturbation theory for ordinary differential equations. J Differ Equ 31(1):53–98

Forde NR, Izhaky D, Woodcock GR, Wuite GJL, Bustamante C (2002) Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by escherichia coli RNA polymerase. Proc Nat Acad Sci 99(18):11682–11687

Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81(25):2340–2361

Gorin G, Wang M, Golding I, Heng X (2020) Stochastic simulation and statistical inference platform for visualization and estimation of transcriptional kinetics. PLoS ONE 15(3):e0230736

Halpern KB, Tanami S, Landen S, Chapal M, Szlak L, Hutzler A, Nizhberg A, Itzkovitz S (2015) Bursty gene expression in the intact mammalian liver. Mol Cell 58(1):147–156

Heng X, Skinner SO, Sokac AM, Golding I (2016) Stochastic kinetics of nascent RNA. Phys Rev Lett 117(12):128101

Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. J Math Biol 54(1):1–26

Jones CKRT (1995) Geometric singular perturbation theory. In: Dynamical systems. Springer, pp 44–118

Klumpp S, Hwa T (2008) Stochasticity and traffic jams in the transcription of ribosomal RNA: intriguing role of termination and antitermination. Proc Nat Acad Sci 105(47):18159–18164

Larson DR, Zenklusen D, Bin W, Chao JA, Singer RH (2011) Real-time observation of transcription initiation and elongation on an endogenous yeast gene. Science 332(6028):475–478

Lenstra TL, Rodriguez J, Chen H, Larson DR (2016) Transcription dynamics in living cells. Ann Rev Biophys 45:25–47

Öcal K, Grima R, Sanguinetti G (2019) Parameter estimation for biochemical reaction networks using Wasserstein distances. J Phys A Math Theor 53(3):034002

Peccoud J, Ycart B (1995) Markovian modeling of gene-product synthesis. Theor Popul Biol 48(2):222–234

Popović N, Marr C, Swain PS (2016) A geometric analysis of fast-slow models for stochastic gene expression. J Math Biol 72(1–2):87–122

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. PLoS Biol 4(10):e309

Raj A, Van Den Bogaard P, Rifkin SA, Van Oudenaarden A, Tyagi S (2008) Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods 5(10):877–879

Rodriguez J, Ren G, Day CR, Zhao K, Chow CC, Larson DR (2019) Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity. Cell 176(1–2):213–226

Sanchez A, Golding I (2013) Genetic determinants and cellular constraints in noisy gene expression. Science 342(6163):1188–1193

Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. Proc Nat Acad Sci 105(45):17256–17261

Skinner SO, Xu H, Nagarkar-Jaiswal S, Freire PR, Zwaka TP, Golding I (2016) Single-cell analysis of transcription kinetics across the cell cycle. eLife 5:e12175

Smith S, Cianci C, Grima R (2017) Macromolecular crowding directs the motion of small molecules inside cells. J R Soc Interface 14(131):20170047

Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F (2011) Mammalian genes are transcribed with widely different bursting kinetics. Science 332(6028):472–474

Tiina R, Antti H, Shannon H, Olli Y-H, AndreS R (2010) Effects of transcriptional pausing on gene expression dynamics. PLoS Comput Biol 6(3):e1000704

Veerman F, Marr C, Popović N (2018) Time-dependent propagators for stochastic models of gene expression: an analytical method. J Math Biol 77(2):261–312

Warren PB, Tănase-Nicola S, ten Wolde PR (2006) Exact results for noise power spectra in linear biochemical reaction networks. J Chem Phys 125(14):144904

Zechner C, Ruess J, Krenn P, Pelet S, Peter M, Lygeros J, Koeppl H (2012) Moment-based inference predicts bimodality in transient gene expression. Proc Nat Acad Sci 109(21):8340–8345

Zenklusen D, Larson DR, Singer RH (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. Nat Struct Mol Biol 15(12):1263