Multi-scale bursting in stochastic gene expression

Zhixing Cao,[1,2,*] Tatiana Filatova,[2,3] Diego A. Oyarzún,[2,4] and Ramon Grima[2,†]

[1] *The Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, PR China*
[2] *School of Biological Sciences, the University of Edinburgh, United Kingdom*
[3] *School of Mathematics, the University of Edinburgh, United Kingdom*
[4] *School of Informatics, the University of Edinburgh, United Kingdom*

(Dated: July 27, 2019)

Transcriptional bursting is a major source of noise in gene expression. Motivated by recent experiments, we study a model including slow burst initiation and termination, and fast RNA polymerase recruitment and pause release. We show that the time-dependent distribution of mRNA numbers is accurately approximated by a telegraph model with a Michaelis-Menten like dependence of the effective transcription rate on polymerase abundance. We also show that gene dosage compensation, a common feature of mammalian gene expression, is an emergent property of our stochastic model.

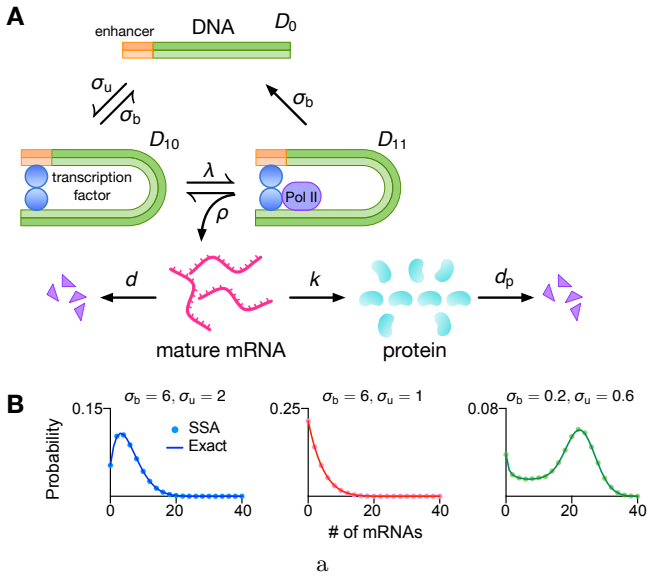
PACS numbers: 87.10.Mn, 02.50.-r, 82.39.-k, 87.17.Aa

There is widespread evidence that mammalian genes are expressed in bursts: infrequent periods of transcriptional activity that produce a large number of messenger RNA (mRNA) transcripts within a short period of time [1–3]. This is in contrast to constitutive expression where mRNAs are produced in random, uncorrelated events, with a time-independent probability [4]. The size and frequency of transcriptional bursts affect the magnitude of temporal fluctuations in mRNA and protein content of a cell, and thus constitute an important source of intracellular noise [5].

A large number of studies have sought to elucidate the mechanisms leading to bursting and by constructing simple stochastic models that can explain the data. The simplest of these models is the telegraph model whereby (i) a gene is in two states, an ON state where mRNA is expressed and an OFF state where there is no expression. (ii) mRNA degrades in the cytoplasm. These first-order reactions are effective since each encapsulates the effect of a large number of underlying biochemical reactions. The chemical master equation of this model has been solved exactly to obtain the probability distribution of mRNA numbers as a function of time [6]. For parameter conditions consistent with bursty expression, the steady-state distribution is well approximated by a negative binomial that fits some of the experimental data [7].

Recent studies have suggested modifications to the telegraph model to include details of polymerase dynamics including its recruitment and release from the paused state [8, 9]. These changes are necessary to explain the presence of multiple timescales in the bursty expression of mammalian promoters. In this letter, we present the first detailed study of this model. We derive an exact steady-state solution and, by mapping it onto an effective telegraph model, we obtain an approximate time-dependent solution. The theory allows us to tease apart the transcriptional and post-transcriptional mechanisms that have the largest impact on mRNA fluctuations, and provide an intuitive explanation for gene dosage compensation in mammalian cells.

Model. We consider a stochastic multi-scale transcriptional bursting model (recently introduced in [9] and henceforth referred to as the multi-scale model; see Fig. 1A), whereby a gene fluctuates between three states: two permissive states (D_{10} and D_{11}) and a non-permissive state (D_0). The transition from D_0 to D_{10} (burst initiation) is mediated by transcription factor binding with rate constant σ_u which is reversible with rate constant σ_b (this transition may alternatively represent other processes such as nucleosome remodeling). Subsequently the



a

FIG. 1. (A) Schematic of the stochastic multi-scale transcriptional bursting model. (B) Distribution for mRNA numbers obtained from the generating function in Eq. (2) agrees with SSA predictions. The kinetic parameters are $\rho = 60$, $\lambda = 40$, $d = 1$, and the rest of the parameters are indicated in each panel.

binding of RNA polymerase II (Pol II) to D_{10} with rate constant λ (which is proportional to Pol II abundance) leads to D_{11}. This represents a state in which Pol II is paused. The polymerase is released from this state with rate constant ρ leading to the production of an mRNA molecule (denoted as M) and the unbinding of polymerase which returns the gene to state D_{10}. In the paused state D_{11}, both the polymerase and the transcription factor can unbind from the gene and lead to the non-permissive state D_0 (burst termination). Both reversible switches operate at different timescales (hours versus minutes) with $\max\{\sigma_\mathrm{b}, \sigma_\mathrm{u}\} \ll \min\{\rho, \lambda\}$, leading to multi-scale transcriptional bursting [8, 9]. The mRNA transcripts are translated into protein (denoted P) with rate constant k. Both the mRNA and proteins decay with rate constants d and d_p, respectively. All reactions are first-order, characterized by exponentially distributed waiting times between successive reactions.

The reaction $D_{11} \to D_{10} + M$ is an effective description of the reactions: $D_{11} \to D_{10} + N$, $N \to M$ where N is nascent mRNA, i.e pre-spliced mRNA. The reaction $N \to M$ is often modeled with a deterministic time delay [10, 11] but theory shows that it can be modeled with a stochastic exponential time delay provided the timescale of nascent mRNA production (ρ^{-1}) is much smaller than the timescale governing the transitions between permissive and non-permissive states ($(\sigma_\mathrm{u} + \sigma_\mathrm{b})^{-1}$) [10]. Also, no explicit nascent mRNA description is needed provided that it is short lived compared to mature mRNA. Since these two conditions are physiologically realistic in many cases, we choose to ignore detailed modeling of nascent mRNA dynamics and model the direct production of mature mRNA with an exponentially distributed time delay.

Exact solution. Let $P_\theta(n,t)$ ($\theta = 0, 10, 11$) denote the probability of a cell being in state D_θ with n mRNAs at time t (arguments n and t are hereafter omitted for brevity). The dynamics of probability P_θ are described by the set of coupled master equations

$$\begin{aligned}
\partial_t P_0 =& (\mathbb{E}^1 - 1) d n P_0 - \sigma_\mathrm{u} P_0 + \sigma_\mathrm{b}(P_{10} + P_{11}),\\
\partial_t P_{10} =& (\mathbb{E}^1 - 1) d n P_{10} - (\sigma_\mathrm{b} + \lambda) P_{10} + \sigma_\mathrm{u} P_0 + \rho \mathbb{E}^{-1} P_{11}, \quad (1)\\
\partial_t P_{11} =& (\mathbb{E}^1 - 1) d n P_{11} - (\rho + \sigma_\mathrm{b}) P_{11} + \lambda P_{10},
\end{aligned}$$

where the step operator $\mathbb{E}^i$ acts on a general function $g(n)$ as $\mathbb{E}^i g(n) = g(n+i)$ [12]. Defining $G(u) = \sum_\theta \sum_n (u+1)^n P_\theta(n)$ and using the generating function method, we solve Eq. (1), yielding the exact steady-state solution for $G(u)$ in terms of the generalized hypergeometric function

$$G(u) = {}_1F_2\left(\frac{\sigma_\mathrm{u}}{d}; \frac{\gamma_1}{d}, \frac{\gamma_2}{d}; \frac{\rho\lambda}{d^2} u\right), \qquad (2)$$

with $\gamma_1 = \sigma_\mathrm{b} + \sigma_\mathrm{u}$ and $\gamma_2 = \sigma_\mathrm{b} + \rho + \lambda$. The exact distribution of mRNA numbers at steady state is then given by $P(n) = \sum_\theta P_\theta(n) = \frac{1}{n!} \frac{\mathrm{d}^n}{\mathrm{d}u^n} G(u)|_{u=-1}$ (See [13] for a closed-form expression). In Fig. 1B we show that

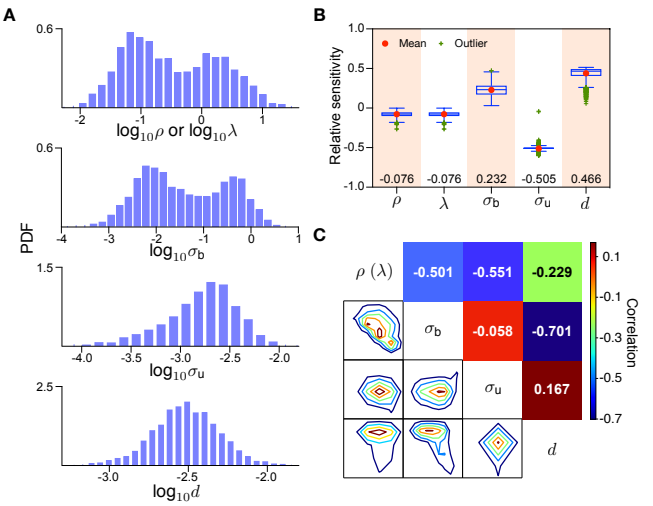


FIG. 2. Relative sensitivity analysis of the coefficient variation $\overline{\mathrm{CV}}$ of mRNA noise over 5 kinetic parameters for 3575 genes of CAST allele data for mouse fibroblast cells. (A) Shows distributions of the 5 kinetic parameters in the dataset (obtained from [3]), where values of ρ or λ are calculated using Eq. (4). (B) Box plots indicate the median (whose values are also shown at bottom), the 25%, 75% quantiles, and mean and outliers of relative sensitivity Λ_p for each parameter. (C) Joint distribution (lower panel) and Pearson correlation (upper panel) of Λ_p for each pair of kinetic parameters suggest that the pairs $(\sigma_\mathrm{b}, \sigma_\mathrm{u})$ and (σ_u, d) are the least dependent pairings.

distributions obtained from Eq. (2) as well as the corresponding modality (a phenotypic signature [14]) are indistinguishable from distributions produced using the stochastic simulation algorithm (SSA) [15]. It can be further shown by perturbation theory [13] that the exact solution Eq. (2) reduces to the generating function of the negative binomial distribution $P(n) = \mathrm{NB}(\frac{\sigma_\mathrm{u}}{d}, \frac{\rho}{\rho+\alpha})$ with $\alpha = \sigma_\mathrm{b}\gamma_2/\lambda$, when ρ, λ and σ_b are much greater than the rest of the parameters.

Sensitivity analysis. The exact solution in Eq. (2) allows us to examine the stochastic properties of the multi-scale model over large swathes of parameter space. We investigate the relative sensitivity of the coefficient of variation of mRNA fluctautions, $\mathrm{CV} = \sqrt{\mathrm{Var}(n)}/\langle n \rangle$, which is typically employed as a measure of the magnitude of transcriptional noise. To this end, we calculate the first two central moments, ($\langle n \rangle$ and $\mathrm{Var}(n)$), from Eq. (2) using $\langle n \rangle = \partial_u G|_{u=0}$ and $\mathrm{Var}(n) = \partial_u^2 G|_{u=0} + \langle n \rangle - \langle n \rangle^2$. The mean and CV are then given by

$$\langle n \rangle = \frac{\sigma_\mathrm{u}\lambda\rho}{d\gamma_1\gamma_2}, \qquad (3a)$$

$$\mathrm{CV}^2 = \frac{1}{\langle n \rangle} + \frac{d}{\sigma_\mathrm{u}} \cdot \frac{\gamma_1}{\gamma_1 + d} \cdot \frac{\gamma_2}{\gamma_2 + d}. \qquad (3b)$$

Note that since the parameters ρ and λ appear sym-

metrically in Eq. (3) and also due to the unavailability of experimental data, for simplicity we enforce the constraint $\rho = \lambda$. Hence, the relative sensitivity of the quantity $\overline{\mathrm{CV}} = \mathrm{CV}|_{\rho=\lambda}$, which can serve as a gauge of transcriptional noise, is insightful to study and defined as $\Lambda_p = (p/\overline{\mathrm{CV}})\partial\overline{\mathrm{CV}}/\partial p$ for a model parameter p, meaning that 1% change in p leads to a Λ_p% change in $\overline{\mathrm{CV}}$. The parameter values for the sensitivity analysis were sampled from experimental distributions recently inferred for 3575 genes of CAST allele in mouse fibroblasts [3], using the telegraph model. To obtain values for ρ and λ, we equate the mean of the telegraph model (with ON switching rate σ_b, OFF switching rate σ_u, transcription rate ρ_u and degradation rate d) $\langle n\rangle_{\mathrm{tel}} = \sigma_{\mathrm{u}}\rho_{\mathrm{u}}/\gamma_1 d$ with the mean of the multi-scale model (Eq. (3a)) under the constraint $\rho = \lambda$, giving

$$\rho = \rho_{\mathrm{u}}\left(1 + \sqrt{1 + \frac{\sigma_{\mathrm{b}}}{\rho_{\mathrm{u}}}}\right). \tag{4}$$

Distributions for each parameter in the dataset are presented in Fig. 2A and the box plots in Fig. 2B show the relative sensitivity for each parameter. The parameters in order of most sensitive first are σ_u, d, σ_b and $\rho = \lambda$. This order is the same as obtained by ranking parameters according to the inverse of their mean experimental values (the mean of the distributions in Fig. 2A) implying that changes to the CV are most easily accomplished by perturbations to the slowest reactions. Given the vectors Λ_{p_1} and Λ_{p_2} for any pair $p_1 \neq p_2$ and p_1, p_2 in the set $\{\rho, \lambda, \sigma_{\mathrm{b}}, \sigma_{\mathrm{u}}, d\}$ where each entry is a different gene, in Fig. 2C we calculate the Pearson correlation coefficient between the vectors and the corresponding joint distributions. This shows that (σ_{u}, d) is the least dependent pairing and hence they constitute a quasi-orthogonal decomposition of the sensitivity.

Effective telegraph model. The generalized hypergeometric function in Eq. (2), though exact, does not give much biological intuition. Its connection to the conventional telegraph model, which is commonly used for inference of single cell RNA sequencing data [3], is somewhat elusive because it has only a single timescale whereas our model has two. Next we use the first passage time method to reduce our model into an effective telegraph model. To this end, we consider the transcription motif of the multi-scale model, $D_{10} \xrightarrow{\lambda} D_{11} \xrightarrow{\rho} D_{10} + M$, whose corresponding master equations for producing newborn mRNA starting from state D_{10} are

$$\begin{aligned} \partial_t P_{10} &= -\lambda P_{10}, \\ \partial_t P_{11} &= \lambda P_{10} - \rho P_{11}, \\ \partial_t P_{\mathrm{M}} &= \rho P_{11}, \end{aligned} \tag{5}$$

where P_{10}, P_{11} and P_{M} represent the probability of staying in states D_{10}, D_{11} or producing a new mRNA respectively. We remark that the reaction $D_{11} \to D_0$ is absent from the motif due to its relatively small reaction rate σ_{b} compared to ρ and λ. The initial conditions for Eq. (5) are $P_{10}|_{t=0} = 1$, $P_{11}|_{t=0} = P_{\mathrm{M}}|_{t=0} = 0$. Solving for P_{M} in Eq. (5), we can calculate the mean first passage time for mRNA production

$$\langle t_{\mathrm{f}}\rangle = \int_0^\infty t P_f \mathrm{d}t = \frac{\rho + \lambda}{\lambda\rho}, \tag{6}$$

where $P_f = \partial_t P_{\mathrm{M}}$ is the first-passage time distribution [16]. Since the effective transcription rate is the inverse of the mean first passage time, it immediately follows that the effective telegraph model is

$$D_0 \underset{\sigma_{\mathrm{b}}}{\overset{\sigma_{\mathrm{u}}}{\rightleftharpoons}} D_{10} \xrightarrow{\rho_{\mathrm{u}} = \frac{\lambda\rho}{\lambda+\rho}} D_{10} + M, \quad M \xrightarrow{d} \varnothing. \tag{7}$$

Alternatively, one can obtain this result by equating the means of our model Eq. (3a) and of the telegraph model $\langle n\rangle_{\mathrm{tel}} = \rho_{\mathrm{u}}\sigma_{\mathrm{u}}/\gamma_1 d$ and solving for the effective production rate ρ_{u}, giving $\rho_{\mathrm{u}} = \lambda\rho/\gamma_2 \simeq \frac{\lambda\rho}{\lambda+\rho}$ since typically $\rho, \lambda \gg \sigma_{\mathrm{b}}$.

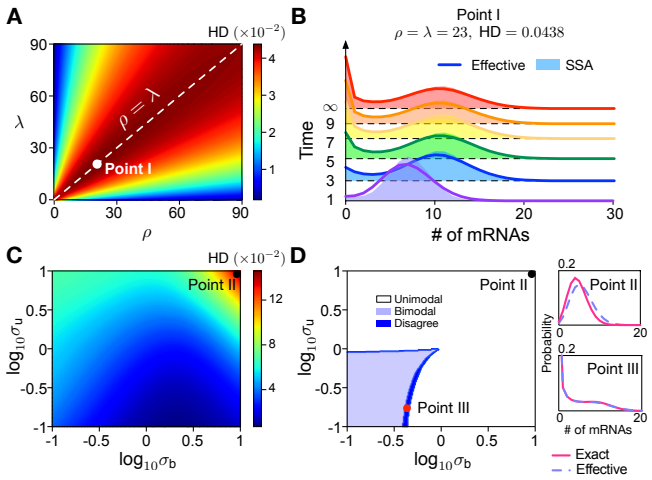


FIG. 3. Effective telegraph model approximation for the distribution of mRNA numbers of the multi-scale model. (A) Heatmap of the Hellinger distance between steady-state distributions of mRNA numbers for the effective telegraph model and the multi-scale model as a function of ρ and λ with parameters $\sigma_{\mathrm{u}} = 0.2$, $\sigma_{\mathrm{b}} = 0.1$ and $d = 1$. The dashed line corresponds to the contour line $\rho = \lambda$. The discrepancy between the two distributions becomes larger as ρ and λ approaches the line $\rho = \lambda$. (B) Shows the time-dependent distributions for Point I in (A) (the point with the largest HD) predicted by the effective model compared to those computed by the SSA for the multi-scale model. (C) Heatmap of HD between the two steady-state distributions as a function of σ_{b} and σ_{u} with $\rho = \lambda = 23$ and $d = 1$. (D) Shows a stochastic bifurcation diagram for the number of modes of the steady-state distributions predicted by the two models. The miniscule dark blue region is where modality predictions of both models disagree thus corroborating the high accuracy of the effective telegraph model approximation. Insets show distributions corresponding to the points marked in (C,D).

In Fig. 3, we show the high accuracy of the effective telegraph model approximation from Eq. (7). In particular, Fig. 3A shows a heatmap of the distance between the distributions of mRNA numbers predicted by the effective telegraph model and the multi-scale model. As a distance measure, we use the Hellinger distance (HD), a Euclidean distance based metric normalized to the interval between 0 and 1. The effective telegraph model is naturally a more accurate description to the multi-scale model when there is one rate limiting step (large difference between ρ and λ) rather than when there are two rate limiting steps ($\rho = \lambda$). Indeed it can be shown using perturbation theory that in the limit of large ρ or large λ, the time-dependent solution of the multi-scale model converges to that of the telegraph model [13].

Since the time-dependent distribution of the telegraph model is known in closed-form [6, 17], it follows that by the effective model in Eq. (7) we have an approximation for the time-dependent distribution of the multi-scale model too. The accuracy of this approximation is shown in Fig. 3B where it is compared to the time-dependent distributions computed using the SSA for the multi-scale model. The parameters here correspond to those of Point I in Fig. 3A (the largest HD). Differences between the distributions of the two models are negligible except near time $t = 0$. We further investigate how burst initiation and termination rates (σ_u, σ_b) affect the approximation error with a heatmap of HD as a function of σ_u and σ_b (Fig. 3C), and a stochastic bifurcation diagram for the number of modes of the effective telegraph and multi-scale model distributions (Fig. 3D) at steady state. The point of maximum HD in Fig. 3C (Point II) displays distributions that are not that different from each other – see upper right inset of Fig. 3D. The two models display the same number of modes in all regions of parameter space except for a narrow region where modality detection is challenging because the distributions have a broad plateau – see lower right inset of Fig. 3D (Point III). This again confirms the high accuracy of the effective telegraph model approximation.

Connection to refractory model. Besides the telegraph model, another prevalent stochastic transcriptional model is the refractory model [2] (a three-state model, see Fig. 4A left), wherein the burst initiation requires two steps instead of one. To understand the connection between our model and the refractory model, we exactly solve the refractory model for the steady-state distribution of mRNA numbers [13] and similarly map the refractory model onto an effective telegraph model by matching the mean mRNA numbers

$$\langle n\rangle_{\rm ref} = \frac{\lambda\sigma_u\rho_u}{d(\lambda\sigma_u + \lambda\sigma_b + \sigma_u\sigma_b)}, \quad \langle n\rangle_{\rm tel} = \frac{\rho_u\bar{\sigma}_u}{d(\sigma_b + \bar{\sigma}_u)}.$$

leading to an effective burst initiation rate $\bar{\sigma}_u = \frac{\sigma_u\lambda}{\sigma_u+\lambda}$ and the corresponding effective model shown in Fig. 4A right.

We then compare the steady-state distributions of the refractory model and its effective telegraph model. A heatmap of HD quantifying their distributional difference and a modality diagram (marked as black lines) of the two distributions are illustrated in Fig. 4B. Both the regions of high HD and Region 2 where only the telegraph model predicts bimodality are significantly large; also Region 1 where both predict bimodality is small. This shows that the refractory model, in general, is not well approximated by the telegraph model, particularly the latter's probability for low mRNA numbers is not accurate – see Fig. 4C. Given the telegraph model's excellent approximation to the multi-scale model, it is clear that the multi-scale model and refractory model can be distinguished.

Gene dosage compensation. It has been observed that after replication some genes do not exhibit a two-fold increase in mRNA numbers, a counter-intuitive phenomenon known as gene dosage compensation [10]. Here we show that this property naturally follows from our

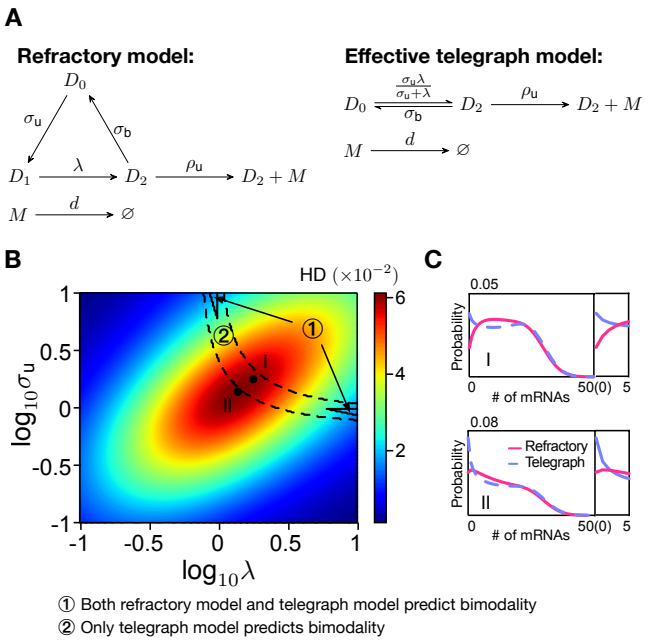


FIG. 4. Effective telegraph model approximation for the refractory model. (A) Schematics illustrating the refractory model and the corresponding effective telegraph model. (B) Shows a heatmap of HD between the steady-state distributions of mRNA numbers predicted by the refractory model and its effective telegraph model, and a bifurcation diagram of the number of modes thereof (marked as black lines) as a function of σ_u and λ with parameters $\sigma_b = 0.8$, $\rho_u = 30$ and $d = 1$. (C) Plots of the distributions for Points I and II in (B), showing significant disagreement in the height of the zero mode of the mRNA number distribution (insets show a zoom at the mode at zero).

multi-scale model and does not require the postulation of new mechanisms. According to the effective telegraph model, the mean of the number of mRNAs is approximately proportional to the effective production rate, i.e. $\langle n\rangle \propto \rho_{\rm u}(\lambda) = \frac{\rho\lambda}{\rho+\lambda}$ which has a Michaelis-Menten like dependency on the polymerase Pol II abundance comprised in the parameter λ. Assuming that the amount of Pol II protein remains constant right after replication, e.g. in cases where replication occurs on a timescale faster than protein turnover, replication causes gene copy number to double but the amount of Pol II per gene is halved. We thus have that the mean number of mRNA before replication is $\langle n\rangle^- \propto \rho_{\rm u}(\lambda)$ while after replication it is $\langle n\rangle^+ \propto 2\rho_{\rm u}(\lambda/2)$. It follows that the fold change upon replication is

$$\eta = \frac{\langle n\rangle^+}{\langle n\rangle^-} = \frac{2(\rho+\lambda)}{2\rho+\lambda} < 2,$$

and therefore our model displays the property of gene dosage compensation. Furthermore assuming that $\rho = \lambda$, our model predicts that the fold change is $\eta = \frac{4}{3}$, a value consistent with experimental measurements for genes $OCT4$ and $Nanog$ in mouse embryonic stem cells (see Fig. 2B in [18]).

Protein dynamics. Finally we extend the multi-scale model to provide analytic distributions for protein numbers. This allows interpretations of single-cell data of protein expression under the classic short-lived mRNA assumption ($d \gg d_{\rm p}$) [19]. Given the network in Fig. 1A, it can be shown [13] that the generating function corresponding to the steady-state distribution of protein numbers is given by

$$G(v) = {}_3F_2(a_1, a_2, a_3; b_1, b_2; bv), \tag{8}$$

with $b_1 = (\sigma_{\rm b} + \sigma_{\rm u})/d_{\rm p}$, $b_2 = (\sigma_{\rm b} + \lambda + \rho)/d_{\rm p}$, the mean translational burst size $b = k/d$, and the parameters a_1, a_2 and a_3 being solutions of

$$\begin{aligned} a_1a_2a_3 &= \sigma_{\rm u}\lambda\rho/d_{\rm p}^3, \\ a_1 + a_2 + a_3 &= b_1 + b_2, \\ a_1a_2 + a_1a_3 + a_2a_3 &= b_1b_2 + \lambda\rho/d_{\rm p}^2. \end{aligned}$$

In the limit of large λ or ρ, we show in [13] that Eq. (8) reduces to the Gaussian hypergeometric function (${}_2F_1$), which was reported in Ref. [19] for the three-stage model of gene expression.

Summary and Discussion. In this letter, we have performed a detailed study of a multi-scale model of bursty gene expression based on recent experimental data from mammalian cells. We have derived simple closed-form expressions for the approximate time evolution of this model and used the theory (i) to understand which reactions contribute mostly to fluctuations, (ii) to show how gene dosage compensation emerges from the special connectivity of the gene states in our model. The simplicity of the moment equations allow the easy inference of rate parameters from single-cell data using maximum likelihood methods [20]. Potential extensions include the impact of cell cycle effects such as binomial partitioning and variability in the cell cycle duration. Use of the recently developed linear mapping approximation [21] appears to be a promising means to extend the analytical solution of the present model to include DNA-protein interactions.

Z.C. gratefully acknowledges support of the UK Research Councils Synthetic Biology for Growth programme and of the BBSRC, EPSRC and MRC (B/M018040/1) and careful proofreading by J. Holehouse. R.G. acknowledges support from BBSRC grant no. BB/M025551/1. D.O. acknowledges support from the Human Frontier Science Program (grant no. RGY076/2015).

* edward.cao@ed.ac.uk / zcao@ecust.edu.cn
† ramon.grima@ed.ac.uk

[1] K. B. Halpern *et al.*, Mol. Cell. **58**, 147 (2015).
[2] D. M. Suter *et al.*, Science **332**, 472 (2011).
[3] A. J. Larsson *et al.*, Nature **565**, 251 (2019).
[4] D. Zenklusen, D. R. Larson, and R. H. Singer, Nat. Struct. Mol. Biol. **15**, 1263 (2008).
[5] A. Sanchez and I. Golding, Science **342**, 1188 (2013).
[6] J. Peccoud and B. Ycart, Theor. Popul. Biol. **48**, 222 (1995).
[7] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, PLOS Biol. **4**, e309 (2006).
[8] K. Tantale *et al.*, Nat. Commun. **7**, 12248 (2016).
[9] C. R. Bartman *et al.*, Mol. Cell. **73**, 519 (2019).
[10] H. Xu, S. O. Skinner, A. M. Sokac, and I. Golding, Phys. Rev. Lett. **117**, 128101 (2016).
[11] A. Senecal *et al.*, Cell Rep. **8**, 75 (2014).
[12] N. G. Van Kampen, *Stochastic processes in physics and chemistry* (Elsevier, Amsterdam, 1992).
[13] See Supplemental Information for detailed derivations.
[14] P. Thomas, N. Popović, and R. Grima, Proc. Natl. Acad. Sci. USA **111**, 6994 (2014).
[15] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[16] S. Redner, *A guide to first-passage processes* (Cambridge University Press, 2001).
[17] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Phys. Rev. E **79**, 031911 (2009).
[18] S. O. Skinner *et al.*, Elife **5**, e12175 (2016).
[19] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. USA **105**, 17256 (2008).
[20] Z. Cao and R. Grima, J. Royal Soc. Interface **16**, 20180967 (2019).
[21] Z. Cao and R. Grima, Nat. Commun. **9**, 3305 (2018).

Supplemental Information for: Multi-scale bursting in stochastic gene expression

Zhixing Cao,[1,2,*] Tatiana Filatova,[2,3] Diego A. Oyarzún,[2,4] and Ramon Grima[2,†]

[1] *The Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, PR China*
[2] *School of Biological Sciences, the University of Edinburgh, United Kingdom*
[3] *School of Mathematics, the University of Edinburgh, United Kingdom*
[4] *School of Informatics, the University of Edinburgh, United Kingdom*

(Dated: July 27, 2019)

EXACT STEADY-STATE SOLUTION FOR THE MARGINAL mRNA DISTRIBUTION OF THE MULTI-SCALE MODEL

To solve Eq. (1) in the main text, we use the generating function method and define $G_\theta(z) = \sum_n z^n P_\theta(n)$ for $\theta = 0, 10, 11$ so that Eq. (1) in the main text can be recast as a set of coupled partial differential equations

$$\partial_t G_0 + d(z-1)\partial_z G_0 = -\sigma_\mathrm{u} G_0 + \sigma_\mathrm{b} G_{10} + \sigma_\mathrm{b} G_{11}, \tag{S1a}$$
$$\partial_t G_{10} + d(z-1)\partial_z G_{10} = \rho z G_{11} - (\sigma_\mathrm{b} + \lambda) G_{10} + \sigma_\mathrm{u} G_0, \tag{S1b}$$
$$\partial_t G_{11} + d(z-1)\partial_z G_{11} = -\rho G_{11} - \sigma_\mathrm{b} G_{11} + \lambda G_{10}, \tag{S1c}$$

wherein the variable z is dropped for brevity. In order to solve Eq. (S1) in the limit of long times, we set $\partial_t G_\theta = 0$, solve G_{10} from Eq. (S1c) as a function of G_{11}, and combine the yielded result to solve G_0 from Eq. (S1b) as a function of G_{11} so that Eq. (S1a) consequently becomes a differential equation with G_{11} being the only variable

$$d^3u^2\partial_u^3 G_{11} + (3d + \gamma_1 + \gamma_2)d^2u\partial_u^2 G_{11} + [(d+\gamma_1)(d+\gamma_2) - \rho\lambda u]d\partial_u G_{11} - (d+\sigma_\mathrm{u})\rho\lambda G_{11} = 0, \tag{S2}$$

with $u = z - 1$, $\gamma_1 = \sigma_\mathrm{b} + \sigma_\mathrm{u}$ and $\gamma_2 = \rho + \lambda + \sigma_\mathrm{b}$. By defining a new variable $x = \rho\lambda u/d^2$, Eq. (S2) can be further simplified to

$$x^2\partial_x^3 G_{11} + \Big(1 + \frac{\gamma_1 + d}{d} + \frac{\gamma_2 + d}{d}\Big)x\partial_x^2 G_{11} + \Big(\frac{\gamma_1 + d}{d}\frac{\gamma_2 + d}{d} - x\Big)\partial_x G_{11} - \frac{\sigma_u + d}{d} G_{11} = 0,$$

which is in the canonical form of the differential equation for the generalized hypergeometric function

$$x^2\partial_x^3 f(x) + (1 + b_1 + b_2)x\partial_x^2 f(x) + (b_1 b_2 - x)\partial_x f(x) - a_1 f(x) = 0,$$

admitting the solution $f(x) = C_1 F_2(a_1; b_1, b_2; x)$ with C being an integration constant. Hence, the solution for G_{11} is

$$G_{11} = C \cdot {}_1F_2\left(\frac{\sigma_\mathrm{u} + d}{d}; \frac{\gamma_1 + d}{d}, \frac{\gamma_2 + d}{d}; \frac{\rho\lambda}{d^2}u\right). \tag{S3}$$

On the other hand, summing Eqs. (S1a)-(S1c) and denoting $G = \sum_\theta G_\theta$, one can get $\partial_u G = \rho G_{11}/d$, which together with Eq. (S3) leads to

$$G(u) = C_2 \cdot {}_1F_2\left(\frac{\sigma_\mathrm{u}}{d}; \frac{\sigma_\mathrm{b} + \sigma_\mathrm{u}}{d}, \frac{\sigma_\mathrm{b} + \rho + \lambda}{d}; \frac{\rho\lambda}{d^2}u\right).$$

Note that in the last step we made use of the general relation $\partial_z\, {}_1F_2(a; b, c; z) = \frac{a}{bc} \cdot {}_1F_2(a+1; b+1, c+1; z)$. The integration constant C_2 is found to be 1 by using the normalization condition $G(0) = 1$. Furthermore, the marginal probability of finding n mRNAs in a cell is

$$P(n) = \frac{1}{n!}\frac{\mathrm{d}^n G(u)}{\mathrm{d}u^n}\bigg|_{u=-1} = \frac{1}{n!}\left(\frac{\rho\lambda}{d^2}\right)\frac{(\frac{\sigma_\mathrm{u}}{d})_n}{(\frac{\sigma_\mathrm{b}+\sigma_\mathrm{u}}{d})_n(\frac{\sigma_\mathrm{b}+\rho+\lambda}{d})_n}\,{}_1F_2\left(\frac{\sigma_\mathrm{u}}{d} + n; \frac{\sigma_\mathrm{b}+\sigma_\mathrm{u}}{d} + n, \frac{\sigma_\mathrm{b}+\rho+\lambda}{d} + n; -\frac{\rho\lambda}{d^2}\right),$$

where $(\cdot)_n$ is the Pochhammer symbol.

ANALYTIC DISTRIBUTION FOR mRNA NUMBERS WHEN ρ, λ AND $\sigma_{\rm b}$ ARE LARGE

Given the large values of ρ, λ and $\sigma_{\rm b}$, we implement the following parametrization

$$\sigma_{\rm b} \mapsto \sigma_{\rm b}\delta, \qquad \rho \mapsto \rho\delta, \qquad \lambda \mapsto \lambda\delta,$$

where δ is a large real number.

By means of the method of characteristics, solving Eq. (S1) is tantamount to seeking a solution to the ODE system

$$\begin{aligned}
&\partial_s t = 1 \quad \Rightarrow \quad t = s \\
&\partial_s z = d(z-1) \quad \Rightarrow \quad z - 1 = r\mathrm{e}^{ds} \\
&\partial_s G = \rho\delta(z-1)G_{11}, && \text{(S4a)}\\
&\partial_s G_{10} = \rho\delta z G_{11} - \sigma_{\rm b}\delta G_{10} - \lambda\delta G_{10} + \sigma_{\rm u}(G - G_{10} - G_{11}), && \text{(S4b)}\\
&\partial_s G_{11} = -\rho\delta G_{11} - \sigma_{\rm b}\delta G_{11} + \lambda\delta G_{10}. && \text{(S4c)}
\end{aligned}$$

Dividing δ on both sides of Eqs. (S4a)–(S4c), one obtains a singular system consisting of

$$\begin{cases}
\epsilon\partial_s G = \rho(z-1)G_{11}, \\
\epsilon\partial_s G_{10} = \rho z G_{11} - \sigma_{\rm b} G_{10} - \lambda G_{10} + \epsilon\sigma_{\rm u}(G - G_{10} - G_{11}), \\
\epsilon\partial_s G_{11} = -\rho G_{11} - \sigma_{\rm b} G_{11} + \lambda G_{10},
\end{cases} \tag{S5}$$

with $\epsilon = 1/\delta \simeq 0$. Expanding G, G_{10} and G_{11} in Eq. (S5) as a series in powers of ϵ,

$$G = G^{(0)} + \epsilon G^{(1)} + \mathcal{O}(\epsilon^2), \qquad G_{10} = G_{10}^{(0)} + \epsilon G_{10}^{(1)} + \mathcal{O}(\epsilon^2), \qquad G_{11} = G_{11}^{(0)} + \epsilon G_{11}^{(1)} + \mathcal{O}(\epsilon^2),$$

and matching the orders of ϵ, we have

$$\text{Order of } \epsilon^0 : \begin{cases}
\rho(z-1)G_{11}^{(0)} = 0 & \Rightarrow \quad G_{11}^{(0)} = 0 \\
\rho z G_{11}^{(0)} - \sigma_{\rm b} G_{10}^{(0)} - \lambda G_{10}^{(0)} = 0 & \Rightarrow \quad G_{10}^{(0)} = 0
\end{cases}$$

and

$$\text{Order of } \epsilon^1 : \begin{cases}
\partial_s G^{(0)} = \rho(z-1)G_{11}^{(1)} & \\
\partial_s G_{10}^{(0)} = \rho z G_{11}^{(1)} - \sigma_{\rm b} G_{10}^{(1)} - \lambda G_{10}^{(1)} + \sigma_{\rm u}(G^{(0)} - G_{10}^{(0)} - G_{11}^{(0)}) & \Rightarrow \quad \rho z G_{11}^{(1)} - \sigma_{\rm b} G_{10}^{(1)} - \lambda G_{10}^{(1)} + \sigma_{\rm u} G^{(0)} = 0 \\
\partial_s G_{11}^{(0)} = -\rho G_{11}^{(1)} - \sigma_{\rm b} G_{11}^{(1)} + \lambda G_{10}^{(1)} & \Rightarrow \quad -\rho G_{11}^{(1)} - \sigma_{\rm b} G_{11}^{(1)} + \lambda G_{10}^{(1)} = 0.
\end{cases}$$

Then, we have

$$\partial_s G^{(0)} = -\frac{\rho u \sigma_{\rm u}}{\rho u - \alpha} G^{(0)}$$

where $\alpha = \sigma_{\rm b}\gamma_2/\lambda$ and $u = z - 1 = r\mathrm{e}^{ds}$. Its solution immediately follows as

$$G^{(0)} = C(r)(\rho r\mathrm{e}^{ds} - \alpha)^{-\frac{\sigma_{\rm u}}{d}} \tag{S6}$$

with $C(r)$ being a function of r to be determined from the initial condition. Suppose that the initial condition for this process is $g(u) = G^{(0)}|_{t=0}$, which is known *a priori*. For instance, say the initial distribution of n mRNA molecules is $P(n) = p_n$, then $g(u) = \sum_n p_n(u+1)^n$. Letting s be equal to 0 (or equivalently $t = 0$), it follows $u = r$ and $g(u) = g(r)$, and we can establish the following relation

$$g(r) = C(r)(\rho r - \alpha)^{-\frac{\sigma_{\rm u}}{d}},$$

from which we can solve $C(r)$ as

$$C(r) = g(r)(\rho r - \alpha)^{\frac{\sigma_{\rm u}}{d}}.$$

Substituting the latter back into Eq. (S6) and replacing $r = u\mathrm{e}^{-dt}$, we can calculate the leading-order solution of G from (S6) as

$$G(u) = g(u\mathrm{e}^{-dt})\left(\frac{\rho u\mathrm{e}^{-dt} - \alpha}{\rho u - \alpha}\right)^{\frac{\sigma_\mathrm{u}}{d}}. \tag{S7}$$

At steady state, the leading-order solution in (S7) becomes

$$G(z) = \left(\frac{\alpha}{\alpha - \rho(z-1)}\right)^{\frac{\sigma_\mathrm{u}}{d}}$$

and the corresponding distribution of mRNA numbers is a negative binomial distribution NB($\frac{\sigma_\mathrm{u}}{d}$, $\frac{\rho}{\rho+\alpha}$).

CONVERGENCE TO TELEGRAPH MODEL FOR LARGE ρ

To this end, we parametrize ρ as $\rho \mapsto \rho\delta$, where δ is a large real number. As such, Eq. (S1) can be recast as

$$\partial_t G_0 + d(z-1)\partial_z G_0 = -\sigma_\mathrm{u} G_0 + \sigma_\mathrm{b} G_{10} + \sigma_\mathrm{b} G_{11}, \tag{S8a}$$
$$\partial_t G_{10} + d(z-1)\partial_z G_{10} + (\sigma_\mathrm{b} + \lambda)G_{10} - \sigma_\mathrm{u} G_0 = \rho\delta z G_{11}, \tag{S8b}$$
$$\partial_t G_{11} + d(z-1)\partial_z G_{11} + \sigma_\mathrm{b} G_{11} - \lambda G_{10} = -\rho\delta G_{11}. \tag{S8c}$$

Dividing both sides of Eqs. (S8b)-(S8c) by δ and setting $\epsilon = \delta^{-1}$, we have that

$$\epsilon\Big(\partial_t G_{10} + d(z-1)\partial_z G_{10} + (\sigma_\mathrm{b} + \lambda)G_{10} - \sigma_\mathrm{u} G_0\Big) = \rho z G_{11}, \tag{S9a}$$
$$\epsilon\Big(\partial_t G_{11} + d(z-1)\partial_z G_{11} + \sigma_\mathrm{b} G_{11} - \lambda G_{10}\Big) = -\rho G_{11}. \tag{S9b}$$

Again using the same method as before, we expand G_0, G_{10} and G_{11} in Eqs. (S8a) and (S9) as a series in powers of ϵ, collect the terms for ϵ^0 and ϵ^1 and obtain

$$\text{Order of } \epsilon^0 : \begin{cases} \partial_t G_0^{(0)} + d(z-1)\partial_z G_0^{(0)} = -\sigma_\mathrm{u} G_0^{(0)} + \sigma_\mathrm{b} G_{10}^{(0)} + \sigma_\mathrm{b} G_{11}^{(0)}, \\ \rho z G_{11}^{(0)} = 0, \\ \rho G_{11}^{(0)} = 0, \end{cases} \tag{S10}$$

and

$$\text{Order of } \epsilon^1 : \begin{cases} \partial_t G_{10}^{(0)} + d(z-1)\partial_z G_{10}^{(0)} + (\sigma_\mathrm{b} + \lambda)G_{10}^{(0)} - \sigma_\mathrm{u} G_0^{(0)} = \rho z G_{11}^{(1)}, \\ \partial_t G_{11}^{(0)} + d(z-1)\partial_z G_{11}^{(0)} + \sigma_\mathrm{b} G_{11}^{(0)} - \lambda G_{10}^{(0)} = -\rho G_{11}^{(1)}. \end{cases} \tag{S11}$$

From Eq. (S10), we can solve that $G_{11}^{(0)} = 0$, with which we can further get $\lambda G_{10}^{(0)} = \rho G_{11}^{(1)}$ from Eq. (S11). Given both results, Eqs. (S10) and (S11) can be simplified to

$$\begin{cases} \partial_t G_0^{(0)} + d(z-1)\partial_z G_0^{(0)} = -\sigma_\mathrm{u} G_0^{(0)} + \sigma_\mathrm{b} G_{10}^{(0)}, \\ \partial_t G_{10}^{(0)} + d(z-1)\partial_z G_{10}^{(0)} = \lambda(z-1)G_{10}^{(0)} - \sigma_\mathrm{b} G_{10}^{(0)} + \sigma_\mathrm{u} G_0^{(0)}, \end{cases}$$

which are exactly the generating function equations of the telegraph model (See Eqs. (A2) and (A3) in [1]), thus showing that the multi-scale transcriptional bursting model converges to the telegraph model when $\rho \to \infty$. A similar proof can be constructed to show that the telegraph model is also obtained in the limit $\lambda \to \infty$.

EXACT STEADY-STATE DISTRIBUTION FOR THE REFRACTORY MODEL

Given the reaction scheme illustrated in Fig. 4A in the main text, it follows that the temporal evolution of probability $P_\theta(n)$ of finding n mRNAs and gene state D_θ ($\theta = 0$, 1 or 2) can be described by the following master

equations

$$\begin{cases}\partial_t P_0(n) = (\mathbb{E}^1 - 1)dnP_0(n) + \sigma_\mathrm{b}P_2(n) - \sigma_\mathrm{u}P_0(n),\\ \partial_t P_1(n) = (\mathbb{E}^1 - 1)dnP_1(n) + \sigma_\mathrm{u}P_0(n) - \lambda P_1(n),\\ \partial_t P_2(n) = (\mathbb{E}^1 - 1)dnP_2(n) + (\mathbb{E}^{-1} - 1)\rho_\mathrm{u}P_2(n) + \lambda P_1(n) - \sigma_\mathrm{b}P_2(n).\end{cases}$$

The corresponding generating function equations are given by

$$\partial_t G_0 + d(z-1)\partial_z G_0 = \sigma_\mathrm{b}G_2 - \sigma_\mathrm{u}G_0, \tag{S12a}$$
$$\partial_t G_1 + d(z-1)\partial_z G_1 = \sigma_\mathrm{u}G_0 - \lambda G_1, \tag{S12b}$$
$$\partial_t G_2 + d(z-1)\partial_z G_2 = \rho_\mathrm{u}(z-1)G_2 + \lambda G_1 - \sigma_\mathrm{b}G_2, \tag{S12c}$$

where $G_\theta = \sum_n z^n P_\theta(n)$. We intend to solve Eqs. (S12) at steady state and thus set $\partial_t G_\theta = 0$. Then, we solve G_1 as a function of G_2 from Eq. (S12c), subsequently substitute it into Eq. (S12b) and solve G_0 as a function of G_2. Following that, Eq. (S12a) becomes an ordinary differential equation with G_2 being the only variable to be solved

$$\begin{aligned}u^2\partial_u^3 G_2 + (3 + \tilde\lambda + \tilde\sigma_\mathrm{b} + \tilde\sigma_\mathrm{u} - \tilde\rho_\mathrm{u}u)u\partial_u^2 G_2 + [1 + \tilde\sigma_\mathrm{b} + \tilde\sigma_\mathrm{u} + \tilde\sigma_\mathrm{b}\tilde\sigma_\mathrm{u} - \tilde\rho_\mathrm{u}(3 + \tilde\sigma_\mathrm{u})u + \tilde\lambda(1 + \tilde\sigma_\mathrm{b} + \tilde\sigma_\mathrm{u} - \tilde\rho_\mathrm{u}u)]\partial_u G_2\\ -(1 + \tilde\lambda)(1 + \tilde\sigma_\mathrm{u})\tilde\rho_\mathrm{u}G_2 = 0,\end{aligned} \tag{S13}$$

where $\tilde\rho_\mathrm{u}$, $\tilde\lambda$, $\tilde\sigma_\mathrm{b}$ and $\tilde\sigma_\mathrm{u}$ are the kinetic parameters normalized with respect to d and $u = z - 1$. Eq. (S13) is the canonical form of the differential equation for the generalized hypergeometric function ${}_2F_2$, admitting the solution

$$G_2(u) = C \cdot {}_2F_2(\tilde\lambda + 1, \tilde\sigma_\mathrm{u} + 1; \beta_1 - \beta_2 + 1, \beta_1 + \beta_2 + 1; \tilde\rho_\mathrm{u}u) \tag{S14}$$

where C is an integration constant, and β_1 and β_2 denote

$$\beta_1 = \frac{\tilde\sigma_\mathrm{u} + \tilde\sigma_\mathrm{b} + \tilde\lambda}{2}, \qquad \beta_2 = \frac{1}{2}\sqrt{\tilde\lambda^2 - 2\tilde\lambda\tilde\sigma_\mathrm{b} + \tilde\sigma_\mathrm{b}^2 - 2\tilde\lambda\tilde\sigma_\mathrm{u} - 2\tilde\sigma_\mathrm{b}\tilde\sigma_\mathrm{u} + \tilde\sigma_\mathrm{u}^2}.$$

Summing Eqs. (S12) leads to $\partial_u G = \partial_u(\sum_\theta G_\theta) = \tilde\rho_\mathrm{u}G_2$, one can obtain G from Eq. (S14) in the form of the generalized hypergeometric function:

$$G(u) = C_2 \cdot {}_2F_2\left(\tilde\lambda, \tilde\sigma_\mathrm{u}; \beta_1 - \beta_2, \beta_1 + \beta_2; \tilde\rho_\mathrm{u}u\right), \tag{S15}$$

and C_2 is found to be 1 by the normalization condition $G(0) = 1$. Eq. (S15) together with $P(n) = \frac{\mathrm{d}^n G}{n!\mathrm{d}u^n}|_{u=-1}$ defines the distribution of mRNA numbers for the refractory model in steady-state conditions. A similar but more general solution was reported in [2].

ANALYTIC MARGINAL DISTRIBUTION FOR PROTEIN NUMBERS FOR THE MULTI-SCALE MODEL IN THE LIMIT OF FAST mRNA DECAY

From the reaction scheme illustrated in Fig. 1A in the main text, one can write down the following master equations describing the time evolution of the probability $P_\theta(n, m)$ of finding n mRNAs, m proteins and gene state D_θ ($\theta = 0, 10, 11$) in a cell:

$$\begin{cases}\partial_t P_0(n,m) = &d(n+1)P_0(n+1,m) - dnP_0(n,m) + d_\mathrm{p}(m+1)P_0(n,m+1) - d_\mathrm{p}mP_0(n,m)\\ &+ knP_0(n,m-1) - knP_0(n,m) - \sigma_\mathrm{u}P_0(n,m) + \sigma_\mathrm{b}P_{10}(n,m) + \sigma_\mathrm{b}P_{10}(n,m),\\ \partial_t P_{10}(n,m) = &d(n+1)P_{10}(n+1,m) - dnP_{10}(n,m) + d_\mathrm{p}(m+1)P_{10}(n,m+1) - d_\mathrm{p}mP_{10}(n,m)\\ &+ knP_{10}(n,m-1) - knP_{10}(n,m) + \sigma_\mathrm{u}P_0(n,m) - (\sigma_\mathrm{b} + \lambda)P_{10}(n,m) + \rho P_{11}(n-1,m),\\ \partial_t P_{11}(n,m) = &d(n+1)P_{11}(n+1,m) - dnP_{11}(n,m) + d_\mathrm{p}(m+1)P_{11}(n,m+1) - d_\mathrm{p}mP_{11}(n,m)\\ &+ knP_{11}(n,m-1) - knP_{11}(n,m) + \lambda P_{10}(n,m) - (\rho + \sigma_\mathrm{b})P_{11}(n,m).\end{cases} \tag{S16}$$

By defining $G_\theta = \sum_n \sum_m {z_\mathrm{m}}^n {z_\mathrm{p}}^m P_\theta(n,m)$, solving Eq. (S16) is tantamount to seeking solutions to the set of differential equations

$$\begin{cases}\partial_t G_0 + [d(z_\mathrm{m} - 1) - k(z_\mathrm{p} - 1)z_\mathrm{m}]\partial_{z_\mathrm{m}} G_0 + d_\mathrm{p}(z_\mathrm{p} - 1)\partial_{z_\mathrm{p}} G_0 = -\sigma_\mathrm{u}G_0 + \sigma_\mathrm{b}G_{10} + \sigma_\mathrm{b}G_{11},\\ \partial_t G_{10} + [d(z_\mathrm{m} - 1) - k(z_\mathrm{p} - 1)z_\mathrm{m}]\partial_{z_\mathrm{m}} G_{10} + d_\mathrm{p}(z_\mathrm{p} - 1)\partial_{z_\mathrm{p}} G_{10} = \sigma_\mathrm{u}G_0 - (\sigma_\mathrm{b} + \lambda)G_{10} + \rho z_\mathrm{m}G_{11},\\ \partial_t G_{11} + [d(z_\mathrm{m} - 1) - k(z_\mathrm{p} - 1)z_\mathrm{m}]\partial_{z_\mathrm{m}} G_{11} + d_\mathrm{p}(z_\mathrm{p} - 1)\partial_{z_\mathrm{p}} G_{11} = \lambda G_{10} - (\rho + \sigma_\mathrm{b})G_{11}.\end{cases} \tag{S17}$$

By means of the method of characteristics, Eq. (S17) is equivalently represented as

$$\partial_s t = 1, \qquad \partial_s z_{\rm m} = d(z_{\rm m}-1) - k(z_{\rm p}-1)z_{\rm m}, \qquad \partial_s z_{\rm p} = d_{\rm p}(z_{\rm p}-1),$$

and

$$\begin{cases} \partial_s G_0 = -\sigma_{\rm u} G_0 + \sigma_{\rm b} G_{10} + \sigma_{\rm b} G_{11}, \\ \partial_s G_{10} = \sigma_{\rm u} G_0 - (\sigma_{\rm b}+\lambda) G_{10} + \rho z_{\rm m} G_{11}, \\ \partial_s G_{11} = \lambda G_{10} - (\rho+\sigma_{\rm b}) G_{11}. \end{cases}$$

Assuming that mRNA decays much faster than protein such that $\partial_s z_{\rm m} \simeq 0$ [3], we get that

$$z_{\rm m} = \frac{1}{1-bv}, \qquad \text{and} \qquad v = z_{\rm p} - 1, \tag{S18}$$

and $b = k/d$ is the mean translational burst size. Using Eq. (S18) we can reduce Eq. (S17) to

$$v\partial_v G_0 = -\tilde{\sigma}_{\rm u} G_0 + \tilde{\sigma}_{\rm b} G_{10} + \tilde{\sigma}_{\rm b} G_{11}, \tag{S19a}$$

$$v\partial_v G_{10} = \tilde{\sigma}_{\rm u} G_0 - (\tilde{\sigma}_{\rm b}+\tilde{\lambda}) G_{10} + \frac{\tilde{\rho}}{1-bv} G_{11}, \tag{S19b}$$

$$v\partial_v G_{11} = \tilde{\lambda} G_{10} - (\tilde{\rho}+\tilde{\sigma}_{\rm b}) G_{11}, \tag{S19c}$$

where $\tilde{\sigma}_{\rm b}$, $\tilde{\sigma}_{\rm u}$, $\tilde{\rho}$ and $\tilde{\lambda}$ are kinetic parameters normalized with respect to protein degradation rate $d_{\rm p}$. It follows from summing Eqs. (S19) that

$$G_{11} = \frac{(1-bv)\partial_v G}{\tilde{\rho} b}.$$

Using the definitions $b_1 = \tilde{\sigma}_{\rm b} + \tilde{\sigma}_{\rm u}$ and $b_2 = \tilde{\sigma}_{\rm b} + \tilde{\lambda} + \tilde{\rho}$, Eqs. (S19) can be simplified to

$$(1-bv)v^2\partial_v^3 G + [1+b_1+b_2-bv(3+b_1+b_2)]v\partial_v^2 G + \{b_1 b_2 - bv[(1+b_1)(1+b_2)+\tilde{\lambda}\tilde{\rho}]\}\partial_v G - b\tilde{\sigma}_{\rm u}\tilde{\lambda}\tilde{\rho} G = 0,$$

which admits a solution

$$G(v) = {}_3F_2(a_1, a_2, a_3; b_1, b_2; bv), \tag{S20}$$

with a_1, a_2 and a_3 being roots of

$$\begin{cases} a_1 a_2 a_3 = \tilde{\sigma}_{\rm u}\tilde{\lambda}\tilde{\rho}, \\ a_1 + a_2 + a_3 = b_1 + b_2, \\ a_1 a_2 + a_1 a_3 + a_2 a_3 = b_1 b_2 + \tilde{\lambda}\tilde{\rho}. \end{cases}$$

Hence summarizing, Eq. (S20) and $P(m) = \frac{{\rm d}^m G(v)}{m! {\rm d} v^m}|_{v=-1}$ define the steady-state distribution of protein numbers, which is

$$P(m) = \frac{b^m}{m!}\frac{(a_1)_n (a_2)_n (a_3)_n}{(b_1)_n (b_2)_n} {}_3F_2(a_1+n, a_2+n, a_3+n; b_1+n, b_2+n; -b),$$

given that mRNA is short-lived.

Next we will show the solution Eq. (S20) converges to the Gaussian hypergeometric function (${}_2F_1$) for the three-stage gene expression model [3] when ρ is large. To this end, we parameterize $\tilde{\rho}$ in Eqs. (S19b)-(S19c) as $\tilde{\rho} \mapsto \tilde{\rho}\delta$ where δ is a large number. Dividing both sides of Eqs. (S19b)-(S19c) by δ, we have

$$\epsilon\big(v\partial_v G_{10} - \tilde{\sigma}_{\rm u} G_0 + (\tilde{\sigma}_{\rm b}+\tilde{\lambda}) G_{10}\big) = \frac{\tilde{\rho}}{1-bv} G_{11}, \tag{S21a}$$

$$\epsilon\big(v\partial_v G_{11} - \tilde{\lambda} G_{10} + \tilde{\sigma}_{\rm b} G_{11}\big) = -\tilde{\rho} G_{11}, \tag{S21b}$$

where $\epsilon = 1/\delta \simeq 0$. Again similarly, we expand G_0, G_{10} and G_{11} in Eqs. (S19a) and (S21) as a series in powers of ϵ, collect the terms for ϵ^0 and ϵ^1 and obtain

$$\text{Order of } \epsilon^0 : \begin{cases} v\partial_v G_0^{(0)} = -\tilde{\sigma}_{\rm u} G_0^{(0)} + \tilde{\sigma}_{\rm b} G_{10}^{(0)} + \tilde{\sigma}_{\rm b} G_{11}^{(0)}, \\ \dfrac{\tilde{\rho}}{1-bv} G_{11}^{(0)} = 0, \\ \tilde{\rho} G_{11}^{(0)} = 0, \end{cases} \tag{S22}$$

and

$$\text{Order of } \epsilon^1 : \begin{cases} v\partial_v G_{10}^{(0)} - \tilde{\sigma}_{\rm u} G_0^{(0)} + (\tilde{\sigma}_{\rm b} + \tilde{\lambda}) G_{10}^{(0)} = \dfrac{\tilde{\rho}}{1-bv} G_{11}^{(1)}, \\ v\partial_v G_{11}^{(0)} - \tilde{\lambda} G_{10}^{(0)} + \tilde{\sigma}_{\rm b} G_{11}^{(0)} = -\tilde{\rho} G_{11}^{(1)}, \end{cases} \tag{S23}$$

From Eqs. (S22), we get $G_{11}^{(0)} = 0$, which is used to reduce Eqs. (S23) and the first equation in Eq. (S22) to

$$\begin{cases} v\partial_v G_0^{(0)} = -\tilde{\sigma}_{\rm u} G_0^{(0)} + \tilde{\sigma}_{\rm b} G_{10}^{(0)}, \\ v\partial_v G_{10}^{(0)} = \tilde{\sigma}_{\rm u} G_0^{(0)} - \tilde{\sigma}_{\rm b} G_{10}^{(0)} + \dfrac{\tilde{\lambda} bv}{1-bv} G_{10}^{(0)}, \end{cases} \tag{S24}$$

Note that Eq. (S24), which is the leading order of Eqs. (S19), is exactly the same as the generating functions of the three-stage gene expression model reported in [3] (See Eqs. (68)-(69) in SI thereof). By means of similar arguments, one can show the reduction of our model when λ is large.

* edward.cao@ed.ac.uk / zcao@ecust.edu.cn
† ramon.grima@ed.ac.uk

[1] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Phys. Rev. E **79**, 031911 (2009).
[2] T. Zhou and J. Zhang, SIAM J. Appl. Math. **72**, 789 (2012).
[3] V. Shahrezaei and P. S. Swain, Proc. Natl. Acad. Sci. USA **105**, 17256 (2008).