Predicting variability in nascent RNA from transcription initiation kinetics

Juraj Szavits-Nossan* and Ramon Grima†
School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, United Kingdom
(Dated: March 30, 2022)

Nascent (actively transcribed) RNA accurately reflects transcriptional kinetics. Models predicting its fluctuation signature are difficult to solve because they are inherently non-Markovian—this is because elongation of nascent RNA takes approximately a fixed amount of time. Here, we derive an exact relationship between the distribution of the times between successive initiation events and the steady-state distribution of nascent RNA numbers. We verify the accuracy of the theory using live cell imaging data in prokaryotic cells and using simulations of eukaryotic transcription.

Transcription in single cells occurs in bursts whose size and timing is random [1]. Intuitively, a burst originates from rapid mRNA transcription when a promoter briefly switches on. The experimental distribution of mature mRNA numbers can often be easily fitted using the exact solution of a simple Markov model of gene expression, called the telegraph model [2–4]. This model describes the promoter switching between two states of activity and inactivity, the production of mature mRNA occurring one molecule at a time whilst in the active state, and the degradation of mature mRNA via a first-order reaction. This fitting procedure leads to a set of kinetic parameters that provides a fluctuation signature for each gene. However, in recent years doubts have arisen about the validity of this characterization of gene expression, principally because mature (cellular) mRNA does not provide a direct read-out of transcription [5, 6]—its fluctuations are strongly influenced by various post-transcriptional sources such as splicing, export and DNA replication.

In order to circumvent such issues, it has been proposed that gene expression can be more accurately understood by studying nascent RNA fluctuations, i.e. variability in the number of RNA polymerase (RNAp) molecules that are actively involved in the elongation phase of transcription. By its very definition, this is a direct read-out of transcription. The numbers of actively transcribing RNAp can be estimated directly from nascent single-cell sequencing methods [7] or more commonly using single molecule fluorescence in situ hybridization (sm-FISH) where intronic probes specifically label (non-spliced) nascent RNA [6, 8].

The construction and solution of models of nascent RNA dynamics is an outstanding theoretical challenge. The simplest choice is to describe elongation as a single step reaction. However, such models are difficult to solve because they are inherently non-Markovian—this is because removal of nascent RNA approximately occurs a fixed amount of time after initiation, leading to memory effects over the duration of elongation. Consequently, models of this type have been solved only for very simple cases [5, 9–13] which lack a description of the multistep process of initiation as has been elucidated by experiments [14].

In this Letter, we derive an exact relationship between the probability distribution of the initiation times (the times between successive initiation events) and the steady-state distribution of nascent RNA numbers, under the assumption that the fluctuations in the elongation time are negligible. Importantly, we put no restrictions on the initiation time distribution or the model that it comes from, except that the kinetic rates do not change over time. The procedure is illustrated in Fig. 1 for a kinetic model of eukaryotic transcription initiation. The accuracy of our theory is verified by stochastic simulations and also using experimental data from live cell imaging.

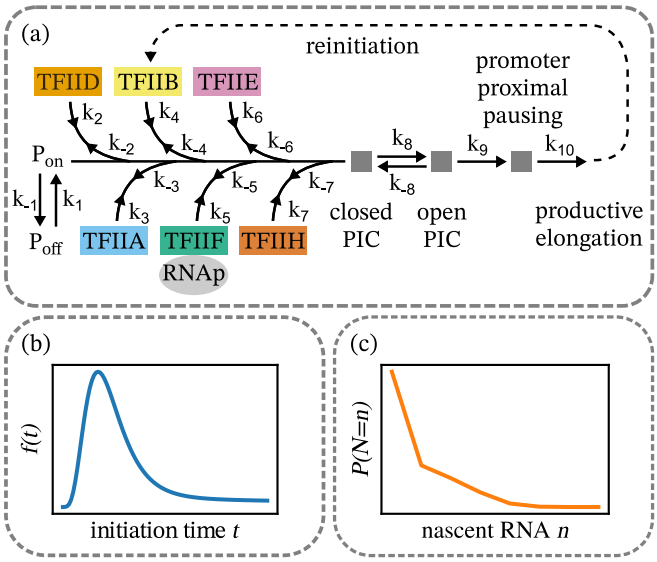

FIG. 1. A schematic of our theoretical approach. (a) A stepwise model of eukaryotic transcription initiation based on [14, 15] (see later in the main text for full details of the model). (b) Probability density $f(t)$ of the initiation time calculated from Eq. (10) for the stepwise model in (a). (c) Stationary distribution of the nascent RNA, calculated from the initiation time distribution $f(t)$ in (b) using Eq. (6).

Distribution of the number of nascent RNA. Transcription is divided into initiation, elongation and termination phases. Initiation is a complex multistep process which starts by the binding of an RNAp and ends with the production of a nascent transcript, i.e. when the RNAp

starts productive elongation. After a fixed time T, we assume that the elongation and termination phases finish, leading to a mature transcript.

This process can be interpreted as a renewal process [16]. A key starting point of our approach is the consideration of the probability density function of the initiation times, i.e. the times between successive initiation events, which we denote by $f(t)$. The assumptions here are that $f(t)$ does not change over time and that there is no transcriptional memory between initiation events, which is supported by experimental evidence [17]. For now, we assume that $f(t)$ and its Laplace transform $f^*(s)$ are known, and later we consider specific examples.

Let $T_1, T_2, T_3 \dots$ be a sequence of times at which initiation events occur, measured from some reference time point $t = 0$, and let $t_1 = T_1$, $t_2 = T_2 - T_1$, $t_3 = T_3 - T_2, \dots$ be the initiation times. We denote the probability density function of t_1 by $f_1(t_1)$, which may differ from $f(t_1)$, unless an initiation event occurred exactly at $t = 0$. From the definition of the initiation times, it follows that the time of the i-th initiation is given by $T_i = \sum_{j=1}^{i} t_j$. Since all t_j are independent, the probability density function of T_i, $k_i(T_i)$, is given by $k_i(T_i) = (f_1 * f^{*(i-1)})(T_i)$, where $*$ stands for convolution and $f^{*(i-1)}$ is the $(i-1)$-fold convolution of f.

Next, we compute the probability distribution of the number of initiation events $N_I(t)$ that occurred up to time t. Note that $N_I(t) = n$ is equivalent to $T_n \leq t < T_{n+1}$. Hence, the probability $P(N_I(t) = n)$ is given by

$$\begin{aligned} P(N_I(t) = n) &= P(T_n \leq t < T_{n+1}), \\ &= K_n(t) - K_{n+1}(t), \end{aligned} \tag{1}$$

where $K_n(t) = \int_0^t dt'\ k_n(t')$ for $n \geq 1$ is the cumulative distribution function of T_n, and $K_0(t) = 1$. From here, it follows that the number of RNA polymerases $N(t)$ actively engaged in elongation at time t is equal to

$$N(t) = \begin{cases} N_I(t), & t \leq T, \\ N_I(t) - N_I(t-T), & t > T, \end{cases} \tag{2}$$

where T is the elongation time.

In order to find the probability distribution of $N(t)$ for $t > T$, we first need to find the probability density function $f_{t_0}(\tau)$ of the forward recurrence time τ, where τ is defined as the time until the next initiation event measured from a fixed time point t_0 (we will later set $t_0 = t - T$). The probability density function of τ is given by

$$f_{t_0}(\tau) = f(t_0 + \tau) + \int_0^{t_0} dt' h(t_0 - t') f(t' + \tau), \tag{3}$$

where $h(t) = \sum_{n=1}^{\infty} k_n(t)$ is the probability to observe an initiation event in the time interval $[t, t + dt]$ [16]. In renewal theory, $h(t)$ is called the renewal density because it is equal to $d\langle N_I(t)\rangle/dt$, where $\langle N_I(t)\rangle$ is the mean of $N_I(t)$. The Laplace transform of $h(t)$ takes a rather simple form in terms of $f^*(s)$, namely $h^*(s) = f^*(s)/[1 - f^*(s)]$. The probability $P(N(t) = n)$ is given by

$$P(N(t) = n) = \begin{cases} 1 - F_{t-T}(T), & n = 0, \\ (f_{t-T} * f^{*(n-1)})(T), & n \geq 1, \end{cases} \tag{4}$$

where $F_{t-T}(\tau)$ is the cumulative distribution function of the forward recurrence time measured from $t - T$.

Eq. (4) simplifies in the steady state limit in which $t_0 \to \infty$. One can show that for large t, $h(t) = 1/\mu$, where μ is the mean initiation time, which follows from a Tauberian theorem that connects large-t behaviour of $h(t)$ with small-s behaviour of $h^*(s)$ [18]. Assuming $\lim_{t\to\infty} f(t) = 0$, the expression for $f_{t_0}(\tau)$ simplifies for $t_0 \to \infty$ to

$$f_\infty(\tau) = \lim_{t_0 \to \infty} f_{t_0}(\tau) = \frac{1 - F(\tau)}{\mu}, \tag{5}$$

where $F(t) = \int_0^t dt'\ f(t')$ is the cumulative distribution function of the initiation time. Since the Laplace transform of $f_\infty(\tau)$ is given by $f^*_\infty(s) = (1 - f^*(s))/(\mu s)$, the Laplace transform $P^*(n, s)$ of the stationary probability distribution $P(N = n)$ with respect to the elongation time T reads,

$$\begin{aligned} P^*(n, s) &= \int_0^\infty dT\ \mathrm{e}^{-Ts} P(N = n) \\ &= \begin{cases} \frac{\mu s - 1 + f^*(s)}{\mu s^2}, & n = 0 \\ \frac{[1 - f^*(s)]^2 [f^*(s)]^{n-1}}{\mu s^2}, & n \geq 1. \end{cases} \end{aligned} \tag{6}$$

Eq. (6) is our main result. It connects the Laplace transform $P^*(n, s)$ to the Laplace transform $f^*(s)$. Later on, we will show how to find $f^*(s)$ for a stepwise model of transcription initiation. We will show that $f^*(s)$ is a rational function of s, which means that $P^*(n, s)$ can be inverted using methods for partial fraction decomposition [19]. The moments of $P(N = n)$ can be computed from the probability generating function $G(z) = \sum_{n=0}^{\infty} z^n P(N = n)$, whose Laplace transform $G^*(z, s)$ with respect to the elongation time T is given by

$$G^*(z, s) = \frac{1}{s} + \frac{(z - 1)[1 - f^*(s)]}{\mu s^2 [1 - z f^*(s)]}. \tag{7}$$

For example, the mean $\langle N \rangle = T/\mu$, and the variance $\mathrm{Var}[N]$ is given by

$$\mathrm{Var}[N] = \mathcal{L}^{-1}\left\{ \frac{1 + f^*(s)}{\mu s^2 [1 - f^*(s)]} \right\}(T) - \left(\frac{T}{\mu}\right)^2, \tag{8}$$

where $\mathcal{L}^{-1}\{\dots\}(T)$ is the inverse Laplace transform of the expression in the curly brackets, evaluated at T.

Models of transcription initiation. Our main result, Eq. (6), requires that $f^*(s)$ is known. Here, we find $f^*(s)$

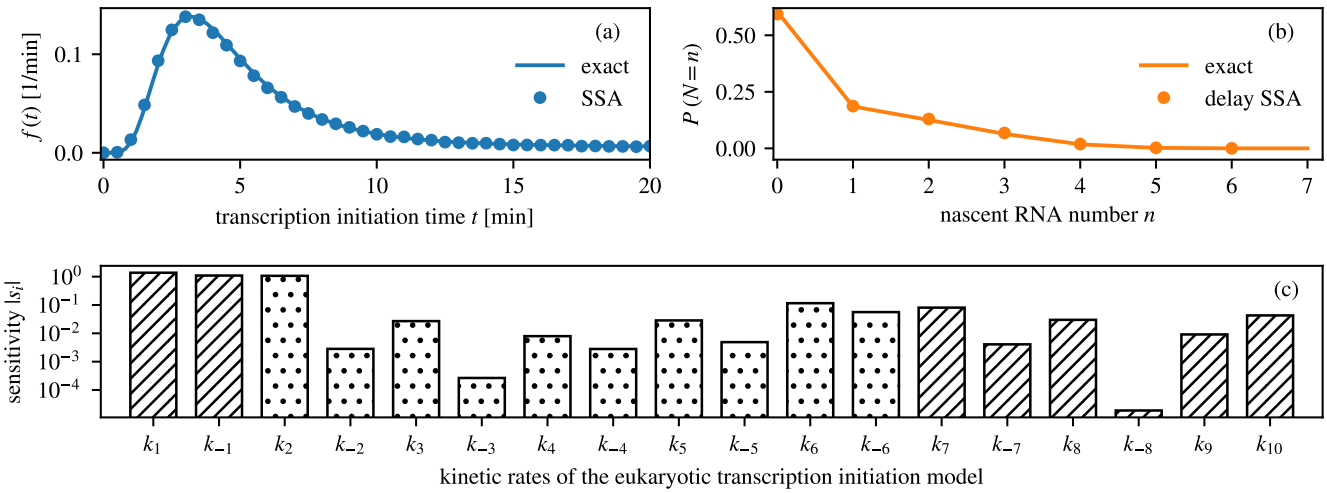

FIG. 2. (a) Probability density function $f(t)$ of the initiation time for the model of eukaryotic transcription initiation illustrated in Fig. 1(a) with reinitiation from the on state ($K = 2$). Solid line is the theoretical prediction from Eq. (10), and points are from stochastic simulations (SSA). (b) Probability distribution of the nascent RNA, obtained from Eq. (6), and compared to stochastic simulations (delay SSA) performed using `DelaySSAToolkit.jl` package in Julia [20, 21]. (c) Absolute value of the sensitivity coefficient $|s_i|$ of the Fano factor FF_N of the nascent RNA distribution (see text for definition). Rates are labelled according to Fig. 1 and Eq. (9). Bars with diagonal lines have $s_i < 0$, whereas bars with dots have $s_i > 0$.

for a generic stepwise model of initiation described by the following reaction scheme,

$$S_1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} S_2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \cdots \underset{k_{-M+1}}{\overset{k_{M-1}}{\rightleftharpoons}} S_M \xrightarrow{k_M} S_K + N, \tag{9}$$

where $K \in [1, M]$. For this model, $f(t)$ and $f^*(s)$ read

$$f(t) = k_M(\exp(\mathbf{A}t))_{MK}, \; f^*(s) = k_M(s\mathbf{I} - \mathbf{A})^{-1}_{MK}, \tag{10}$$

where $\mathbf{A}$ is a tridiagonal matrix defined as

$$\begin{aligned} &A_{11} = -k_1, \; A_{ii} = -(k_{-i+1} + k_i), \; i = 2, \ldots, M, \\ &A_{ii+1} = k_{-i}, \; A_{i+1i} = k_i, \; i = 1, \ldots, M-1, \end{aligned} \tag{11}$$

$\exp(\mathbf{A}t)$ is the matrix exponential and $\mathbf{I}$ is the $M \times M$ identity matrix. We note that $f^*(s)$ can be computed analytically, because $s\mathbf{I} - \mathbf{A}$ is a tridiagonal matrix, for which the inverse is known explicitly [22]. It also follows from Eqs. (10) and (11) that $f^*(s)$ is a rational function of s, hence $f(t)$ can be found from $f^*(s)$ by the partial fraction decomposition.

The reaction scheme in Eq. (9) includes the Poisson process ($S_1 \to S_1 + N$) and the telegraph process ($S_1 \rightleftharpoons S_2 \to S_2 + N$) that have been studied before, and for which the stationary nascent RNA distribution has been computed analytically by solving the delay chemical master equation [5, 9, 12]. We derive these results again using Eq. (6) in Supplemental Material [21]. As a new result, we calculate the nascent RNA distribution for the fully irreversible process with all forward rates equal to λ, for which the initiation time distribution is an Erlang distribution—this distribution is ubiquitous in cellular biology and naturally arises from Markov multistep processes [23],

$$P(N = n) = \frac{e^{-\lambda T}}{M} \left\{ \sum_{j=0}^{N_+ - 1} (N_+ - j) \frac{(\lambda T)^j}{j!} - 2 \sum_{j=0}^{N_0 - 1} (N_0 - j) \frac{(\lambda T)^j}{j!} + \sum_{j=0}^{N_- - 1} (N_- - j) \frac{(\lambda T)^j}{j!} \right\}, \tag{12}$$

where $N_+ = M(n+1)$, $N_0 = Mn$, $N_- = M(n-1)$ [24]. The derivation of Eq. (12) is presented in Supplemental Material [21].

As our next example, we consider a stepwise model of eukaryotic transcription initiation, illustrated in Fig. 1(a) [14, 15] and described by Eq. (9) with ten states ($M = 10$). P_{on} (S_1) and P_{off} (S_2) are the on and off states of the promoter, respectively. General transcription factors and RNA polymerase bind the promoter in the following order: TFIID (S_3), TFIIA (S_4), TFIIB (S_5), TFIIF and RNAp (S_6), TFIIE (S_7) and TFIIH, resulting in the closed preinitiation complex (PIC, S_8). The TFIIH unwinds the promoter DNA, creating an open PIC (S_9) that begins the elongation. In metazoans, the elongating RNA polymerase pauses shortly after the initiation (S_{10}) [25, 26]. The RNAp is eventually released into productive elongation, clearing the promoter for the next round of initiation. Kinetic parameters for this model have been collected from the literature [1, 27–34] and adjusted for the gene PLEC in mouse, for which the on and off switching rates have been measured experimentally [21]. We assumed that the reinitiation starts from the

on state ($K = 2$). Alternatively, TFIID and TFIIA may remain bound until a new round of initiation [35]—this case ($K = 4$) is considered in Supplemental Material [21].

Figs. 2(a) and 2(b) show the initiation time distribution predicted by Eq. (10) and the corresponding nascent RNA distribution predicted by Eq. (6), respectively. Both distributions agree well with the ones obtained by stochastic simulations. The initiation time distribution has a peak at a non-zero value and a long tail due to the off rate k_{-1} being much larger than the on rate k_1 for this gene [1]. The nascent distribution has a prominent peak at zero, followed by a slow, nearly linear decrease. Nascent RNA distributions of similar shape have been observed experimentally [6].

In Fig. 2(c), we show the relative sensitivity coefficient [36] $s_i = (k_i/\mathrm{FF}_N)\partial \mathrm{FF}_N/\partial k_i$ of the Fano factor $\mathrm{FF}_N = \mathrm{Var}[N]/\langle N\rangle$ of the nascent RNA distribution, computed for each of the eighteen model parameters k_i using Eq. (8). A value x of s_i means that a 1% change in the value of k_i causes a x% change in the value of FF_N. According to Fig. 2(c), FF_N is most sensitive to changes in the on and off rates, as well as changes in the binding rate of TFIID, which indeed have been identified previously as rate-limiting steps in transcription initiation [37, 38]. However, we note that the value of s_i will likely change for a different choice of parameters. For example, we get markedly different values of s_i if we assume that TFIID and TFIIA remain bound at the promoter until reinitiation (Fig. S1(c) in Supplemental Material [21]).

Comparison with experimental data. So far, we have assumed a kinetic model of transcription initiation, and from there derived the initiation time distribution. If the initiation times are measured experimentally, then we can fit a function $f_{\mathrm{fit}}(t)$ to the measured distribution and then use $f_{\mathrm{fit}}(t)$ to predict the nascent RNA distribution $P(N = n)$.

One such experiment was reported in Ref [39], in which transcription of a target gene was followed in live *Escherichia coli* cells using MS2 tagging, one initiation event at a time. The target gene was controlled by the *tetA* promoter, which was induced by anhydrotetracycline (aTc) at a concentration of 15 ng/ml at two temperatures, 24°C and 37°C. The probability density function of the initiation times was measured and fitted to a hypoexponential distribution with three parameters (Fig. 3(a) and Fig. S2(a) in Supplemental Material [21]). The Laplace transform of this distribution is given by

$$f^*_{\mathrm{fit}}(s) = \frac{r_1 r_2 r_3}{(s + r_1)(s + r_2)(s + r_3)}. \tag{13}$$

Fig. 3(b) shows the distribution of the fraction of cells with a given number of mRNA, $M(t)$, measured at $t_1 = 1$ h after the induction by aTc at 24°C, compared to the theoretical prediction. Because the MS2-tagged mRNA did not degrade during the experiment, we assumed that the number of mRNA at time t_1, $M(t_1)$, was equal to

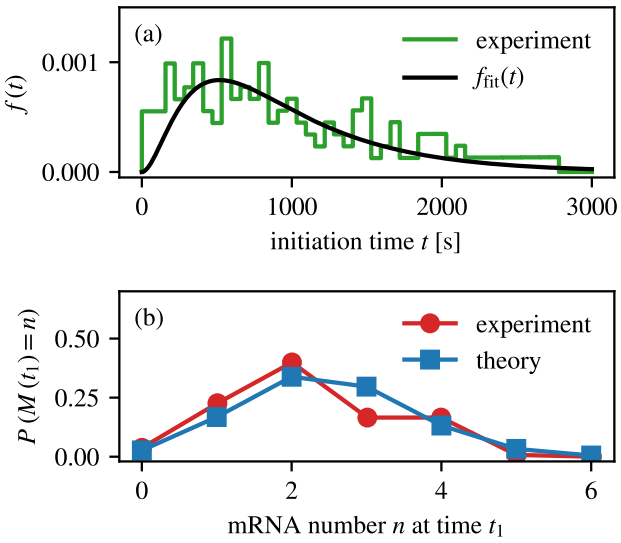

FIG. 3. (a) Distribution of the initiation times measured in Ref. [39] at 24°C. Black line is a fit to the hypoexponential distribution with three parameters, $r_1 = 1/620$ s^{-1}, $r_2 = 1/240$ s^{-1} and $r_3 = 1/115$ s^{-1}, obtained by inverting $f^*_{\mathrm{fit}}(s)$ in Eq. (13). (b) Distribution of the fraction of cells with a given number of mRNA molecules measured 1 h following induction by aTc at 24°C, compared to the theoretical prediction.

the number of initiation events between $t_0 = 20$ min and $t_1 - T$, $N_I(t_1 - T) - N_I(t_0)$, where $t_0 = 20$ min was the time at which full induction was reached, and T was the elongation time. The elongation time was only tens of seconds, hence $t_1 - T \approx t_1$. The theoretical distribution $P(M(t_1) = n)$ was calculated by inverting the Laplace transform $[1 - f^*_{t_0}(s)]/s$ for $n = 0$ and $f^*_{t_0}(s)[f^*_{\mathrm{fit}}(s)]^{n-1}$ for $n \geq 1$, evaluated at $t_1 - t_0 = 40$ min, whereby we assumed $f_{t_0}(\tau) = f_\infty(\tau) = [1 - F_{\mathrm{fit}}(\tau)]/\mu$. We emphasize that no fitting parameters were necessary other than those of $f_{\mathrm{fit}}(t)$. The results at 37°C are also in good agreement with the experimental data (see Fig. S2(b) in Supplemental Material [21]). A very good agreement between experiment and theory suggests that the initiation time distribution carried enough information to correctly predict the distribution of the accumulated mRNA.

Conclusion. We have presented a theory that predicts the distribution of nascent RNA from the initiation time distribution and the elongation time. The theory assumes that the elongation time is fixed. Transcription models of this kind have been solved previously for simple models of initiation kinetics, whereas our results are applicable to broad models of initiation kinetics. The theory also allows for a model-free prediction of the nascent RNA distribution, provided the initiation time distribution is measured experimentally.

This work was supported by a Leverhulme Trust research award (RPG-2020-327).

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

[1] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, Mammalian genes are transcribed with widely different bursting kinetics, Science **332**, 472 (2011).
[2] J. Peccoud and B. Ycart, Markovian modeling of gene-product synthesis, Theoretical Population Biology **48**, 222 (1995).
[3] A. J. Larsson, P. Johnsson, M. Hagemann-Jensen, L. Hartmanis, O. R. Faridani, B. Reinius, Å. Segerstolpe, C. M. Rivera, B. Ren, and R. Sandberg, Genomic encoding of transcriptional burst kinetics, Nature **565**, 251 (2019).
[4] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, Stochastic mrna synthesis in mammalian cells, PLoS Biol **4**, e309 (2006).
[5] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Stochastic kinetics of nascent rna, Phys. Rev. Lett. **117**, 128101 (2016).
[6] S. O. Skinner, H. Xu, S. Nagarkar-Jaiswal, P. R. Freire, T. P. Zwaka, and I. Golding, Single-cell analysis of transcription kinetics across the cell cycle, eLife **5**, 1 (2016).
[7] G.-J. Hendriks, L. A. Jung, A. J. Larsson, M. Lidschreiber, O. A. Forsman, K. Lidschreiber, P. Cramer, and R. Sandberg, Nasc-seq monitors rna synthesis in single cells, Nature communications **10**, 1 (2019).
[8] Y. Wan, D. G. Anastasakis, J. Rodriguez, M. Palangat, P. Gudla, G. Zaki, M. Tandon, G. Pegoraro, C. C. Chow, M. Hafner, *et al.*, Dynamic imaging of nascent rna reveals general principles of transcription dynamics and stochastic splice site selection, Cell **184**, 2878 (2021).
[9] L. F. Lafuerza and R. Toral, Exact solution of a stochastic protein dynamics model with delayed degradation, Phys. Rev. E **84**, 1 (2011).
[10] S. Choubey, J. Kondev, and A. Sanchez, Deciphering transcriptional dynamics in vivo by counting nascent rna molecules, PLOS Computational Biology **11**, 1 (2015).
[11] S. Choubey, J. Kondev, and A. Sanchez, Distribution of initiation times reveals mechanisms of transcriptional regulation in single cells, Biophysical Journal **114**, 2072 (2018).
[12] Q. Jiang, X. Fu, S. Yan, R. Li, W. Du, Z. Cao, F. Qian, and R. Grima, Neural network aided approximation and parameter inference of non-markovian models of gene expression, Nature communications **12**, 1 (2021).
[13] S. Braichenko, J. Holehouse, and R. Grima, Distinguishing between models of mammalian gene expression: telegraph-like models versus mechanistic models, Journal of The Royal Society Interface **18**, 20210510 (2021).
[14] S. Sainsbury, C. Bernecky, and P. Cramer, Structural basis of transcription initiation by rna polymerase ii, Nature reviews Molecular cell biology **16**, 129 (2015).
[15] R. E. Kingston and M. R. Green, Modeling eukaryotic transcriptional activation, Current Biology **4**, 325 (1994).
[16] D. Cox, *Renewal theory* (Methuen, London, 1967).
[17] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, Real-time observation of transcription initiation and elongation on an endogenous yeast gene, Science **332**, 475 (2011).
[18] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation*, Encyclopedia of Mathematics and its Applications (Cambridge University Press, 1987).
[19] H. Kung and D. Tong, Fast algorithms for partial fraction decomposition, SIAM Journal on Computing **6**, 582 (1977).
[20] X. Fu, X. Zhou, D. Gu, Z. Cao, and R. Grima, Delayssatoolkit.jl: stochastic simulation of reaction systems with time delays in julia, bioRxiv 10.1101/2022.01.21.477236 (2022).
[21] See Supplemental Material for derivations, additional figures, model parameters, and computer programs for solving the eukaryotic transcription model.
[22] Y. Huang and W. F. McColl, Analytical inversion of general tridiagonal matrices, Journal of Physics A: Mathematical and General **30**, 7919 (1997).
[23] R. Perez-Carrasco, C. Beentjes, and R. Grima, Effects of cell cycle variability on lineage and population measurements of messenger rna abundance, Journal of the Royal Society Interface **17**, 20200360 (2020).
[24] We have used a convention according to which a sum in which the upper bound is lower than the lower bound is equal to zero.
[25] K. Adelman and J. T. Lis, Promoter-proximal pausing of rna polymerase ii: emerging roles in metazoans, Nature Reviews Genetics **13**, 720 (2012).
[26] I. Jonkers, H. Kwak, and J. T. Lis, Genome-wide dynamics of pol ii elongation and its interplay with promoter proximal pausing, chromatin, and exons, eLife **3**, e02407 (2014).
[27] B. Hoopes, J. LeBlanc, and D. Hawley, Kinetic analysis of yeast tfiid-tata box complex formation suggests a multi-step pathway., Journal of Biological Chemistry **267**, 11539 (1992).
[28] Z. Zhang, B. P. English, J. B. Grimm, S. A. Kazane, W. Hu, A. Tsai, C. Inouye, C. You, J. Piehler, P. G. Schultz, L. D. Lavis, A. Revyakin, and R. Tjian, Rapid dynamics of general transcription factor tfiib binding during preinitiation complex assembly revealed by single-molecule analysis, Genes & Development **30**, 2106 (2016).
[29] C. A. Weideman, R. C. Netter, L. R. Benjamin, J. J. McAllister, L. A. Schmiedekamp, R. A. Coleman, and B. Pugh, Dynamic interplay of tfiia, tbp and tata dna11edited by t. richmond, Journal of Molecular Biology **271**, 61 (1997).
[30] G. A. Rosen, I. Baek, L. J. Friedman, Y. J. Joo, S. Buratowski, and J. Gelles, Dynamics of rna polymerase ii and elongation factor spt4/5 recruitment during activator-dependent transcription, Proceedings of the National Academy of Sciences **117**, 32348 (2020).
[31] I. Baek, L. J. Friedman, J. Gelles, and S. Buratowski, Single-molecule studies reveal branched pathways for activator-dependent assembly of rna polymerase ii preinitiation complexes, Molecular Cell **81**, 3576 (2021).
[32] L. Friedman and J. Gelles, Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation, Cell **148**, 679 (2012).
[33] M. S. Buckley, H. Kwak, W. R. Zipfel, and J. T. Lis, Kinetics of promoter pol ii on hsp70 reveal stable pausing and key insights into its regulation, Genes & Development **28**, 14 (2014).
[34] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer, In vivo dynamics of rna polymerase ii transcription, Nature Structural &

Molecular Biology **14**, 796 (2007).

[35] L. Zawel, K. P. Kumar, and D. Reinberg, Recycling of the general transcription factors during rna polymerase ii transcription., Genes & Development **9**, 1479 (1995).

[36] O. Wolkenhauer, P. Wellstead, K.-H. Cho, and B. Ingalls, Sensitivity analysis: from model parameters to system behaviour, Essays in Biochemistry **45**, 177 (2008).

[37] J. Colgan and J. L. Manley, Tfiid can be rate limiting in vivo for tata-containing, but not tata-lacking, rna polymerase ii promoters., Genes & Development **6**, 304 (1992).

[38] C. R. Bartman, N. Hamagami, C. A. Keller, B. Giardine, R. C. Hardison, G. A. Blobel, and A. Raj, Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation, Molecular Cell **73**, 519 (2019).

[39] A.-B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, and A. S. Ribeiro, Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells , Nucleic Acids Research **40**, 8472 (2012).