Steady-state distributions of nascent RNA for general initiation mechanisms

Juraj Szavits-Nossan* and Ramon Grima†
School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JH, United Kingdom
(Dated: July 17, 2022)

Fluctuations in the number of nascent RNA accurately reflect transcriptional activity. However, mathematical models predicting their distributions are difficult to solve analytically due to their non-Markovian nature stemming from transcriptional elongation. Here we circumvent this problem by deriving an exact relationship between the steady-state distribution of nascent RNA and the distribution of initiation times, which can be computed for any general initiation mechanism described by a set of first-order reactions. We test our theory using simulations and live cell imaging data.

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 fitted using the exact solution of a simple Markov model of gene expression, called the telegraph model [2–5]. 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. However, in recent years, doubts have arisen about the validity of this kinetic description, principally because mature (cellular) mRNA does not provide a direct read-out of transcription [6, 7]—the fluctuations of mRNA numbers in the cell are strongly influenced by various post-transcriptional events such as splicing, nuclear 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 RNAPs can be estimated directly from nascent single-cell sequencing methods [8] or more commonly using single molecule fluorescence in situ hybridization (sm-FISH) where intronic probes specifically label (non-spliced) nascent RNA [7, 9]. Fitting of these numbers using stochastic models that account for RNAP dynamics can help us more accurately understand which regulatory steps in transcription are tuned to achieve required mRNA expression levels [10, 11].

Real-time observation of transcription in *vivo* has revealed that initiation is a stochastic process, whereas elongation and termination are fairly deterministic [12]. Transcription can be thus modelled by a stochastic, multistep process describing initiation, followed by a deterministic, single-step reaction describing elongation and termination. Despite this simplification, models of this type are particularly difficult to solve because they are inherently non-Markovian. So far, only two such models have been solved analytically: one in which the promoter is always active [13, 14], and the other in which the promoter switches between two states of activity and inactivity (Fig. 1(b)) [6]. A few other, more realistic models have been studied, but only the first two moments of the nascent RNA distribution have been obtained analytically [15–17]. None of these models, however, account for the complex multistep process of initiation that has been elucidated by decades of biochemical research [18].

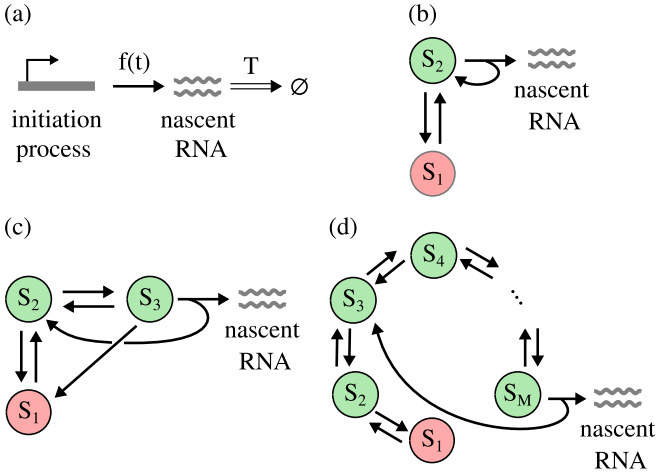

FIG. 1. (a) Nascent RNA is produced by an initiation process at time intervals that are independent and identically distributed (iid) random variables with an arbitrary probability density function $f(t)$. Elongation and termination are modelled deterministically, i.e. they take a fixed amount of time T to finish. (b)-(d) Examples of the initiation processes that can be studied by this framework, in the order of increasing complexity. S_1, S_2, etc are gene states. Red circles denote the off states. (b) The telegraph process [2]. (c) A three-state process that accounts for the binding of RNAP [19]. (d) A stepwise process of eukaryotic transcription that accounts for the binding of general transcription factors and RNAP [18, 20], and the promoter proximal pausing of RNAP in metazoans [21, 22].

In this Letter, we develop a general framework that allows us to find nascent RNA distribution for models with stochastic, multistep initiation, under the assumption of deterministic elongation and termination. The significance of this result is that it applies to any initiation process that produces nascent RNA at time intervals that

are independent and identically distributed (iid) random variables with an arbitrary probability density function $f(t)$ (Fig. 1(a)). This is true for any stochastic process described by first-order reactions with time-independent kinetic rates, some of which are shown in Fig. 1. We argue that while transcription initiation includes invariably many bimolecular steps, these are well approximated by pseudo first-order reactions because transcriptional machinery is generally abundant; furthermore, any variations of the rates are over timescales that are much longer than the elongation time and hence can be ignored [23]. For such processes, we solve a first passage time problem to get an analytic expression for $f(t)$. The steady-state nascent RNA distribution is then computed using renewal theory, which generalizes the Poisson process by allowing for a non-exponential $f(t)$. We verify our theory by stochastic simulations and using experimental results for the transcription kinetics in *Escherichia coli* [24].

Waiting time distribution between successive nascent RNA production events. Transcription is divided into initiation, elongation and termination. Initiation is a complex multistep process which starts by the binding of an RNAP at the promoter, together with several transcription factors, and ends when the RNAP escapes the promoter and starts productive elongation. During elongation, RNAP traverses the gene and copies its sequence into a nascent RNA. The termination occurs at the gene end, leading to a mature RNA transcript.

We consider a stochastic initiation process consisting of M gene states labelled by $S_1, \ldots, S_M$. The gene switches between the states according to first-order reactions

$$S_i \underset{k_{ji}}{\overset{k_{ij}}{\rightleftharpoons}} S_j,\ i,j = 1,\ldots,M,\ i \neq j,\ S_M \xrightarrow{k_M} S_K + N. \quad (1)$$

The last reaction describes production of nascent RNA (N), after which the initiation process starts again from the state S_K (K takes a fixed value between 1 and M). Eq. 1 is a general description of the initiation process using first-order reactions. Particular cases are obtained by removing some reactions, which is equivalent to setting the rates of those reactions to zero.

We are interested in computing the probability density function (pdf) of the waiting time between successive events of nascent RNA production. This is a first passage time problem that can be solved by replacing the last reaction in Eq. 1 with $S_M \to A$, where A is an absorbing state, i.e. once the process reaches A, it stops. The pdf $f(t)$ is then equal to $k_M P_M(t)$, where $P_M(t)$ is the probability that the gene is in state M at time t, given that it was in state K at time 0, and $k_M dt$ is the probability that the transition $S_M \to A$ occurs in the time interval $[t, t+dt)$. The probability $P_M(t)$ can be found by solving the following master equation,

$$\frac{dP_i}{dt} = \sum_{j=1}^{M} k_{ji} P_j - \left(\sum_{j=1, j\neq i}^{M} k_{ij} + \delta_{iM} k_M \right) P_i, \quad (2)$$

with the initial condition $P_i(0) = \delta_{iK}$. The pdf $f(t)$ is given by

$$f(t) = k_M \left(\exp(\mathbf{L}t)\right)_{MK}, \quad (3)$$

where the matrix elements of $\mathbf{L}$ read

$$L_{ij} = k_{ij},\ i \neq j,\ L_{ii} = -\sum_{j=1, j\neq i}^{M} k_{ij} - \delta_{iM} k_M. \quad (4)$$

A central function in renewal theory is the Laplace transform of the pdf $f(t)$, denoted by $f^*(s)$. Formally,

$$f^*(s) = k_M P_M^*(s) = k_M (s\mathbf{I} - \mathbf{L})^{-1}_{MK}, \quad (5)$$

where $P_M^*(s)$ is the Laplace transform of $P_M(t)$ and $\mathbf{I}$ is the $M \times M$ identity matrix. Figs. 2(a)-(c) show $f(t)$ for the initiation processes depicted in Fig. 1, computed from Eq. (5) and compared to the results of stochastic simulations. For the stepwise model of initiation in Fig. 1(d), $f^*(s)$ can be computed analytically for any number of states M, because the matrix $\mathbf{L}$ is tridiagonal, for which the inverse is known explicitly [25]. The derivation of $f^*(s)$ for all three models in Fig. 1 is presented in Supplemental Material [26].

Distribution of nascent RNA in the steady state. Our approach for solving the general problem in Fig. 1(a) is based on recognizing that the number of nascent RNA production events that have occurred up to time t, denoted by $N_I(t)$, constitutes a renewal process [27]. A renewal process generalizes the Poisson process by allowing for non-exponentially distributed waiting times between successive events. As in the Poisson process, the waiting times in a renewal process are mutually independent. In transcription, this assumption is supported by experimental evidence [12].

We denote by T_i the time of the i-th nascent RNA production event, measured from some reference time point $t = 0$. For simplicity, we assume that a nascent RNA production event occurred immediately before $t = 0$. If t_i denotes the waiting time between i-th and $(i-1)$-th nascent RNA production events, $t_i = T_i - T_{i-1}$, where $T_0 = 0$, then $T_i = \sum_{j=1}^{i} t_j$. Since t_j are iid random variables, the pdf of T_i is the i-fold convolution of f, $k_i(t) = f^{*i}(t)$.

The probability of $N_I(t) = n$ is equivalent to the probability of $T_n \leq t < T_{n+1}$, which is given by

$$\begin{aligned} P(N_I(t) = n) &= P(T_n \leq t < T_{n+1}), \\ &= K_n(t) - K_{n+1}(t), \end{aligned} \quad (6)$$

where $K_n(t) = \int_0^t dt'\ k_n(t')$ for $n \geq 1$ is the cumulative distribution function of T_n, and $K_0(t) = 1$. Since the total time of elongation and termination is fixed, the number of nascent RNA actively engaged in transcription at time t is equal to

$$N(t) = \begin{cases} N_I(t), & t \leq T, \\ N_I(t) - N_I(t-T), & t > T. \end{cases} \quad (7)$$

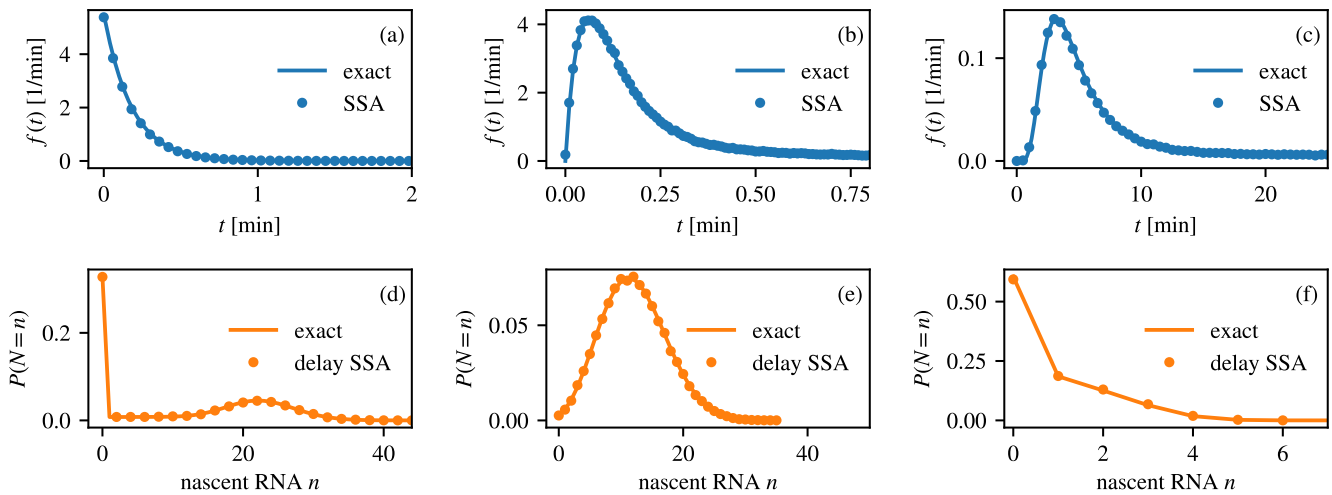

FIG. 2. (a) to (c) Probability density function $f(t)$ of the waiting time between successive nascent RNA production events for the initiation processes depicted in Fig. 1(b) to 1(d), respectively. Solid line is the theoretical prediction from Eq. (3), and points are from stochastic simulations (SSA). (b) The respective probability distribution of the nascent RNA, obtained from Eq. (9), and compared to stochastic simulations (delay SSA) performed using `DelaySSAToolkit.jl` package in Julia [26, 28]. Model parameters are listed in Supplemental Material [26].

For $t \leq T$, the probability $P(N(t) = n)$ is given by Eq. (6). 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 $t_0 = t - T$). In the steady state, $t_0 \to \infty$ and the expression for $f_{t_0}(\tau)$ simplifies to

$$f_\infty(\tau) = \lim_{t_0 \to \infty} f_{t_0}(\tau) = \frac{1 - F(\tau)}{\mu}, \tag{8}$$

where $F(t) = \int_0^t dt' \; f(t')$ is the cumulative distribution function of the waiting times between successive nascent RNA production events, and μ is the mean waiting time [27]. Let $P(N = n)$ denote the probability to find n nascent RNA on the gene in the steady state. Because nascent RNA production events are mutually independent, the Laplace transform $P^*(n, s)$ of $P(N = n)$ with respect to the elongation and termination time T is given by

$$\begin{aligned} P^*(n, s) &= \int_0^\infty dT \; \mathrm{e}^{-Ts} P(N = n) \\ &= \begin{cases} \frac{\mu s - 1 + f^*(s)}{\mu s^2}, & n = 0, \\ \frac{[1 - f^*(s)]^2 [f^*(s)]^{n-1}}{\mu s^2}, & n \geq 1, \end{cases} \end{aligned} \tag{9}$$

where we have used the fact that the Laplace transform of $f_\infty(\tau)$ is given by $f^*_\infty(s) = (1 - f^*(s))/(\mu s)$.

Eq. (9) is our main result. It connects the Laplace transform of $P(N = n)$ to the Laplace transform of $f(t)$ for *all* initiation processes described by Eq. (2). For these processes, $f^*(s)$ is a rational function of s, meaning that $P(N = n)$ can be computed from Eq. (9) by partial fraction decomposition, for which many methods are available [29]. 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 T is given by

$$G^*(z, s) = \frac{1}{s} + \frac{(z - 1)[1 - f^*(s)]}{\mu s^2 [1 - z f^*(s)]}. \tag{10}$$

For example, the mean number of nascent RNA is equal to T/μ, and the variance is given by

$$\mathrm{Var}[N] = \mathcal{L}^{-1} \left\{ \frac{1 + f^*(s)}{\mu s^2 [1 - f^*(s)]} \right\} (T) - \left(\frac{T}{\mu} \right)^2, \tag{11}$$

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

Eq. (9) allows us to compute nascent RNA distribution analytically without using the complicated delay chemical master equation. In Supplemental Material, we demonstrate this for the Poisson process, the telegraph process and the fully irreversible process with equal forward rates λ [26]. In the last case, the waiting time distribution is an Erlang distribution, which is an important distribution that is found ubiquitously in cellular biology [30]. The nascent RNA distribution for this case reads

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

where $N_+ = M(n+1)$, $N_0 = Mn$, $N_- = M(n-1)$ [31].

Figs. 2(d)-(f) show the nascent RNA distribution for the initiation processes depicted in Fig. 1, computed from (9) and compared with the results of stochastic simulations. In particular, results in Figs. 2 (c) and (f) were obtained for a ten-state initiation process ($M = 10$) as a canonical model of eukaryotic transcription initiation [18, 20] using kinetic parameters collected from the literature [1, 32–39]. This model includes the on and off switching of the promoter, the binding and unbinding of six general transcription factors (IID, IIA, IIB, IIF, IIE and IIH) and RNAP, the unwinding of the double-stranded DNA, and the promotor proximal pausing of RNAP in metazoans [21, 22]. We assumed that the reinitiation starts from the on state ($K = 2$), whereas in Supplemental Material we consider the case in which transcription factors IID and IIA remain bound for a faster reinitiation ($K = 4$) [40, 41]. Sensitivity analysis for $K = 2$ revealed that the variability in nascent RNA is most sensitive to changes in the on and off rates, and the binding rate of TFIID, which indeed have been identified previously as rate-limiting steps in transcription initiation [19, 42].

Comparison with experimental data. In Ref. [24], transcription kinetics of a target gene were followed in live *Escherichia coli* cells, one transcription event at a time. The target gene was controlled by the *tetA* promoter, which was induced (turned on) by anhydrotetracycline (aTc) at a concentration of 15 ng/ml at two temperatures, 24°C and 37°C. The waiting time distribution between successive mature RNA production events was measured experimentally and fitted to a hypoexponential distribution $f_{\rm fit}(t)$ with rates r_1, r_2 and r_3 (Fig. 3(a)). The Laplace transform of this distribution is given by

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

Fig. 3(b) shows the distribution of the fraction of cells with a given number of mRNA, $M(t)$, measured at $t = t_1 = 1$ h after the induction by aTc at 24°C. This fraction is in a very good agreement with our theoretical prediction, which was computed as follows. The time of the full induction by aTC was estimated to 20 min [24]. We assumed that no transcription occurred before t_0, and that the pdf of the waiting time until the first transcription event was $f_{\rm fit}$. Because mature RNA did not degrade during the experiment, we assumed that the number of mature RNA measured at t_1, $M(t_1)$, equalled the number of individual transcription events between t_0 and $t_1 - T$, where T was the elongation and termination time. The latter was only tens of seconds, hence we approximated $t_1 - T$ by t_1. The theoretical distribution $P(M(t_1) = n)$ was calculated from Eq. (6) by inverting the Laplace transform $[1 - f^*_{\rm fit}(s)][f^*_{\rm fit}]^n/s$, evaluated at $t_1 - t_0 = 40$ min. We stress that no fitting parameters were used other than those of $f_{\rm fit}(t)$. The results at 37°C are also in good agreement with the experimental data (see Fig. S2(b) in Supplemental Material [26]). A very good agreement between the experiment and theory suggests that the knowledge of the waiting time distribution was sufficient to correctly predict the distribution of the accumulated mRNA.

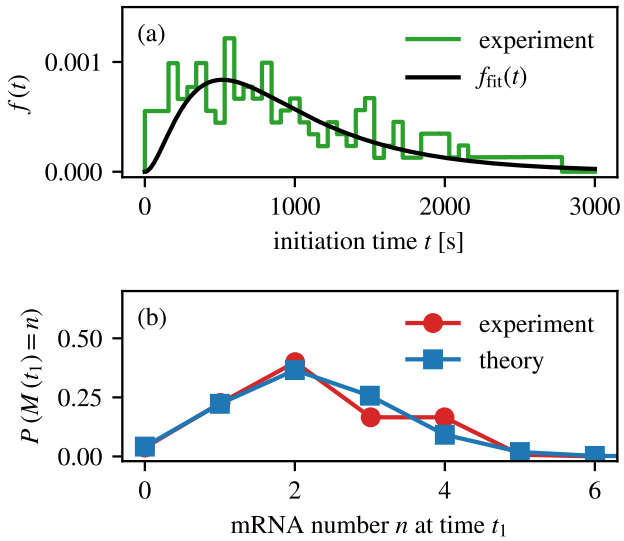

FIG. 3. (a) Waiting time distribution between successive mature RNA production events measured in Ref. [24] at 24°C. Black line is a fit to the hypoexponential distribution with three parameters, $r_1 = 1/620$ s^{-1}, $r_2 = 1/240$ s^{-1} and $r_3 = 1/115$ s^{-1}, obtained by inverting $f^*_{\rm fit}(s)$ 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 obtained from Eq. (6) using $f^*_{\rm fit}(s)$ defined in Eq. (13).

Conclusion. We have presented a general framework that allows us to predict the distribution of nascent RNA from the distribution of waiting times between successive nascent RNA production events, assuming the time of elongation and termination to be fixed. The significance of our solution is that it applies to any initiation mechanism that is modelled by first-order reactions with time-independent rates. The theory also allows for a model-free prediction of the nascent RNA distribution, provided the waiting time distribution is measured experimentally. Important directions for future work include assessing the importance of stochastic events in transcription elongation, such as RNAP collisions, ubiquitous pausing and premature termination, which have not been accounted for in this framework.

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] L. Ham, R. D. Brackston, and M. P. H. Stumpf, Extrinsic noise and heavy-tailed laws in gene expression, Phys. Rev. Lett. **124**, 108101 (2020).
[6] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Stochastic kinetics of nascent rna, Phys. Rev. Lett. **117**, 128101 (2016).
[7] 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).
[8] 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).
[9] 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).
[10] H. Xu, L. A. Sepúlveda, L. Figard, A. M. Sokac, and I. Golding, Combining protein and mrna quantification to decipher transcriptional regulation, Nature Methods **12**, 739 (2015).
[11] L. A. Sepúlveda, H. Xu, J. Zhang, M. Wang, and I. Golding, Measurement of gene regulation in individual cells reveals rapid switching between promoter states, Science **351**, 1218 (2016).
[12] 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).
[13] L. F. Lafuerza and R. Toral, Exact solution of a stochastic protein dynamics model with delayed degradation, Phys. Rev. E **84**, 051121 (2011).
[14] 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).
[15] S. Choubey, J. Kondev, and A. Sanchez, Deciphering transcriptional dynamics in vivo by counting nascent rna molecules, PLOS Computational Biology **11**, 1 (2015).
[16] S. Choubey, J. Kondev, and A. Sanchez, Distribution of initiation times reveals mechanisms of transcriptional regulation in single cells, Biophysical Journal **114**, 2072 (2018).
[17] 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).
[18] S. Sainsbury, C. Bernecky, and P. Cramer, Structural basis of transcription initiation by rna polymerase ii, Nature reviews Molecular cell biology **16**, 129 (2015).
[19] 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).
[20] R. E. Kingston and M. R. Green, Modeling eukaryotic transcriptional activation, Current Biology **4**, 325 (1994).
[21] K. Adelman and J. T. Lis, Promoter-proximal pausing of rna polymerase ii: emerging roles in metazoans, Nature Reviews Genetics **13**, 720 (2012).
[22] 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).
[23] Many eukaryotic genes have initiation rates that change over time due to their dependence on the cell volume [43]. However, these changes occur on a timescale (the cell cycle duration) that is much longer than the elongation time.
[24] 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).
[25] Y. Huang and W. F. McColl, Analytical inversion of general tridiagonal matrices, Journal of Physics A: Mathematical and General **30**, 7919 (1997).
[26] See Supplemental Material at [URL will be inserted by publisher] for additional derivations, figures, tables with model parameters, and computer programs for solving the transcription models from Fig. 1.
[27] D. Cox, *Renewal theory* (Methuen, London, 1967).
[28] 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).
[29] H. Kung and D. Tong, Fast algorithms for partial fraction decomposition, SIAM Journal on Computing **6**, 582 (1977).
[30] 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).
[31] 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.
[32] 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).
[33] 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).

[34] 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).

[35] 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).

[36] I. Baek, L. J. Friedman, J. Gelles, and S. Buratowski, Single-molecule studies reveal branched pathways for activator-dependent assembly of rna polymerase ii pre-initiation complexes, Molecular Cell **81**, 3576 (2021).

[37] L. Friedman and J. Gelles, Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation, Cell **148**, 679 (2012).

[38] 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).

[39] 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).

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

[41] D. Yean and J. D. Gralla, Transcription reinitiation rate: A potential role for TATA box stabilization of the TFIID:TFIIA:DNA complex, Nucleic Acids Research **27**, 831 (1999).

[42] 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).

[43] X.-M. Sun, A. Bowman, M. Priestman, F. Bertaux, A. Martinez-Segura, W. Tang, C. Whilding, D. Dormann, V. Shahrezaei, and S. Marguerat, Size-dependent increase in rna polymerase ii initiation rates mediates gene expression scaling with cell size, Current Biology **30**, 1217 (2020).