Nascent RNA kinetics: Transient and steady state behavior of models of transcription

Sandeep Choubey
FAS Center for Systems Biology and Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

(Received 23 September 2017; revised manuscript received 2 December 2017; published 9 February 2018; corrected 15 February 2018)

Regulation of transcription is a vital process in cells, but mechanistic details of this regulation still remain elusive. The dominant approach to unravel the dynamics of transcriptional regulation is to first develop mathematical models of transcription and then experimentally test the predictions these models make for the distribution of mRNA and protein molecules at the individual cell level. However, these measurements are affected by a multitude of downstream processes which make it difficult to interpret the measurements. Recent experimental advancements allow for counting the nascent mRNA number of a gene as a function of time at the single-cell level. These measurements closely reflect the dynamics of transcription. In this paper, we consider a general mechanism of transcription with stochastic initiation and deterministic elongation and probe its impact on the temporal behavior of nascent RNA levels. Using techniques from queueing theory, we derive exact analytical expressions for the mean and variance of the nascent RNA distribution as functions of time. We apply these analytical results to obtain the mean and variance of nascent RNA distribution for specific models of transcription. These models of initiation exhibit qualitatively distinct transient behaviors for both the mean and variance which further allows us to discriminate between them. Stochastic simulations confirm these results. Overall the analytical results presented here provide the necessary tools to connect mechanisms of transcription initiation to single-cell measurements of nascent RNA.

DOI: 10.1103/PhysRevE.97.022402

I. INTRODUCTION

Transcription is a multistep process that starts with the binding of general transcription factors (TFs) to a region of DNA, called the promoter [1]. Through the action of these TFs, RNA polymerase (RNAP) is recruited to the promoter. This RNAP molecule subsequently initiates transcription of the corresponding gene, resulting in the synthesis of a mRNA. The process of transcription initiation is regulated in response to environmental and intracellular cues. How this regulation occurs *in vivo* remains one of the fundamental questions in regulatory biology.

Transcription initiation is inherently stochastic [2–4], which leads to variability in levels of mRNA [3,5–8] and protein [9–13] across a population of genetically identical cells. Theoretical [14–16] and experimental [6,13,17] studies have tried to connect the mean and variability at mRNA and protein levels with models of transcription initiation. However, interpreting the distributions of mRNA and protein molecules is challenging because they are affected by other stochastic processes downstream of transcription initiation, such as the maturation time of fluorescent reporters [18], mRNA transport [19], splicing [20], and small RNA regulation [21], all of which can create variability in the levels of mRNA and protein.

An alternative to counting mRNA and protein molecules is to count the number of nascent RNAs (or the number of transcribing polymerases) at the single-cell level. Recent experimental advancements allow for counting nascent RNAs associated with a gene of interest [22–25]. Nascent RNAs are a more direct readout of transcription initiation dynamics compared to mRNA and protein levels [23,26]. For instance, electron micrograph (EM) [27–29] images enable counting of the RNAP molecules engaged in transcribing a gene, while nascent RNAs can be counted by fluorescently labeling them using fluorescence *in situ* hybridization (FISH) [25,30]. These approaches give snapshots of transcription, but experiments that observe transcriptional dynamics in single cells using MS2 tags provide complete time traces corresponding to individual transcription sites [22,31,32]. The EM and FISH methods provide the steady state nascent RNA measurements but MS2 measurements carry information about both the transient and steady state regimes. Multiple studies have examined how the different models of transcription initiation generate steady state nascent RNA levels [23,26,33], However, the transient behaviors of these models for the nascent RNA distribution remain relatively unexplored.

In this manuscript we use analytical tools from queueing theory and show how to compute the mean and variance of the transient distribution of nascent RNAs for a general mechanism of transcription initiation and deterministic elongation. We apply these results to obtain the mean and variance of nascent RNAs for specific and well-studied models of transcription. We show that these different models of initiation show qualitatively distinct transient behaviors for both the mean and variance, which enables us to discriminate between these models. We confirm these analytical results using Gillespie simulations [34]. Overall our results provide the necessary tools to infer the dynamics of transcription initiation as well as extract the associated kinetic parameters from single-cell nascent RNA measurements.

II. MODEL

To connect mechanisms of transcription initiation with the experimentally measured transient nascent RNA distributions,

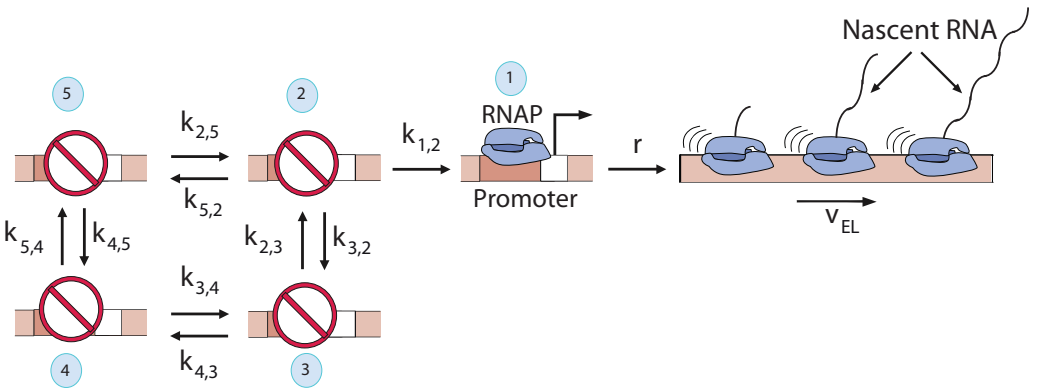

FIG. 1. Model: An arbitrary promoter architecture with N number of promoter states is considered. Here we show a model with five different states. The different promoter states are 1–5. The rate of transition from the ith to the jth state is given by k_{ji}. The rate of initiation from the first state is r. The rate of initiation from every other promoter states is zero. After initiation, each polymerase molecule traverses the gene at a speed v_{EL}, where L is the length of the gene. Time taken to traverse the gene is, therefore, $T = L/v_{\mathrm{EL}}$.

we consider a model of transcriptional dynamics with a general scheme of initiation mechanism followed by a deterministic elongation process. We consider the promoter dynamics to model the transcription initiation process. As shown in Fig. 1, the promoter transitions between different states as different transcription factors bind and fall off the promoter region, thereby modulating the rate of initiation. We assume that a promoter can exist in N different states. The transition rate of going from the ith to the jth state is k_{ji}. The rate of transcription initiation from one of these promoter states (in which transcription can occur) is r and from the rest of the states is zero. We further assume that after initiating transcription, every polymerase molecule moves along the gene at a constant speed v_{EL} [23,26,35]. Consequently, for a gene of length L, the time it takes for each polymerase to traverse the gene is $T = \frac{L}{v_{\mathrm{EL}}}$ [33,36]. Although transcription elongation is typically more complicated and involves pausing and backtracking of polymerases along the gene, for a vast number of genes our assumption of deterministic elongation is justified [23,26]. For a discussion on this assumption see the Appendix.

In this section, we lay out the basic mathematical framework to obtain the analytical expressions for the mean and variance of the nascent RNA distribution as functions of time, for this model of transcription.

In order to tackle this problem, we take a cue from the renewal processes in queueing theory [37]. Some theoretical studies have applied queueing theory to explore the impact of transcription and translation on the statistical properties of mRNA and protein distributions across isogenic populations [16,38,39]. Queueing theory deals with the mathematical analysis of the waiting lines formed by customers randomly arriving at a service station and staying in the station until they receive service from a number of servers. Diverse types of queueing process can be defined based on the following four features: (a) the arrival time distribution of the customers; (b) the distribution of the number of customers, i.e., the batches in every arrival; (c) the service time distribution; and (d) the number of available servers. For if the arrival and service time distributions are independent and identically distributed (iid), i.e., these two processes are renewal processes, then for these models closed form analytical expressions for the moments of the number of busy servers can be obtained as functions of time [37]. Our model of transcription can be mapped to a queueing process in the following way: Individual transcription initiation events are analogous to the arrival of customers with a batch size of 1, the transcription elongation process of each polymerase molecule is akin to customers being serviced, and finally since the polymerases move independently of each other, the number of servers in the equivalent queueing model is infinite. Since in our model RNAPs initiate transcription from only one of the promoter states, the distribution of times between successive initiation events is independent and identically distributed and hence resembles a renewal process (for a discussion on renewal processes, see the Appendix). Given its characteristics our model can be described by the well-studied $G/D/\infty$ system in the queueing theory literature. In this framework, the symbol G corresponds to the general waiting time distribution, D stands for deterministic service time, and "∞" stands for an infinite number of servers. The $G/D/\infty$ system has been analyzed in detail in previous work in queueing theory [40]. By exploiting this mapping, we extract exact analytical expressions for the mean and variance of the nascent RNA distribution for a gene of interest.

The first step towards computing mean and variance of the nascent RNA distribution is to obtain an expression for the waiting time distribution between successive initiation events. We calculate the probability $q(x)$ that two transcription initiation events are separated by time x; i.e., if one initiation event happens at time $x = 0$, the next initiation event occurs at time x later, between x and $x + dx$. We employ a master equation approach to compute the probability P_i for the promoter to be in the ith state at time x without having initiated any transcription event between 0 and x. The master equation for the time evolution of P_i is given by

$$\frac{d}{dx}P_i = \sum_{j=1}^{N}[k_{ji}P_j - k_{ij}P_i] - r_i P_i. \tag{1}$$

To compute the probability that two successive initiation events are separated by a time x, we consider the only active state m from which transcription initiation occurs. The probability that two initiation events are separated by time x,

and the second initiation event happens at time x between x and $x + dx$ is given by

$$q(x) = r P_m(x). \tag{2}$$

Here $q(x)$ is a product of the probability $P_m(x)$ that the promoter is in the mth state at time x and the probability rdx that an RNAP molecule initiates transcription from this state between time x and $x + dx$. For details of the calculation see Ref. [41] and the Appendix.

Next, we obtain an analytical expression for the renewal function for our model. Renewal function $R(t)$ is a key quantity in the field of queueing theory [37], and is defined as the expectation value of the number of arrivals of the customers at the service station (equivalent to the initiation events for our model) observed up to a time t [37]. For a detailed discussion on the renewal function, see the Appendix.

To obtain the renewal function we consider the renewal equation [37] since the renewal function $R(t)$ satisfies the renewal equation, given by

$$R(t) = Q(t) + \int_0^t R(t - x)q(x)dt. \tag{3}$$

where $Q(t)$ is the cumulative distribution function of waiting times, defined as $Q(t) = \int_0^t q(x)dx$. To obtain the renewal function we need to solve the renewal equation. Details of the calculation of the renewal function are given in the Appendix.

Following the work of Liu *et al.* [40], we find the mean $M(t)$ and variance $\mathrm{Var}(t)$ of the nascent RNA distribution as a function of time, given by

$$\begin{aligned} M(t) &= R(t), \ \ t \leqslant T \\ &= R(t) - R(t - T), \ \ t > T, \\ \mathrm{Var}(t) &= 2\int_0^t R(t - y)dR(y) + R(t) - [R(t)]^2, \ \ t \leqslant T \\ &= 2\int_{t-T}^t R(t - y)dR(y) + M(t)[1 - M(t)], \ \ t > T. \end{aligned} \tag{4}$$

Both the mean and variance are piecewise-defined functions with two domains in time defined by less or greater than the time (T) that a polymerase molecule takes to traverse a gene.

In the steady state, when time $t \to \infty$, we get the following expressions for the mean (M) and variance (Var):

$$\begin{aligned} M &= \lambda T, \\ \mathrm{Var} &= 2\lambda \int_0^T R(t)dt + \lambda T(1 - \lambda T). \end{aligned} \tag{5}$$

Here λ is the long-time average of the number of initiation events per unit time. This quantity is a product of the steady state probability of the promoter being in the mth state and the rate of initiation from this state, r. For details see the Appendix.

The procedure described above can be applied to promoters with any number of states one of which is transcriptionally active.

III. DIFFERENT MODELS OF TRANSCRIPTION INITIATION EXHIBIT DISTINCT TEMPORAL BEHAVIORS FOR THE NASCENT RNA DISTRIBUTION

To gain mechanistic insights into the dynamics of transcription initiation, we apply our analytical results to three well-studied models of initiation: (a) Poisson initiation [13,41], (b) ON-OFF initiation [26,42], and (c) two-step initiation [38,43], as shown in Fig. 2. These three models have been established as the canonical models of initiation due to their simplicity and their ability to successfully capture the mechanisms of transcription initiation of many genes, both in prokaryotes and eukaryotes [6,26,44]. The key finding of this section is that these three models predict qualitatively different behaviors at the nascent RNA level, for the mean and the variance as functions of time.

A. Poisson initiation model

In the Poisson initiation model, the promoter remains in an active state and initiates transcription at a constant rate r, as shown in Fig. 2(a). This is a well-studied model of transcription and fully captures the initiation mechanism of many constitutively expressed genes in various organisms [44]. We compute the transient and steady state mean and variance of the distribution of nascent RNAs using the results in the previous section.

The analytical expressions for the mean, $M(t)$, and variance, $\mathrm{Var}(t)$, of the nascent RNA distribution using Eq. (4) are given by

$$\begin{aligned} M(t) &= rt, \ \ t \leqslant T \\ &= rT, \ \ t > T, \\ \mathrm{Var}(t) &= rt, \ \ t \leqslant T \\ &= rT, \ \ t > T. \end{aligned} \tag{6}$$

In the transient regime, the mean goes linearly as a function of time for $t \leqslant T$ [Fig. 2(a)], where T is the time a single RNAP takes to traverse the entire length of the gene of interest. At a time greater than T, the mean is constant and is given by the product of the initiation rate and T. Similarly, the variance also grows linearly in time for $t \leqslant T$. At any time greater than T, the variance becomes constant and equals the value of the mean.

We can obtain the steady state mean and variance of the nascent RNA distribution, by taking the limit of time $t \to \infty$,

$$\begin{aligned} M &= \lim_{t\to\infty} M(t) = rT, \\ \mathrm{Var} &= \lim_{t\to\infty} \mathrm{Var}(t) = rT. \end{aligned} \tag{7}$$

We arrive at the same results for the steady state mean and the variance from Eq. (5). It must be noted that the Fano factor which is defined as the ratio of the variance and mean is independent of time and is 1 for this model, as shown in Fig. 2(a). Hence, this model can serve as a point of reference while discussing more complicated initiation models, as will be discussed in the next section.

B. ON-OFF initiation model

Next, we consider the ON-OFF model of initiation. This model has been established as the "hydrogen atom" model of transcriptional regulation due to its effectiveness in explaining "bursty" gene expression observed for a wide range of genes [45]. As shown in Fig. 2(b), in this model, the promoter switches between a transcriptionally active (ON) and an inactive (OFF) state. The rate of switching from the ON state to the OFF state is k_{OFF}, and from the OFF to ON state is k_{ON}. The

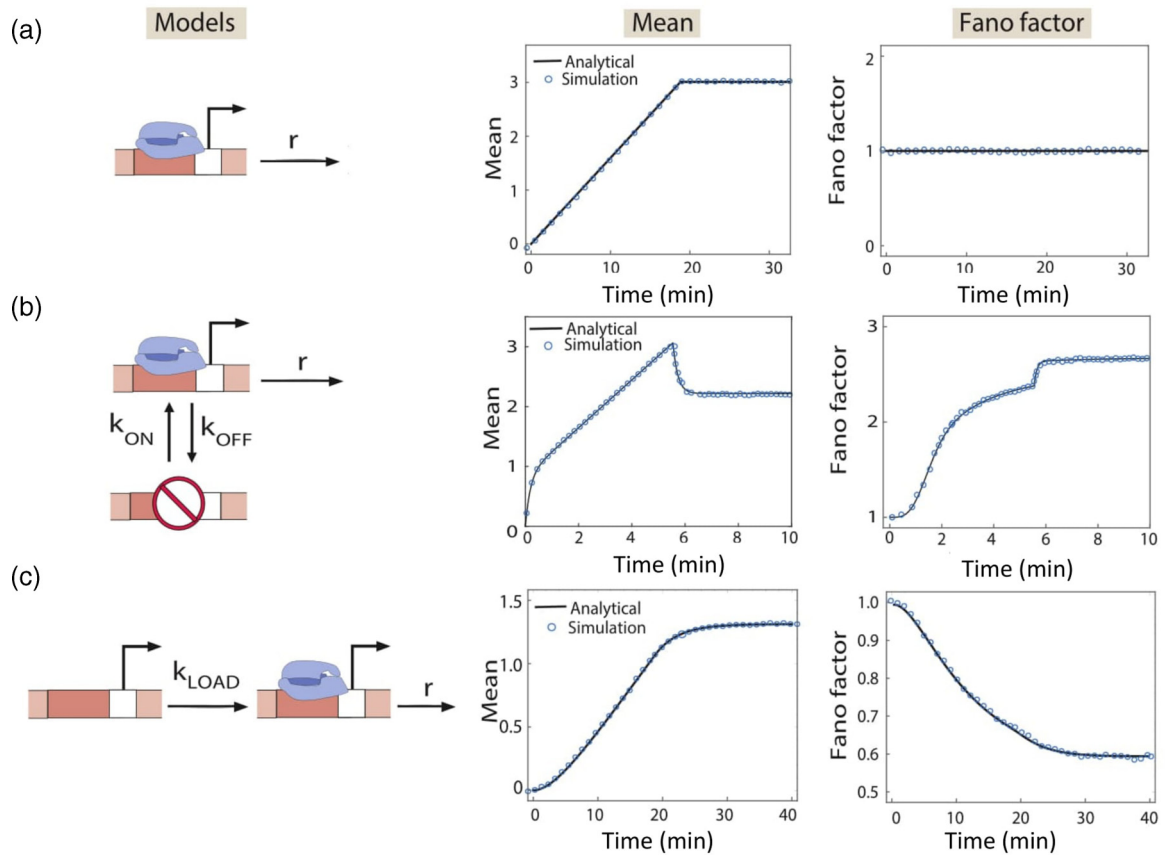

FIG. 2. (a) Poisson initiation model. The promoter always remains active and initiates transcription at a rate r. Using our analytical results, we explore the mean and the Fano factor as functions of time, for this model. We also use Gillespie simulations to confirm the analytical results. To make these plots, we use the following parameters: $r = 0.16/\text{min}$, $v_{\text{EL}} = 0.8\,\text{kb/min}$, and $L = 15\,\text{kb}$, i.e., $T = 18.75\,\text{min}$. These parameters are characteristic of various genes (such as PDR5, MDN1, etc.) in yeast, as reported in [6]. (b) ON-OFF initiation model. The promoter switches between two states: an active and an inactive one. The rate of switching from the active state to the inactive state is k_{OFF}, and from the inactive to the active state is k_{ON}. From the active state, transcription initiation occurs at a rate r. Mean and Fano factor profiles: From the analytical expressions we have obtained, we explore the mean and Fano factor as functions of time. Simulations results are also shown. To illustrate this point, we use the following: $k_{\text{ON}} = 0.435/\text{min}$, $k_{\text{OFF}} = 5$, $k_{\text{INI}} = 5/\text{min}$, $L = 4436\,\text{bps}$, and $v_{\text{EL}} = 0.8\,\text{kb/min}$, which are characteristic of the PDR5 promoter in yeast, as reported in data published in [6]. (c) Two-step model of initiation. Initiation happens in two sequential steps: the formation of the preinitiation complex at the promoter occurs with rate k_{LOAD} followed by RNA polymerase escaping the promoter leading to an initiation event at rate r. Predictions for the mean and Fano factor profiles are shown as functions of time. Simulations results are shown. For the two-step model, we use $k_{\text{LOAD}} = 0.14/\text{min}$, $r = 0.14/\text{min}$, and $v_{\text{EL}} = 0.8\,\text{kb/min}$, characteristic of the MDN1 promoter, which we find by analyzing the data reported in [25].

rate of transcription initiation in the ON state is r and in the OFF state is zero. The ON and OFF states might correspond to a free promoter state and another bound by a repressor, or a nucleosome, respectively.

The analytical formulas for the transient and steady state mean of the nascent RNA distribution are given by

$$\begin{aligned} M(t) &= -Ae^{-Ct} + Bt + A, \quad t \leqslant T \\ &= Ae^{-Ct}[-1 + e^{TC}] + BT, \quad t > T, \\ M &= \frac{k_{\text{ON}} r T}{(k_{\text{ON}} + k_{\text{OFF}})}, \end{aligned} \tag{8}$$

where

$$A = \frac{r k_{\text{OFF}}}{(k_{\text{ON}} + k_{\text{OFF}})^2}, \quad B = \frac{r k_{\text{ON}}}{(k_{\text{ON}} + k_{\text{OFF}})}, \quad C = (k_{\text{ON}} + k_{\text{OFF}}).$$

See the Appendix for the detailed derivation. Interestingly, the mean nascent RNA number displays three regimes in the transient phase. Initially the mean increases just as it would for a Poisson initiation model with initiation rate r, since in this regime the promoter remains in the ON state before switching to the OFF state. In the second regime, the mean level increases with a smaller slope as the promoter starts to switch between the two states. In the third regime, the mean level increases and overshoots the steady state level, unlike the Poisson model of initiation, as shown in Fig. 2(b). After time T, the mean decreases and eventually reaches the steady state value. The amount by which the mean overshoots is given by $A - Ae^{-CT}$, which is a function of the different kinetic rates. This signature can be used to discriminate the Poisson and ON-OFF models of initiation.

Next, we compute the transient and steady state expressions of the variance, given by

$$\begin{aligned}\mathrm{Var}(t) &= B^2T^2 + 2A^2e^{-tC}(-1+e^{tC}-tC)\\ &\quad + \frac{4AB(-1+e^{-tC}+tC)}{C}\\ &\quad + A - Ae^{-tC} + Bt - (A - Ae^{-tC} + Bt)^2,\ \ t \leqslant T,\\ &= B^2T^2 + 2A^2e^{-tC}(-1+e^{TC}-TC)\\ &\quad + \frac{ABe^{-(t+T)C}(e^{tC}+e^{TC})(1+e^{TC}(-1+TC))}{C}\\ &\quad - e^{2tC}(A(-1+e^{TC}) + Be^{tC}T)\\ &\quad \times (A(-1+e^{TC}) + e^{tC}(-1+BT)),\ \ t > T.\end{aligned}$$

$$\mathrm{Var} = M\begin{bmatrix}1 + \frac{2rk_{\mathrm{OFF}}}{(k_{\mathrm{ON}}+k_{\mathrm{OFF}})^2}\\ +\frac{2rk_{\mathrm{OFF}}}{(k_{\mathrm{ON}}+k_{\mathrm{OFF}})^3}\left(\frac{e^{-(k_{\mathrm{ON}}+k_{\mathrm{OFF}})T}-1}{T}\right)\end{bmatrix}. \tag{9}$$

See the Appendix for the detailed derivation. We use these expressions to compute the Fano factor, and monitor its behavior as a function of time, as shown in Fig. 2(b). Initially at small times, the promoter mostly remains in the ON state, resulting in a Fano factor of unity. However, as time t increases, the Fano factor goes up and eventually saturates.

C. Two-step initiation model

In this model, transcription initiation proceeds in two sequential steps. In the first step, a RNAP molecule binds to an empty promoter with a probability of k_{LOAD} per unit time, followed by transcription initiation at a rate r, in the second step, as shown in Fig. 2(c). In eukaryotes, the first step of RNAP binding could correspond to the assembly of the transcriptional machinery at the promoter region. Recent experimental studies have shown that for many promoters in yeast [26] and *Escherichia coli* [42], initiation commences through two sequential steps.

We compute the transient $[M(t)]$ and steady state (M) mean of the nascent RNA distribution for this model, using Eqs. (4) and (5). The expressions are given by

$$\begin{aligned}M(t) &= \frac{A}{B^2}(e^{-Bt}-1) + \frac{A}{B}t,\ \ t \leqslant T\\ &= \frac{Ae^{-Bt}[1-e^{BT}+BTe^{Bt}]}{B^2},\ \ t > T,\\ M &= \frac{AT}{B},\end{aligned} \tag{10}$$

where $A = rk_{\mathrm{LOAD}}$, and $B = r + k_{\mathrm{LOAD}}$.

The mean increases as a function of time and eventually reaches the steady state value, as shown in Fig. 2(c).

Next, using Eqs. (4) and (5), we calculate the transient $[\mathrm{Var}(t)]$ and steady state (Var) variance, which are given by

$$\begin{aligned}\mathrm{Var}(t) &= \frac{A}{B^2}(e^{-Bt}-1) + \frac{A}{B}t - \left[\frac{A(e^{-Bt}-1)}{B^2} + \frac{At}{B}\right]^2 - \frac{A^2e^{-Bt}[6+2Bt-e^{Bt}(6-4Bt+B^2t^2)]}{B^4},\ \ t \leqslant T\\ &= \frac{Ae^{-Bt}[1-e^{BT}+BTe^{Bt}]\left[1-\frac{Ae^{-Bt}(1-e^{BT}+BTe^{Bt})}{B^2}\right]}{B^2} - \frac{A^2e^{-Bt}\begin{bmatrix}4+2^{B(t-T)}+2BT+2e^{BT}(-2+BT)\\ -e^{Bt}(2-2BT+B^2T^2)\end{bmatrix}}{B^4},\ \ t > T,\\ \mathrm{Var} &= M\begin{bmatrix}1-\frac{AT}{B}+\frac{A}{TB^2}\\ \times\left(\frac{2-2e^{-BT}}{B} - 2T + BT^2\right)\end{bmatrix}.\end{aligned} \tag{11}$$

The Fano factor (ratio of the variance and mean) is plotted as a function of time, as shown in Fig. 2(c). The steady state value of the Fano factor goes below 1, as expected [26].

For all the three models, we confirm the analytical results for the mean and Fano factors of the nascent RNA distribution using Gillespie simulations [34]. One of the crucial outcomes of our theoretical analysis is that we can discriminate between the Poisson, ON-OFF, and two-step initiation models by tracking the temporal behavior of the nascent RNA distributions.

IV. DISCUSSION

A key challenge in regulatory biology is to discriminate between the mechanisms of transcription *in vivo*. Recent years have seen an explosion of precision measurements at the single-cell level [3,5–8]. By keeping pace with these experimental studies, theoretical models [14,15,26,38,41,46–53] have been developed to unravel the dynamics of transcription. The aim of these studies has been to explicitly test the validity of these models and refine our understanding.

In this manuscript, we focus on the impact of transcription initiation dynamics on the temporal behavior of nascent RNA levels in individual cells, which can be experimentally measured [22,54]. We propose a general model of transcription and for this model compute the exact analytical expressions for the mean and variance of nascent RNA distributions as functions of time. These analytical expressions capture the model behavior of both the transient and steady state regimes. We illustrate the utility of the theory presented here by applying it to three well-known models of initiation, namely, the Poisson, ON-OFF, and two-step models. At the level of both the mean and Fano factor of the nascent RNA level, these three models make distinct predictions, which allows us to discriminate between them.

The analytical framework developed here offers a way to tap into measured distributions of nascent RNAs. The first

two moments of nascent RNA distribution obtained from a population of cells as functions of time can be fitted to a mathematical model that incorporates the stochastic kinetics of transcription. The extracted fitting parameters can be interpreted as representative of the kinetic properties of stochastic gene expression such as burst size, burst frequency, average transcription rate, etc.

It is increasingly becoming clear that different genes are regulated in various ways [14,31,41]. Although the Poisson, two-step, and ON-OFF models are good representatives of transcription initiation mechanisms for many genes, complex promoters with more than two states are widely prevalent [41]. While combinatorial control of transcription initiation by multiple species of transcription factors is commonplace, even promoters that are regulated by a single transcription factor, such as bacterial promoters can have multiple states [41,50,55]. Mammalian genes that show bursty gene expression consist of three promoter states, two of which are inactive and one is active [56]. Our analytical results can be easily applied to these above-mentioned models of transcription to develop a better understanding of transcriptional dynamics.

In the model presented here, it is implicitly assumed that the transcriptional machinery such as RNA polymerases and transcription factors are abundant in the cell and are not limiting. This assumption breaks down if the numbers of these transcriptional resources are limiting [57–59]. This will alter the waiting time distribution between successive initiation events. However, our framework would still be applicable in such cases, as long as we can compute the exact distribution of initiation times for the specific scenario under consideration.

In conclusion, the theoretical analysis presented here offers the necessary tools to connect mechanisms of transcription initiation with single-cell nascent RNA data and gain mechanistic insights into the dynamics of initiation.

ACKNOWLEDGMENTS

We are indebted to Jane Kondev, Alvaro Sanchez, Hernan Garcia, and Jeff Gelles for numerous discussions, which have shaped our understanding of transcriptional regulation. We would also like to thank Clarissa Scholes, Shashank Sekhar, and the anonymous reviewers for detailed and valuable comments on the paper.

APPENDIX: ANALYTICAL EXPRESSIONS FOR THE MEAN AND VARIANCE OF THE NASCENT RNA DISTRIBUTION

1. Waiting time distribution between successive initiation events

The first step towards computing the mean and variance of the nascent RNA distribution for a general model of transcription, as shown in Fig. 1, is to obtain an expression for the waiting time distribution between successive initiation events. We calculate the probability $q(x)$ that two transcription initiation events are separated by time x; i.e., if one initiation event happens at time $x = 0$, the next initiation event occurs at time x later, between x and $x + dx$. We employ a master equation approach to compute the probability P_i for the promoter to be in the ith state at time x without initiating any transcription event between 0 and x. The master equation for the time evolution of P_i is given by

$$\frac{d}{dx}P_i = \sum_{j=1}^{N}[k_{ji}P_j - k_{ij}P_i] - r_i P_i. \tag{A1}$$

The above set of linear equations can be expressed as matrices, as follows:

$$\frac{d}{dx}\boldsymbol{P} = [\boldsymbol{K} - \boldsymbol{R}]\boldsymbol{P}. \tag{A2}$$

The vector $\boldsymbol{P} = (P_1, P_2, \ldots, P_N)$, contains probabilities for all possible promoter states at time x, without having made a transcript up to time x. Matrix $\boldsymbol{K}$ contains rates of transitioning between the different promoter states; k_{ij} is the rate at which the promoter switches from the jth state to the ith state. $\boldsymbol{R}$ contains transcription initiation rates from all the different states. The solution to this matrix equation is given by

$$\boldsymbol{P}(x) = e^{[\boldsymbol{K}-\boldsymbol{R}]x}\boldsymbol{P}(x = 0). \tag{A3}$$

To find $\boldsymbol{P}$, we need to specify the initial condition for $\boldsymbol{P}$, i.e., $\boldsymbol{P}(x = 0)$. In our model, $\boldsymbol{P}(x = 0)$ is given by a column vector whose entries depend on the details of the transcription initiation models we study. This will be demonstrated when we explore specific examples in the ensuing sections. We can solve the above equation by finding the eigenvalues and eigenvectors of the $\boldsymbol{K}$-$\boldsymbol{R}$ matrix appearing in the exponent of the exponential function. In the ensuing sections, we will show the solutions of this equation for a few specific cases.

The probability that two initiation events are separated by time x, and the second initiation event happens at time x between x and $x + dx$ is given by

$$q(x) = r P_m(x). \tag{A4}$$

$q(x)$ is a product of the probability P_m that the promoter is in the mth state at time x (and no other transcription initiation events have happened before this time), and the probability rdx that an RNAP molecule initiates transcription from this state between time x and $x + dx$. It must be noted that transcription initiation happens only from the mth state.

The waiting time distribution between successive initiation events is an important random variable in the field of queueing theory [37] as will be clear in the next section.

2. Renewal function: Expectation value of the number of initiation events within a given period

We next obtain an analytical expression for the renewal function for our model. Renewal function is a key quantity in the field of queueing theory [37] since most of the available analytical results are obtained using the renewal function [37].

To introduce the idea of renewal function to the readers, we briefly introduce the key concepts of renewal theory. We reproduce some of the known results in renewal theory which can be found in the classic text on probability theory by Ross and Pekoz [37]. It is well known that for a Poisson counting process where an event occurs with a constant rate, the times between successive events are independent and identically distributed (iid) exponential random variables [37]. This can be generalized for other counting processes for which the times between successive events are independent

and identically distributed with an arbitrary distribution. Such counting processes are called renewal processes. One example of renewal process is the lightbulb example [37]. This example deals with the lifespan of a lightbulb. Let us assume that we have one lightbulb in a room and we turn it on. The bulb is kept turned on until it runs out. When the bulb runs out we switch on another bulb. We assume that the time it takes to switch on the lightbulb is zero. If the life span of one lightbulb does not affect the life span of another, the life spans of different bulbs are identically distributed. Hence this process, $\{N(t),\ t \geqslant 0\}$, is a renewal process where $N(t)$ represents the number of lightbulbs that have failed by time t.

The mean of the distribution $P[N(t)]$ of the number of renewal events in time t is given by $R(t)$. The function $R(t)$ is known as the *renewal function*. It is a key quantity in the renewal theory literature due to its usefulness in obtaining different statistical properties of the system such as the distribution of the number of renewals. Let us assume that the distribution of waiting times between successive arrival events is given by $q(x)$, where two arrival events are separated by time x; i.e., if one arrival event happens at time 0, the next arrival occurs at time x later, between x and $x + dx$. There is a one-to-one correspondence between the cumulative distribution of interarrival distributions $Q(t) = \int_0^t q(x)dx$. and the renewal functions $R(t)$, which is given by

$$R(t) = Q(t) + \int_0^t R(t-x)q(x)dx. \tag{A5}$$

In order to obtain the renewal function, we need to solve the renewal equation. By taking a Laplace transform of both sides of the renewal equation, Eq. (A5), we arrive at the following relationship:

$$L_R(s) = \frac{L_x(s)}{s[1 - L_x(s)]}. \tag{A6}$$

Here $L_R(s)$ and $L_x(s)$ are given by

$$L_x(s) = \int_0^\infty q(x)e^{-sx}dx, \quad \text{and} \quad L_R(s) = \int_0^\infty R(t)e^{-st}dt. \tag{A7}$$

The strategy here is to first compute the Laplace transform $L_x(s)$ of the waiting time distribution given by Eq. (A4). Substituting the expression of $L_x(s)$ in Eq. (A6), we obtain an expression for $L_R(s)$, which is the Laplace transform of the renewal function. To obtain the renewal function, we need to compute the inverse Laplace transform of $L_R(s)$, which is given by

$$R(t) = \frac{1}{2\pi i} \lim_{T\to\infty} \int_{\gamma - iT}^{\gamma + iT} e^{st} L_R(s)ds. \tag{A8}$$

With the renewal function at hand, we can compute the first two moments of the nascent RNA distribution as functions of time using known results in the queueing theory literature [40].

In a fine piece of work, Liu *et al.* [40] computed the moments of the number of busy servers in a queue for a wide range of models defined by the $G/D/\infty$ system. From the set of analytical results that they obtained, we find the mean $M(t)$ of the nascent RNA distribution as a function of time is given by

$$\begin{aligned} M(t) &= R(t),\ t \leqslant T \\ &= R(t) - R(t-T),\ t > T. \end{aligned} \tag{A9}$$

The mean is a piecewise-defined function with two domains in time defined by less or greater than time T. The same is true for the variance $\mathrm{Var}(t)$, given by

$$\begin{aligned} \mathrm{Var}(t) &= 2\int_0^t R(t-y)dR(y) + R(t) - [R(t)]^2,\ t \leqslant T \\ &= 2\int_{t-T}^t R(t-y)dR(y) + M(t)[1 - M(t)],\ t > T. \end{aligned} \tag{A10}$$

In the steady state, when time $t \to \infty$, we get the following expressions for the mean(M) and variance (Var):

$$M = \lambda T, \tag{A11}$$

$$\mathrm{Var} = 2\lambda \int_0^T R(t)dt + \lambda T(1 - \lambda T). \tag{A12}$$

Here λ is the long-time average of the number of initiation events per unit time. This quantity is a product of the steady state probability of the promoter being in the mth state and the rate of initiation from this state, r, given by

$$\lambda = r\, {P_m}^{ss}. \tag{A13}$$

Here ${P_m}^{ss}$ is the steady state probability for the promoter to be in the first state. Steady state probabilities are given by the following equation:

$$\boldsymbol{K}\boldsymbol{P} = 0, \tag{A14}$$

where the normalization condition to be satisfied is $\sum_{m=1}^{N} {P_m}^{ss} = 1$. The procedure described above can be applied to promoters with any number of states, one of which is transcriptionally active. In the proceeding sections, we demonstrate how to use the results developed here to find analytical expressions for the first two moments of nascent RNA distribution for some of the well-studied models [41] of transcription.

3. Poisson initiation model

In the Poisson initiation model, the promoter remains in an active state and initiation happens at a constant rate r, as shown in Fig. 2(a).

To compute the transient and steady state mean and variance of the number of nascent RNAs along a gene of interest using the results obtained in the previous section, first we compute the waiting time distribution between successive initiation events. Using Eq. (A1), we write down the master equation for the probability $P(x)$ that no initiation event has occurred between time 0 and x,

$$\frac{d}{dx}P = -rP. \tag{A15}$$

The solution to this master equation is given by

$$P(x) = e^{-rx}. \tag{A16}$$

We can obtain the waiting time distribution using Eq. (A4),

$$q(x) = re^{-rx}. \tag{A17}$$

Next, we compute the renewal function using the procedure described in the previous section.

Following Eq. (A7), the Laplace transform $L_x(s)$ for the waiting time distribution is given by

$$L_x(s) = \int_0^\infty r e^{-rx} e^{-st} dx. \tag{A18}$$

This integral can be very easily computed using standard results, available in *Mathematica*. $L_x(s)$ is given by

$$L_x(s) = \frac{r}{s+r}. \tag{A19}$$

We compute $L_R(s)$, using Eq. (A6),

$$L_R(s) = \frac{r}{s^2}. \tag{A20}$$

Using the expression of $L_R(s)$, we can obtain an expression for the renewal function $R(t)$ from Eq. (A8). The inverse Laplace transform can be computed using standard results from *Mathematica* and is given by

$$R(t) = rt. \tag{A21}$$

The functional form of the renewal function can be easily intuited. In this model, the promoter is assumed to initiate transcription with a constant probability r per unit time. Hence within a given duration from 0 to t, the number of renewals is rt.

Being armed with the renewal function we can obtain closed form analytical expressions for the mean and variance as functions of time. Using Eq. (A9), we find the mean nascent RNA number $M(t)$ is given by

$$\begin{aligned} M(t) &= rt, \ \ t \leqslant T \\ &= rT, \ \ t > T. \end{aligned} \tag{A22}$$

As expected, in the transient regime, the mean goes linearly as a function of time for $t \leqslant T$. At any time greater than T, the mean is constant and is simply given by the product of the initiation rate and T (the time a single RNAP takes to traverse the gene of interest), as shown in Fig. 2(b).

The variance as a function of time can be obtained from Eq. (A10). Using simple algebra, we arrive at an expression for the variance [Var(t)],

$$\begin{aligned} \mathrm{Var}(t) &= rt, \ \ t \leqslant T \\ &= rT, \ \ t > T. \end{aligned} \tag{A23}$$

Like the mean, variance also grows linearly in time for $t \leqslant T$. At any time greater than T, the variance becomes constant and equals the value of the mean. In the steady state, i.e., in the limit $t \to \infty$ we find

$$\mathrm{Var} = \lim_{t\to\infty} \mathrm{Var}(t) = rT. \tag{A24}$$

We can arrive at the same results for the mean and the variance from Eqs. (A11) and (A12).

4. ON-OFF initiation model

Next, we consider the ON-OFF model. As shown in Fig. 2(b), in this model the promoter switches between a transcriptionally active (ON) and an inactive (OFF) state. The rate of transcription initiation in the ON state is r and in the OFF state is zero. To compute the mean and variance for this model as functions of time we consider the waiting time distribution $q(x)$ between successive initiation events. We monitor the probability $q(x)$ of two transcription initiation events being separated by time x. The probabilities of the promoter being in the ON and OFF states at time x are $P_{\mathrm{ON}}(x)$ and $P_{\mathrm{OFF}}(x)$, respectively. To compute the probability distributions $P_{\mathrm{ON}}(x)$ and $P_{\mathrm{OFF}}(x)$, we write down the master equation as employed for the one-step initiation model, by monitoring all the possible ways that lead to changes in the fraction of time the promoter stays in the ON or the OFF state. Following Eq. (A1), the time evolution equations for the probability $P_{\mathrm{ON}}(x)$ and $P_{\mathrm{OFF}}(x)$ are given by

$$\frac{d}{dx} P_{\mathrm{ON}}(x) = k_{\mathrm{ON}} P_{\mathrm{OFF}}(x) - (r + k_{\mathrm{OFF}}) P_{\mathrm{ON}}(x), \tag{A25}$$

$$\frac{d}{dx} P_{\mathrm{OFF}}(x) = -k_{\mathrm{ON}} P_{\mathrm{OFF}}(x) + k_{\mathrm{OFF}} P_{\mathrm{ON}}(x). \tag{A26}$$

In the matrix form, the above set of equations can be written as

$$\frac{d}{dt}\mathrm{P} = [\mathrm{K} - \mathrm{R}]\mathrm{P}. \tag{A27}$$

Here $\boldsymbol{P}$ is given by $\boldsymbol{P} = (P_{\mathrm{ON}}, P_{\mathrm{OFF}})$. $\mathrm{K} = \begin{pmatrix} -k_{\mathrm{OFF}} & k_{\mathrm{ON}} \\ k_{\mathrm{OFF}} & -k_{\mathrm{ON}} \end{pmatrix}$, $\mathrm{R} = \begin{pmatrix} -r & 0 \\ 0 & 0 \end{pmatrix}$ are the rate matrices.

Since transcription initiation happens only in the ON state, the initial conditions are $P_{\mathrm{ON}}(x = 0) = 1$, $P_{\mathrm{OFF}}(x = 0) = 0$.

By computing the eigenvalues and eigenvectors of the exponential function and with a bit of algebra, we find

$$q(x) = A_1 k_1 e^{-k_1 x} + A_2 k_2 e^{-k_2 x}, \tag{A28}$$

where the constants k_1, k_2, A_1, and A_2 are given by

$$k_{1,2} = \frac{1}{2}[k_{\mathrm{ON}} + k_{\mathrm{OFF}} + r \pm \sqrt{(k_{\mathrm{ON}} + k_{\mathrm{OFF}} + r)^2 - 4 r k_{\mathrm{ON}}}], \tag{A29}$$

$$A_1 = \frac{r - k_2}{k_1 - k_2}, \tag{A30}$$

$$A_2 = 1 - A_1. \tag{A31}$$

Next, following the protocol as before, we take the Laplace transforms of the renewal equation. After a bit of algebra using Eq. (A6), we get

$$L_R(s) = \frac{s(A_1 k_1 + A_2 k_2) + k_1 k_2}{s^2[s + (k_1 + k_2 - A_1 k_1 - A_2 k_2)]}. \tag{A32}$$

Taking the inverse Laplace transform, we obtain an expression $R(t)$ for the renewal function,

$$R(t) = -A e^{-Ct} + Bt + A, \tag{A33}$$

where

$$A = \frac{r k_{\mathrm{OFF}}}{(k_{\mathrm{ON}} + k_{\mathrm{OFF}})^2}, \quad B = \frac{r k_{\mathrm{ON}}}{(k_{\mathrm{ON}} + k_{\mathrm{OFF}})},$$
$$C = (k_{\mathrm{ON}} + k_{\mathrm{OFF}}).$$

With the renewal function at hand, we can calculate the transient expressions for the first two moments of the nascent RNA distribution.

The mean nascent RNA number $M(t)$ is given by

$$\begin{aligned} M(t) &= -Ae^{-Ct} + Bt + A, \ \ t \leqslant T \\ &= Ae^{-Ct}[-1 + e^{TC}] + BT, \ \ t > T. \end{aligned} \tag{A34}$$

In the limit of $t \to \infty$, we acquire the steady state expression of the mean M, given by

$$M = \frac{k_{\rm ON} r T}{(k_{\rm ON} + k_{\rm OFF})}. \tag{A35}$$

Next, we compute the variance $\mathrm{Var}(t)$ as a function of time. The variance is given by

$$\begin{aligned} \mathrm{Var}(t) &= B^2T^2 + 2A^2e^{-tC}(-1 + e^{tC} - tC) + \frac{4AB(-1 + e^{-tC} + tC)}{C} + A - Ae^{-tC} + Bt - (A - Ae^{-tC} + Bt)^2, \ \ t \leqslant T, \\ &= B^2T^2 + 2A^2e^{-tC}(-1 + e^{TC} - TC) + \frac{ABe^{-(t+T)C}(e^{tC} + e^{TC})(1 + e^{TC}(-1 + TC))}{C} \\ &\quad - e^{2tC}(A(-1 + e^{TC}) + Be^{tC}T)(A(-1 + e^{TC}) + e^{tC}(-1 + BT)), \ \ t > T. \end{aligned} \tag{A36}$$

In the $t \to \infty$ limit, we get the steady state expression for the variance. This expression has been reported before [26].

$$\mathrm{Var} = M\left[1 + \frac{2rk_{\rm OFF}}{(k_{\rm ON}+k_{\rm OFF})^2} + \frac{2rk_{\rm OFF}}{(k_{\rm ON}+k_{\rm OFF})^3}\left(\frac{e^{-(k_{\rm ON}+k_{\rm OFF})T}-1}{T}\right)\right]. \tag{A37}$$

We use these expressions to compute the Fano factor, and monitor its behavior as a function of time, as shown in Fig. 2(b).

5. Two-step initiation model

In this model, transcription initiation proceeds in two sequential steps. In the first step, a RNAP molecule binds to an empty promoter with a probability of $k_{\rm LOAD}$ per unit time, followed by the escape of the promoter bound RNAP molecule at a rate r in the second step, as shown in Fig. 2(c). Using Eq. (A1), we write down the master equations for the probability of the promoter to be in the empty $[P_1(x)]$ and RNAP bound $[P_2(x)]$ state after time x, given by

$$\frac{d}{dx}P_1(x) = -k_{\rm LOAD}P_1(x) + rP_2(x), \tag{A38}$$

$$\frac{d}{dx}P_2(x) = -rP_2(x) + k_{\rm LOAD}P_1(x). \tag{A39}$$

We can solve the above set of master equations using the matrix method, as shown in the first section. Transcription initiation happens only in the empty promoter state $[P_1(x)]$ since after every initiation event the promoter goes back to the empty state. Hence the initial conditions are given by $P_1(x=0) = 1$, $P_2(x=0) = 0$. Using Eq. (A4), we can easily compute the waiting time distribution $q(x)$, which is given by

$$q(x) = \frac{rk_{\rm LOAD}}{k_{\rm LOAD} - r}(e^{-rx} - e^{-k_{\rm LOAD}x}). \tag{A40}$$

Next, using Eq. (A6), we obtain the Laplace transforms of the renewal function $L_R(s)$, given by

$$L_R(s) = \frac{A}{B^2}\left[\frac{1}{s+B} - \frac{1}{s} + \frac{B}{s^2}\right], \tag{A41}$$

where $A = rk_{\rm LOAD}$ and $B = r + k_{\rm LOAD}$.

Taking the inverse Laplace transform [Eq. (A8)], we obtain an expression $R(t)$ for the renewal function,

$$R(t) = \frac{A}{B^2}(e^{-Bt} - 1) + \frac{A}{B}t. \tag{A42}$$

With the renewal function at hand, we can easily calculate the mean $M(t)$ and variance $\mathrm{Var}(t)$ of the nascent RNA distribution, using Eqs. (A9) and (A10).

The mean $M(t)$ as a function of time is given by

$$\begin{aligned} M(t) &= \frac{A}{B^2}(e^{-Bt} - 1) + \frac{A}{B}t, \ \ t \leqslant T \\ &= \frac{Ae^{-Bt}[1 - e^{BT} + BTe^{Bt}]}{B^2}, \ \ t > T. \end{aligned} \tag{A43}$$

The mean increases as a function of time and asymptotically reaches the steady state value of $M = \frac{AT}{B}$, as shown in Fig. 2(c).

Next, using Eq. (A10), we obtain the variance $\mathrm{Var}(t)$ as a function of time, given by

$$\begin{aligned} \mathrm{Var}(t) &= \frac{A}{B^2}(e^{-Bt} - 1) + \frac{A}{B}t - \left[\frac{A(e^{-Bt} - 1)}{B^2} + \frac{At}{B}\right]^2 - \frac{A^2e^{-Bt}[6 + 2Bt - e^{Bt}(6 - 4Bt + B^2t^2)]}{B^4}, \ \ t \leqslant T \\ &= \frac{Ae^{-Bt}[1 - e^{BT} + BTe^{Bt}]\left[1 - \frac{Ae^{-Bt}(1-e^{BT}+BTe^{Bt})}{B^2}\right]}{B^2} - \frac{A^2e^{-Bt}\left[\begin{matrix}4 + 2^{B(t-T)} + 2BT + 2e^{BT}(-2 + BT) \\ -e^{Bt}(2 - 2BT + B^2T^2)\end{matrix}\right]}{B^4}, \ \ t > T. \end{aligned} \tag{A44}$$

We obtain the steady state variance of the nascent RNA level in the limit $t \to \infty$, given by

$$\text{Var} = M\left[\begin{array}{l} 1 - \frac{AT}{B} + \frac{A}{TB^2} \\ \times\left(\frac{2-2e^{-BT}}{B} - 2T + BT^2\right) \end{array}\right]. \tag{A45}$$

The Fano factor (ratio of the variance and mean) is plotted as a function of time shown in Fig. 2(c). The steady state value of the Fano factor goes below 1, as expected [26].

6. Limitations of the model

Our model assumes that transcription initiation time scales are much slower compared to the elongation time scale. Consequently, RNAPs do not interfere with each other while moving along the gene. This approximation is reasonable for all but the strongest promoters such as ribosomal promoters characterized by very fast initiation [60,61]. In a previous work [26], we demonstrated this explicitly using numerical simulations which included a detailed model of transcription elongation that involved excluded-volume interaction between adjacent RNAP molecules, as well as pausing of RNAPs along the gene [4,8]. We showed that the mean and variance of steady state nascent RNA distributions based on a simple model of transcription with stochastic initiation and deterministic elongation, and the simulation results based on a more realistic model of elongation that incorporates traffic jams and pausing of RNAPs, only start to break down when the initiation time scales become comparable to the elongation time scales. For a detailed discussion on this issue, please see Ref. [26].

In our manuscript, we monitor the RNA polymerase number on a gene as a function of time. Since most of the existing experimental methods, such as single-molecule FISH and MS2, count nascent RNAs, we use nascent RNA counts as a proxy for the RNAP counts. However, while considering experiments that count nascent RNAs it is important to be mindful of the fact that the number of RNAP molecules along a gene is not necessarily equal to the nascent RNA counts. In our previous work, we have discussed this point in detail [26]. Transcribing polymerase molecules have partial nascent transcripts attached to them depending on how far along the gene they have moved (as indicated in Fig. 1). In single-molecule FISH experiments, the RNA sequence that is targeted by the fluorescent probes determines if these transcripts are detected or not. Probes against the $5'$ end detect transcripts early on, while probes against the $3'$ end will detect only almost finished transcripts [62,63]. However, if there is a way to correctly extract the RNAP number distribution from nascent RNA intensity, our model can accurately transform these data into information about the transcriptional dynamics. Moreover, recent experiments involving MS2 tags have shown that the nascent RNA count is a good proxy for nascent RNA counts [22,64].

It is important to notice the limitations of the mathematical framework developed here in this manuscript. Our goal here is to test specific hypotheses about biological mechanisms by comparing our model predictions with the data. However, like all quantitative models, our model would be most informative when there is a discrepancy between the model predictions and experimental data. This would offer the opportunity to discard wrong assumptions we make about the dynamics of transcription. Also, such a scenario would suggest that we either need to include more factors to the model or come up with a different model.

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th ed. (Garland Science, New York, 2002).

[2] J. M. Raser and E. K. O'Shea, Science **309**, 2010 (2005).

[3] B. Munsky, G. Neuert, and A. van Oudenaarden, Science **336**, 183 (2012).

[4] A. Raj and A. van Oudenaarden, Cell **135**, 216 (2008).

[5] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, PLoS Biol. **4**, e309 (2006).

[6] D. Zenklusen, D. R. Larson, and R. H. Singer, Nat. Struct. Mol. Biol. **15**, 1263 (2008).

[7] G.-W. Li and X. S. Xie, Nature **475**, 308 (2011).

[8] A. M. Femino, F. S. Fay, K. Fogarty, and R. H. Singer, Science **280**, 585 (1998).

[9] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Science **329**, 533 (2010).

[10] T. Lionnet, Mol. Syst. Biol. **9**, 710 (2013).

[11] J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie, Science **311**, 1600 (2006).

[12] L. Cai, N. Friedman, and X. S. Xie, Nature **440**, 358 (2006).

[13] A. Raj and A. van Oudenaarden, Annu. Rev. Biophys. **38**, 255 (2009).

[14] Á. Sánchez and J. Kondev, Proc. Natl. Acad. Sci. U.S.A. **105**, 5081 (2008).

[15] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. U.S.A. **105**, 17256 (2008).

[16] N. Kumar, A. Singh, and R. V. Kulkarni, PLoS Comput. Biol. **11**, e1004292 (2015).

[17] D. L. Jones, R. C. Brewster, and R. Phillips, Science **346**, 1533 (2014).

[18] G. Q. Dong and D. R. McMillen, Phys. Rev. E **77**, 021908 (2008).

[19] A. Singh and P. Bokes, Biophys. J. **103**, 1087 (2012).

[20] E. Melamud and J. Moult, Nucleic Acids Res. **37**, 4873 (2009).

[21] C. Baker, T. Jia, and R. V. Kulkarni, Phys. Rev. E **85**, 061915 (2012).

[22] H. G. Garcia, M. Tikhonov, A. Lin, and T. Gregor, Curr. Biol. **23**, 2140 (2013).

[23] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Phys. Rev. Lett. **117**, 128101 (2016).

[24] D. R. Larson, R. H. Singer, and D. Zenklusen, Trends Cell Biol. **19**, 630 (2009).

[25] S. J. Gandhi, D. Zenklusen, T. Lionnet, and R. H. Singer, Nat. Struct. Mol. Biol. **18**, 27 (2011).

[26] S. Choubey, J. Kondev, and A. Sanchez, PLoS Comput. Biol. **11**, e1004345 (2015).

[27] S. L. French, Y. N. Osheim, F. Cioci, M. Nomura, and A. L. Beyer, Mol. Cell. Biol. **23**, 1558 (2003).

[28] S. J. Anderson, M. L. Sikes, Y. Zhang, S. L. French, S. Salgia, A. L. Beyer, M. Nomura, and D. A. Schneider, J. Biol. Chem. **286**, 18816 (2011).
[29] S. Quan, N. Zhang, S. French, and C. L. Squires, J. Bacteriol. **187**, 1632 (2005).
[30] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, Science **332**, 475 (2011).
[31] K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J.-M. Victor, M.-C. Robert, S. Capozi, R. Chouaib, V. Bäcker, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk, and E. Bertrand, Nat. Commun. **7**, 12248 (2016).
[32] T. Muramoto, D. Cannon, M. Gierliński, A. Corrigan, G. J. Barton, and J. R. Chubb, Proc. Natl. Acad. Sci. U.S.A. **109**, 7350 (2012).
[33] A. Senecal, B. Munsky, F. Proux, N. Ly, F. E. Braye, C. Zimmer, F. Mueller, and X. Darzacq, Cell Rep. **8**, 75 (2014).
[34] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[35] L. Rosenfeld, E. Kepten, S. Yunger, Y. Shav-Tal, and Y. Garini, Phys. Rev. E **92**, 032715 (2015).
[36] D. L. Bentley, Nat. Rev. Genet. **15**, 163 (2014).
[37] S. M. Ross and E. A. Pekoz, *A Second Course in Probability* (pekozbooks, Boston, 2007).
[38] T. Jia and R. V. Kulkarni, Phys. Rev. Lett. **106**, 058102 (2011).
[39] V. Elgart, T. Jia, and R. V. Kulkarni, Phys. Rev. E **82**, 021901 (2010).
[40] L. Liu, B. R. K. Kashyap, and J. G. C. Templeton, J. Appl. Probab. **27**, 671 (1990).
[41] A. Sanchez, S. Choubey, and J. Kondev, Methods (San Diego, Calif.) **62**, 13 (2013).
[42] A.-B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, and A. S. Ribeiro, Nucleic Acids Res. **40**, 8472 (2012).
[43] T. B. Kepler and T. C. Elston, Biophys. J. **81**, 3116 (2001).
[44] A. Sanchez and I. Golding, Science **342**, 1188 (2013).
[45] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, Cell **123**, 1025 (2005).
[46] V. Elgart, T. Jia, A. T. Fenley, and R. Kulkarni, Phys. Biol. **8**, 046001 (2011).
[47] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Phys. Rev. E **79**, 031911 (2009).
[48] R. D. Dar, B. S. Razooky, A. Singh, T. V. Trimeloni, J. M. McCollum, C. D. Cox, M. L. Simpson, and L. S. Weinberger, Proc. Natl. Acad. Sci. U.S.A. **109**, 17454 (2012).
[49] A. Singh, B. Razooky, C. D. Cox, M. L. Simpson, and L. S. Weinberger, Biophys. J. **98**, L32 (2010).
[50] A. Sanchez, H. G. Garcia, D. Jones, R. Phillips, and J. Kondev, PLoS Comput. Biol. **7**, e1001100 (2011).
[51] M. Thattai and A. van Oudenaarden, Genetics **167**, 523 (2004).
[52] J. Jędrak and A. Ochab-Marcinek, Phys. Rev. E **94**, 032401 (2016).
[53] J. Zhang and T. Zhou, Biophys. J. **106**, 479 (2014).
[54] T. Lucas, T. Ferraro, B. Roelens, J. De Las Heras Chanes, A. M. Walczak, M. Coppey, and N. Dostatni, Curr. Biol. **23**, 2135 (2013).
[55] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips, Curr. Opin. Genet. Dev. **15**, 125 (2005).
[56] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, Science **332**, 472 (2011).
[57] M. Rydenfelt, R. S. Cox, H. Garcia, and R. Phillips, Phys. Rev. E **89**, 012702 (2014).
[58] R. C. Brewster, F. M. Weinert, H. G. Garcia, D. Song, M. Rydenfelt, and R. Phillips, Cell **156**, 1312 (2014).
[59] D. Das, S. Dey, R. C. Brewster, and S. Choubey, PLoS Comput. Biol. **13**, e1005491 (2017).
[60] S. Klumpp and T. Hwa, Proc. Natl. Acad. Sci. U.S.A. **105**, 18159 (2008).
[61] M. Dobrzyński and F. J. Bruggeman, Proc. Natl. Acad. Sci. U.S.A. **106**, 2583 (2009).
[62] L. S. Churchman and J. S. Weissman, Nature **469**, 368 (2011).
[63] M. T. Shawn and C. Little, Cell **154**, 789 (2013).
[64] J. P. Bothma, H. G. Garcia, E. Esposito, G. Schlissel, T. Gregor, and M. Levine, Proc. Natl. Acad. Sci. U.S.A. **111**, 10598 (2014).

Correction: The fourth sentence of the abstract contained a misspelled word and has been corrected.