Joint distribution of nuclear and cytoplasmic mRNA levels in stochastic models of gene expression: analytical results and parameter inference

Yiling Wang,[1] Juraj Szavits-Nossan,[2] Zhixing Cao,[1,3,*] and Ramon Grima[2,†]

[1] *State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China*

[2] *School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3BF, U. K.*

[3] *Department of Chemical Engineering, Queen's University, Kingston, Canada K7L 3N6*

(Dated: April 29, 2024)

Stochastic models of gene expression are routinely used to explain large variability in measured mRNA levels between cells. These models typically predict the distribution of the total mRNA level per cell but ignore compartment-specific measurements which are becoming increasingly common. Here we construct a two-compartment model that describes promoter switching between active and inactive states, transcription of nuclear mRNA and its export to the cytoplasm where it decays. We obtain an analytical solution for the joint distribution of nuclear and cytoplasmic mRNA levels in steady-state conditions. Based on this solution, we build an efficient and accurate parameter inference method which is orders of magnitude faster than conventional methods.

Gene expression is an inherently noisy process [1–4] resulting in significant variation of mRNA and protein numbers between cells. Stochastic models have been studied to shed light on the biophysical processes leading to this variability [5–7] and to fit data and thereby estimate the parameter values specific to each gene [8–11].

The vast majority of models of stochastic gene expression are based on the telegraph model whereby a gene is in one of two states (active and inactive), mRNA is produced from the active state and subsequently decays [12, 13]. The mRNA in this model has often been interpreted as the total mRNA in a cell. However, nowadays mRNA levels can be measured simultaneously in both the nucleus and the cytoplasm [14–17]. Hence more refined stochastic models of gene expression are needed which can fit this type of data and to also understand how the spatial partitioning of molecular systems may play a role in the reduction of noise [18].

Models involving more than one type of mRNA, i.e. describing all or a subset of nascent, mature nuclear and mature cytoplasmic mRNA levels, have been constructed and solved either numerically [15, 19] or analytically for various special cases of interest [20–27]. The analytical solutions provide expressions for the moments or the marginal distributions of mRNA levels except Ref [20] which additionally provides the joint distribution of nuclear and cytoplasmic mRNA under the assumption of instantaneous transcriptional bursting, i.e. a gene is active for an infinitesimally short time and in this time a discrete number of mRNAs (a burst) are produced. This approximation is debatable for eukaryotic genes which spend many minutes in the active state [28].

In this paper, we consider a general stochastic model of gene expression and derive the joint distribution for mRNA levels measured at two different points of the mRNA lifecycle—these two types of mRNA can be interpreted as nascent and mature nuclear mRNA or unspliced and spliced mRNA or (mature) nuclear and cytoplasmic mRNA. We verify the accuracy of the solution using Monte Carlo simulations and finally use it to devise an accurate and computationally efficient method for the inference of gene-specific kinetic parameters from commonly available data.

We consider a stochastic model of nuclear and cytoplasmic mRNA fluctuations which summarises the known biology. This is described by the following reaction system,

$$G \underset{\sigma_{\rm on}}{\overset{\sigma_{\rm off}}{\rightleftharpoons}} G^*, \quad G \xrightarrow{\rho} G + N, \quad N \overset{\tau}{\Rightarrow} M, \quad M \xrightarrow{d} \varnothing. \quad (1)$$

A cartoon of the model is shown in Fig. 1. A promoter switches between two states, an active state G and an inactive state G^*, with rates $\sigma_{\rm off}$ and $\sigma_{\rm on}$. Transcription occurs only from the state G leading to nuclear mRNA (N) production at a rate ρ. After a deterministic time de-

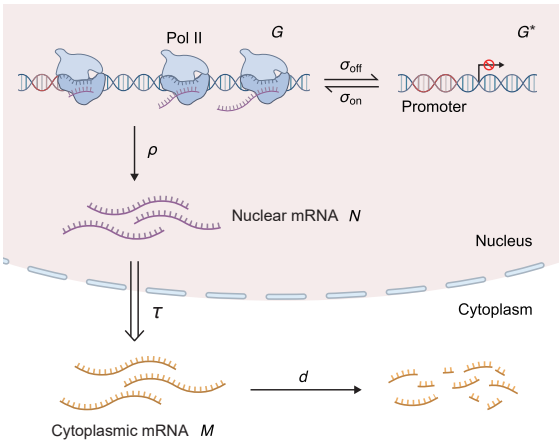

FIG. 1. Illustration of a two-compartment model of mRNA dynamics corresponding to the reaction system in Eq. (1). A promoter switches between an inactive and an active state. Nuclear mRNA is transcribed from the active state and after a time τ it is exported to the cytoplasm where it degrades.

lay τ, nuclear mRNAs are then exported to the cytoplasm thus becoming cytoplasmic mRNA M. These molecules then degrade at a rate d. In accordance with previous work [29], we do not model the degradation of nuclear mRNA because bona fide coding transcripts are specifically protected from RNA decay enzymes in the nucleus [30]. Note that while some studies suggest a three-state model rather than a two-state model for transcription in mammalian cells [31], the marginal mRNA distributions of these competing models cannot be easily distinguished from each other [32] and hence here we only focus on the two-state model.

The probabilistic description of the reaction system (1) is given by the delay master equation:

$$\begin{aligned}\partial_t P_0(n,m,t) &= \sigma_{\text{off}} P_1(n,m,t) - \sigma_{\text{on}} P_0(n,m,t)\\ &+ \rho \mathbf{1}_{t-\tau}\left(\mathbb{E}^{(0,-1)} - \mathbb{E}^{(-1,0)}\right)\bar{P}_0(n,\tau) Q_1(m,t-\tau)\\ &+ d\left(\mathbb{E}^{(0,1)} - 1\right) m P_0(n,m,t), \qquad (2a)\\ \partial_t P_1(n,m,t) &= \sigma_{\text{on}} P_0(n,m,t) - \sigma_{\text{off}} P_1(n,m,t)\\ &+ \rho \mathbf{1}_{t-\tau}\left(\mathbb{E}^{(0,-1)} - \mathbb{E}^{(-1,0)}\right)\bar{P}_1(n,\tau) Q_1(m,t-\tau)\\ &+ d\left(\mathbb{E}^{(0,1)} - 1\right) m P_1(n,m,t)\\ &+ \rho\left(\mathbb{E}^{(-1,0)} - 1\right) P_1(n,m,t), \qquad (2b)\end{aligned}$$

where $P_\varphi(n,p,t)$ is the probability of observing n nuclear mRNAs and m cytoplasmic mRNAs in a cell when a promoter is inactive ($\varphi = 0$) or active ($\varphi = 1$). Here, $\mathbf{1}_{t-\tau} = 1$ if and only if $t \geq \tau$. The step operator $\mathbb{E}^{(i,j)}$ which acts on a general function $f(n,m)$ is defined as $\mathbb{E}^{(i,j)} f(n,m) = f(n+i, m+j)$. Note that the delay master equation [33, 34] is not the standard master equation [35], because the presence of a deterministic export time means that the model is non-Markovian. A derivation of Eq. (2) can be found in the Supplementary Materials (SM).

The probability $\bar{P}_\varphi(n,\tau)$ in Eq. (2) is the solution to the master equation of the reaction system

$$G \underset{\sigma_{\text{on}}}{\overset{\sigma_{\text{off}}}{\rightleftharpoons}} G^*, \quad G \xrightarrow{\rho} G + N,$$

which is initiated with the gene in the active state G and zero nuclear mRNAs. By contrast, the probability $Q_1(m,t-\tau)$ in Eq. (2) corresponds to the reactions of the telegraph model

$$G \underset{\sigma_{\text{on}}}{\overset{\sigma_{\text{off}}}{\rightleftharpoons}} G^*, \quad G \xrightarrow{\rho} G + M, \quad M \xrightarrow{d} \varnothing,$$

with the same initial conditions of Eq. (1). Specifically, these two probabilities are the solutions to the following compact master equation:

$$\begin{aligned}\partial_t H_0(l,t) &= \sigma_{\text{off}} H_1(l,t) + d\mathbf{1}_Q(\mathbb{E}^1 - 1) l H_0(l,t)\\ &- \sigma_{\text{on}} H_0(l,t), \qquad (3a)\\ \partial_t H_1(l,t) &= -\sigma_{\text{off}} H_1(l,t) + d\mathbf{1}_Q(\mathbb{E}^1 - 1) l H_1(l,t)\\ &+ \rho(\mathbb{E}^{-1} - 1) H_1(l,t) + \sigma_{\text{on}} H_0(l,t), \qquad (3b)\end{aligned}$$

where $\mathbb{E}^i f(l) = f(l+i)$. The initial conditions for solving $\bar{P}_\varphi(n,\tau)$ are $H_\varphi(l,0) = 0$ for any l and $\varphi = 0,1$ except $H_1(0,0) = 1$. By setting $\mathbf{1}_Q = 0$ and solving Eq. (3) for a time τ, the probability $\bar{P}_\varphi(n,\tau)$ is indeed the solution at time τ, i.e., $\bar{P}_\varphi(n,\tau) = H_\varphi(n,\tau)$ ($l = n$). By contrast, solving $Q_1(m,t-\tau)$ requires the initial conditions $H_\varphi(m,0) = \sum_n P_\varphi(n,m,0)$ and $\mathbf{1}_Q = 1$. Then, the probability $Q_1(m,t-\tau)$ is yielded as the solution to Eq. (3) at time $t-\tau$, i.e., $Q_1(m,t-\tau) = H_1(m,t-\tau)$.

Next, we show how to solve Eq. (2) in steady-state conditions by means of the generating function method. By defining the generating functions $G_\varphi(u_1,u_2,t) = \sum_{n,m}(u_1+1)^n(u_2+1)^m P_\varphi(n,m,t)$, $G^\tau_{N\varphi}(u_1) = \sum_n (u_1+1)^n \bar{P}_\varphi(n,\tau)$ and $G_{M\varphi}(u_2,t) = \sum_m (u_2+1)^m Q_\varphi(m,t)$, Eq. (2) simplifies to

$$\begin{aligned}\partial_t G_0 =& \rho(u_2 - u_1) G^\tau_{N0} G_{M1} - d u_2 \partial_{u_2} G_0\\ &+ \sigma_{\text{off}} G_1 - \sigma_{\text{on}} G_0, \qquad (4a)\\ \partial_t G_1 =& \rho(u_2 - u_1) G^\tau_{N1} G_{M1} - d u_2 \partial_{u_2} G_1\\ &+ \sigma_{\text{on}} G_0 - \sigma_{\text{off}} G_1 + \rho u_1 G_1. \qquad (4b)\end{aligned}$$

For compactness of notation, the arguments t, u_1 and u_2 are hereafter omitted. Similarly, for solving $G^\tau_{N\varphi}$, Eq. (3) becomes

$$\begin{aligned}\partial_t G_{N0} =& \sigma_{\text{off}} G_{N1} - \sigma_{\text{on}} G_{N0}\\ \partial_t G_{N1} =& \sigma_{\text{on}} G_{N0} - \sigma_{\text{off}} G_{N1} + \rho u_1 G_{N1},\end{aligned}$$

which admits the solution

$$\begin{aligned}G^\tau_{N0} =& \frac{\sigma_{\text{off}}}{d\theta}\left(e^{-v_2\tau} - e^{-v_1\tau}\right), \qquad (5a)\\ G^\tau_{N1} =& \frac{v_1 - \sigma_{\text{on}}}{d\theta} e^{-v_1\tau} - \frac{v_2 - \sigma_{\text{on}}}{d\theta} e^{-v_2\tau}, \qquad (5b)\end{aligned}$$

where $v_{1,2} = d(r \pm \theta - 1)/2$, $r = 1 - x_1 + (\sigma_{\text{off}} + \sigma_{\text{on}})/d$, $\theta = [(x_1 - \sigma_{\text{off}}/d - \sigma_{\text{on}}/d)^2 + 4\sigma_{\text{on}} x_1/d]^{\frac{1}{2}}$ and $x_1 = \rho u_1/d$. By imposing the stationary condition, G_{M1} becomes the steady-state solution to the generating function form of Eq. (3), which was earlier reported in Ref. [36]. Specifically, $G_{M1} = \alpha M(1 + \sigma_{\text{on}}/d, 1 + \sigma_{\text{off}}/d + \sigma_{\text{on}}/d, x_2)$ with $\alpha = \sigma_{\text{on}}/(\sigma_{\text{off}} + \sigma_{\text{on}})$, $x_2 = \rho u_2/d$ and $M(a,b,x)$ being the Kummer function.

By combining Eqs. (4a) and (4b) in the steady-state limit, we obtain the following ordinary differential equation for the joint probability generating function $G = G_0 + G_1$,

$$\begin{aligned}&d^2 u_2^2 \partial^2_{u_2} G + d u_2(\sigma_{\text{on}} + \sigma_{\text{off}} - \rho u_1 + d)\partial_{u_2} G\\ &- \sigma_{\text{on}} \rho u_1 G = -\rho u_1 I_0 + (\sigma_{\text{on}} + \sigma_{\text{off}}) I + d u_2 \partial_{u_2} I, \qquad (6)\end{aligned}$$

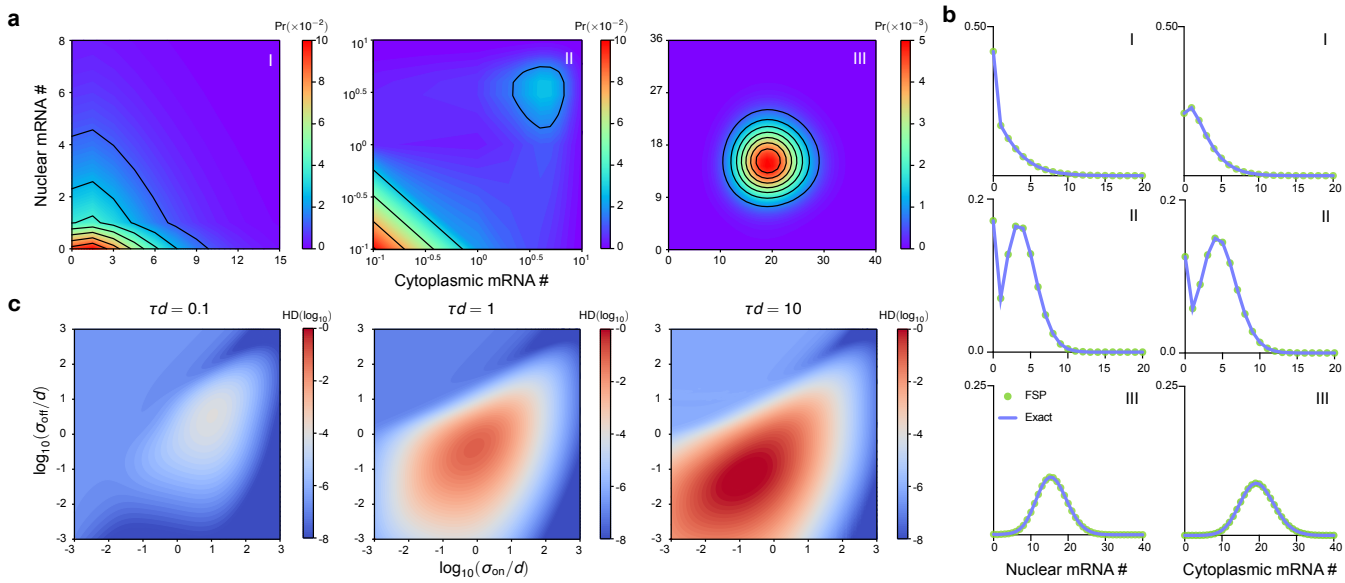

FIG. 2. (a) Comparison of the analytical joint distribution of nuclear and cytoplasmic mRNA numbers in the non-Markovian Model I (reaction system (1)) predicted by Eqs. (7) and (8)—shown by a color heatmap—with the joint distribution computed by direct numerical solution of the delay master equation, Eq. (2), using the finite state projection method (FSP) shown by black contours. The excellent match between the contours and the density plot provides a verification of the exactness of the analytical joint distribution. The kinetic parameters for the three cases are: (I) $\rho = 120$, $\sigma_{\rm off} = 100$, $\sigma_{\rm on} = 2$, $\tau = 0.8$, $d = 1$; (II) $\rho = 5$, $\sigma_{\rm off} = 0.041$, $\sigma_{\rm on} = 0.2$, $\tau = 0.8$, $d = 1$; (III) $\rho = 20$, $\sigma_{\rm off} = 0.1$, $\sigma_{\rm on} = 4.5$, $\tau = 0.8$, $d = 1$. (b) Marginal distributions of nuclear and cytoplasmic mRNA levels corresponding to the joint distributions of the three parameter cases. (c) Hellinger distance (HD) between the joint distributions of Models I (reaction system (1)) and II (reaction system (9)) as a function of the normalised gene state switching rates $\sigma_{\rm on}/d$ and $\sigma_{\rm off}/d$, and the normalised export time τ/d (ρ is fixed to 5).

where $I_0 = \rho(u_2 - u_1)G^{\tau}_{N0}G_{M1}$, $I_1 = \rho(u_2 - u_1)G^{\tau}_{N1}G_{M1}$ and $I = I_0 + I_1$. The solution to this equation takes a simple form in terms of two Kummer functions,

$$G = \frac{v_1 e^{-v_2\tau} - v_2 e^{-v_1\tau}}{d\theta} M\left(\frac{\sigma_{\rm on}}{d}, \frac{\sigma_{\rm on}+\sigma_{\rm off}}{d}, \frac{\rho u_2}{d}\right) + \rho u_1 \frac{e^{-v_2\tau} - e^{-v_1\tau}}{d\theta} G_{M1}. \quad (7)$$

The solutions of the conditional generating functions G_0 and G_1 can be found in SM. Finally, the joint distribution solution is constructed using the relation

$$P(n,m) = \frac{1}{n!m!}\frac{\partial^n}{\partial u_1^n}\frac{\partial^m}{\partial u_2^m} G\Big|_{u_1=-1,u_2=-1}. \quad (8)$$

which can be efficiently calculated by a symbolic computation program.

In Fig. 2a, we verify the accuracy of the exact solution given by Eqs. (7) and (8) by comparing it with the joint distribution that is obtained by solving Eq. (2) directly using the finite state projection (FSP) method [37], a standard commonly used method for studying stochastic reaction systems. In Fig. 2b we also show a comparison of the marginal distributions of nuclear and cytoplasmic mRNA computed using our solution and FSP. We find that an excellent match between the two is achieved for all kinetic parameters considered, which cover all shapes of the marginal distributions that are possible in telegraph-like models [38]. Note that the theoretical marginal nuclear mRNA distribution can be obtained from Eqs. (7) and (8) by setting $u_2 = 0$ while the marginal cytoplasmic mRNA distribution follows from the same equation by setting $u_1 = 0$.

Thus far we have exactly solved the probabilistic description of the reaction system (1) which we shall henceforth refer to as Model I. An alternative, competing model is one in which the export reaction is modelled by a first-order reaction:

$$G \underset{\sigma_{\rm on}}{\overset{\sigma_{\rm off}}{\rightleftharpoons}} G^*, \quad G \xrightarrow{\rho} G + N, \quad N \xrightarrow{\gamma} M, \quad M \xrightarrow{d} \varnothing. \quad (9)$$

In this model, versions of which have been considered by various studies [29, 39], the export time is exponentially distributed and hence the dynamics are Markovian—we shall refer to this model as Model II. To meaningfully compare the two models, we set $\gamma = 1/\tau$ so that mean nuclear export time is the same in both. In Fig. 2c we show the Hellinger distance between the steady-state joint distribution of Model I computed using Eqs. (7) and (8) and the steady-state joint distribution of Model II which is calculated using FSP since an analytical solution for this model is missing. We find that the differences between the two models are very small when either $d\tau$ is small (nuclear export is fast compared to cytoplasmic degradation) or expression is constitutive $\sigma_{\rm on} \gg \sigma_{\rm off}$ or expression is bursty $\sigma_{\rm off} \gg \sigma_{\rm on}$. An analysis of experimental data in Ref. [29] shows that about 80% of genes

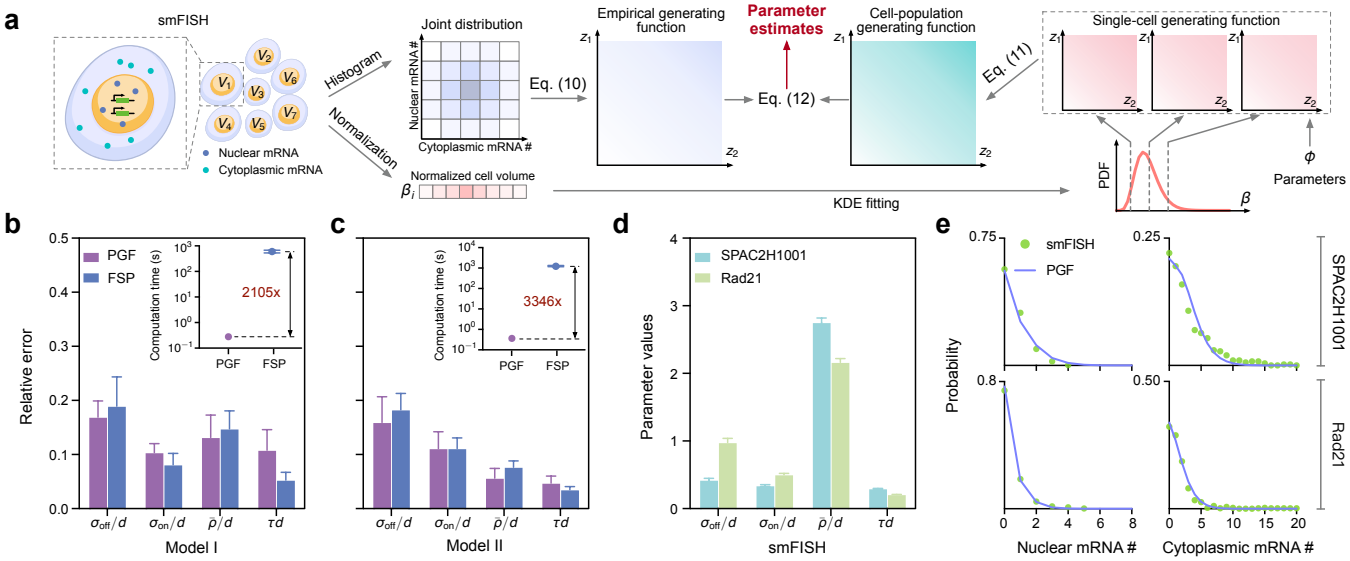

FIG. 3. (a) Illustration of all the steps involved in the probability generating function (PGF) based inference method. (b) Comparison of the estimation accuracy and computation efficiency of the inference method based on the PGF and the standard maximum likelihood method based on FSP. Synthetic nuclear and cytoplasmic mRNA data is generated by Monte Carlo methods simulating Model I (b) and Model II (c) for 2000 cells, and the normalized cell volume β of each cell is sampled from the distribution Gamma(7.5,0.11). There are 10 sets of true values of the four kinetic parameters randomly generated for both (b) and (c) (see SM). In the PGF method, Eqs. (10)–(12) are employed to perform the inference both in (b) and (c). By contrast, FSP directly solves the single-copy joint distribution from the delay master equation Eq. (2) in (b) and the Markovian master equation corresponding to the reaction system (9) in (c). These are subsequently convoluted to compute the single-cell (two independent copies) joint distribution and averaged over the cell volume distribution to obtain the cell-population joint distribution, i.e. the likelihood. The truncated state space size for FSP is 45×45 in (b) and 50×50 in (c). The accuracy of the two methods, as measured by the relative error, is comparable for all parameters though their computation times are considerably different: 0.278s (PGF) and 585.09s (FSP) in (b), and 0.364s (PGF) and 1217.48s (FSP) in (c) when the inference algorithm is run on a MacBook Pro with an 11 core Apple M3 Pro chip and 18G RAM. (d) Parameter estimates based on smFISH data of gene SPAC2H1001 and Rad21 in fission yeast. (e) Constructed marginal distributions of nuclear and cytoplasmic mRNA based on parameter estimates inferred by PGF. Bars: standard error of the mean (SEM).

in mammalian cells have $d\tau < 1$ (see SM) and hence suggests that in most cases, independent of the actual mechanism describing nuclear export, our solution given by Eqs. (7) and (8) is accurate.

Next, we use the exact generating function solution given by Eq. (7) to devise an efficient and accurate method to estimate the parameter values of Model I or Model II from gene expression data. We first benchmarked the accuracy of our method using simulated data, since in this case the parameter values are known. To closely mimic the experimental data collected from mammalian cells, we account for extrinsic noise due to the coupling of transcription to cell volume [40, 41] by choosing the transcription rate ρ to be proportional to the normalized cell volume β (cell volume divided by the average cell volume), i.e., the transcription reaction in Models I and II reads $G \xrightarrow{\bar{\rho}\beta} G + N$. This implies that the burst size (ρ/σ_{off}) increases with the cell volume, as has been shown by various experimental studies in eukaryotic cells [42, 43]. Furthermore, since eukaryotic cells have multiple copies of each gene, for simplicity we assume that each cell has two independent copies [44]. Given a parameter set and a volume distribution, one can then simulate either Model I using the delay stochastic simulation algorithm [45] or Model II using the stochastic simulation algorithm [46] to generate n_c tuples of nuclear-cytoplasmic mRNA and cell volume observations (n_i, m_i, β_i) which mimics experimental measurements from n_c cells. Our parameter inference method then consists of (i) constructing an empirical generating function from the data and (ii) finding the parameter set which minimizes the distance between the latter function and the theoretical generating function Eq. (7) modified to account for extrinsic noise and multiple gene copies. The empirical generating function is computed as

$$\bar{G}(z_1, z_2) = \sum_{i=1}^{n_c} \frac{1}{n_c} z_1^{n_i} z_2^{m_i}. \tag{10}$$

The model's generating function solution (modified to account for extrinsic noise due to volume and two independent gene copies) is given by

$$G(z_1, z_2, \phi) = \int_0^\infty G^2(z_1, z_2, \phi, \beta)\hat{p}(\beta)\mathrm{d}\beta, \tag{11}$$

where $z_1 = u_1 + 1$, $z_2 = u_2 + 1$ and the parameter vector is $\phi = \{\bar{\rho}, \sigma_{\text{on}}, \sigma_{\text{off}}, \tau\}$. Note that the distribution of

the normalized cell volume $\hat{p}(\beta)$ is estimated by fitting β_i with a kernel density estimation model (KDE) [47]. Finally, the inferred kinetic parameters are obtained by solving the optimization problem:

$$\phi = \arg\min_{\phi} \int_0^1 \int_0^1 |G(z_1, z_2, \phi) - \bar{G}(z_1, z_2)|^2 \mathrm{d}z_1 \mathrm{d}z_2. \tag{12}$$

In Fig. 3a we show an illustration of all the steps involved in the probability generating function (PGF) based inference method. In Fig. 3b we show by the purple bars the relative errors of this inference method, given as input 10 different datasets obtained by simulation of Model I (see SM for parameter value details). The errors are small and comparable in size to those obtained by maximizing the likelihood (the joint probability distribution solution of the CME) numerically computed using the FSP method [15, 37] and shown by the blue bars. However, the PGF method is remarkably faster, showing a speedup of a few thousand times over FSP. In Fig. 3c we show the same comparison but using simulated data from Model II (see SM for for parameter value details).

Finally, we applied the PGF inference method to infer parameters from single cell measurements of cell size, nuclear and cytoplasmic mRNA levels for the fission yeast genes, SPAC2H1001 and Rad21 [17]. The mRNA levels were measured using single-molecule fluorescent in situ hybridization (smFISH) in mononucleated cells in which each gene has two copies. The parameter estimates are shown in Fig. 3d where the error bars are computed using bootstrapping (see SM for details). For both genes, we find evidence of timescale separation between two sets of processes: switching between the active and inactive states is slow compared to degradation, while the synthesis of mRNA and nuclear export are fast. Fig. 3e confirms that the experimental marginal distributions are well fit by the marginal distributions constructed using the estimated parameters.

In summary, we have presented an exact solution of the steady-state joint distribution of nuclear and cytoplasmic mRNA levels in a two-compartment model of gene expression. Unlike previous solutions which are only valid for active promoter states that are infinitesimally short [20], the present solution is valid over all of the parameter space which makes it ideal for accurate parameter estimation from experimental data. A promising direction of future research work is the extension of the derivation of the joint distribution of nuclear and cytoplasmic mRNA to more complex models of gene expression with general initiation mechanisms involving an arbitrary number of gene states.

Acknowledgment.— Z.C. acknowledges support from NSFC Grant (62073137), Shanghai Action Plan for Technological Innovation Grants (22ZR1415300, 22511104000, 23S41900500), the Natural Science and Engineering Research Council of Canada's (NSERC's) Discovery Grant (RGPIN-2024-06015) and Shanghai Center of Biomedicine Development. R.G. and J.S.-N. acknowledge support from a Leverhulme Trust grant (RPG-2020-327). We thank Xinyue Liu and Dr. Jie Zhou for graphical illustration.

* z.cao@queensu.ca
† ramon.grima@ed.ac.uk

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, Science **297**, 1183 (2002).
[2] W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins, Noise in eukaryotic gene expression, Nature **422**, 633 (2003).
[3] J. Rodriguez, G. Ren, C. R. Day, K. Zhao, C. C. Chow, and D. R. Larson, Intrinsic dynamics of a human gene reveal the basis of expression heterogeneity, Cell **176**, 213 (2019).
[4] A. Sanchez and I. Golding, Genetic determinants and cellular constraints in noisy gene expression, Science **342**, 1188 (2013).
[5] N. Kumar, T. Platini, and R. V. Kulkarni, Exact distributions for stochastic gene expression models with bursting and feedback, Physical Review Letters **113**, 268105 (2014).
[6] N. Friedman, L. Cai, and X. S. Xie, Linking stochastic dynamics to population distribution: an analytical framework of gene expression, Physical Review Letters **97**, 168302 (2006).
[7] L. Ham, R. D. Brackston, and M. P. Stumpf, Extrinsic noise and heavy-tailed laws in gene expression, Physical Review Letters **124**, 108101 (2020).
[8] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl, Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings, Nature Methods **11**, 197 (2014).
[9] 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).
[10] X. Fu, H. P. Patel, S. Coppola, L. Xu, Z. Cao, T. L. Lenstra, and R. Grima, Quantifying how post-transcriptional noise and gene copy number variation bias transcriptional parameter inference from mRNA distributions, Elife **11**, e82493 (2022).
[11] Z. Kilic, M. Schweiger, C. Moyer, D. Shepherd, and S. Pressé, Gene expression model inference from snapshot RNA data using bayesian non-parametrics, Nature Computational Science **3**, 174 (2023).
[12] J. Peccoud and B. Ycart, Markovian modeling of gene-product synthesis, Theoretical Population Biology **48**, 222 (1995).
[13] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, Stochastic mRNA synthesis in mammalian cells, PLoS Biology **4**, e309 (2006).
[14] C. Xia, J. Fan, G. Emanuel, J. Hao, and X. Zhuang, Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression, Proceedings of the National

Academy of Sciences **116**, 19490 (2019).
[15] B. Munsky, G. Li, Z. R. Fox, D. P. Shepherd, and G. Neuert, Distribution shapes govern the discovery of predictive models for gene regulation, Proceedings of the National Academy of Sciences **115**, 7533 (2018).
[16] G. Li and G. Neuert, Multiplex RNA single molecule FISH of inducible mrnas in single yeast cells, Scientific Data **6**, 94 (2019).
[17] D. E. Weidemann, J. Holehouse, A. Singh, R. Grima, and S. Hauf, The minimal intrinsic stochasticity of constitutively expressed eukaryotic genes is sub-poissonian, Science Advances **9**, eadh5138 (2023).
[18] T. Stoeger, N. Battich, and L. Pelkmans, Passive noise filtering by cellular compartmentalization, Cell **164**, 1151 (2016).
[19] 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**, e12175 (2016).
[20] A. Singh and P. Bokes, Consequences of mRNA transport on stochastic variability in protein levels, Biophysical Journal **103**, 1087 (2012).
[21] M. Sturrock, S. Li, and V. Shahrezaei, The influence of nuclear compartmentalisation on stochastic dynamics of self-repressing gene expression, Journal of Theoretical Biology **424**, 55 (2017).
[22] G. Gorin and L. Pachter, Modeling bursty transcription and splicing with the chemical master equation, Biophysical Journal **121**, 1056 (2022).
[23] G. Gorin, S. Yoshida, and L. Pachter, Assessing markovian and delay models for single-nucleus RNA sequencing, Bulletin of Mathematical Biology **85**, 114 (2023).
[24] T. Filatova, N. Popovic, and R. Grima, Statistics of nascent and mature RNA fluctuations in a stochastic model of transcriptional initiation, elongation, pausing, and termination, Bulletin of Mathematical Biology **83**, 1 (2021).
[25] T. Filatova, N. Popović, and R. Grima, Modulation of nuclear and cytoplasmic mRNA fluctuations by time-dependent stimuli: Analytical distributions, Mathematical Biosciences **347**, 108828 (2022).
[26] J. Szavits-Nossan and R. Grima, Mean-field theory accurately captures the variation of copy number distributions across the mRNA life cycle, Physical Review E **105**, 014410 (2022).
[27] C. Shi, X. Yang, J. Zhang, and T. Zhou, Stochastic modeling of the mRNA life process: A generalized master equation, Biophysical Journal **122**, 4023 (2023).
[28] Time spent in the active state can be estimated from data in table 1 of [48], since the mean active time is the inverse of σ_b.
[29] K. B. Halpern, I. Caspi, D. Lemze, M. Levy, S. Landen, E. Elinav, I. Ulitsky, and S. Itzkovitz, Nuclear retention of mRNA in mammalian tissues, Cell Reports **13**, 2653 (2015).
[30] S. Bresson and D. Tollervey, Surveillance-ready transcription: nuclear RNA decay as a default fate, Open Biology **8**, 170270 (2018).
[31] 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).
[32] Z. Cao, T. Filatova, D. A. Oyarzún, and R. Grima, A stochastic model of gene expression with polymerase recruitment and pause release, Biophysical Journal **119**, 1002 (2020).
[33] A. Leier and T. T. Marquez-Lago, Delay chemical master equation: direct and closed-form solutions, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **471**, 20150049 (2015).
[34] 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**, 2618 (2021).
[35] N. G. Van Kampen, Stochastic processes in physics and chemistry, Vol. 1 (Elsevier, 1992).
[36] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Stochasticity of gene products from transcriptional pulsing, Physical Review E **79**, 031911 (2009).
[37] B. Munsky and M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, The Journal of Chemical Physics **124** (2006), 10.1063/1.2145882.
[38] F. Jiao, Q. Sun, M. Tang, J. Yu, and B. Zheng, Distribution modes and their corresponding parameter regions in stochastic gene transcription, SIAM Journal on Applied Mathematics **75**, 2396 (2015).
[39] M. M. Hansen, R. V. Desai, M. L. Simpson, and L. S. Weinberger, Cytoplasmic amplification of transcriptional noise generates substantial cell-to-cell variability, Cell Systems **7**, 384 (2018).
[40] C. Jia and R. Grima, Coupling gene expression dynamics to cell size dynamics and cell cycle events: Exact and approximate solutions of the extended telegraph model, Iscience **26** (2023), 10.1016/j.isci.2022.105746.
[41] P. Thomas and V. Shahrezaei, Coordination of gene expression noise with cell size: analytical results for agent-based models of growing cell populations, Journal of the Royal Society Interface **18**, 20210274 (2021).
[42] O. Padovan-Merhar, G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone, S. W. Foley, A. R. Wu, L. S. Churchman, A. Singh, and A. Raj, Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms, Molecular Cell **58**, 339 (2015).
[43] S. Berry and L. Pelkmans, Mechanisms of cellular mrna transcript homeostasis, Trends in Cell Biology **32**, 655 (2022).
[44] Q. Deng, D. Ramsköld, B. Reinius, and R. Sandberg, Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells, Science **343**, 193 (2014).
[45] X. Cai, Exact stochastic simulation of coupled chemical reactions with delays, The Journal of Chemical Physics **126** (2007), 10.1063/1.2710253.
[46] D. T. Gillespie, Stochastic simulation of chemical kinetics, Annual Review of Physical Chemistry **58**, 35 (2007).
[47] Y.-C. Chen, A tutorial on kernel density estimation and recent advances, Biostatistics & Epidemiology **1**, 161 (2017).
[48] Z. Cao and R. Grima, Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells, Proceedings of the National Academy of Sciences **117**, 4682 (2020).
[49] Nist digital library of mathematical functions, `https://dlmf.nist.gov/`, Release 1.1.10 of 2023-06-15.

Supplemental Material

In this supplement, we first derive the delay master equations for the non-Markovian model of the reaction system (1). Then, we analytically solve the master equations for the generating function method conditional on the gene state. Finally, we present technical details of the generating function based method which is used to estimate kinetic parameters from experimental mRNA data characterised by both intrinsic and extrinsic noise.

I. Derivation of the delay master equations

Consider the delayed mRNA export model, which consists of five reactions:

$$G \underset{\sigma_{\text{on}}}{\overset{\sigma_{\text{off}}}{\rightleftharpoons}} G^*, \qquad G \xrightarrow{\rho} G + N, \qquad N \overset{\tau}{\Rightarrow} M, \qquad M \xrightarrow{d} \varnothing. \tag{S1}$$

The gene can switch between active G and inactive G^* states, transcribes nuclear mRNA N while in the active state which subsequently is exported after a time τ from the nucleus to the cytoplasm thus becoming cytoplasmic mRNA M. Finally, the cytoplasmic mRNA is degraded via a first-order reaction at a rate constant d.

To derive the delay master equations describing the stochastic dynamics, we want to obtain the probability $P(\varphi, n, m, t+\Delta t)$, i.e., the probability of observing n nuclear mRNAs and m cytoplasmic mRNAs at time $t+\Delta t$ given that the gene state is φ. Note that the inactive and active gene states are denoted by $\varphi = 0$ and $\varphi = 1$, respectively. There are four sets of events which contribute to this probability in the time interval $(t, t+\Delta t)$:

(A) a nuclear mRNA transcript is produced while the gene is active;

(B) either the gene state changes or a cytoplasmic mRNA is degraded;

(C) a nuclear mRNA transcript is exported to the cytoplasm and becomes a cytoplasmic mRNA;

(D) no reaction occurs.

In the infinitesimal limit $\Delta t \to 0$, the probability of event (A) simply follows the conventional propensity associated with a firing of the transcription reaction:

$$P(1, n, m, t+\Delta t; 1, n-1, m, t) = \rho \Delta t P(1, n-1, m, t), \tag{S2}$$

where $P(\varphi, n, m, t; \varphi', n', m', t')$ is the joint probability of having gene state φ, n nuclear mRNAs and m cytoplasmic mRNAs at time t and gene state φ', n' nuclear mRNAs and m' cytoplasmic mRNAs at time t'.

Similarly, the joint probabilities of the reactions associated with event (B) are

$$\begin{aligned} P(0, n, m, t+\Delta t; 1, n, m, t) =& \sigma_{\text{off}} \Delta t P(1, n, m, t), \\ P(1, n, m, t+\Delta t; 0, n, m, t) =& \sigma_{\text{on}} \Delta t P(0, n, m, t), \\ P(\varphi, n, m, t+\Delta t; \varphi, n, m+1, t) =& d(m+1) P(\varphi, n, m+1, t). \end{aligned} \tag{S3}$$

The probability of event (C) occurring requires a careful consideration of the history of the process: (i) $(1, n', m', t) \to (1, n'+1, m', t-\tau+\Delta t)$; (ii) $(1, n'+1, m', t-\tau+\Delta t) \to (\varphi, n+1, m-1, t)$; (iii) $(\varphi, n+1, m-1, t) \to (\varphi, n, m, t+\Delta t)$, where n' and m' are the number of nuclear and cytoplasmic RNAs at time $t-\tau$. The elementary probability of event (C) can hence be decomposed into

$$\begin{aligned} & P(\varphi, n, m, t+\Delta t; \varphi, n+1, m-1, t; 1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) \\ =& P(1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) && \text{(Part I)} \\ & \times P(\varphi, n+1, m-1, t | 1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) && \text{(Part II)} \\ & \times P(\varphi, n, m, t+\Delta t | \varphi, n+1, m-1, t; 1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) && \text{(Part III)} \end{aligned} \tag{S4}$$

The probability of Part III is 1 since every production event is followed by an export event a time τ later, i.e.,

$$P(\text{Part III}) = 1. \tag{S5}$$

According to Eq. (S2), the probability of Part I is

$$P(\text{Part I}) = \rho\Delta t P(1, n', m', t-\tau). \tag{S6}$$

The two "+1"s of nuclear mRNAs in the probability of Part II represent a nuclear mRNA was born in time $(t-\tau, t-\tau+\Delta t)$ and stays in the system till time t, and notably does not participate in any reaction in time $(t-\tau+\Delta t, t)$. Thus, we have

$$P(\varphi, n+1, m-1, t|1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) = P(\varphi, n, m-1, t|1, n', m', t-\tau+\Delta t). \tag{S7}$$

Since all nuclear mRNAs born before time $t-\tau$ (of which the number is n') are exported from nucleus to cytoplasm, the number of nuclear mRNAs (n) at time t is independent of n' and m', but only depends on the gene state at time $t-\tau$. Besides, the number of cytoplasmic mRNAs ($m-1$) at time t is determined by the number of nuclear mRNAs out of n' exported to cytoplasm and the number of cytoplasm RNA out of m' degraded in the time interval $(t-\tau+\Delta t, t)$. As such, we have

$$P(\varphi, n, m-1, t|1, n', m', t-\tau+\Delta t) = P(\varphi, n, \bullet, t|1, 0, 0, t-\tau+\Delta t)P(\bullet, \bullet, m-1, t|1, n', m', t-\tau+\Delta t), \tag{S8}$$

where the symbol $\bullet$ stands for the marginalization over the variable in the same position. For instance,

$$P(\varphi, n, \bullet, t|1, 0, 0, t-\tau+\Delta t) = \sum_m P(\varphi, n, m, t|1, 0, 0, t-\tau+\Delta t). \tag{S9}$$

It follows that by means Eqs. (S4)-(S9) that the total probability of event (C) is given by

$$\begin{aligned}
&P(\varphi, n, m, t+\Delta t; \varphi, n+1, m-1, t) \\
=&\sum_{n',m'} P(\varphi, n, m, t+\Delta t; \varphi, n+1, m-1, t; 1, n'+1, m', t-\tau+\Delta t; 1, n', m', t-\tau) \\
=&\rho\Delta t P(\varphi, n, \bullet, t|1, 0, 0, t-\tau+\Delta t) \sum_{n',m'} P(\bullet, \bullet, m-1, t|1, n', m', t-\tau+\Delta t)P(1, n', m', t-\tau).
\end{aligned} \tag{S10}$$

Note that here we used the law of total probability. See Fig. S1 for an illustration of the history of event (C).

The probability that event (D) occurs is obtained by subtracting all the probabilities of one reaction occurring from $P(\varphi, n, m, t+\Delta t)$. Specifically, we have

$$\begin{aligned}
P(\varphi, n, m, t+\Delta t; \varphi, n, m, t) =& P(\varphi, n, m, t) - \delta_\varphi P(\varphi, n+1, m, t+\Delta t; \varphi, n, m, t) - P(1-\varphi, n, m, t+\Delta t; \varphi, n, m, t) \\
&- P(\varphi, n-1, m+1, t+\Delta t; \varphi, n, m, t) - P(\varphi, n, m+1, t+\Delta t; \varphi, n, m, t),
\end{aligned} \tag{S11}$$

where δ_φ is Kronecker delta, and $\delta_\varphi = 1$ if and only if $\varphi = 1$.

By means of Eqs. (S2), (S3), (S10) and (S11), it can be shown that

$$\begin{cases}
\begin{aligned}
P(0, n, m, t+\Delta t) =& \sigma_{\text{off}}\Delta t P(1, n, m, t) + d(m+1)\Delta t P(0, n, m+1, t) \\
&+ \rho\Delta t P(0, n, \bullet, t|1, 0, 0, t-\tau+\Delta t) \sum_{n',m'} P(\bullet, \bullet, m-1, t|1, n', m', t-\tau+\Delta t)P(1, n', m', t-\tau) \\
&+ [P(0, n, m, t) - \sigma_{\text{on}}\Delta t P(0, n, m, t) - dm\Delta t P(0, n, m, t) \\
&- \rho\Delta t P(0, n-1, \bullet, t|1, 0, 0, t-\tau+\Delta t) \sum_{n',m'} P(\bullet, \bullet, m, t|1, n', m', t-\tau+\Delta t)P(1, n', m', t-\tau)], \\
P(1, n, m, t+\Delta t) =& \sigma_{\text{on}}\Delta t P(0, n, m, t) + \rho\Delta t P(1, n-1, m, t) + d(m+1)\Delta t P(1, n, m+1, t) \\
&+ \rho\Delta t P(1, n, \bullet, t|1, 0, 0, t-\tau+\Delta t) \sum_{n',m'} P(\bullet, \bullet, m-1, t|1, n', m', t-\tau+\Delta t)P(1, n', m', t-\tau) \\
&+ [P(1, n, m, t) - \sigma_{\text{off}}\Delta t P(1, n, m, t) - \rho\Delta t P(1, n, m, t) - dm\Delta t P(1, n, m, t) \\
&- \rho\Delta t P(1, n-1, \bullet, t|1, 0, 0, t-\tau+\Delta t) \sum_{n',m'} P(\bullet, \bullet, m, t|1, n', m', t-\tau+\Delta t)P(1, n', m', t-\tau)].
\end{aligned}
\end{cases} \tag{S12}$$

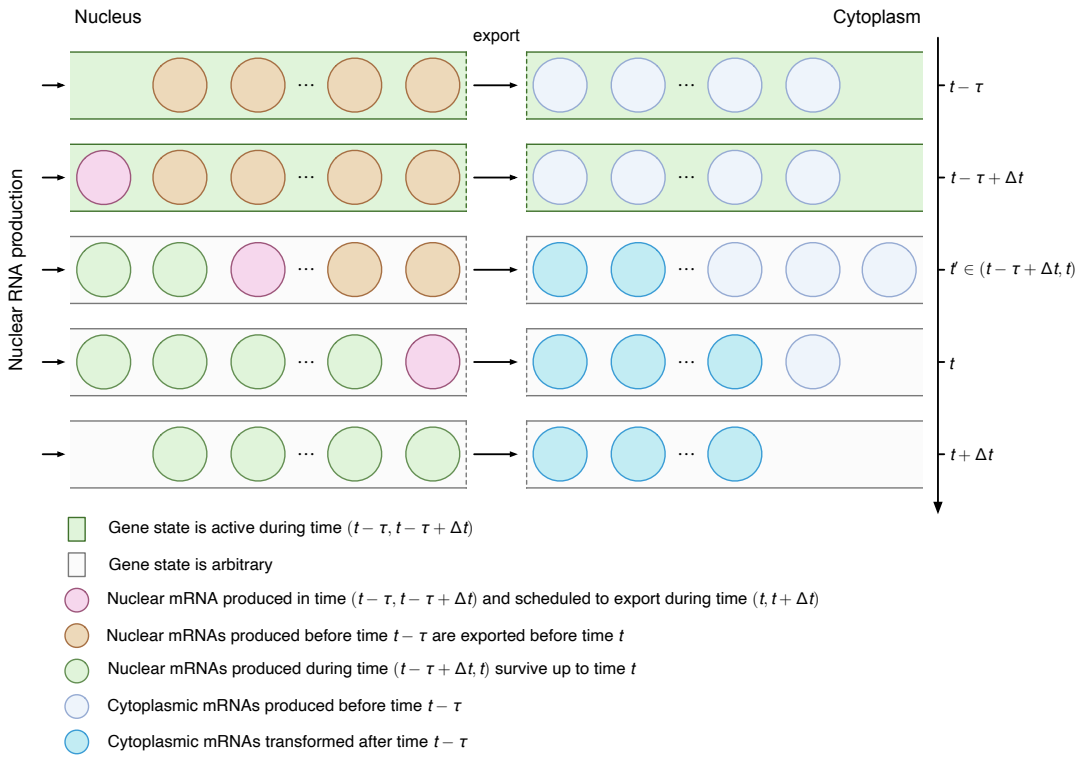

FIG. S1. Illustration of the chain of events culminating in the export of a nuclear mRNA in Model I (refer to event (C)). Let's assume there are n' nuclear mRNAs (brown balls) and m' cytoplasmic mRNAs (purple balls) produced before time $t-\tau$. In the next time step $t-\tau+\Delta t$, there is a newborn nuclear mRNA (red ball). During $(t-\tau+\Delta t, t)$, the first n' brown mRNAs are progressively exported and turned into cytoplasmic mRNAs (blue balls) and in the meanwhile n new nuclear mRNAs (green balls) are produced. Since each nuclear RNA stays for a fixed time τ, the red RNA definitely leaves in the time interval $(t, t+\Delta t)$. Note that during the time interval $(t-\tau+\Delta t, t)$, the red RNA does not participate in any dynamics of the system which explains Eq. (S7) and the dynamics of the green mRNAs follows Eq. (S14) which is a simple telegraph model without decay from $(0, \tau)$.

Dividing both sides of Eq. (S12) by Δt and letting $\Delta t \to 0$, the delay master equation of the delayed RNA export model is obtained:

$$\begin{cases} \dfrac{\mathrm{d}P(0,n,m,t)}{\mathrm{d}t} = \sigma_{\mathrm{off}}P(1,n,m,t) - \sigma_{\mathrm{on}}P(0,n,m,t) + d(m+1)P(0,n,m+1,t) - dmP(0,n,m,t) \\ \qquad + \rho[P(0,n,\bullet,t|1,0,0,t-\tau)P(\bullet,\bullet,m-1,t;1,\bullet,\bullet,t-\tau) - P(0,n-1,\bullet,t|1,0,0,t-\tau)P(\bullet,\bullet,m,t;1,\bullet,\bullet,t-\tau)], \\ \dfrac{\mathrm{d}P(1,n,m,t)}{\mathrm{d}t} = \sigma_{\mathrm{on}}P(0,n,m,t) - \sigma_{\mathrm{off}}P(1,n,m,t) + d(m+1)P(1,n,m+1,t) - dmP(1,n,m,t) + \rho P(1,n-1,m,t) - \rho P(1,n,m,t) \\ \qquad + \rho[P(1,n,\bullet,t|1,0,0,t-\tau)P(\bullet,\bullet,m-1,t;1,\bullet,\bullet,t-\tau) - P(1,n-1,\bullet,t|1,0,0,t-\tau)P(\bullet,\bullet,m,t;1,\bullet,\bullet,t-\tau)], \end{cases} \tag{S13}$$

in which the following simplification step is used

$$\sum_{n',m'} P(\bullet,\bullet,m-1,t|1,n',m',t-\tau+\Delta t)P(1,n',m',t-\tau) = P(\bullet,\bullet,m-1,t;1,\bullet,\bullet,t-\tau).$$

Next we seek solutions to the two probabilities, i.e., $P(\varphi,n,\bullet,t|1,0,0,t-\tau)$ and $P(\bullet,\bullet,m-1,t;1,\bullet,\bullet,t-\tau)$. The n nuclear mRNAs of the two conditional probabilities $P(\varphi,n,\bullet,t|1,0,0,t-\tau)$ for $\varphi = 0,1$ in Eq. (S13) are produced during time $(t-\tau,t)$, and the pertinent dynamics are that of a reaction system composed of the first three reactions in Eq. (S1), i.e., $G \rightleftharpoons G^*$ and $G \to G+N$. Specifically, we have

$$\boxed{\begin{cases} \dfrac{\mathrm{d}\bar{P}_0(n,t)}{\mathrm{d}t} = -\sigma_{\mathrm{on}}\bar{P}_0(n,t) + \sigma_{\mathrm{off}}\bar{P}_1(n,t), \\ \dfrac{\mathrm{d}\bar{P}_1(n,t)}{\mathrm{d}t} = \sigma_{\mathrm{on}}\bar{P}_0(n,t) - \sigma_{\mathrm{off}}\bar{P}_1(n,t) + \rho[\bar{P}_1(n-1,t) - \bar{P}_1(n,t)], \end{cases}} \tag{S14}$$

initiated at $\bar{P}_0(n,0)=0$ for any n, $\bar{P}_1(n,0)=1$ for $n=0$ and equal to 0 otherwise, as well as $P(\varphi,n,\bullet,t|1,0,0,t-\tau)=\bar{P}_\varphi(n,\tau)$ for $\varphi=0,1$.

By summing over all possible n on both sides of Eq. (S13), we obtain

$$\begin{cases} \dfrac{\mathrm{d}P(0,m,t)}{\mathrm{d}t} =\sigma_{\mathrm{off}}P(1,m,t)-\sigma_{\mathrm{on}}P(0,m,t)+d(m+1)P(0,m+1,t)-dmP(0,m,t) \\ \qquad\qquad +\rho[P(m-1,t;1,\bullet,\bullet,t-\tau)-P(m,t;1,\bullet,\bullet,t-\tau)]\times P(0,\bullet,\bullet,t|1,0,0,t-\tau), \\ \dfrac{\mathrm{d}P(1,m,t)}{\mathrm{d}t} =\sigma_{\mathrm{on}}P(0,m,t)-\sigma_{\mathrm{off}}P(1,m,t)+d(m+1)P(1,m+1,t)-dmP(1,m,t) \\ \qquad\qquad +\rho[P(m-1,t;1,\bullet,\bullet,t-\tau)-P(m,t;1,\bullet,\bullet,t-\tau)]\times P(1,\bullet,\bullet,t|1,0,0,t-\tau). \end{cases} \tag{S15}$$

Summing up the two equations in Eq. (S15) leads to

$$\frac{\mathrm{d}P(m,t)}{\mathrm{d}t}=d(m+1)P(m+1,t)-dmP(m,t)+\rho[P(m-1,t;1,\bullet,\bullet,t-\tau)-P(m,t;1,\bullet,\bullet,t-\tau)], \tag{S16}$$

where $P(m,t)=\sum_\varphi P_\varphi(m,t)$. Given the nature of the delayed export, the dynamics of cytoplasmic mRNAs is equivalent to that of a conventional telegraph model a τ time earlier. The telegraph model encompasses the following set of reactions

$$G \underset{\sigma_{\mathrm{on}}}{\overset{\sigma_{\mathrm{off}}}{\rightleftharpoons}} G^*, \qquad G \xrightarrow{\rho} G+M, \qquad M \xrightarrow{d} \varnothing,$$

of which the governing master equations are

$$\boxed{\begin{cases} \dfrac{\mathrm{d}Q_0(m,t)}{\mathrm{d}t}=\sigma_{\mathrm{off}}Q_1(m,t)-\sigma_{\mathrm{on}}Q_0(m,t)+d(m+1)Q_0(m+1,t)-dmQ_0(m,t), \\ \dfrac{\mathrm{d}Q_1(m,t)}{\mathrm{d}t}=-\sigma_{\mathrm{off}}Q_1(m,t)+\sigma_{\mathrm{on}}Q_0(m,t)+d(m+1)Q_0(m+1,t)-dmQ_0(m,t)+\rho[Q_1(m-1,t)-Q_1(m,t)]. \end{cases}} \tag{S17}$$

Also by summing up the two equations in Eq. (S17), we have that

$$\frac{\mathrm{d}Q(m,t)}{\mathrm{d}t}=d(m+1)Q(m+1,t)-dmQ(m,t)+\rho[Q_1(m-1,t)-Q_1(m,t)], \tag{S18}$$

where $Q(m,t)=\sum_\varphi Q_\varphi(m,t)$. Noted that $P(m,t)=Q(m,t-\tau)$, and by comparing Eqs. (S16) and (S18), we conclude that for $t>\tau$

$$P(m,t;1,\bullet,\bullet,t-\tau)=Q_1(m,t-\tau).$$

By replacing the two conditional probabilities in Eq. (S13) with $\bar{P}_\varphi(n,\tau)$ and $Q_1(m,t-\tau)$, Eq. (S13) becomes

$$\boxed{\begin{cases} \dfrac{\mathrm{d}P(0,n,m,t)}{\mathrm{d}t} =\sigma_{\mathrm{off}}P(1,n,m,t)-\sigma_{\mathrm{on}}P(0,n,m,t)+d(m+1)P(0,n,m+1,t)-dmP(0,n,m,t) \\ \qquad\qquad +\rho\mathbf{1}_{t-\tau}[\bar{P}_0(n,\tau)Q_1(m-1,t-\tau)-\bar{P}_0(n-1,\tau)Q_1(m,t-\tau)], \\ \dfrac{\mathrm{d}P(1,n,m,t)}{\mathrm{d}t} =\sigma_{\mathrm{on}}P(0,n,m,t)-\sigma_{\mathrm{off}}P(1,n,m,t)+d(m+1)P(1,n,m+1,t)-dmP(1,n,m,t)+\rho P(1,n-1,m,t)-\rho P(1,n,m,t) \\ \qquad\qquad +\rho\mathbf{1}_{t-\tau}[\bar{P}_1(n,\tau)Q_1(m-1,t-\tau)-\bar{P}_1(n-1,\tau)Q_1(m,t-\tau)], \end{cases}} \tag{S19}$$

which together with Eqs. (S14) and (S17) fully characterizes the stochastic dynamics of the delayed export model in Eq. (S1). Note that in Eq. (S19), $\mathbf{1}_{t-\tau}=1$ if and only if $t>\tau$.

II. Solution of the delay master equations

In this section, we solve Eq. (S19) in steady state conditions by means of the generating function method. By defining generating functions $G_\varphi(u_1,u_2,t)=\sum_{n,m}(u_1+1)^n(u_2+1)^mP(\varphi,n,m,t)$, $G_{N\varphi}(u_1,t)=\sum_n(u_1+1)^n\bar{P}_\varphi(n,t)$

and $G_{M\varphi}(u_2,t)=\sum_m (u_2+1)^m Q_\varphi(m,t)$, Eqs. (S14), (S17) and (S19) are transformed into

$$\begin{cases}\partial_t G_0 =\sigma_{\text{off}} G_1-\sigma_{\text{on}} G_0-du_2\partial_{u_2}G_0+\rho(u_2-u_1)G^\tau_{N0}G_{M1},\\ \partial_t G_1 =-\sigma_{\text{off}} G_1+\sigma_{\text{on}} G_0-du_2\partial_{u_2}G_1+\rho u_1 G_1+\rho(u_2-u_1)G^\tau_{N1}G_{M1},\end{cases} \tag{S20}$$

$$\begin{cases}\partial_t G_{N0} =\sigma_{\text{off}} G_{N1}-\sigma_{\text{on}} G_{N0},\\ \partial_t G_{N1} =-\sigma_{\text{off}} G_{N1}+\sigma_{\text{on}} G_{N0}+\rho u_1 G_{N1},\end{cases} \tag{S21}$$

and

$$\begin{cases}\partial_t G_{M0} =\sigma_{\text{off}} G_{M1}-\sigma_{\text{on}} G_{M0}+du_2\partial_{u_2}G_{M0},\\ \partial_t G_{M1} =-\sigma_{\text{off}} G_{M1}+\sigma_{\text{on}} G_{M0}+du_2\partial_{u_2}G_{M1}+\rho u_2 G_{M1},\end{cases} \tag{S22}$$

where the arguments u_1, u_2 and t are suppressed for brevity, and G^τ_{N0} is short for $G_{N0}(u_1,\tau)$. Prior to solving Eq. (S20), we need to find the solutions to Eqs. (S21) and (S22).

The exact solution to Eq. (S22) has been reported in Ref. [36]. At steady state, Eq. (S22) admits the solution

$$G_{M0}=\frac{\sigma_{\text{off}}}{\sigma_{\text{off}}+\sigma_{\text{on}}}\mathbf{M}_{-1,0},\qquad G_{M1}=\frac{\sigma_{\text{on}}}{\sigma_{\text{off}}+\sigma_{\text{on}}}\mathbf{M}_{0,0}, \tag{S23}$$

where $x_2=\rho u_2/d$ and the Kummer function

$$\mathbf{M}_{i,j}:=M\left(1+\frac{\sigma_{\text{on}}}{d}+i,1+\frac{\sigma_{\text{off}}+\sigma_{\text{on}}}{d}+j,x_2\right).$$

Note that Eq. (S21) consists of two ordinary differential equations (ODEs). We solve G_{N1} as a function of G_{N0} from the first equation in Eq. (S21) and insert the solution into the second one to obtain a second-order ODE. Using initial conditions $G_{N0}=0$ and $G_{N1}=1$ at time $t=0$ we then solve the second-order ODE at time τ to obtain

$$G^\tau_{N0}=\frac{\sigma_{\text{off}}}{d\theta}\exp\left[-\tau d\left(\frac{r-\theta-1}{2}\right)\right]-\frac{\sigma_{\text{off}}}{d\theta}\exp\left[-\tau d\left(\frac{r+\theta-1}{2}\right)\right], \tag{S24}$$

with $x_1=\rho u_1/d$, $\alpha=\sigma_{\text{on}}/(\sigma_{\text{off}}+\sigma_{\text{on}})$, $r=1+(\sigma_{\text{off}}+\sigma_{\text{on}})/d-x_1$ and

$$\theta=\sqrt{\left(\frac{\sigma_{\text{off}}+\sigma_{\text{on}}}{d}-x_1\right)^2+\frac{4\sigma_{\text{on}}x_1}{d}}.$$

The probability associated with the active gene state immediately follows

$$G^\tau_{N1}=\left(\frac{r+\theta-1}{2\theta}-\frac{\sigma_{\text{on}}}{d\theta}\right)\exp\left[-\tau d\left(\frac{r+\theta-1}{2}\right)\right]-\left(\frac{r-\theta-1}{2\theta}-\frac{\sigma_{\text{on}}}{d\theta}\right)\exp\left[-\tau d\left(\frac{r-\theta-1}{2}\right)\right]. \tag{S25}$$

Letting $\partial_t G_0=\partial_t G_1=0$, we solve G_1 as a function of G_0 from the first equation in Eq. (S20), and have that

$$G_1=[\sigma_{\text{on}}G_0+du_2\partial_{u_2}G_0-\rho(u_2-u_1)G^\tau_{N0}G_{m1}]/\sigma_{\text{off}}. \tag{S26}$$

Summing up the two equations in Eq. (S20) leads to

$$du_2\partial_{u_2}(G_0+G_1)=\rho u_1 G_1+\rho(u_2-u_1)(G^\tau_{N0}+G^\tau_{N1})G_{M1}. \tag{S27}$$

Inserting Eqs. (S23) and (S26) into Eq. (S27), we then get an equation that only depends on G_0:

$$\begin{aligned}&u_2^2\partial_{u_2}^2G_0+\left(1+\frac{\sigma_{\text{off}}+\sigma_{\text{on}}}{d}-x_1\right)\partial_{u_2}G_0-\frac{\sigma_{\text{on}}}{d}x_1G_0\\ =&\underbrace{(u_2-u_1)x_2G^\tau_{N0}\partial_{u_2}G_{M1}+x_2G^\tau_{N0}G_{M1}+(x_2-x_1)(\sigma_{\text{off}}/d-x_1)G^\tau_{N0}G_{M1}+(x_2-x_1)\sigma_{\text{off}}G^\tau_{N1}G_{M1}/d}_{\mathcal{A}}.\end{aligned} \tag{S28}$$

By means of Eqs. (13.3.3), (13.3.4) and (13.3.15) in [49], the term $\mathcal{A}$ simplifies to

$$\begin{aligned}\mathcal{A} =& \frac{\alpha(1+\sigma_{\text{on}}/d)(x_2-x_1)x_2}{1+(\sigma_{\text{off}}+\sigma_{\text{on}})/d} G^{\tau}_{N0}\mathbf{M}_{1,1} + \alpha x_2 G^{\tau}_{N0}\mathbf{M}_{0,0} + \alpha(x_2-x_1)(\sigma_{\text{off}}/d - x_1)G^{\tau}_{N0}\mathbf{M}_{0,0} + \alpha(x_2-x_1)\sigma_{\text{off}}G^{\tau}_{N1}\mathbf{M}_{0,0}/d \\ =& \alpha(x_2-x_1)(1+\sigma_{\text{on}}/d)G^{\tau}_{N0}\mathbf{M}_{1,0} + \alpha(x_2-x_1)\left[\left(\frac{\sigma_{\text{off}}-\sigma_{\text{on}}}{d} - x_1\right)G^{\tau}_{N0} + \frac{\sigma_{\text{off}}}{d}G^{\tau}_{N1}\right]\mathbf{M}_{0,0} + \alpha x_1 G^{\tau}_{N0}\mathbf{M}_{0,0}.\end{aligned} \tag{S29}$$

It can be ascertained that Eq. (S28) is a linear inhomogeneous second-order ODE. Therefore, the solution to Eq. (S28) is a combination of solutions to the general ODEs of the following form

$$x^2 f'' + Axf' + Bf = Hx^i M(a,b,cx),$$

where A, B, H, a, b, c are constants and $i = 0, 1$. The general solution is given by

$$f = c_1 x^{\frac{1}{2}(1-A-\lambda)} + c_2 x^{\frac{1}{2}(1-A+\lambda)} + \frac{Hx^i}{(i+\ell_-)\lambda}{}_2F_2\left(a, i+\ell_-; b, 1+i+\ell_-; cx\right) - \frac{Hx^i}{(i+\ell_+)\lambda}{}_2F_2\left(a, i+\ell_+; b, 1+i+\ell_+; cx\right)$$

with $\lambda = \sqrt{1-2A+A^2-4B}$, $\ell_- = (A-1-\lambda)/2$ and $\ell_+ = (A-1+\lambda)/2$. By defining the hypergeometric function

$$\mathbf{F}^{\pm}_{i,j,k,l} := {}_2F_2\left(1+i+\frac{\sigma_{\text{on}}}{d}, \frac{1+r\pm\theta}{2}+j; 1+k+\frac{\sigma_{\text{on}}+\sigma_{\text{off}}}{d}, \frac{3+r\pm\theta}{2}+l; x_2\right),$$

the general solution to Eq. (S28) reads

$$\begin{aligned}G_0 =& k_A x_2\left(\frac{2}{1+r-\theta}\mathbf{F}^{-}_{1,0,0,0} - \frac{2}{1+r+\theta}\mathbf{F}^{+}_{1,0,0,0}\right) + k_B x_2\left(\frac{2}{1+r-\theta}\mathbf{F}^{-}_{0,0,0,0} - \frac{2}{1+r+\theta}\mathbf{F}^{+}_{0,0,0,0}\right) \\ &- k_A x_1\left(\frac{2}{r-\theta-1}\mathbf{F}^{-}_{1,-1,0,-1} - \frac{2}{r+\theta-1}\mathbf{F}^{+}_{1,-1,0,-1}\right) - k_C x_1\left(\frac{2}{r-\theta-1}\mathbf{F}^{-}_{0,-1,0,-1} - \frac{2}{r+\theta-1}\mathbf{F}^{+}_{0,-1,0,-1}\right) + c_1 u_2^{\frac{1-r-\theta}{2}} + c_2 u_2^{\frac{1-r+\theta}{2}}\end{aligned} \tag{S30}$$

where

$$k_A = \frac{\alpha(d+\sigma_{\text{on}})G^{\tau}_{N0}}{d\theta}, \qquad k_B = \frac{\alpha[(\sigma_{\text{off}}-\sigma_{\text{on}}-dx_1)G^{\tau}_{N0} + \sigma_{\text{off}}G^{\tau}_{N1}]}{d\theta}, \qquad k_C = k_B - \alpha G^{\tau}_{N0}/\theta,$$

and constants c_1 and c_2 remain to be determined.

Since $1-r-\theta < 0$ and $u_2 \in [-1, 0]$, the term $u_2^{\frac{1-r-\theta}{2}}$ becomes ill-defined when u_2 approaches 0 from the negative real axis, thereby indicating $c_1 = 0$. Additionally, $(1-r+\theta)/2$ can hardly be an integer, and the generating function is a polynomial of only integer powers, thereby suggesting $c_2 = 0$.

By plugging Eq. (S30) into Eq. (S26), we obtain an equation for G_1 of the form

$$\begin{aligned}G_1 =& k_A k_1^- x_2\mathbf{F}^{-}_{1,0,0,0} + k_A k_{23}^- x_2^2\mathbf{F}^{-}_{2,1,1,1} - k_A k_1^+ x_2\mathbf{F}^{+}_{1,0,0,0} - k_A k_{23}^+ x_2^2\mathbf{F}^{+}_{2,1,1,1} + k_B k_1^- x_2\mathbf{F}^{-}_{0,0,0,0} + k_B k_{13}^- x_2^2\mathbf{F}^{-}_{1,1,1,1} - k_B k_1^+ x_2\mathbf{F}^{+}_{0,0,0,0} \\ &- k_B k_{13}^+ x_2^2\mathbf{F}^{+}_{1,1,1,1} + k_A k_2^-\mathbf{F}^{-}_{1,-1,0,-1} - k_A k_{21}^- x_1 x_2\mathbf{F}^{-}_{2,0,1,0} + k_A k_2^+\mathbf{F}^{+}_{1,-1,0,-1} + k_A k_{21}^+ x_2\mathbf{F}^{+}_{2,0,1,0} + k_C k_2^-\mathbf{F}^{-}_{0,-1,0,-1} - k_C k_{11}^- x_2\mathbf{F}^{-}_{1,0,1,0} \\ &+ k_C k_2^+\mathbf{F}^{+}_{0,-1,0.-1} + k_C k_{11}^+ x_1 x_2\mathbf{F}^{+}_{1,0,1,0} - \alpha d(x_2-x_1)G^{\tau}_{N0}\mathbf{M}_{0,0}/\sigma_{\text{off}},\end{aligned} \tag{S31}$$

with

$$k_1^{\pm} = \frac{2(d+\sigma_{\text{on}})}{\theta\sigma_{\text{off}}(1+r\pm\theta)}, \qquad k_2^{\pm} = \frac{2\sigma_{\text{on}}x_1}{\sigma_{\text{off}}(r\pm\theta-1)}, \qquad k_{ij}^{\pm} = \frac{2d(id+\sigma_{\text{on}})}{\theta\sigma_{\text{off}}(d+\sigma_{\text{on}}+\sigma_{\text{off}})(j+r\pm\theta)}.$$

III. Parameter inference using the generating function solution

Consider a sample of n_c cells in steady-state conditions. Specifically, in each cell i, we measure its nuclear mRNA counts (n_i), cytoplasmic mRNA counts (m_i) and cellular volume (V_i). The empirical distribution for the cell population is

$$\hat{P}(n,m) = \frac{1}{n_c}\sum_{i=1}^{n_c}\mathbf{1}_{n=n_i, m=m_i}.$$

where $\mathbf{1}_{n=n_i,m=m_i} = 1$ if and only if $n = n_i$ and $m = m_i$. The empirical generating function for the cell population is then computed as

$$\bar{G}(z_1, z_2) = \sum_{n,m} \hat{P}(n,m) z_1^n z_2^m = \sum_{i=1}^{n_c} \frac{1}{n_c} z_1^{n_i} z_2^{m_i}, \tag{S32}$$

where $z_1 = u_1 + 1$ and $z_2 = u_2 + 1$. The normalized volume for cell i is given by

$$\beta_i = \frac{V_i}{\sum_i V_i / n_c}. \tag{S33}$$

Because of extrinsic noise due to the coupling of the transcription rate to cell volume, a common property of many eukaryotic cells [42, 43], the model in Eq. (S1) is further modified to

$$G \underset{\sigma_{\text{on}}}{\overset{\sigma_{\text{off}}}{\rightleftharpoons}} G^*, \qquad G \xrightarrow{\bar{\rho}\beta} G + N, \qquad N \overset{\tau}{\Rightarrow} M, \qquad M \xrightarrow{d} \varnothing,$$

where β is a random variable sampled from an unknown distribution $p(\beta)$, representing the transcriptional extrinsic noise. The unknown distribution is estimated by using the kernel density estimation method [47] and the estimate is denoted as $\hat{p}(\beta)$. The kinetic parameters $\phi = \{\bar{\rho}, \sigma_{\text{off}}, \sigma_{\text{on}}, \tau\}$ are the ones which we will infer. The joint distribution of nuclear and cytoplasmic mRNA counts (n, m) for cell i is given by the model is $P(n, m, \phi, \beta_i)$, and the corresponding generating function for a single copy gene in cell i is $G(z_1, z_2, \phi, \beta_i)$. Assuming each cell has two independent copies of each gene, the generating function for the cell population reads

$$G(z_1, z_2, \phi) = \int_0^\infty G^2(z_1, z_2, \phi, \beta) \hat{p}(\beta) \mathrm{d}\beta. \tag{S34}$$

To reduce computation effort, we utilize the Gauss quadrature method and transform the integral variable, thereby Eq. (S34) becomes

$$G(z_1, z_2, \phi) = \frac{\beta_{\max} - \beta_{\min}}{2} \int_{-1}^{1} G^2(z_1, z_2, \phi, \beta_x) p(\beta_x) \mathrm{d}x, \tag{S35}$$

where

$$\beta_x = \frac{\beta_{\max} - \beta_{\min}}{2} x + \frac{\beta_{\max} + \beta_{\min}}{2}.$$

Here $\beta_{\max} = \max_i \beta_i$, $\beta_{\min} = \min_i \beta_i$ and x is the new integral variable. The numerical approximation to Eq. (S35) is then

$$G(z_1, z_2, \phi) \approx \frac{\beta_{\max} - \beta_{\min}}{2} \sum_{j=1}^{N_\beta} w_j G^2(z_1, z_2, \theta, \beta_{x_j}) p(\beta_{x_j}), \tag{S36}$$

where x_j is the j-th integral point of the N_β-th order Gauss quadrature approximation, and w_j is the corresponding integral weight, which is calculated based on Legendre polynomials and automatically computed by the Julia command `gausslegendre`.

To infer the kinetic parameters ϕ, we aim to minimize the loss function in the form of the L^2 norm

$$J_{\text{PGF}}(\phi) = \int_0^1 \int_0^1 \left| G(z_1, z_2, \phi) - \bar{G}(z_1, z_2) \right|^2 \mathrm{d}z_1 \mathrm{d}z_2. \tag{S37}$$

To reduce the computational load, we also apply the Gauss quadrature method to numerically compute integrals in Eq. (S37)

$$J_{\text{PGF}}(\phi) \approx \frac{1}{4} \sum_{j=1}^{N_y} \sum_{k=1}^{N_y} w_j w_k \left| G\left(\frac{y_{1,j}+1}{2}, \frac{y_{2,k}+1}{2}, \phi\right) - \bar{G}\left(\frac{y_{1,j}+1}{2}, \frac{y_{2,k}+1}{2}\right) \right|^2 \tag{S38}$$

where $y_{1,j} \in [-1, 1]$ and $y_{2,k} \in [-1, 1]$ are the j-th and k-th integral points of the N_y-th order Gauss quadrature approximation, and w_j, w_k are the corresponding integral weights given by the command `gausslegendre`.

Finally, we employ the Nelder-Mead algorithm for minimizing $J_{\mathrm{PGF}}(\varphi)$. This is because: (i) it is more robust against the selection of initial conditions, particularly compared to gradient-based optimization algorithms; (ii) the computation of the gradients $\partial_\phi J_{\mathrm{PGF}}(\phi)$ is rather time demanding. The optimization is implemented in the package `Optim.jl` in Julia.

The numerical recipe of the generating function based inference method is summarized in Algorithm 1.

Algorithm 1 Generating function based parameter inference

Input: Number of cells (n_c), the counts of nuclear and cytoplasmic mRNAs (n_i, m_i) and cellular size (V_i)
Output: Kinetic parameters $\phi = \{\bar{\rho}, \sigma_{\mathrm{off}}, \sigma_{\mathrm{on}}, \tau\}$.

```
Calculate the normalized cellular volume β_i by using Eq. (S33)
Use KDE to estimate p̂(β) according to the data {β_i}
Generate Gauss quadrature points and weights (x_j, w_j) for β by the command gausslegendre
Generate Gauss quadrature points and weights (y_{1,j}, y_{2,k}, w_j, w_k) for z_1, z_2 by the command gausslegendre
Compute the empirical generating function by using Eq. (S32)
Initialize the inferred parameters φ
while Threshold reached do
    Compute the generating function for the inferred parameters φ by using Eq. (S36)
    Compute the loss function J_PGF(φ) by using Eq. (S38)
    Update the inferred parameters φ by the Nelder-Mead optimization algorithm
end while
return Kinetic parameters φ = {ρ̄, σ_off, σ_on, τ}
```

IV. Parameter inference using the finite state projection method

In this section, we describe the technical details of the inference protocol based on the finite state projection (FSP). The Markovian export model consists of reactions

$$G \underset{\sigma_{\mathrm{on}}}{\overset{\sigma_{\mathrm{off}}}{\rightleftharpoons}} G^*, \qquad G \xrightarrow{\bar{\rho}\beta} G + N, \qquad N \xrightarrow{\gamma = 1/\tau} M, \qquad M \xrightarrow{d} \varnothing.$$

The corresponding master equation is

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} P_0(n,m,\beta,t) = \bar{\rho}\beta \left(\mathbb{E}^{(-1,0)} - 1\right) P_0(n,m,\beta,t) + \gamma \left(\mathbb{E}^{(-1,1)} - 1\right) n P_0(n,m,\beta,t) \\ \qquad\qquad + d \left(\mathbb{E}^{(0,1)} - 1\right) m P_0(n,m,\beta,t) + \sigma_{\mathrm{on}} P_1(n,m,\beta,t) - \sigma_{\mathrm{off}} P_0(n,m,\beta,t), \\ \dfrac{\mathrm{d}}{\mathrm{d}t} P_1(n,m,\beta,t) = \gamma \left(\mathbb{E}^{(-1,1)} - 1\right) n P_1(n,m,\beta,t) + d \left(\mathbb{E}^{(0,1)} - 1\right) m P_1(n,m,\beta,t) \\ \qquad\qquad - \sigma_{\mathrm{on}} P_1(n,m,\beta,t) + \sigma_{\mathrm{off}} P_0(n,m,\beta,t), \end{cases} \tag{S39}$$

for a single cell with a given β. The step operator $\mathbb{E}^{(i,j)}$ acts on a general function $f(n_1, n_2)$ as $\mathbb{E}^{(i,j)} f(n_1, n_2) = f(n_1 + i, n_2 + j)$ [35]. By imposing the stationary condition, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t} P_\varphi(n,m,\beta,t) = 0,$$

for $\varphi = 0, 1$ and setting $P_\varphi(n,m,\beta,t) = 0$ for either $n \geq N$ or $m \geq N$, where N is a relatively large number, the single-cell stationary distribution $P_\varphi(n,m,\beta|\phi)$ for given β and ϕ can be computed by solving the $2(N+1)^2$ algebraic equations of Eq. (S39) simultaneously. Using a similar approach, the stationary distribution for the cell population is computed using

$$P_\varphi(n,m|\phi) \approx \frac{\beta_{\max} - \beta_{\min}}{2} \sum_{j=1}^{N_\beta} w_j P^{*2}(n,m,\beta_{x_j}|\phi) p(\beta_{x_j}), \tag{S40}$$

for the same tuple (β_{x_j}, w_j) in Eq. (S36). The probability $P^{*2}(n,m,\beta_{x_j}|\phi)$ is the convolution of the probability $P(n,m,\beta_{x_j}|\phi)$.

Given the n_c tuples of counts $\mathcal{D} = \{(n_i, m_i)\}$, the total likelihood of all observations is the product over every tuple, i.e.,

$$\mathcal{L}(\mathcal{D}|\phi) = \prod_{i=1}^{n_c} P(n_i, m_i|\phi).$$

The inference of the kinetic parameters proceeds by minimizing the negative log likelihood, that is

$$J_{\text{FSP}}(\phi) = -\sum_{i=1}^{n_c} \log P(n_i, m_i|\phi). \tag{S41}$$

The numerical recipe for inference using FSP is summarized in Algorithm 2, and the optimization details are the same as those of the inference method using the generating function.

Algorithm 2 Finite state projection (FSP) based parameter inference

Input: Number of cells (n_c), the counts of nuclear and cytoplasmic RNAs (n_i, m_i) and cellular length (V_i)
Output: Kinetic parameters $\phi = \{\bar{\rho}, \sigma_{\text{off}}, \sigma_{\text{on}}, \tau\}$.

```
Calculate the normalized cellular volume β_i by using Eq. (S33)
Use KDE to estimate p̂(β) according to the data {β_i}
Generate Gauss quadrature points and weights (x_j, w_j) for β by the command gausslegendre
Initialize the inferred parameters φ
while Threshold reached do
    Compute the single-copy probability P(n, m, β_{x_j}|φ) for the inferred parameters φ and β_{x_j}
    Compute the single-cell probability P^{*2}(n, m, β_{x_j}|φ) by convolutions
    Compute the probability P(n, m, |φ) for the cell population by using Eq. (S40)
    Compute the loss function J_FSP(φ) by using Eq. (S41)
    Update the inferred parameters φ by the Nelder-Mead optimization algorithm
end while
return Kinetic parameters φ = {ρ̄, σ_off, σ_on, τ}
```

V. Estimate the range of export time

According to Eq. (7), the steady-state means of nuclear and cytoplasmic mRNA counts are

$$\langle n \rangle = \frac{\rho \sigma_{\text{on}} \tau}{\sigma_{\text{off}} + \sigma_{\text{on}}}, \qquad \langle m \rangle = \frac{\rho \sigma_{\text{on}}}{d(\sigma_{\text{on}} + \sigma_{\text{off}})},$$

which indicates that $\langle n \rangle / \langle m \rangle = \tau d$. To estimate τd, we extracted the averaged numbers of nuclear and cytoplasmic mRNAs in mouse MIN6 cells and Liver cells from Table S2 of Ref. [29]. By excluding genes of which either the mean number of nuclear or cytoplasmic mRNAs is zero, we found that approximately 80% genes are characterised by $\tau d < 1$ (see Fig. S2).

VI. Bootstrapping of experimental data

In the smFISH dataset [17], there are 503 and 537 mononucleated cells for which the numbers of nuclear and cytoplasmic mRNAs have been measured for the genes SPAC2H1001 and Rad21, respectively. The cell volume for each cell is also reported. Note that we specifically chose these genes from the dataset because their mRNA Fano factor (variance divided by the mean of mRNA number fluctuations) is greater than 1, a property of both the non-Markovian and Markovian models of gene expression considered in this paper. To quantify the uncertainty of parameter estimates obtained by the PGF based inference method, we utilize the method of bootstrapping. Specifically, given n_c tuples of nuclear and cytoplasmic mRNA counts (n_i, m_i), we produce a histogram of the n_c tuples, thus yielding an estimate of the joint distribution $\hat{P}(n, m)$. Then, another batch of n_c tuples are randomly sampled from the original dataset (allowing for repetition) and these are fed to the PGF based inference method to estimate the kinetic parameters. This step is repeated 10 times for each gene. We calculate the standard error of mean (SEM) for the 10 estimates of each kinetic parameter, which is reported by the error bar in Fig. 3d.

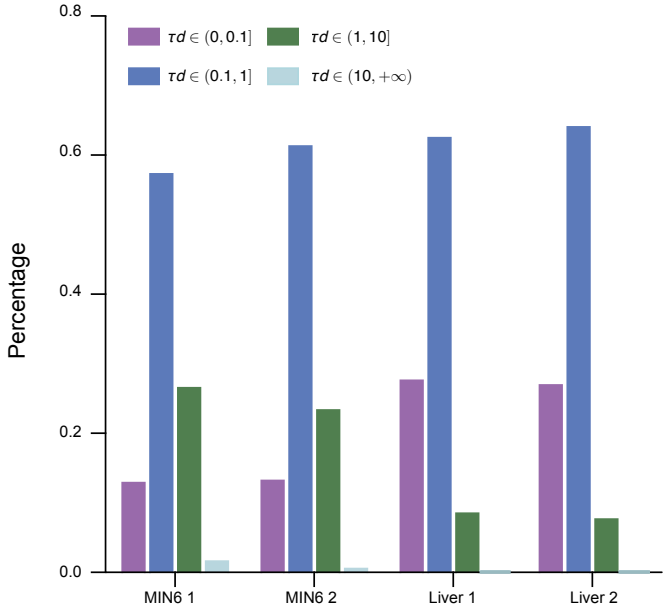

FIG. S2. Distributions of τd estimated from mRNA data measured in two samples of MIN6 cells and liver cells.

VII. Code and data availability

All codes and data have been deposited at https://github.com/edwardcao3026/pgf-inf.

TABLE S1. Inferred and true kinetic parameters from simulated data—data summarised in Fig. 3b. The true parameters are sampled from $\bar{\rho}/d \in [0, 80]$, $\sigma_{\text{on}}/d \in [0, 2]$ and $\sigma_{\text{off}}/d \in [0, 40]$, — these are roughly in the range estimated for various types of eukaryotic cells in Table I of Ref. [48]. The parameter $d\tau$ is sampled from $d\tau \in [0, 1]$ since about 80% of genes fall in this range according to Fig. S2.

	PGF				FSP				TRUE			
	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd
Set 1	24.29	13.17	0.25	0.57	18.36	8.69	0.23	0.58	25.01	11.87	0.22	0.57
Set 2	29.19	20.53	0.49	0.66	24.75	19.33	0.53	0.63	27.14	19.48	0.52	0.61
Set 3	32.28	16.63	0.40	0.73	25.91	14.09	0.42	0.67	27.14	15.04	0.43	0.61
Set 4	26.11	7.50	0.35	0.19	32.86	9.29	0.35	0.20	39.14	10.04	0.31	0.17
Set 5	56.19	8.08	0.07	0.45	36.64	7.05	0.08	0.34	40.62	8.84	0.09	0.32
Set 6	74.48	37.82	0.92	0.26	52.06	22.88	0.85	0.27	73.26	30.41	0.79	0.28
Set 7	23.23	9.37	0.56	0.64	23.94	11.58	0.62	0.62	26.67	12.13	0.61	0.58
Set 8	39.73	22.60	1.41	0.20	58.11	40.48	1.61	0.18	44.07	28.13	1.43	0.18
Set 9	3.90	0.11	0.38	0.60	3.79	0.08	0.32	0.59	4.06	0.16	0.42	0.61
Set 10	4.52	0.25	0.19	0.58	4.26	0.22	0.18	0.57	4.44	0.23	0.17	0.57

TABLE S2. Inferred and true kinetic parameters from simulated data—data summarised in Fig. 3c. The true parameters are sampled from $\bar{\rho}/d \in [0, 80]$, $\sigma_{\text{on}}/d \in [0, 2]$ and $\sigma_{\text{off}}/d \in [0, 40]$ — these are roughly in the range estimated for various types of eukaryotic cells in Table I of Ref. [48]. The parameter $d\tau$ is sampled from $d\tau \in [0, 1]$ since about 80% of genes fall in this range according to Fig. S2..

	PGF				FSP				TRUE			
	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd
Set 1	4.53	0.16	0.12	0.51	5.04	0.21	0.13	0.53	5.14	0.21	0.13	0.52
Set 2	8.18	0.01	0.03	0.28	8.57	0.03	0.06	0.28	8.51	0.04	0.09	0.28
Set 3	4.79	0.17	0.17	0.83	5.75	0.30	0.21	0.84	5.75	0.31	0.21	0.84
Set 4	7.42	0.05	0.13	0.74	7.78	0.06	0.15	0.72	7.74	0.06	0.14	0.73
Set 5	30.86	15.14	0.48	0.23	38.63	15.52	0.42	0.20	36.05	14.83	0.42	0.18
Set 6	11.11	0.02	0.11	0.87	11.78	0.04	0.14	0.86	11.78	0.04	0.13	0.85
Set 7	6.26	5.30	1.03	0.70	9.16	6.16	0.82	0.64	10.01	6.48	0.81	0.64
Set 8	26.90	11.47	0.18	0.28	42.99	9.89	0.13	0.29	33.26	8.14	0.14	0.27
Set 9	6.00	5.85	1.08	0.70	7.96	6.00	0.84	0.66	9.89	7.57	0.82	0.68
Set 10	4.38	0.34	0.11	0.58	5.18	0.54	0.14	0.55	4.86	0.51	0.13	0.54

TABLE S3. Inferred kinetic parameters based on experimental smFISH data. The ten sets are obtained by bootstrapping the original data and repeating inference for each. Results are summarised in Fig. 3d.

	SPAC2H1001				Rad21			
	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd	$\bar{\rho}/d$	σ_{off}/d	σ_{on}/d	τd
Experiment	2.74	0.41	0.34	0.29	2.10	0.87	0.47	0.20
Set 1	2.93	0.54	0.40	0.27	1.82	0.85	0.56	0.20
Set 2	2.95	0.44	0.35	0.27	2.33	1.19	0.55	0.21
Set 3	2.79	0.45	0.33	0.30	2.26	0.92	0.45	0.22
Set 4	2.62	0.36	0.30	0.30	2.11	0.78	0.42	0.21
Set 5	2.97	0.52	0.38	0.30	2.02	0.74	0.44	0.21
Set 6	2.20	0.21	0.23	0.33	2.01	0.95	0.52	0.20
Set 7	2.75	0.50	0.39	0.32	2.03	0.79	0.44	0.21
Set 8	2.88	0.38	0.31	0.28	2.34	1.29	0.62	0.20
Set 9	2.73	0.41	0.33	0.29	2.33	1.07	0.51	0.21
Set 10	2.68	0.42	0.36	0.30	2.41	1.30	0.56	0.20