Preprint

Exact solution of a three-stage model of stochastic gene expression model including cell-cycle dynamics

Yiling Wang,[1, a)] Zhenhua Yu,[1, a)] Ramon Grima,[2, b)] and Zhixing Cao[1, 3, c)]
[1)] *Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China*
[2)] *School of Biological Sciences, the University of Edinburgh, Max Born Crescent, Edinburgh, EH9 3BF, Scotland, United Kingdom*
[3)] *Department of Chemical Engineering, Queen's University, Kingston, Canada K7L 3N6*

(Dated: 29 August 2023)

The classical three-stage model of stochastic gene expression predicts the statistics of single cell mRNA and protein number fluctuations as a function of the rates of promoter switching, transcription, translation, degradation and dilution. While this model is easily simulated, its analytical solution remains an unsolved problem. Here we modify this model to explicitly include cell-cycle dynamics and then derive an exact solution for the time-dependent joint distribution of mRNA and protein numbers. We show large differences between this model and the classical model which captures cell-cycle effects implicitly via effective first-order dilution reactions. In particular we find that the Fano factor of protein numbers calculated from a population snapshot measurement are underestimated by the classical model whereas the correlation between mRNA and protein can be either over- or underestimated, depending on the timescales of mRNA degradation and promoter switching relative to the mean cell-cycle duration time.

I. INTRODUCTION

Over the past two decades, experiments have shown that gene expression is inherently noisy[1–6] manifesting most visibly in the large variability of mRNA and protein numbers between cells. Understanding the origin of this noise and how its magnitude can be tuned by changes to the rate parameter values are important questions that can be addressed experimentally but also using mathematical modelling. Mathematical models of stochastic gene expression can be grouped into three classes: (i) those predicting only the statistics of mRNA numbers; (ii) those predicting only the statistics of protein numbers; (iii) those predicting both types of statistics including the correlation between mRNA and proteins. Of course the latter type is the most realistic and useful, however the other two are much more commonly found in the literature because of their simplicity and their use in interpreting data of only one of these two gene products, e.g. single-cell RNA sequencing[7,8] and fluorescent labeling of proteins[9,10]. We next briefly review the large variety of models in each of these categories.

The simplest model of type (i) is the simple-birth death process whereby the birth reaction models transcription of mRNA and the death process models its subsequent decay via degradation and dilution. This is easily solved exactly and predicts a Poisson distribution in steady-state conditions[11]. However, a significant body of experiments suggests that the Fano factor (FF, the variance divided by the mean) for mRNA numbers is often substantially higher than that of a Poisson distribution[4,7,12]. To reconcile this discrepancy, a revision is necessary to the simple birth-death model which involves assuming that mRNA production occurs in bursts with a random size sampled from a geometric distribution; this can be solved exactly leading to a negative binomial distribution for mRNA numbers in steady-state conditions (which is well approximated by a gamma distribution when the mean mRNA numbers are not too small)[12–14]. However both of these models cannot explain the bimodality of distributions for mRNA numbers observed in some experiments[15]. A model that can explain bimodality and reduces to the birth-death or bursty birth-death models under certain conditions is the telegraph model[16,17]. This assumes that a promoter can switch between inactive and active states, and transcription occurs only from the latter; this model has been solved exactly in steady-state and in time, yielding an mRNA distribution in terms of hypergeometric functions. The commonest model of type (ii) is the bursty birth-death model which predicts steady-state negative binomial distributions for protein numbers (which is well approximated by a gamma distribution when the mean protein number is high)[18,19]. The main difference from the same model when used to predict mRNA distributions is the origin of bursting. For mRNA, bursts of expression occur when the promoter alternates between long transcriptional pauses and short intense periods of transcriptional activity. For proteins, bursts can arise from promoter switching and separately from the rapid translation of short-lived mRNA (transcriptional and translational bursting)[20]. Extensions of models of type (i) and (ii) to include an enhanced level of biological realism such as regular cell division events, gene dosage compensation, DNA replication, concentration homeostasis, cell-size homeostasis and cell-to-cell variation in parameters, have also been developed and solved exactly or approximately under appropriate assumptions[13,21–30].

[a)] These authors contributed equally.
[b)] Electronic mail: ramon.grima@ed.ac.uk
[c)] Electronic mail: z.cao@queensu.ca

The classical models of type (iii) are the two-stage and three-stage models of gene expression[20]. The two-stage model has a simple birth-death process for mRNA production and degradation augmented with two additional first-order reactions describing protein translation and degradation/dilution. The joint distribution of mRNA and protein numbers of this model has been solved exactly in steady-state using the generating function method[31]; the time-dependent joint distribution has been obtained by invoking the partitioning property of Poisson processes[32]. A more realistic model is the three-stage model[3] which is the two-stage model with the birth-death model for mRNA noise replaced by the telegraph model. This model accounts for promoter switching, transcription, degradation and translation, and hence can be seen as the simplest model to capture the essential processes involved in the central dogma. While the moments of the gene products can be easily obtained[33] (because the moment equations for any system of first-order reactions are closed) there are no exact analytical solutions for this model's marginal protein distributions and the joint distribution of mRNA and protein numbers (note that the marginal mRNA distribution is known and it is the same as that of the telegraph model). However even if these could be obtained, the issue remains that this model suffers from some obvious shortcomings, the most prominent of which is the fact that it does not explicitly model cell-cycle dynamics; for example it cannot capture the observation that typically in each cell, protein numbers increase between birth and cell division, followed by a sharp dip just after cell division[34]. This means that it is not clear how to relate the steady-state distribution predictions of the three-stage model to those from a population snapshot measurement, i.e. one that estimates the mRNA and protein in each cell at a point in time for a population of dividing cells with unsynchronized cell-cycles.

In this paper, we overcome these limitations. We extend the three-stage model to include a cell-cycle description and exactly solve for the time-dependent joint distribution of mRNA and protein numbers, thus providing a first dynamic stochastic description of both gene products within each cell-cycle and also across generations for two cases: a population of synchronized cells and the more typical population of unsynchronized cells where cell-age varies between cells. The population snapshot distribution of this model converges to a steady-state after a some generations; we numerically compare these to the steady-state solution from the classical three-stage model. In some regions of parameter space, we find large differences between the two models in their predictions of the protein Fano factor and of the correlation between mRNA and protein, thus highlighting the importance of an explicit cell-cycle description.

II. RESULTS

A. Models of stochastic gene expression

We consider three models of gene expression with increasing degrees of biological detail.

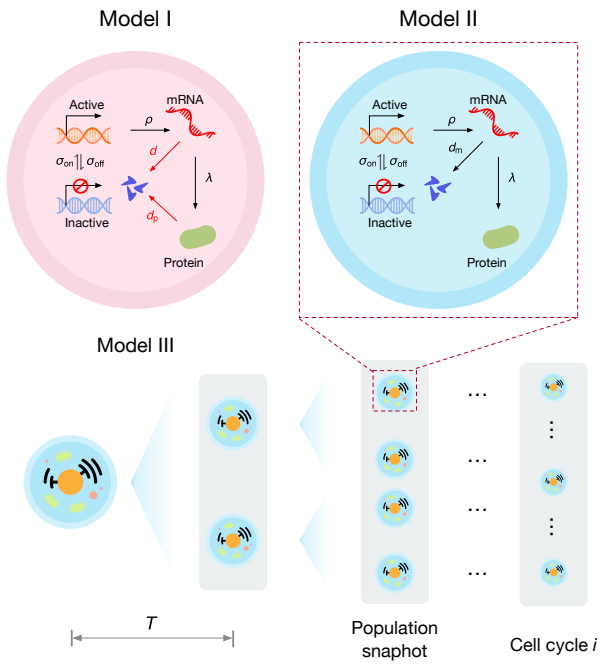


FIG. 1. Illustration of three stochastic gene expression models. Model I: the classical three-stage model of stochastic gene expression that provides an effective description of promoter switching between active and inactive states, transcription, translation, mRNA decay and protein dilution. Model II: same as Model I except it does not describe protein dilution. This is a more realistic description of time-dependent gene expression within a cell-cycle than Model I since there is no protein dilution prior to cell division. Model III: this extends Model II by assuming that that a cell-cycle is of length T and that at cell division the mother cell's mRNAs and proteins are randomly partitioned between its daughter cells. It provides a description of time-dependent gene expression within and across cell-cycles and is hence the most realistic model of the three.

Model I

The classical three-stage model of stochastic gene expression (illustrated in the top left corner of Fig. 1) is described by the following reaction scheme

$$D_0 \underset{\sigma_{\rm on}}{\overset{\sigma_{\rm off}}{\rightleftharpoons}} D_1, \quad D_0 \xrightarrow{\rho} D_0 + M, \quad M \xrightarrow{\lambda} M + P, \\ M \xrightarrow{d} \varnothing, \quad P \xrightarrow{d_{\rm p}} \varnothing. \tag{1}$$

Here a promoter switches between two states: an active state D_0 and an inactive state D_1. The transitions between the two states occur at rates $\sigma_{\rm off}$ and $\sigma_{\rm on}$. Transcription occurs only from the state D_0 leading to mRNA (M) production at a rate ρ. Proteins (P) are then translated from mRNA at a rate λ. mRNAs decay at a rate d (due to degradation and dilution), and protein dilution (due to cell division) takes place at a rate $d_{\rm p}$. Note that we cannot ignore active mRNA degradation because the lifetime of mRNA in mammalian cells is a fraction of the cell-cycle duration whereas for proteins the opposite

is true: their lifetime due to degradation often exceeds the cell-cycle duration and hence degradation can be ignored in our model. For e.g. in mouse fibroblasts, the median lifetime of mRNA is 9 hours, the median lifetime of proteins is 46 hours and the cell-cycle duration is 27.5 hours[35]. All reactions in Eq. (1) are first-order and to be understood as being effective, i.e. describing the overall effect of thousands of elementary reactions enabling transcription, translation and decay processes inside a cell. The chemical master equation of this model has not been exactly solved; in particular an inspection of the generating function equations shows that the problematic term is that associated with the effective dilution reaction $P \to \varnothing$.

Model II

In view of this issue we consider a simpler model (illustrated in the top right corner of Fig. 1) described by the following reaction scheme

$$D_0 \underset{\sigma_{\rm on}}{\overset{\sigma_{\rm off}}{\rightleftharpoons}} D_1, \quad D_0 \xrightarrow{\rho} D_0 + M, \quad M \xrightarrow{\lambda} M + P, \quad M \xrightarrow{d_{\rm m}} \varnothing.$$

This is the same as Model I but without the protein dilution reaction; as well the mRNA decay rate here means specifically that due to degradation (has no contribution from dilution due to cell division) and hence we label it as d_m instead of d. As we shall see in the next section this is more mathematically tractable than Model I; however also it is a more biologically realistic model of time-dependent gene expression *within a single cell-cycle* (from birth to cell division) since there is no dilution prior to cell division.

Model III

Of course, generally we are not only interested in intra cell-cycle gene expression but also across many cell-cycles. To describe this case we extend Model II by assuming that at the end of the cell-cycle, mRNA and protein molecules are randomly partitioned with probability 1/2 between two daughter cells. Specifically, just after cell division, the probability of a daughter cell having $n_{\rm m}$ mRNAs and $n_{\rm p}$ proteins, conditional on the mother cell having $n'_{\rm m}$ mRNAs and $n'_{\rm p}$ proteins, is given by

$$P(n_{\rm m}, n_{\rm p}|n'_{\rm m}, n'_{\rm p}) = \binom{n'_{\rm m}}{n_{\rm m}}\binom{n'_{\rm p}}{n_{\rm p}} 2^{-n'_{\rm m}-n'_{\rm p}}. \tag{2}$$

While being a simple model of the complex processes occurring at cell division, this is clearly far more realistic than the first-order protein dilution reaction in Model I since dilution now occurs at a point in time rather than continuously. Model III is illustrated at the bottom of Fig. 1.

The probabilistic description of Model II is given by the CME

$$\begin{cases} \dfrac{\rm d}{{\rm d}t}P_0(n_{\rm m}, n_{\rm p}, t) = \rho\left(\mathbb{E}^{(-1,0)} - 1\right) P_0(n_{\rm m}, n_{\rm p}, t) + \lambda\left(\mathbb{E}^{(0,-1)} - 1\right) n_{\rm m} P_0(n_{\rm m}, n_{\rm p}, t) \\ \qquad\qquad + d_{\rm m}\left(\mathbb{E}^{(1,0)} - 1\right) n_{\rm m} P_0(n_{\rm m}, n_{\rm p}, t) + \sigma_{\rm on} P_1(n_{\rm m}, n_{\rm p}, t) - \sigma_{\rm off} P_0(n_{\rm m}, n_{\rm p}, t), \\ \dfrac{\rm d}{{\rm d}t}P_1(n_{\rm m}, n_{\rm p}, t) = \lambda\left(\mathbb{E}^{(0,-1)} - 1\right) n_{\rm m} P_1(n_{\rm m}, n_{\rm p}, t) + d_{\rm m}\left(\mathbb{E}^{(1,0)} - 1\right) n_{\rm m} P_1(n_{\rm m}, n_{\rm p}, t) \\ \qquad\qquad - \sigma_{\rm on} P_1(n_{\rm m}, n_{\rm p}, t) + \sigma_{\rm off} P_0(n_{\rm m}, n_{\rm p}, t), \end{cases} \tag{3}$$

where $P_\varphi(n_{\rm m}, n_{\rm p}, t)$ is the probability of observing $n_{\rm m}$ mRNAs and $n_{\rm p}$ proteins in a cell when a promoter is active ($\varphi = 0$) or inactive ($\varphi = 1$). The step operator $\mathbb{E}^{(i,j)}$ acts on a general function $f(n_1, n_2)$ as $\mathbb{E}^{(i,j)} f(n_1, n_2) = f(n_1 + i, n_2 + j)$[36]. For compactness of notation, the time argument t is hereafter omitted. Next, we show how to solve Eq. (3) by means of the generating function method.

By defining the generating function $G_\varphi = \sum_{n_m, n_p} z_1^{n_{\rm m}} z_2^{n_{\rm p}} P_\varphi(n_{\rm m}, n_{\rm p})$, Eq. (3) can be recast as the following set of partial differential equations

$$\partial_t G_0 = \rho w_1 G_0 - d_{\rm m} w_1 \partial_{w_1} G_0 + \lambda(w_1 + 1) w_2 \partial_{w_1} G_0 - \sigma_{\rm off} G_0 + \sigma_{\rm on} G_1, \tag{4a}$$

$$\partial_t G_1 = -d_{\rm m} w_1 \partial_{w_1} G_1 + \lambda(w_1 + 1) w_2 \partial_{w_1} G_1 + \sigma_{\rm off} G_0 - \sigma_{\rm on} G_1, \tag{4b}$$

where $w_1 = z_1 - 1$ and $w_2 = z_2 - 1$. For ease of notation, the arguments of the generating function w_1 and w_2 are omitted. Once the generating function equations are solved, the conditional joint distributions $P_\varphi(n_{\rm m}, n_{\rm p})$ in Eq. (3) can be constructed using the relation

$$P_\varphi(n_{\rm m}, n_{\rm p}) = \frac{1}{n_{\rm m}! n_{\rm p}!} \frac{\partial^{n_{\rm m}}}{\partial w_1^{n_{\rm m}}} \frac{\partial^{n_{\rm p}}}{\partial w_2^{n_{\rm p}}} G_\varphi \bigg|_{w_1=-1, w_2=-1}. \tag{5}$$

We note that because the evaluation of high-order symbolic derivatives can be troublesome, a more practical approach is needed. We first symbolically expand the generating function as a Taylor series of w_1 using the Mathematica command `Series` and then perform a Taylor series expansion to the $n_{\rm m}^{\rm th}$ coefficient for w_2 using the Mathematica command `NSeries`. The probability $P_\varphi(n_{\rm m}, n_{\rm p})$ is then simply given by the $n_{\rm p}^{\rm th}$ coefficient. A

similar approach has previously been used for the rapid evaluation of marginal distributions from the generating function[37,38].

We start by using Eq. (4a) to write G_1 as a function of G_0:

$$G_1 = \frac{1}{\sigma_{\rm on}}\left[\partial_t G_0 + (\sigma_{\rm off} - \rho w_1)G_0 + (d_{\rm m} w_1 - \lambda w_2 - \lambda w_1 w_2)\partial_{w_1} G_0\right]. \tag{6}$$

Summing up Eqs. (4a) and (4b), and plugging Eq. (6) into the resultant equation, we have a parabolic partial differential equation that depends only on G_0:

$$\begin{aligned}&\partial^2_{tt} G_0 + 2(d_{\rm m} w_1 - \lambda w_2 - \lambda w_1 w_2)\partial^2_{tw_1} G_0 + (d_{\rm m} w_1 - \lambda w_2 - \lambda w_1 w_2)^2 \partial^2_{w_1 w_1} G_0 + (\sigma_{\rm off} + \sigma_{\rm on} - \rho w_1)\partial_t G_0 \\ &\quad + (d_{\rm m} w_1 - \lambda w_2 - \lambda w_1 w_2)(d_{\rm m} + \sigma_{\rm off} + \sigma_{\rm on} - \rho w_1 - \lambda w_2)\partial_{w_1} G_0 - \rho(d_{\rm m} w_1 - \lambda w_2 - \lambda w_1 w_2 + \sigma_{\rm on} w_1) G_0 = 0.\end{aligned} \tag{7}$$

Note that there is no partial derivative with respect to w_2 in Eq. (7), which suggests that the solutions to Eq. (7) live in a space spanned by the coordinates (t, w_1). To solve Eq. (7), we convert the coordinates (t, w_1) to a new set of coordinates (u, v) as

$$\begin{aligned}u &= \ln(d_{\rm m} w_1 - \lambda w_2 - \lambda w_1 w_2) - (d_{\rm m} - \lambda w_2)t, \\ v &= d_{\rm m} w_1 - \lambda w_2 - \lambda w_1 w_2.\end{aligned} \tag{8}$$

This is a standard coordinate-conversion technique for solving parabolic PDEs[39]. The following relations between the derivatives of the two sets of coordinates hold

$$\partial_t u = -s, \qquad \partial_{w_1} u = \frac{s}{v}, \qquad \partial_t v = 0, \qquad \partial_{w_1} v = s, \tag{9}$$

where $s = d_{\rm m} - \lambda w_2$. We use the chain rule and Eq. (9) to convert the first- and second-order derivatives of G_0 under the coordinates (t, w_1) to those under the new coordinates:

$$\begin{cases}\partial_t G_0 = \partial_u G_0 \cdot \partial_t u + \partial_v G_0 \cdot \partial_t v = -s\partial_u G_0, \\ \partial_{w_1} G_0 = \partial_u G_0 \cdot \partial_{w_1} u + \partial_v G_0 \cdot \partial_{w_1} v = \dfrac{s}{v}\partial_u G_0 + s\partial_v G_0, \\ \partial^2_{tt} G_0 = \partial_t(-s\partial_u G_0) = \partial_u(-s\partial_u G_0)\cdot\partial_t u + \partial_v(-s\partial_u G_0)\cdot\partial_t v = s^2\partial^2_{uu} G_0, \\ \partial^2_{tw_1} G_0 = \partial_{w_1}(-s\partial_u G_0) = \partial_u(-s\partial_u G_0)\cdot\partial_{w_1} u + \partial_v(-s\partial_u G_0)\cdot\partial_{w_1} v = -\dfrac{s^2}{v}\partial^2_{uu} G_0 - s^2\partial^2_{uv} G_0, \\ \partial^2_{w_1 w_1} G_0 = \partial_{w_1}\left(\dfrac{s}{v}\partial_u G_0 + s\partial_v G_0\right) = \partial_u\left(\dfrac{s}{v}\partial_u G_0 + s\partial_v G_0\right)\partial_{w_1} u + \partial_v\left(\dfrac{s}{v}\partial_u G_0 + s\partial_v G_0\right)\partial_{w_1} v \\ \qquad = \dfrac{s^2}{v^2}\partial^2_{uu} G_0 + \dfrac{2s^2}{v}\partial^2_{uv} G_0 - \dfrac{s^2}{v^2}\partial_u G_0 + s^2\partial^2_{vv} G_0.\end{cases} \tag{10}$$

By using these equations, the parabolic PDE Eq. (7) further reduces to a second-order ODE

$$\begin{aligned}&v^2\partial^2_{vv} G_0 + v\left(1 - \frac{d_{\rm m}\rho + v\rho}{s^2} + \frac{\rho + \sigma_{\rm off} + \sigma_{\rm on}}{s}\right)\partial_v G_0 \\ &- \rho\left(\frac{v - \sigma_{\rm on}}{s^2} + \frac{d_{\rm m}\sigma_{\rm on} + v\sigma_{\rm on}}{s^3}\right) G_0 = 0.\end{aligned} \tag{11}$$

Note the coefficient of G_0 in Eq. (11) is v-dependent, which prevents us formulating the solution as standard solutions of some canonical second-order ODEs. To circumvent this issue, we replace G_0 in Eq. (11) with

$$G_0 = v^k \tilde{G}_0, \tag{12}$$

which simplifies Eq. (11) to

$$v\partial^2_{vv}\tilde{G}_0 + \left(r - \frac{\rho v}{s^2}\right)\partial_v \tilde{G}_0 - \frac{\rho}{s^2}\left(\frac{r + \theta + 1}{2}\right)\tilde{G}_0 = 0, \tag{13}$$

where r, θ and k are defined as

$$\begin{aligned}\theta &= \frac{d_{\mathrm{m}}\rho}{s^2} - \frac{\rho + \sigma_{\mathrm{off}} - \sigma_{\mathrm{on}}}{s},\\ r &= 1 + \sqrt{4\left(\frac{\rho d_{\mathrm{m}}\sigma_{\mathrm{on}}}{s^3} - \frac{\rho\sigma_{\mathrm{on}}}{s^2}\right) + \left(\theta - \frac{2\sigma_{\mathrm{on}}}{s}\right)^2},\\ k &= \frac{1}{2}\left(\theta - \frac{2\sigma_{\mathrm{on}}}{s} + r - 1\right).\end{aligned} \tag{14}$$

Changing to a description in terms of the variable $x = \rho v/s^2$, Eq. (13) further simplifies to

$$x\partial^2_{xx}\tilde{G}_0 + (r - x)\partial_x\tilde{G}_0 - \frac{1}{2}(r + \theta + 1)\tilde{G}_0 = 0, \tag{15}$$

which is the canonical form of Kummer's equation (see Eq. (13.2.1) in Ref. 40). This suggests the general solution to Eq. (15) is

$$\tilde{G}_0 = f(u)y_2 + g(u)y_1, \tag{16}$$

where y_1 and y_2 have the form:

$$\begin{cases} y_1 = M\left(\frac{r+\theta+1}{2}, r, x\right),\\ y_2 = x^{1-r}M\left(\frac{-r+\theta+3}{2}, 2-r, x\right).\end{cases}$$

Note that M is the confluent hypergeometric function, also known as Kummer's function, and the functions f and g only depend on u, and are to be determined from the initial conditions. It immediately follows from Eqs. (12) and (16) that the general form of G_0 is

$$G_0 = v^k[f(h)y_2 + g(h)y_1]. \tag{17}$$

Note that the functions f and g depend on u in Eq. (16), while they depend on $h = \exp(u)$ in Eq. (17). Since h is a function of u only and f and g are undetermined, we slightly abuse the f-g notations in Eq. (17) to facilitate the following calculations.

Next we solve for G_1 as a function of f, g, y_1 and y_2. This can be done via Eq. (6), and it requires the calculation of $\partial_t G_0$ and $\partial_{w_1} G_0$. Combined with $u = \ln v - st$ and $h = v\exp(-st)$, the two derivatives can be calculated according to Eq. (10) as follows

$$\begin{cases} \partial_t G_0 = -v^k sh[f'(h)y_2 + g'(h)y_1],\\ \partial_{w_1} G_0 = v^{k-1}sk[f(h)y_2 + g(h)y_1]\\ \qquad + v^{k-1}sh[f'(h)y_2 + g'(h)y_1]\\ \qquad + v^k[f(h)\partial_{w_1}y_2 + g(h)\partial_{w_1}y_1].\end{cases} \tag{18}$$

Plugging Eqs. (17) and (18) into Eq. (6), we find

$$\begin{aligned} G_1 = \frac{v^k}{\sigma_{\mathrm{on}}}\big[&(\sigma_{\mathrm{off}} - \rho w_1)(f(h)y_2 + g(h)y_1)\\ &+ sk(f(h)y_2 + g(h)y_1)\\ &+ vf(h)\partial_{w_1}y_2 + vg(h)\partial_{w_1}y_1\big].\end{aligned} \tag{19}$$

By using one of the derivative relations of Kummer's function, i.e. Eq. (13.3.15) in Ref. 40, the terms $\partial_{w_1}y_1$ and $\partial_{w_1}y_2$ in Eq. (19) can be expressed in a form that is free of derivatives over Kummer's function, which easies their numerical evaluation:

$$\begin{cases} \partial_{w_1}y_1 = \frac{\rho}{s}\left[\frac{\epsilon_a + 1}{r}M(\epsilon_a + 2, r + 1, x)\right],\\ \partial_{w_1}y_2 = \frac{\rho}{s}\big[(1-r)x^{-r}M(\epsilon_b + 1, 2 - r, x)\\ \qquad + \left(\frac{\epsilon_b + 1}{2 - r}\right)x^{1-r}M(\epsilon_b + 2, 3 - r, x)\big].\end{cases} \tag{20}$$

Here the two quantities ϵ_a and ϵ_b are defined as

$$\epsilon_a = \frac{r + \theta - 1}{2}, \qquad \epsilon_b = \frac{-r + \theta + 1}{2}. \tag{21}$$

Hence it can be shown (Appendix A) that Eq. (19) simplifies to

$$G_1 = -\frac{sv^k}{\sigma_{\mathrm{on}}}\left[g(h)\epsilon_b M(\epsilon_a, r, x) + f(h)x^{1-r}\epsilon_a M(\epsilon_b, 2 - r, x)\right]. \tag{22}$$

All that remains to obtain a full solution to Model II is to determine f and g in Eqs. (17) and (22). We assume that the initial distribution solutions of the CME Eq. (3), conditioned on the promoter state, are given by $p_0(n_{\mathrm{m}}, n_{\mathrm{p}})$ and $p_1(n_{\mathrm{m}}, n_{\mathrm{p}})$. One can immediately deduce that the corresponding initial conditions for Eqs. (4a)-(4b) are given by

$$g_\varphi(w_1, w_2) = \sum_{n_{\mathrm{m}}}\sum_{n_{\mathrm{p}}}(w_1 + 1)^{n_{\mathrm{m}}}(w_2 + 1)^{n_{\mathrm{p}}}p_\varphi(n_{\mathrm{m}}, n_{\mathrm{p}}),$$

for $\varphi = 0$ or 1. At time $t = 0$, the following relation also holds

$$h = \exp(u) = \exp(\ln v - st) = v\exp(-st) = v.$$

Then, using Eqs. (17) and (19) one can establish the link between the initial conditions $g_0(w_1, w_2)$, $g_1(w_1, w_2)$, and the functions f and g:

$$\begin{cases} v^k[f(v)y_2 + g(v)y_1] = g_0(w_1, w_2),\\ \frac{v^k}{\sigma_{\mathrm{on}}}(\sigma_{\mathrm{off}} - \rho w_1 + sk)[f(v)y_2 + g(v)y_1]\\ \qquad + \frac{v^{k+1}}{\sigma_{\mathrm{on}}}[f(v)\partial_{w_1}y_2 + g(v)\partial_{w_1}y_1] = g_1(w_1, w_2).\end{cases} \tag{23}$$

Solving these equations simultaneously leads us to

$$\begin{cases} f(v) = -\dfrac{g_0(w_1,w_2)v^{-k}\epsilon_b}{(r-1)x^{1-r}\mathrm{e}^x} M\left(\epsilon_a, r, x\right) \\ \qquad\quad - \dfrac{g_1(w_1,w_2)v^{-k}\sigma_{\mathrm{on}}}{s(r-1)x^{1-r}\mathrm{e}^x} M\left(\epsilon_a+1, r, x\right), \\ g(v) = \dfrac{g_0(w_1,w_2)v^{-k}\epsilon_a}{(r-1)\mathrm{e}^x} M\left(\epsilon_b, 2-r, x\right) \\ \qquad\quad + \dfrac{g_1(w_1,w_2)v^{-k}\sigma_{\mathrm{on}}}{s(r-1)\mathrm{e}^x} M\left(\epsilon_b+1, 2-r, x\right), \end{cases} \tag{24}$$

with the condition $h = v$ when $t = 0$ (doe detailed steps of the calculation see Appendix B).

Now, f and g in Eq. (24) are functions of the variable h (see Eq. (17)); it is also the case that $h = v\exp(-st)$ and $h = v$ at time $t = 0$. Hence to obtain the solution of f and g for any time t, we have to replace v by h, i.e., $v \mapsto v\exp(-st)$ and $x \mapsto x\exp(-st)$. Since the coordinate w_1 is dependent on v (w_2 is not a coordinate), we also need to seek a similar mapping $\mathcal{T}_t$ for w_1. From Eq. (8), it is known that

$$w_1 = \frac{v + \lambda w_2}{d_{\mathrm{m}} - \lambda w_1}. \tag{25}$$

Further applying the mapping $v \mapsto v\exp(-st)$, we obtain the mapping $\mathcal{T}_t$

$$\mathcal{T}_t w_1 = \frac{(d_{\mathrm{m}} w_1 - \lambda w_2 - \lambda w_1 w_2)\mathrm{e}^{-st} + \lambda w_2}{d_{\mathrm{m}} - \lambda w_1}.$$

Applying the three mappings $v \mapsto v\exp(-st)$, $x \mapsto x\exp(-st)$ and $w_1 \mapsto \mathcal{T}_t w_1$ to Eq. (25), and combining with Eqs. (17) and (A1), we finally find the full exact time-dependent solution to Model II, which is summarized below as Eqs. (26) and (28).

$$\mathbf{G}_t(w_1,w_2) = \mathbf{K}_t(w_1,w_2)\mathbf{g}(\mathcal{T}_t w_1, w_2) \tag{26}$$

where $\mathbf{G}_t(w_1,w_2) = [G_0(w_1,w_2,t), G_1(w_1,w_2,t)]^\top$, $\mathbf{g}(\mathcal{T}_t w_1, w_2) = [g_0(\mathcal{T}_t w_1, w_2), g_1(\mathcal{T}_t w_1, w_2)]^\top$, and

$$\mathbf{K}_t(w_1,w_2) = \begin{bmatrix} k_1(w_1,w_2,t) & k_2(w_1,w_2,t) \\ k_3(w_1,w_2,t) & k_4(w_1,w_2,t) \end{bmatrix}, \tag{27}$$

with

$$\begin{aligned}
&k_1(w_1,w_2,t) = \gamma_a\left[-\epsilon_b\gamma_b M(\epsilon_a, r, x\mathrm{e}^{-st})M(\epsilon_b+1, 2-r, x) + \epsilon_a M(\epsilon_b, 2-r, x\mathrm{e}^{-st})M(\epsilon_a+1, r, x)\right], \\
&k_2(w_1,w_2,t) = \frac{\gamma_a\sigma_{\mathrm{on}}}{s}\left[-\gamma_b M(\epsilon_a+1, r, x\mathrm{e}^{-st})M(\epsilon_b+1, 2-r, x) + M(\epsilon_b+1, 2-r, x\mathrm{e}^{-st})M(\epsilon_a+1, r, x)\right], \\
&k_3(w_1,w_2,t) = \frac{\gamma_a s\epsilon_a\epsilon_b}{\sigma_{\mathrm{on}}}\left[\gamma_b M(\epsilon_a, r, x\mathrm{e}^{-st})M(\epsilon_b, 2-r, x) - M(\epsilon_b, 2-r, x\mathrm{e}^{-st})M(\epsilon_a, r, x)\right], \\
&k_4(w_1,w_2,t) = \gamma_a\left[\epsilon_a\gamma_b M(\epsilon_a+1, r, x\mathrm{e}^{-st})M(\epsilon_b, 2-r, x) - \epsilon_b M(\epsilon_b+1, 2-r, x\mathrm{e}^{-st})M(\epsilon_a, r, x)\right], \\
&\gamma_a = \frac{\exp(kst - x\mathrm{e}^{-st})}{r-1}, \qquad \gamma_b = \exp[-st(r-1)], \\
&v = d_{\mathrm{m}} w_1 - \lambda w_2 - \lambda w_1 w_2, \qquad s = d_{\mathrm{m}} - \lambda w_2, \\
&x = \rho v/s^2, \qquad \theta = \frac{d_{\mathrm{m}}\rho}{s^2} - \frac{\rho + \sigma_{\mathrm{off}} - \sigma_{\mathrm{on}}}{s}, \\
&r = 1 + \sqrt{4\left(\frac{\rho d_{\mathrm{m}}\sigma_{\mathrm{on}}}{s^3} - \frac{\rho\sigma_{\mathrm{on}}}{s^2}\right) + \left(\theta - \frac{2\sigma_{\mathrm{on}}}{s}\right)^2}, \qquad k = \frac{1}{2}\left(\theta - \frac{2\sigma_{\mathrm{on}}}{s} + r - 1\right), \\
&\epsilon_a = (r + \theta - 1)/2, \qquad \epsilon_b = (-r + \theta + 1)/2, \\
&\mathcal{T}_t w_1 = \frac{(d_{\mathrm{m}} w_1 - \lambda w_2 - \lambda w_1 w_2)\mathrm{e}^{-st} + \lambda w_2}{d_{\mathrm{m}} - \lambda w_2}.
\end{aligned} \tag{28}$$

As mentioned earlier, the joint distribution of mRNA and protein numbers follow directly from Eq. (5).

Special Case (i)

Consider the system initiated without any mRNA and protein and in the active promoter state. The generating functions for these initial conditions are $g_0(w_1,w_2) = 0$ and $g_1(w_1,w_2) = 1$. The general solution given by Eqs. (26) and (28) then greatly simplifies to

$$\begin{aligned}
&G_0(w_1,w_2,t) = k_2(w_1,w_2,t) \\
&= \frac{\sigma_{\mathrm{on}}\exp(-x\mathrm{e}^{-st})}{s(r-1)}\Big[\mathrm{e}^{kst} M\left(\epsilon_b+1, 2-r, x\mathrm{e}^{-st}\right) M(\epsilon_a+1, r, x) \\
&\quad - \mathrm{e}^{st(k-r+1)} M\left(\epsilon_a+1, r, x\mathrm{e}^{-st}\right) M(\epsilon_b+1, 2-r, x)\Big],
\end{aligned} \tag{29a}$$

$$G_1(w_1,w_2,t) = k_4(w_1,w_2,t)$$
$$= \exp(-x\mathrm{e}^{-st})\left[\frac{\epsilon_b \mathrm{e}^{kst}}{r-1} M\left(\epsilon_b+1, 2-r, x\mathrm{e}^{-st}\right) M(\epsilon_a, r, x) + \frac{\epsilon_a \mathrm{e}^{st(k-r+1)}}{r-1} M\left(\epsilon_a+1, r, x\mathrm{e}^{-st}\right) M(\epsilon_b, 2-r, x)\right]. \tag{29b}$$

Special Case (ii)

The marginal distribution of mRNA numbers conditioned on the promoter state can be obtained by setting w_2 to 0 in Eqs. (26) and (28); as expected, this is exactly equal to the solution of the the well known telegraph model[17] since the protein production and decay reactions in Model II have no influence on mRNA dynamics (for detailed calculations see Appendix C).

Special Case (iii)

Similarly setting w_1 to 0 gives us the marginal protein number distribution conditioned on the promoter state which to the best of our knowledge has not been derived before. This distribution considerably simplifies when promoter-inactivation events occur far more frequently than promoter-activation events, i.e., $\sigma_{\text{off}} \gg \sigma_{\text{on}}$; and that mRNA decays fast, i.e. $d_{\text{m}} \gg 0$. This is indeed the exact solution to the chemical master equation of the reduced reaction system

$$D \xrightarrow{\alpha} D + iP, \tag{30}$$

where i is a random number sampled from the geometric distribution $P(\beta, i) = \beta^i/(1+\beta)^{i+1}$ where $i = 0, 1, 2, ...$ Note that β is the mean burst size. The kinetic parameters are given by

$$\alpha = \frac{\sigma_{\text{on}}\rho}{\sigma_{\text{off}}+\rho}, \qquad \beta = \frac{\lambda(\sigma_{\text{off}}+\rho)}{d_{\text{m}}\sigma_{\text{off}}}. \tag{31}$$

For details of the derivation see Appendix D.

Numerical validation of the exact solution

In Figs. 2 and 5, we show that the exact analytical solutions for the joint and marginal distributions are in excellent agreement with those obtained using the stochastic simulation algorithm (SSA)[41] when there is no bimodality (top row) and when there is bimodality (bottom row).

III. EXACT SOLUTION FOR THE TIME-DEPENDENT JOINT DISTRIBUTION OF MRNA AND PROTEIN NUMBERS OF MODEL III

Next we solve Model III which is an extension of Model II. In particular in Model III we introduce a cell-cycle description whereby at the end of each cell-cycle the gene products are randomly partitioned between daughter cells. For generality, we shall assume that the duration of the nth cell-cycle is T_n.

We start by setting the initial conditions of Model II, g_0 and g_1, to be the two generating functions at the beginning of cell-cycle n. Hence using Eq. (26) we have the time-evolution equations of the two generating functions within cell-cycle n (prior to cell division)

$$\mathbf{G}_t^n(w_1,w_2) = \mathbf{K}_t(w_1,w_2)\mathbf{G}_0^n(\mathcal{T}_t w_1, w_2) \tag{32}$$

where $\mathbf{G}_t^n(w_1,w_2) = [G_0^n(w_1,w_2,t),\ G_1^n(w_1,w_2,t)]^\top$, $G_\varphi^n(w_1,w_2,t)$ is the generating function of promoter state D_φ at cell age $t \in [0, T_n)$ of cell-cycle n, and The next step is to seek a recursive equation between $G_\varphi^{n+1}(w_1,w_2,0)$ and $G_\varphi^n(w_1,w_2,0)$. According to the definition of the generating function, we have

$$G_\varphi^{n+1}(z_1,z_2,0) = \sum_{n_{\text{m}}}\sum_{n_{\text{p}}} z_1^{n_{\text{m}}} z_2^{n_{\text{p}}} P_\varphi^{n+1}(n_{\text{m}}, n_{\text{p}}, 0), \tag{33}$$

where we temporarily use z_1 and z_2 instead of w_1 and w_2 to facilitate the following derivation. The probabilities before and after cell division are linked by the law of total probability

$$P_\varphi^{n+1}(n_{\text{m}}, n_{\text{p}}, 0) = \sum_{n'_{\text{m}}}\sum_{n'_{\text{p}}} P_\varphi^n(n'_{\text{m}}, n'_{\text{p}}, T_n) P(n_{\text{m}}, n_{\text{p}} | n'_{\text{m}}, n'_{\text{p}}).$$

Using the binomial partitioning condition Eq. (2), Eq. (33) can be written as

$$G_\varphi^{n+1}(z_1,z_2,0) = \sum_{n_{\text{m}}}\sum_{n'_{\text{m}}} z_1^{n_{\text{m}}} 2^{-n'_{\text{m}}} \binom{n'_{\text{m}}}{n_{\text{m}}} \times \underbrace{\sum_{n_{\text{p}}}\sum_{n'_{\text{p}}} z_2^{n_{\text{p}}} P_\varphi^n(n'_{\text{m}}, n'_{\text{p}}, T_n) \binom{n'_{\text{p}}}{n_{\text{p}}} 2^{-n'_{\text{p}}}}_{\mathcal{E}}. \tag{34}$$

The term $\mathcal{E}$ can be further simplified to

$$\mathcal{E} = \sum_{n'_{\text{p}}} P_\varphi^n(n'_{\text{m}}, n'_{\text{p}}, T_n) \left(\frac{1}{2}\right)^{n'_{\text{p}}} \sum_{n_{\text{p}}} \binom{n'_{\text{p}}}{n_{\text{p}}} z_2^{n_{\text{p}}} \times 1^{n'_{\text{p}} - n_{\text{p}}}$$
$$= \sum_{n'_{\text{p}}} P_\varphi^n(n'_{\text{m}}, n'_{\text{p}}, T_n) \left(\frac{z_2+1}{2}\right)^{n'_{\text{p}}},$$

where the second step follows from the binomial theorem. By performing the same simplification steps Eq. (34) reduces to

$$G_\varphi^{n+1}(z_1,z_2,0) = \sum_{n'_{\text{m}}}\sum_{n'_{\text{p}}} P_\varphi^n(n'_{\text{m}}, n'_{\text{p}}, T_n) \times \left(\frac{z_2+1}{2}\right)^{n'_{\text{p}}} \left(\frac{z_1+1}{2}\right)^{n'_{\text{m}}} = G_\varphi^n\left(\frac{z_1+1}{2}, \frac{z_2+1}{2}, T_n\right),$$

where the last step uses the definition of the generating function. Switching z_1 and z_2 back to w_1 and w_2, we finally obtain the recursive equation linking the generating function prior to cell division to that after cell division

$$G_\varphi^{n+1}(w_1,w_2,0) = G_\varphi^n\left(\frac{w_1}{2}, \frac{w_2}{2}, T_n\right). \tag{35}$$

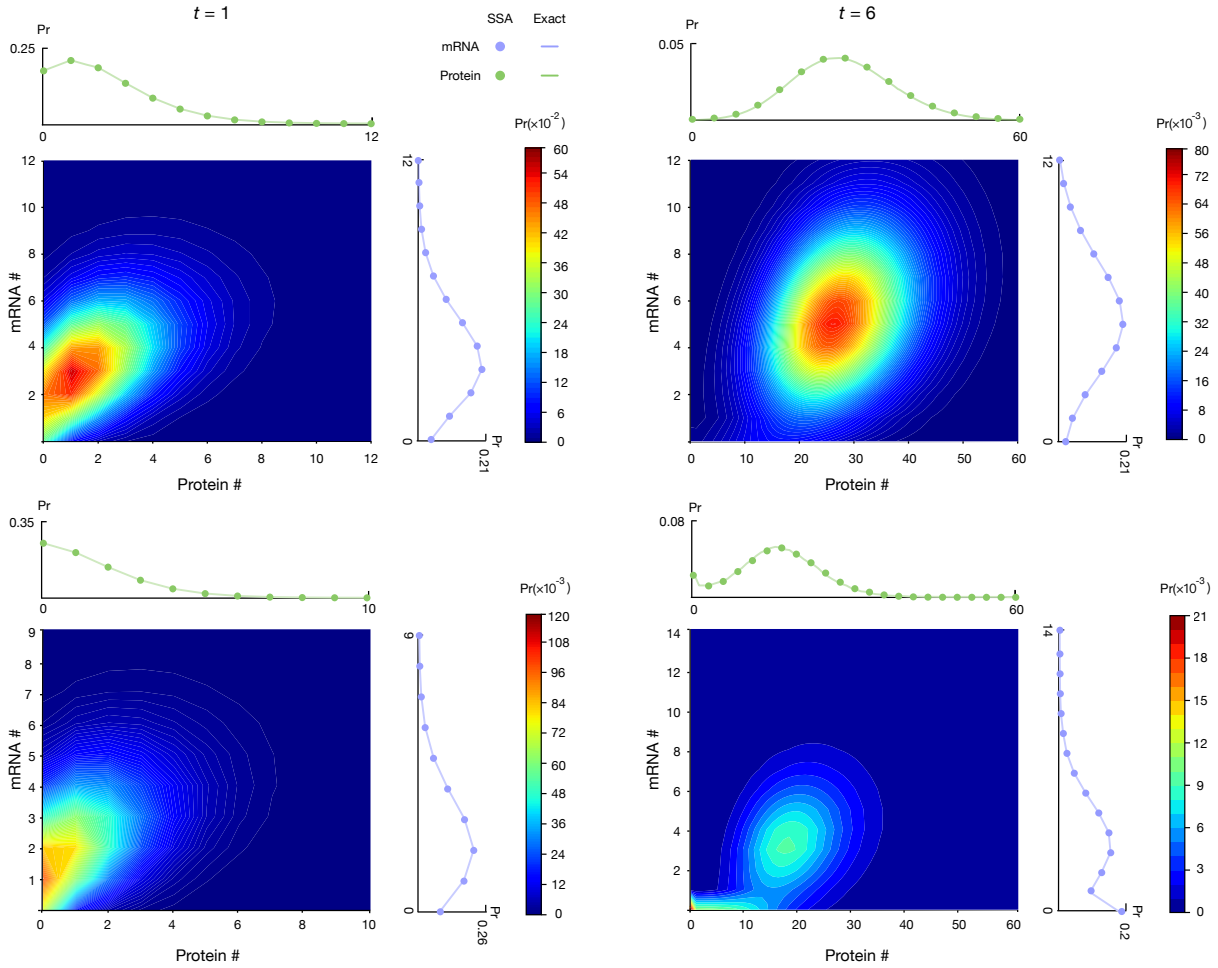


FIG. 2. Comparing the analytical joint and marginal distributions of mRNA and protein numbers in Model II predicted by Eqs. (26) and (28) with those obtained from stochastic simulations using the SSA (the number of independent realizations is 10^7). The heatmaps show the joint distributions of mRNA and protein numbers predicted by the analytical solution at time $t = 1$ and $t = 6$, and they agree well with the predictions given by SSA in Fig. 5. The marginal distributions of mRNA numbers given by the exact solution (solid lines) perfectly match the ones given by the SSA (dots). Similar results are found for the marginal distributions of protein numbers. The kinetic parameters are (top panel) $\rho = 6$, $\lambda = 1$, $\sigma_{\rm off} = 0.1$, $\sigma_{\rm on} = 1$, and $d_{\rm m} = 1$; (bottom panel) $\rho = 4$, $\lambda = 1$, $\sigma_{\rm off} = 0.05$, $\sigma_{\rm on} = 0.02$, and $d_{\rm m} = 1$. Initially the promoter is in the ON state and there are zero mRNA and protein numbers.

In summary, taken together, Eqs. (27), (32) and (35) constitute the exact solution to Model III. We note that in this description, at all points in time, only one cell is tracked, i.e. when a cell divides, only the gene product contents of one of the daughter cells is followed. Hence each independent realisation of the SSA simulating Model III describes a particular lineage; this means that the total number of cells in the population does not change with time. If furthermore $T_n = T$, the population of cells has perfectly synchronized cell-cycles. In Fig. 3, we perform 4 sets of simulation experiments. In each experiment, for simplicity we choose $T_n = T$, i.e. the cell-cycle duration is the same for all cycles; T is however different for each of the four experiments. We compare the marginal distribution of protein counts predicted by our exact solution of Model III and the SSA at different cell ages in the first 4 cell-cycles. As expected, the theoretical predictions and the SSA are found to be in excellent agreement verifying that our theory accurately predicts gene expression within and across cell-cycles. Note also that here we have focused on testing the accuracy of the protein number distribution because the mRNA number distribution for Model III s a special case of an extended version of the telegraph model studied earlier[21]. The theoretical joint distribution of mRNA and protein numbers of Model III can also be directly compared with the SSA and again the two are found to be in excellent agreement (results not shown).

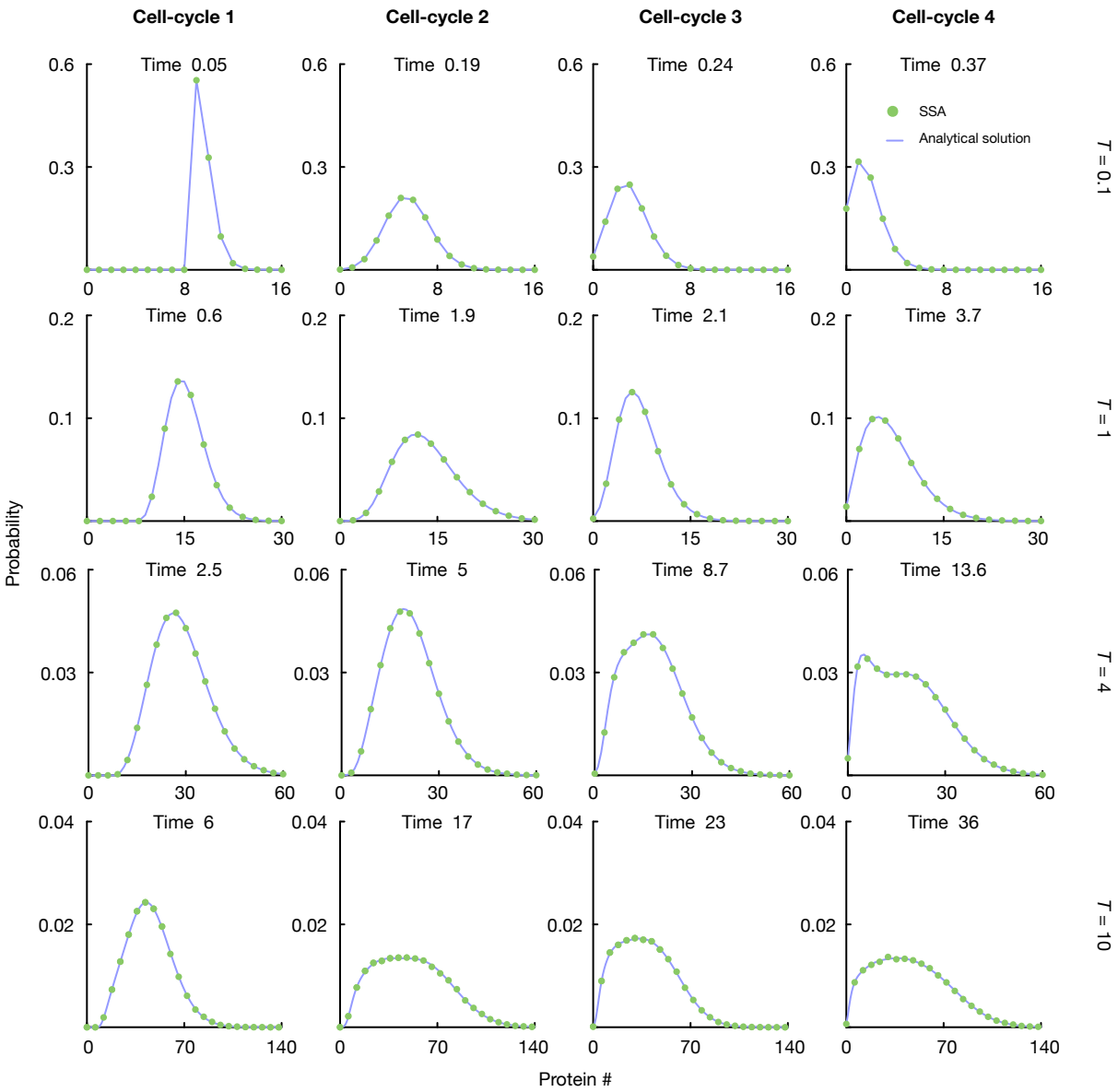


FIG. 3. The marginal distributions of protein numbers in Model III given by theory (Eqs. (27), (32) and (35)) are in excellent agreement with the distributions computed using the SSA at different cell ages and for different cell-cycle duration T. The time shown in the figures is the absolute time measured from the beginning of the first cell-cycle. In all the cases, the system is initiated in the active promoter state D_0 with 4 mRNA and 9 protein molecules. For simplicity, the cell-cycle duration is chosen to be the same for all cell-cycles, i.e. $T_n = T$. The kinetic parameters are $\rho = 2$, $\lambda = 3$, $\sigma_{\rm off} = 0.1$, $\sigma_{\rm on} = 0.1$ and $d_{\rm m} = 1$. We use the modified version of the SSA in Ref. 13, and the number of independent realizations is 10^6.

IV. COMPARISON OF THE STATIONARY GENE PRODUCT NUMBER STATISTICS OF MODELS I & III

Model I

The steady-state statistics of Model I are easy to obtain directly from the moment equations derived from the CME[33]. This is because the moment equations for any system of first-order reactions are closed and hence their solution in steady-state simply amounts to solving a system of simultaneous equations[42]. The means, variances, Fano factors (ratio of variance to mean of the molecule numbers), the covariance of mRNA and protein number fluctuations, and the correlation coefficient are given by:

$$\langle n_{\mathrm{m,I}}\rangle = \frac{\sigma_{\mathrm{on}}\rho}{d(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})}, \tag{36}$$

$$\langle n_{\mathrm{p,I}}\rangle = \frac{\sigma_{\mathrm{on}}\lambda\rho}{dd_{\mathrm{p}}(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})}, \tag{37}$$

$$\begin{aligned}
\mathrm{var}(n_{\mathrm{m,I}}) &= \frac{\rho\sigma_{\mathrm{on}}\left[\rho\sigma_{\mathrm{off}}+d(\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})+(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})^2\right]}{d(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})^2(d+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})},\\
\mathrm{var}(n_{\mathrm{p,I}}) &= \frac{\sigma_{\mathrm{on}}\lambda\rho[(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})(d+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})(d_{\mathrm{p}}+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})(d+d_{\mathrm{p}}+\lambda)+\lambda\rho\sigma_{\mathrm{off}}(d+d_{\mathrm{p}}+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})]}{dd_{\mathrm{p}}(d+d_{\mathrm{p}})(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})^2(d+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})(d_{\mathrm{p}}+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})},\\
\mathrm{FF}(n_{\mathrm{m,I}}) &= 1+\frac{\rho\sigma_{\mathrm{off}}}{(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})(d+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})},\\
\mathrm{FF}(n_{\mathrm{p,I}}) &= 1+\frac{\lambda}{d+d_{\mathrm{p}}}+\frac{\lambda\rho\sigma_{\mathrm{off}}(d+d_{\mathrm{p}}+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})}{(d+d_{\mathrm{p}})(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})(d+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})(d+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})},
\end{aligned} \tag{38}$$

$$\begin{aligned}
\mathrm{cov}(n_{\mathrm{m,I}},n_{\mathrm{p,I}}) &= \frac{\lambda\rho\sigma_{\mathrm{on}}\{(d_{\mathrm{p}}+\sigma_{\mathrm{off}}+\sigma_{\mathrm{on}})[\rho\sigma_{\mathrm{off}}+(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})^2]+d[\rho\sigma_{\mathrm{off}}+d_{\mathrm{p}}(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})+(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})^2]\}}{d(d+d_{\mathrm{p}})(\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})^2(d+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})(d_{\mathrm{p}}+\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}})},\\
\mathrm{corr}(n_{\mathrm{m,I}},n_{\mathrm{p,I}}) &= \mathrm{cov}(n_{\mathrm{m,I}},n_{\mathrm{p,I}})/\sqrt{\mathrm{var}(n_{\mathrm{m,I}})\mathrm{var}(n_{\mathrm{p,I}})}.
\end{aligned} \tag{39}$$

Model III

To compare with the steady-state statistics of Model I, we first need to define what is precisely meant by the steady-state of Model III. We shall consider the common case where measurements of the mRNA and protein numbers for each cell are done at a point in time for a (growing) population of dividing cells with unsynchronized cell-cycles. Furthermore these measurements are done when steady-state growth has been achieved, i.e. the probability that a cell of age t has a given number of mRNA and proteins is independent of which generation it belongs to[43].

The calculations leading to the derivation of the distributions of mRNA and proteins for a population snapshot measurement are complex and will be done in two steps: (i) first we will enforce steady-state growth conditions on Model III, a condition also referred to as the cyclo-stationary condition in the literature[26]. This will lead to a steady-state distribution of gene products for cells of age t where $t \in [0,T]$; (ii) the distributions obtained from the former step will then be integrated over the cell-age distribution in a population thus leading to the final result.

1. Enforcing steady-state growth

According to Eqs. (32) and (35), the generating functions at cell age $t=0$ of two successive cell-cycle are related by

$$\mathbf{G}_0^{n+1}(w_1,w_2) = \mathbf{K}_T(w_1/2,w_2/2)\mathbf{G}_0^{n}(\Omega,w_2/2),$$

with

$$\Omega = \frac{\mathrm{e}^{\left(\frac{\lambda w_2}{2}-d_{\mathrm{m}}\right)T}(\lambda w_1 w_2+2\lambda w_2-2d_{\mathrm{m}}w_1)-2\lambda w_2}{2\lambda w_2-4d_{\mathrm{m}}}.$$

Then, the cyclo-stationary condition suggests that there exists a limit such that $\lim_{n\to\infty}\mathbf{G}_0^n = \mathbf{G}_0^\infty$. It is equivalent to $\mathbf{G}_0^\infty$ satisfying

$$\mathbf{G}_0^\infty(w_1,w_2) = \mathbf{K}_T(w_1/2,w_2/2)\mathbf{G}_0^\infty(\Omega,w_2/2), \tag{40}$$

which means the probability distribution of mRNA and protein numbers at cell age $t=0$ reaches steady state. Once the cyclo-stationary condition is satisfied for $t=0$, the generating functions $\mathbf{G}_t^\infty(w_1,w_2)$ also becomes cyclo-stationary, since $\mathbf{K}_t$ is cell-cycle invariant in

$$\mathbf{G}_t^\infty(w_1,w_2) = \mathbf{K}_t(w_1,w_2)\mathbf{G}_0^\infty(\mathcal{T}_t w_1,w_2). \tag{41}$$

By taking derivative of w_1 on the both sides of Eq. (40) and setting $w_1=w_2=0$, one can get the equations governing the cyclo-stationary means of mRNA numbers at cell age $t=0$

$$\begin{cases}
\langle n_{\mathrm{m,0,III}}\rangle_0 = \dfrac{\mathrm{e}^{-d_{\mathrm{m}}T}}{2}[k_{1,T}\langle n_{\mathrm{m,0,III}}\rangle_0 + k_{2,T}\langle n_{\mathrm{m,1,III}}\rangle_0] \\
\qquad\qquad + \dfrac{\sigma_{\mathrm{on}}k_{1,T}^{(1,0)}+\sigma_{\mathrm{off}}k_{2,T}^{(1,0)}}{\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}}}, \\
\langle n_{\mathrm{m,1,III}}\rangle_0 = \dfrac{\mathrm{e}^{-d_{\mathrm{m}}T}}{2}[k_{3,T}\langle n_{\mathrm{m,0,III}}\rangle_0 + k_{4,T}\langle n_{\mathrm{m,1,III}}\rangle_0] \\
\qquad\qquad + \dfrac{\sigma_{\mathrm{on}}k_{3,T}^{(1,0)}+\sigma_{\mathrm{off}}k_{4,T}^{(1,0)}}{\sigma_{\mathrm{on}}+\sigma_{\mathrm{off}}},
\end{cases} \tag{42}$$

where $\langle n_{\mathrm{m},\varphi,\mathrm{III}}\rangle_t$ is the cyclo-stationary mean mRNA number at promoter state D_φ and cell age t, and we use the following short-hand notations

$$k_{i,t} = k_i(0,0,t), \quad k_{i,t}^{(x,y)} = \partial_{w_1}^x\partial_{w_2}^y k_i(w_1,w_2,t)\big|_{w_1=0,w_2=0}$$

for $i = 1, 2, 3, 4$. The total mean mRNA number at cell age t, $\langle n_{\mathrm{m,III}}\rangle_t$, is then equal to

$$\begin{aligned}\langle n_{\mathrm{m,III}}\rangle_t &= \partial_{w_1}\mathbf{1}\mathbf{G}_t^{\infty}(w_1,w_2)\\ &= \mathrm{e}^{-d_{\mathrm{m}}t}(\langle n_{\mathrm{m,0,III}}\rangle_0 + \langle n_{\mathrm{m,1,III}}\rangle_0)\\ &\quad + \frac{\sigma_{\mathrm{on}}(k_{1,t}^{(1,0)} + k_{3,t}^{(1,0)}) + \sigma_{\mathrm{off}}(k_{2,t}^{(1,0)} + k_{4,t}^{(1,0)})}{\sigma_{\mathrm{on}} + \sigma_{\mathrm{off}}},\end{aligned} \tag{43}$$

where $\mathbf{1} = [1,\ 1]$. Solving for $\langle n_{\mathrm{m,0,III}}\rangle_0$ and $\langle n_{\mathrm{m,1,III}}\rangle_0$ from Eq. (42) and substituting them into Eq. (43), we find an analytical expression for the mean mRNA number at cell age t under cyclo-stationary conditions

$$\langle n_{\mathrm{m,III}}\rangle_t = \frac{\sigma_{\mathrm{on}}\rho(1 + \mathrm{e}^{d_{\mathrm{m}}(T-t)} - 2\mathrm{e}^{d_{\mathrm{m}}T})}{d_{\mathrm{m}}(\sigma_{\mathrm{on}} + \sigma_{\mathrm{off}})(1 - 2\mathrm{e}^{d_{\mathrm{m}}T})}.$$

By taking derivative of w_2 on the both sides of Eqs. (40) and (41) and using the same principles, we can also get the expression for protein ($\langle n_{\mathrm{p,III}}\rangle_t$)

$$\langle n_{\mathrm{p,III}}\rangle_t = \frac{\lambda\rho\sigma_{\mathrm{on}}\left[\mathrm{e}^{d_{\mathrm{m}}(T-t)} + 1 - d_{\mathrm{m}}(t+T) + 2\mathrm{e}^{d_{\mathrm{m}}T}(-1 + d_{\mathrm{m}}(t+T))\right]}{d_{\mathrm{m}}^2(-1 + 2\mathrm{e}^{d_{\mathrm{m}}T})(\sigma_{\mathrm{off}} + \sigma_{\mathrm{on}})},$$

Similarly one can find the second-order raw moments for mRNA numbers and protein numbers under the cyclo-stationary condition $\langle n_{\mathrm{m,III}}^2\rangle_t$, $\langle n_{\mathrm{p,III}}^2\rangle_t$ and $\langle n_{\mathrm{p,III}} n_{\mathrm{m,III}}\rangle_t$. These are too complex to write here (the interested reader is referred to the `Mathematica` notebook).

2. Taking into account the lack of cell-cycle synchronization between cells

Finally we need to take into account the fact that often the cell-cycles of different cells are not synchronized and hence a population snapshot will measure gene product numbers from cells of different ages. The issue is that if we consider stochastic simulations of Model III then if we start from a single cell, at any one time in the future we obtain a population of perfectly synchronized cells due to the deterministic cell-cycle duration. Of course, in reality cell-cycle duration is not deterministic but rather is a random variable, a property that will naturally desynchronize cells. It can be shown that if the cell-cycle duration has some small noise about a mean T then the cell-age distribution is given by $p(t) = 2^{1-t/T}\ln 2/T$ where $t \in [0, T]$[27,43]. Note that this implies that it is more probable to observe cells of a young age (a direct implication of the doubling of the number of cells with age $t = 0$ when cell division occurs).

Hence the observed mean mRNA numbers for a cell population ($\langle n_{\mathrm{m,III}}\rangle$) is given by taking the ensemble average of the cyclo-stationary mean over $p(t)$. This leads to

$$\langle n_{\mathrm{m,III}}\rangle = \int_0^T p(t)\langle n_{\mathrm{m,III}}\rangle_t \mathrm{d}t = \frac{\sigma_{\mathrm{on}}\rho T}{(\sigma_{\mathrm{on}} + \sigma_{\mathrm{off}})(d_{\mathrm{m}}T + \ln 2)}. \tag{44}$$

The stationary mean protein number for the cell population can be similarly computed

$$\langle n_{\mathrm{p,III}}\rangle = \int_0^T p(t)\langle n_{\mathrm{p,III}}\rangle_t \mathrm{d}t = \frac{\lambda\rho\sigma_{\mathrm{on}}T^2}{\ln 2(\sigma_{\mathrm{on}} + \sigma_{\mathrm{off}})(d_{\mathrm{m}}T + \ln 2)}. \tag{45}$$

The variance of mRNA and protein numbers for the cell population are similarly given by

$$\mathrm{var}(n_{\mathrm{m,III}}) = \int_0^T p(t)\langle n_{\mathrm{m,III}}^2\rangle_t \mathrm{d}t - \langle n_{\mathrm{m,III}}\rangle^2, \tag{46}$$

$$\mathrm{var}(n_{\mathrm{p,III}}) = \int_0^T p(t)\langle n_{\mathrm{p,III}}^2\rangle_t \mathrm{d}t - \langle n_{\mathrm{p,III}}\rangle^2, \tag{47}$$

while the covariance of mRNA and protein fluctuations is given by

$$\begin{aligned}\mathrm{cov}(n_{\mathrm{m,III}}, n_{\mathrm{p,III}}) = &\int_0^T p(t)\langle n_{\mathrm{m,III}} n_{\mathrm{p,III}}\rangle_t \mathrm{d}t\\ &- \langle n_{\mathrm{m,III}}\rangle\langle n_{\mathrm{p,III}}\rangle.\end{aligned}$$

The correlation between mRNA and protein is then obtained using

$$\mathrm{corr}(n_{\mathrm{m,III}}, n_{\mathrm{p,III}}) = \frac{\mathrm{cov}(n_{\mathrm{m,III}}, n_{\mathrm{p,III}})}{\sqrt{\mathrm{var}(n_{\mathrm{m,III}})\mathrm{var}(\mathrm{cov}(n_{\mathrm{p,III}})}}. \tag{48}$$

The expressions derived above were validated using stochastic simulations, as follows. As mentioned earlier, a non delta-function cell-age distribution can only arise if the cell-cycle duration is not fixed. To this end, we devise Model IV, one that is the same as Model III except that the cell-cycle duration is equal to the mean T plus small noise which is Erlang distributed (an assumption that is motivated by measurements of several cell types; see Fig.1c in Ref. 27). For a more detailed discussion of Model IV we refer the reader to Appendix E. Simulations of Model IV leads to an exponentially increasing population of cells; at a point in time after several

generations have elapsed, we calculate the means, variances and covariance of mRNA and proteins across the cell population and also the cell age distribution. In Fig. 6 these are compared with those predicted by Model III; excellent agreement is found for three parameter sets thus validating the accuracy of Eqs. (44)-(48).

Comparison of Models I and III

Now we are in a position to compare Models I and III so that we can assess the importance of modelling the cell-cycle explicitly. To ensure a fair model comparison, we first equate the means of the two models, i.e. $\langle n_{\mathrm{m,I}}\rangle = \langle n_{\mathrm{m,III}}\rangle$ and $\langle n_{\mathrm{p,I}}\rangle = \langle n_{\mathrm{p,III}}\rangle$. This equivalence is possible provided the effective mRNA decay rate d and the effective protein dilution rate d_{p} in Model I, are chosen to be

$$d = d_m + \frac{\ln 2}{T}, \qquad d_{\mathrm{p}} = \frac{\ln 2}{T}. \tag{49}$$

Using these effective rates, we compute the Fano factor for proteins in Model I, $\mathrm{FF}(n_{\mathrm{p,I}})$, using Eq. (38). The Fano factor for proteins in Model III, $\mathrm{FF}(n_{\mathrm{p,III}})$, can be computed from Eqs. (45) and (47). We then introduce the ratio R_1 defined by

$$R_1 = \frac{\mathrm{FF}(n_{\mathrm{p,III}})}{\mathrm{FF}(n_{\mathrm{p,I}})}, \tag{50}$$

which quantifies the discrepancy between the two Fano factor predictions provided by Model I and Model III. Similarly, we define the ratio R_2 as

$$R_2 = \frac{\mathrm{corr}(n_{\mathrm{m,III}}, n_{\mathrm{p,III}})}{\mathrm{corr}(n_{\mathrm{m,I}}, n_{\mathrm{p,I}})}, \tag{51}$$

to quantify the discrepancy between the correlation coefficients between mRNA and protein numbers predicted by Model I (Eq. (39)) and Model III (Eq. (48)).

In Fig. 4a and b, we investigate how the ratio R_1 varies across parameter space. The results show that Model I always underestimates the Fano factor of protein number fluctuations. The deviations between the two models is mostly determined by the ratio of the time spent in the ON and OFF states and the cell-cycle duration time ($\sigma_{\mathrm{off}}T$ and $\sigma_{\mathrm{on}}T$, respectively) and to a much lesser extent by the ratio of the mRNA lifetime and the cell-cycle duration time d_mT. Generally, the deviations increase with $\sigma_{\mathrm{on}}T$ and with decreasing $\sigma_{\mathrm{off}}T$ and d_mT. In Fig. 4c and d, we investigate how the ratio R_2 varies across parameter space. The results show that the Model I can either overestimate or underestimate the correlation coefficient. The factor that most strongly determines whether R_2 is less than or greater than 1 is the ratio of the mRNA lifetime and the cell-cycle duration time d_mT; we note that the line $d_mT = 10$ is approximately coincident with $R_2 = 1$.

V. SUMMARY AND DISCUSSION

In this paper, we have extended the classical three-stage model to include a cell-cycle description and solved it exactly using the generating function method. This is a theoretical advance because to the best of our knowledge, presently there is not an analytical solution for the joint distribution of mRNA and protein numbers in a three-state model of gene expression with or without a cell-cycle description. Prior to this work, an exact analytical solution for the joint distribution existed only for the classical two-stage model where there is no promoter switching and no cell-cycle description[31,32] and the marginal distribution of protein numbers in the classical three-stage model was only derived when the timescales of certain reactions are well separated[20].

Because our model has the added advantage of explicitly incorporating cell division whereas the classical three-state model[20] does not, comparing the steady-state population snapshot statistics of the two enabled us to shed light on the effect that cell division has on the predictions of the Fano factor of protein number fluctuations and the correlation coefficient between mRNA and protein number fluctuations. We found there are substantial deviations between the two model predictions because the first-order reaction modelling mRNA and protein dilution in the classical model cannot capture the fluctuations induced by cell division. In particular the classical model underestimates the Fano factor of protein numbers. The same has been previously reported using a model of bursty protein production (where the burst size is sampled from a geometric distribution) and cell division[26]; however this model is a special case of Model III presented in this paper (valid only when certain parameter restrictions apply) and hence the Fano factor analysis presented here is more general. Perhaps more interestingly, we also found that the classical model can either under or overestimate the correlation between mRNA and protein number fluctuations. Roughly speaking, overestimation occurs when the ratio of the cell division time and the mRNA lifetime d_mT is greater than 10 (Fig. 4c and d). This property reflects the intuitive observation that if during most of the cell-cycle the mRNA is roughly in steady-state but the protein number is mostly increasing with time (since removal occurs only at cell division), then the correlation between the two will necessarily be weak; our cell-cycle model captures this but the classical model overestimates the correlation because it predicts that both mRNA and protein reach or approach a steady-state since both decay due to a first-order reaction.

Our model of course also has limitations, the major one being the inherent assumption that protein removal only occurs via dilution and not due to degradation. It is estimated that in mammalian cells about 70% of proteins have a lifetime that is longer than the mean cell-cycle duration[35] – for these proteins, our model is applicable because removal will be dominated by dilution. To obtain a stochastic description for the rest of the proteins,

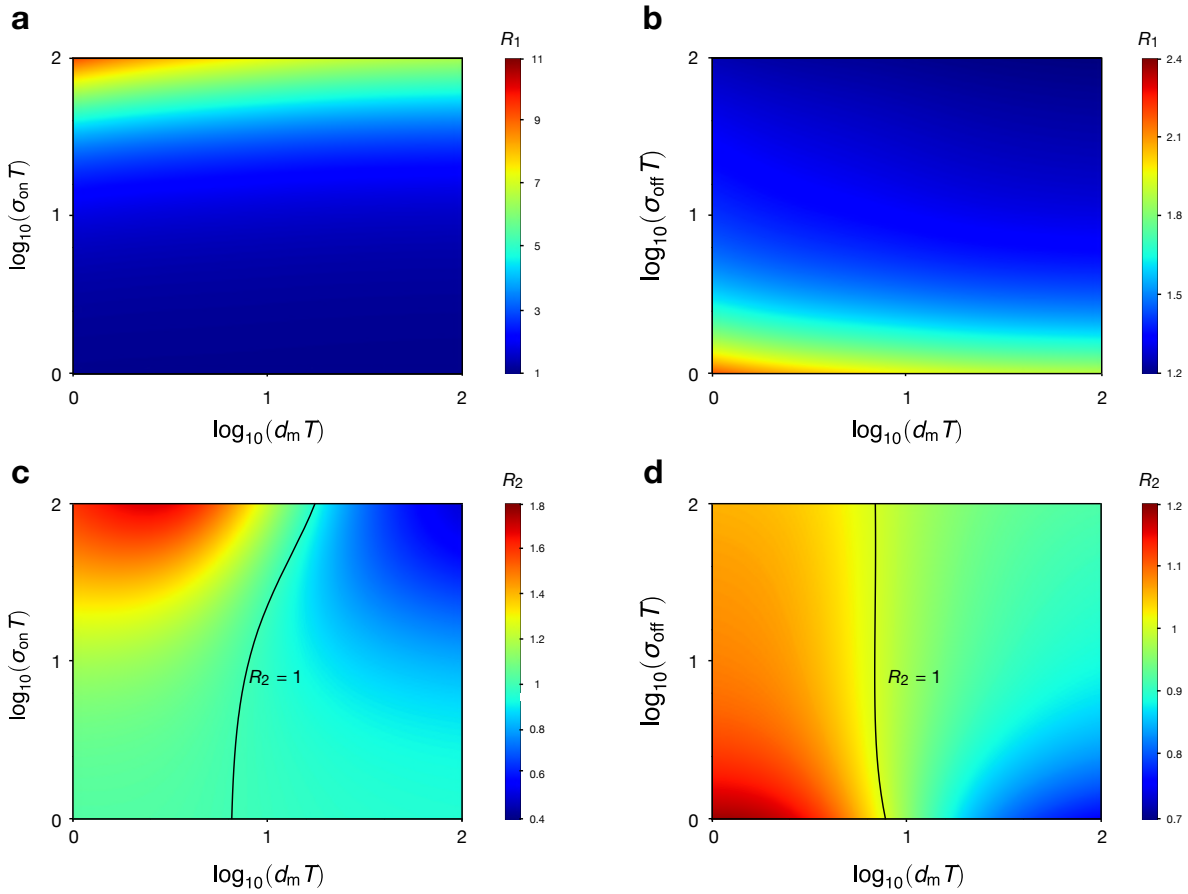


FIG. 4. Comparison of the steady-state Fano factor of protein numbers and of the correlation coefficient between mRNA and protein predicted by Model I and Model III. The ratio R_1 is the Fano factor of Model III divided by that of Model I (Eq. (50)). The ratio R_2 is the correlation coefficient of Model III divided by that of Model I (Eq. (51)). In (a) and (c) we fix the values of the switching off rate $\sigma_{\rm off} = 1.5$ hr^{-1} and the transcription rate $\rho = 7$ hr^{-1} to be the medians reported in Ref. 7, the translation rate $\lambda = 198$ hr$^{-1} \cdot$ mRNA^{-1} to be the median reported in Ref. 35, and the cell-cycle duration $T = 27.5$ hrs to be the value reported in Ref. 35. In (b) and (d) the parameter values are the same, except that now we fix the switching on rate $\sigma_{\rm off} = 0.12$ hr^{-1} to be the median reported in Ref. 7. Note that $R_1 > 1$ all over parameter space implying that Model I always underestimates the Fano factor of protein fluctuations. The ratio R_2 can be greater than or less than 1 implying that Model I can either over or underestimate the correlation coefficient between mRNA and protein numbers.

active protein degradation needs to be included but unfortunately this then makes the analytical solution of the chemical master equation challenging and it is currently unclear how to overcome this issue.

An interesting direction we have not here explored is the implication of our analytical results to parameter inference from experimental data[6,44], i.e. given measurements of single cell mRNA and protein numbers in a population of cells[45], how is inference of rate parameter values (by matching the theoretical and experimental joint distribution via the method of maximum likelihood) using Model III (three-stage model with cell-cycle) different than using Model I (classical three-stage model without cell-cycle)? This is under investigation and we hope to report on it in a separate study.

VI. CODE AND DATA AVAILABILITY

The `Mathematica` notebooks, `Python` codes and relevant data are deposited at `https://github.com/edwardcao3026/three-stage-model.git`.

ACKNOWLEDGMENTS

Z.C. acknowledges support from NSFC No. 62073137, Shanghai Action Plan for Technological Innovation Grant (No. 22ZR1415300, 22511104000, 23S41900500) and Shanghai Center of Biomedicine Development. R.G. acknowledges support from a Leverhulme Trust grant (RPG-2020-327).

Appendix A: Derivation of Eq. (22)

It then follows from Eqs. (19) and (20) that the solution G_1 becomes

$$
\begin{aligned}
G_1 =& \frac{v^k}{\sigma_{\text{on}}} s g(h) \left[\underbrace{-(\epsilon_b + x) M\left(\epsilon_a + 1, r, x\right) + x \left(\frac{\epsilon_a + 1}{r}\right) M\left(\epsilon_a + 2, r + 1, x\right)}_{\mathcal{A}} \right] \\
&+ \frac{v^k}{\sigma_{\text{on}}} s f(h) x^{1-r} \left[\underbrace{-(\epsilon_a + x) M\left(\epsilon_b + 1, 2 - r, x\right) + x \left(\frac{\epsilon_b + 1}{2 - r}\right) M\left(\epsilon_b + 2, 3 - r, x\right)}_{\mathcal{B}} \right].
\end{aligned}
\tag{A1}
$$

In the following steps, we show that Eq. (A1) can be further simplified by making use of recurrence relations of the Kummer function. For the term $\mathcal{A}$, we have

$$
\begin{aligned}
\mathcal{A} &= -(\epsilon_b + x) M(\epsilon_a + 1, r, x) + (\epsilon_a + 1) M(\epsilon_a + 2, r, x) \\
&\quad - (\epsilon_a + 1) M(\epsilon_a + 1, r, x), \\
&= -(\epsilon_a + \epsilon_b + 1 + x) M(\epsilon_a + 1, r, x) \\
&\quad + (\epsilon_a + 1) M(\epsilon_a + 2, r, x), \\
&= -\epsilon_b M(\epsilon_a, r, x).
\end{aligned}
\tag{A2}
$$

Here the first step of Eq. (A2) is achieved by applying Eq. (13.3.4) of Ref. 40 to the second term in $\mathcal{A}$, and the second step re-organizes the expression in the first step, while the third step follows Eq. (13.3.1) of Ref. 40. By similar arguments, we also obtain

$$
\begin{aligned}
\mathcal{B} &= -(\epsilon_a + x) M(\epsilon_b + 1, 2 - r, x) + (\epsilon_b + 1) M(\epsilon_b + 2, 2 - r, x) \\
&\quad - (\epsilon_b + 1) M(\epsilon_b + 1, 2 - r, x), \\
&= -(\epsilon_a + \epsilon_b + 1 + x) M(\epsilon_b + 1, 2 - r, x) \\
&\quad + (\epsilon_b + 1) M(\epsilon_b + 2, 2 - r, x), \\
&= -\epsilon_a M(\epsilon_b, 2 - r, x).
\end{aligned}
\tag{A3}
$$

Finally using Eqs. (A1)-(A3) we obtain the much simpler expression for G_1 given by Eq. (22).

Appendix B: Derivation of Eq. (24)

The left-hand side of both equations in Eq. (23) are linear functions of both f and g, and hence they can be immediately obtained

$$
\begin{cases}
f(v) = \dfrac{g_0(w_1, w_2) v^{-k} \left[(-\sigma_{\text{off}} s - k s^2 + \rho d_{\text{m}} - \rho s + \rho v) y_1 - s v \partial_{w_1} y_1 \right]}{s v (y_1 \partial_{w_1} y_2 - y_2 \partial_{w_1} y_1)} + \dfrac{g_1(w_1, w_2) v^{-k} \sigma_{\text{on}} s y_1}{s v (y_1 \partial_{w_1} y_2 - y_2 \partial_{w_1} y_1)}, \\
g(v) = -\dfrac{g_0(w_1, w_2) v^{-k} \left[(-\sigma_{\text{off}} s - k s^2 + \rho d_{\text{m}} - \rho s + \rho v) y_2 - s v \partial_{w_1} y_2 \right]}{s v (y_1 \partial_{w_1} y_2 - y_2 \partial_{w_1} y_1)} - \dfrac{g_1(w_1, w_2) v^{-k} \sigma_{\text{on}} s y_2}{s v (y_1 \partial_{w_1} y_2 - y_2 \partial_{w_1} y_1)}.
\end{cases}
$$

Note that the term $\mathcal{C} = y_1 \partial_{w_1} y_2 - y_2 \partial_{w_2} y_1$ appears in several parts of the expression; it introduces the reciprocal of the Kummer function into the full solution and may thus hamper the efficient evaluation when we use G_0 and G_1 to obtain $P_0(n_{\text{m}}, n_{\text{p}})$ and $P_1(n_{\text{m}}, n_{\text{p}})$. Therefore, we need to simplify $\mathcal{C}$, which is indeed the Wronskian identity of the two general solutions y_1 and y_2. Introducing $\mathcal{D}$, we find $\mathcal{C}$ simplifies to

$$
\mathcal{C} = \partial_v x \cdot \partial_{w_1} v \cdot \underbrace{(y_1 \partial_x y_2 - y_2 \partial_x y_1)}_{\mathcal{D}} = \frac{\rho}{s} \mathcal{D}.
$$

Utilizing the observations above, one can simplify $\mathcal{D}$ to

$$
\begin{aligned}
\mathcal{D} &= \mathcal{W}\left\{ M\left(\epsilon_a + 1, r, x\right), x^{1-r} M\left(\epsilon_b + 1, 2 - r, x\right) \right\} && \text{Eq. (13.2.4)} \\
&= \mathcal{W}\left\{ \mathbf{M}\left(\epsilon_a + 1, r, x\right), x^{1-r} \mathbf{M}\left(\epsilon_b + 1, 2 - r, x\right) \right\} \Gamma(r) \Gamma(2 - r) && \text{Eq. (13.2.33)} \\
&= \frac{(\sin \pi r) x^{-r} \mathrm{e}^x}{\pi} \frac{\pi}{\sin \pi r} (1 - r) && \text{Eq. (5.5.3)} \\
&= (1 - r) \mathrm{e}^x x^{-r} && \text{Eq. (5.5.1)}
\end{aligned}
\tag{B1}
$$

where $\mathcal{W}\{\cdot, \cdot\}$ stands for the Wronskian identity and $\mathbf{M}(a, b, z) = \Gamma(b) M(a, b, z)$. Note that the equations of

Ref. 40 used in the above simplification steps are described at the end of each step. Following the procedures performed in Eqs. (A2) and (A3), we can use Eq. (B1) to simplify f and g to the form in Eq. (24) under the condition that $h = v$ when $t = 0$.

Appendix C: Marginal distribution of mRNA numbers for Model II

We start by setting $w_2 = 0$ in Eq. (28) which leads to the simplified quantities

$$v = w_1 d_{\mathrm{m}}, \quad s = d_{\mathrm{m}}, \quad x = \frac{\rho w_1}{d_{\mathrm{m}}}, \quad \theta = \frac{\sigma_{\mathrm{on}} - \sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, \quad r = \frac{d_{\mathrm{m}} + \sigma_{\mathrm{on}} + \sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, \quad k = 0, \quad \epsilon_a = \frac{\sigma_{\mathrm{on}}}{d_{\mathrm{m}}},$$
$$\epsilon_b = -\frac{\sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, \quad \gamma_a = \frac{d_{\mathrm{m}}}{\sigma_{\mathrm{on}} + \sigma_{\mathrm{off}}} \exp\left(-\frac{\mathrm{e}^{-d_{\mathrm{m}} t} w_1 \rho}{d_{\mathrm{m}}}\right), \quad \gamma_b = \mathrm{e}^{-(\sigma_{\mathrm{on}} + \sigma_{\mathrm{off}})t}, \quad \mathcal{T}_t w_1 = \mathrm{e}^{-d_{\mathrm{m}} t} w_1.$$

This implies the functions k_1, k_2, k_3 and k_4 in Eq. (28) have the following form

$$\begin{cases} k_1(w_1,t) = \dfrac{\mathrm{e}^{-\bar\rho}}{\Sigma}\left[\mathrm{e}^{-\Sigma t}\sigma_{\mathrm{off}} M\left(1-\dfrac{\sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, 1-\dfrac{\Sigma}{d_{\mathrm{m}}}, \dfrac{w_1\rho}{d_{\mathrm{m}}}\right) M\left(\dfrac{\sigma_{\mathrm{on}}}{d_{\mathrm{m}}}, 1+\dfrac{\Sigma}{d_{\mathrm{m}}}, \bar\rho\right)\right. \\ \qquad \left. +\sigma_{\mathrm{on}} M\left(-\dfrac{\sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, 1-\dfrac{\Sigma}{d_{\mathrm{m}}}, \bar\rho\right) M\left(1+\dfrac{\sigma_{\mathrm{on}}}{d_{\mathrm{m}}}, 1+\dfrac{\Sigma}{d_{\mathrm{m}}}, \dfrac{w_1\rho}{d_{\mathrm{m}}}\right)\right], \\ k_2(w_1,t) = \dfrac{\mathrm{e}^{-\bar\rho}}{\Sigma}\sigma_{\mathrm{on}}\left[-\mathrm{e}^{-\Sigma t} M\left(1-\dfrac{\sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, 1-\dfrac{\Sigma}{d_{\mathrm{m}}}, \dfrac{w_1\rho}{d_{\mathrm{m}}}\right) M\left(1+\dfrac{\sigma_{\mathrm{on}}}{d_{\mathrm{m}}}, 1+\dfrac{\Sigma}{d_{\mathrm{m}}}, \bar\rho\right)\right. \\ \qquad \left. +M\left(1-\dfrac{\sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, 1-\dfrac{\Sigma}{d_{\mathrm{m}}}, \bar\rho\right) M\left(1+\dfrac{\sigma_{\mathrm{on}}}{d_{\mathrm{m}}}, 1+\dfrac{\Sigma}{d_{\mathrm{m}}}, \dfrac{w_1\rho}{d_{\mathrm{m}}}\right)\right], \\ k_3(w_1,t) = \dfrac{\mathrm{e}^{-\bar\rho}}{\Sigma}\sigma_{\mathrm{off}}\left[-\mathrm{e}^{-\Sigma t} M\left(-\dfrac{\sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, 1-\dfrac{\Sigma}{d_{\mathrm{m}}}, \dfrac{w_1\rho}{d_{\mathrm{m}}}\right) M\left(\dfrac{\sigma_{\mathrm{on}}}{d_{\mathrm{m}}}, 1+\dfrac{\Sigma}{d_{\mathrm{m}}}, \bar\rho\right)\right. \\ \qquad \left. +M\left(-\dfrac{\sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, 1-\dfrac{\Sigma}{d_{\mathrm{m}}}, \bar\rho\right) M\left(\dfrac{\sigma_{\mathrm{on}}}{d_{\mathrm{m}}}, 1+\dfrac{\Sigma}{d_{\mathrm{m}}}, \dfrac{w_1\rho}{d_{\mathrm{m}}}\right)\right], \\ k_4(w_1,t) = \dfrac{\mathrm{e}^{-\bar\rho}}{\Sigma}\left[\sigma_{\mathrm{off}} M\left(1-\dfrac{\sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, 1-\dfrac{\Sigma}{d_{\mathrm{m}}}, \bar\rho\right) M\left(\dfrac{\sigma_{\mathrm{on}}}{d_{\mathrm{m}}}, 1+\dfrac{\Sigma}{d_{\mathrm{m}}}, \dfrac{w_1\rho}{d_{\mathrm{m}}}\right)\right. \\ \qquad \left. +\mathrm{e}^{-\Sigma t}\sigma_{\mathrm{on}} M\left(-\dfrac{\sigma_{\mathrm{off}}}{d_{\mathrm{m}}}, 1-\dfrac{\Sigma}{d_{\mathrm{m}}}, \dfrac{w_1\rho}{d_{\mathrm{m}}}\right) M\left(1+\dfrac{\sigma_{\mathrm{on}}}{d_{\mathrm{m}}}, 1+\dfrac{\Sigma}{d_{\mathrm{m}}}, \bar\rho\right)\right]. \end{cases}$$

where $\bar\rho = \rho w_1 \exp(-d_{\mathrm{m}} t)/d_{\mathrm{m}}$ and $\Sigma = \sigma_{\mathrm{on}} + \sigma_{\mathrm{off}}$. For initial conditions of promoter state D_0 and zero mRNA, specifically $g_0(w_1) = 1$ and $g_1(w_1) = 0$, one can straightforwardly verify that the solution becomes perfectly the same as Eqs. (13) and (14) in Ref. 37 (whereby ρ_b should be set to 0) according to $G_0(w_1,t) = k_1(w_1,t)$ and $G_1(w_1,t) = k_3(w_1,t)$.

Appendix D: Marginal distribution of proteins numbers for Model II under timescale separation

We will here show the reduction of the solution Eq. (29) under the conditions $\sigma_{\mathrm{off}} \gg \sigma_{\mathrm{on}}$ and $\rho/\sigma_{\mathrm{off}}$ being some finite quantity. Specifically, for some finite $\tilde\sigma_{\mathrm{off}}$ and $\tilde\rho$, we posit the existence of a variable δ such that $\sigma_{\mathrm{off}} = \delta\tilde\sigma_{\mathrm{off}}$, $\rho = \delta\tilde\rho$ and $\delta \to \infty$.

To achieve so, we will decompose the problem into showing the following five subproblems

(a) $\lim_{\delta\to\infty} \exp[st(k-r+1)] = 0$,

(b) $\lim_{\delta\to\infty} \frac{\sigma_{\mathrm{on}}}{s(r-1)} = 0$,

(c) $\lim_{\delta\to\infty} \frac{\epsilon_b}{r-1} = -1$,

(d) $\lim_{\delta\to\infty} \frac{\epsilon_a}{r-1} = 0$,

(e) $\lim_{\delta\to\infty} \exp(kst)$ exists.

For (a), it is known that $w_2 \in [-1, 0]$ since $z_2 \in [0, 1]$. One immediately has

$$s = d_{\mathrm{m}} - \lambda w_2 \geq d_{\mathrm{m}},$$

and further deduce from Eq. (14) that

$$\theta = \frac{\rho}{s}\left(\frac{d_{\mathrm{m}}}{s} - 1\right) - \frac{\sigma_{\mathrm{off}} - \sigma_{\mathrm{on}}}{s} \leq -\frac{\sigma_{\mathrm{off}} - \sigma_{\mathrm{on}}}{s}.$$

Since $\sigma_{\rm off} \gg \sigma_{\rm on}$ and $\sigma_{\rm off} \to \infty$ as $\delta \to \infty$, we conclude that $\theta \to -\infty$ as $\delta \to \infty$. More specifically, $\theta \sim \mathcal{O}(\delta)$. It is also known from Eq. (14) that

$$\begin{aligned} k - r + 1 &= \frac{\theta - 2\sigma_{\rm on}s^{-1} - \sqrt{\theta^2 + 4\sigma_{\rm off}\sigma_{\rm on}s^{-2}}}{2} \\ &< \frac{\theta - 2\sigma_{\rm on}s^{-1} - |\theta|}{2} < \theta, \end{aligned}$$

which indicates that $k - r + 1 \to -\infty$ as $\delta \to \infty$. Equivalently, it means $\lim_{\delta\to\infty} \exp[st(k - r + 1)] = 0$.

For (b), since $r > 1$ and $s > 0$, we have

$$0 < \frac{\sigma_{\rm on}}{s(r-1)} = \frac{\sigma_{\rm on}}{\sqrt{\theta^2 s^2 + 4\sigma_{\rm off}\sigma_{\rm on}}} < \frac{\sigma_{\rm on}}{\theta s}.$$

As $\delta \to \infty$, $\sigma_{\rm on}/(\theta s) \to 0$, which suggests that

$$\lim_{\delta\to\infty} \frac{\sigma_{\rm on}}{s(r-1)} = 0.$$

For (c), it follows from Eq. (31) that

$$\begin{aligned} \epsilon_a &= \frac{1}{2}\left(\theta + \sqrt{\theta^2 + 4\frac{\sigma_{\rm off}\sigma_{\rm on}}{s^2}}\right), \\ \epsilon_b &= \frac{1}{2}\left(\theta - \sqrt{\theta^2 + 4\frac{\sigma_{\rm off}\sigma_{\rm on}}{s^2}}\right), \\ r - 1 &= \sqrt{\theta^2 + 4\frac{\sigma_{\rm off}\sigma_{\rm on}}{s^2}}, \end{aligned} \tag{D1}$$

which further leads to

$$\lim_{\delta\to\infty} \frac{\epsilon_b}{r-1} = \lim_{\delta\to\infty} \frac{1}{2}\left(-\frac{1}{\sqrt{1 + 4\sigma_{\rm off}\sigma_{\rm on}s^{-2}\theta^{-2}}} - 1\right) = -1,$$

where the last step uses $\theta \sim \mathcal{O}(\delta)$ and $\sigma_{\rm off} \sim \mathcal{O}(\delta)$.

For (d), it is known from Eq. (D1) that

$$\frac{\epsilon_a}{r-1} = -\frac{\sigma_{\rm off}\sigma_{\rm on}}{s^2\epsilon_b(r-1)}.$$

Thus, we have

$$\begin{aligned} \lim_{\delta\to\infty} \frac{\epsilon_a}{r-1} &= \lim_{\delta\to\infty} -\frac{\sigma_{\rm off}\sigma_{\rm on}}{s^2(r-1)^2} \cdot \frac{r-1}{\epsilon_b} = \lim_{\delta\to\infty} \frac{\sigma_{\rm off}\sigma_{\rm on}}{s^2(r-1)^2} \\ &= \lim_{\delta\to\infty} \frac{1}{\theta^2 s^2/(\sigma_{\rm off}\sigma_{\rm on}) + 4} = 0, \end{aligned}$$

since $\theta \sim \mathcal{O}(\delta)$ and $\sigma_{\rm off} \sim \mathcal{O}(\delta)$.

For (e), we first show that k is a finite quantity under the condition $\delta \to \infty$. By using Eqs. (14) and (D1), k can be written as

$$\begin{aligned} k = \epsilon_a - \sigma_{\rm on}s^{-1} &= \frac{\sigma_{\rm on}}{s}\left(-\frac{\sigma_{\rm off}}{\theta s} - 1\right) \\ &= \frac{\sigma_{\rm on}}{s}\left(\frac{d_{\rm m}\rho s^{-1} - \rho + \sigma_{\rm on}}{-d_{\rm m}\rho s^{-1} + \rho + \sigma_{\rm off} - \sigma_{\rm on}}\right). \end{aligned}$$

Hence, we conclude that

$$\lim_{\delta\to\infty} k = \frac{\sigma_{\rm on}}{s}\left(\frac{d_{\rm m}\rho s^{-1} - \rho}{-d_{\rm m}\rho s^{-1} + \rho + \sigma_{\rm off}}\right).$$

Based on the results (a)-(e), the solutions G_0 and G_1 reduce to

$$\begin{cases} G_0(w_1, w_2, t) = 0, \\ G_1(w_1, w_2, t) = \exp\left[\left(\dfrac{d_{\rm m}\rho s^{-1} - \rho}{-d_{\rm m}\rho s^{-1} + \rho + \sigma_{\rm off}}\right)\sigma_{\rm on}t\right] \\ \qquad\qquad \times \exp(-xe^{-st}) \\ \qquad\qquad \times M(\epsilon_b + 1, 2 - r, xe^{-st})M(\epsilon_a, r, x). \end{cases} \tag{D2}$$

Next, we will show how Eq. (D2) can be further reduced under the condition that $d_{\rm m}$ and λ are very large, and $\lambda/d_{\rm m}$ is finite. Specifically, $d_{\rm m} \sim \lambda \sim \mathcal{O}(\Delta)$ and $\Delta \to \infty$. As such, it immediately follows that $s \sim \mathcal{O}(\Delta)$ and

$$\lim_{\Delta\to\infty} x = \lim_{\Delta\to\infty} \frac{\rho v}{s^2} = \lim_{\Delta\to\infty} \frac{\rho(d_{\rm m}w_1 - \lambda w_2 - \lambda w_1 w_2)}{(d_{\rm m} - \lambda w_2)^2} = 0.$$

Therefore, $xe^{-st} \to 0$ as $\Delta \to \infty$. It is suggested by Eq. (14) that r is finite. And so are ϵ_a and ϵ_b, according to Eq. (21). Then, by expanding Kummer functions in Eq. (D2), we obtain that

$$\begin{aligned} M(\epsilon_b + 1, 2 - r, xe^{-st}) &= 1 + o(\Delta^{-1}), \\ M(\epsilon_a, r, x) &= 1 + o(\Delta^{-1}). \end{aligned}$$

These lead to the reduced solution

$$G = G_0 + G_1 = \exp\left[\frac{\sigma_{\rm on}w_2 t\lambda\rho}{d_{\rm m}\sigma_{\rm off} - w_2\lambda(\sigma_{\rm off} + \rho)}\right]. \tag{D3}$$

Finally, we show Eq. (D3) is the solution of the chemical master equation of the reaction system Eq. (30). To this end, we analytically solve the dynamics of this reaction system – instead of directly solving the corresponding generating function equation, we adopt a shortcut by using the generating-function property of a compound process (Supplementary Note 3 in Ref. 46).

We are interested in the distribution of protein numbers $n_{\rm p}(t)$ to the system Eq. (30). Clearly, $n_{\rm p}(t)$ is a random integer variable, and is determined by two factors – the number of "packages" $I(t)$ (where a package stands for a protein-production event occurring before time t) and the number of proteins $X_i \sim \text{Geom}(\frac{\beta}{1+\beta})$ in each package i. Therefore, $n_{\rm p}(t)$ can be presented in the form:

$$n_{\rm p}(t) = \sum_{i=1}^{I(t)} X_i,$$

thereby constituting a compound process. The package number $I(t)$ is determined by the pure birth process with rate α, and hence its generating function is

$$G_I(z, t) = \exp[\alpha t(z - 1)].$$

The generating function of X_i is

$$G_X(z, t) = \frac{1}{1 - b(z - 1)}.$$

Hence, the generating function for $n_{\rm p}(t)$ becomes

$$G_{n_{\rm p}}(z,t) = G_I[G_X(z,t),t] = \exp\left[\frac{\alpha t\beta(z-1)}{1-\beta(z-1)}\right].$$

Finally substituting the expressions for α and β from Eq. (31) into the equation above, we find it leads the same equation as Eq. (D3), thereby completing the proof.

Appendix E: Description of Model IV

Model IV is illustrated in Fig. 6a. We divide the cell-cycle into N stages; the reactions in each stage are the same as Model II. The rate of transitioning between the stages is k and cell division (and binomial partitioning of gene products) occurs when the stage changes from N to 1. The cell-cycle duration is hence a sum of N exponential variables with mean $1/k$, i.e. the cell-cycle duration is Erlang distributed with mean T and coefficient of variation $1/\sqrt{N}$. Note that the cell-cycle stages are not to be understood as the actual biological cell-cycle phases (G_1, S, G_2, M); rather these are introduced as a natural means to model an Erlang distributed cell-cycle duration where the parameter N can be directly estimated from experimental measurements of the cell-cycle duration distribution[27].

Simulations of this model starting from one or more cells leads to an exponentially growing population of cells. After some generations the system reaches a steady-state, in the sense that the population distribution of gene products becomes invariant with time. From a snapshot measurement, we compute the moments of gene products which we compare with those of Model III in Fig. 6b. The cell-age distribution is also computed and it is found to be in good agreement with the cell-age distribution used in calculations of Model III (Fig. 6c).

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," Science **297**, 1183–1186 (2002).

[2] W. J. Blake, M. Kærn, C. R. Cantor, and J. J. Collins, "Noise in eukaryotic gene expression," Nature **422**, 633–637 (2003).

[3] J. M. Raser and E. K. O'Shea, "Control of stochasticity in eukaryotic gene expression," Science **304**, 1811–1814 (2004).

[4] 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**, 213–226 (2019).

[5] A. Sanchez and I. Golding, "Genetic determinants and cellular constraints in noisy gene expression," Science **342**, 1188–1193 (2013).

[6] 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).

[7] 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**, 251–254 (2019).

[8] A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, and S. A. Teichmann, "The technology and biology of single-cell RNA sequencing," Mol. Cell **58**, 610–620 (2015).

[9] Y. Tanouchi, A. Pai, H. Park, S. Huang, N. E. Buchler, and L. You, "Long-term growth data of escherichia coli at a single-cell level," Sci. Data **4**, 1–5 (2017).

[10] A. N. Kapanidis, A. Lepore, and M. El Karoui, "Rediscovering bacteria through single-molecule imaging in living cells," Biophys. J. **115**, 190–202 (2018).

[11] N. Maheshri and E. K. O'Shea, "Living with noisy genes: how cells function reliably with inherent variability in gene expression," Annu. Rev. Biophys. Biomol. Struct. **36**, 413–434 (2007).

[12] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, "Stochastic mRNA synthesis in mammalian cells," PLoS Biol **4**, e309 (2006).

[13] Z. Cao and R. Grima, "Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells," Proc Natl Acad Sci USA **117**, 4682–4692 (2020).

[14] J. Paulsson, O. G. Berg, and M. Ehrenberg, "Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation," Proc Natl Acad Sci USA **97**, 7148–7153 (2000).

[15] Z. S. Singer, J. Yong, J. Tischler, J. A. Hackett, A. Altinok, M. A. Surani, L. Cai, and M. B. Elowitz, "Dynamic heterogeneity and DNA methylation in embryonic stem cells," Mol. Cell **55**, 319–331 (2014).

[16] J. Peccoud and B. Ycart, "Markovian modeling of gene-product synthesis," Theor. Popul. Biol. **48**, 222–234 (1995).

[17] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, "Stochasticity of gene products from transcriptional pulsing," Phys. Rev. E **79**, 031911 (2009).

[18] A. Singh and P. Bokes, "Consequences of mRNA transport on stochastic variability in protein levels," Biophys. J. **103**, 1087–1096 (2012).

[19] N. Friedman, L. Cai, and X. S. Xie, "Linking stochastic dynamics to population distribution: an analytical framework of gene expression," Phys. Rev. Lett. **97**, 168302 (2006).

[20] V. Shahrezaei and P. S. Swain, "Analytical distributions for stochastic gene expression," Proc Natl Acad Sci USA **105**, 17256–17261 (2008).

[21] 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).

[22] C. Jia and R. Grima, "Frequency domain analysis of fluctuations of mRNA and protein copy numbers within a cell lineage: theory and experimental validation," Phys. Rev. X. **11**, 021032 (2021).

[23] C. Jia, A. Singh, and R. Grima, "Concentration fluctuations in growing and dividing cells: Insights into the emergence of concentration homeostasis," PLoS Comput. Biol. **18**, e1010574 (2022).

[24] L. Ham, R. D. Brackston, and M. P. Stumpf, "Extrinsic noise and heavy-tailed laws in gene expression," Phys. Rev. Lett. **124**, 108101 (2020).

[25] J. Jedrak, M. Kwiatkowski, and A. Ochab-Marcinek, "Exactly solvable model of gene expression in a proliferating bacterial cell population with stochastic protein bursts and protein partitioning," Phys. Rev. E **99**, 042416 (2019).

[26] C. H. Beentjes, R. Perez-Carrasco, and R. Grima, "Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics," Phys. Rev. E **101**, 032403 (2020).

[27] R. Perez-Carrasco, C. Beentjes, and R. Grima, "Effects of cell cycle variability on lineage and population measurements of messenger RNA abundance," J. R. Soc. Interface **17**, 20200360 (2020).

[28] I. G. Johnston and N. S. Jones, "Closed-form stochastic solutions for non-equilibrium dynamics and inheritance of cellular components over many cell divisions," Proc. Math. Phys. Eng. Sci. **471**, 20150050 (2015).

[29] M. Soltani and A. Singh, "Effects of cell-cycle-dependent expression on random fluctuations in protein levels," R. Soc. Open Sci. **3**, 160578 (2016).

[30] M. Soltani, C. A. Vargas-Garcia, D. Antunes, and A. Singh, "Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes," PLoS Comput. Biol. **12**, e1004972 (2016).

[31] P. Bokes, J. R. King, A. T. Wood, and M. Loose, "Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression," J. Math. Biol. **64**, 829–854

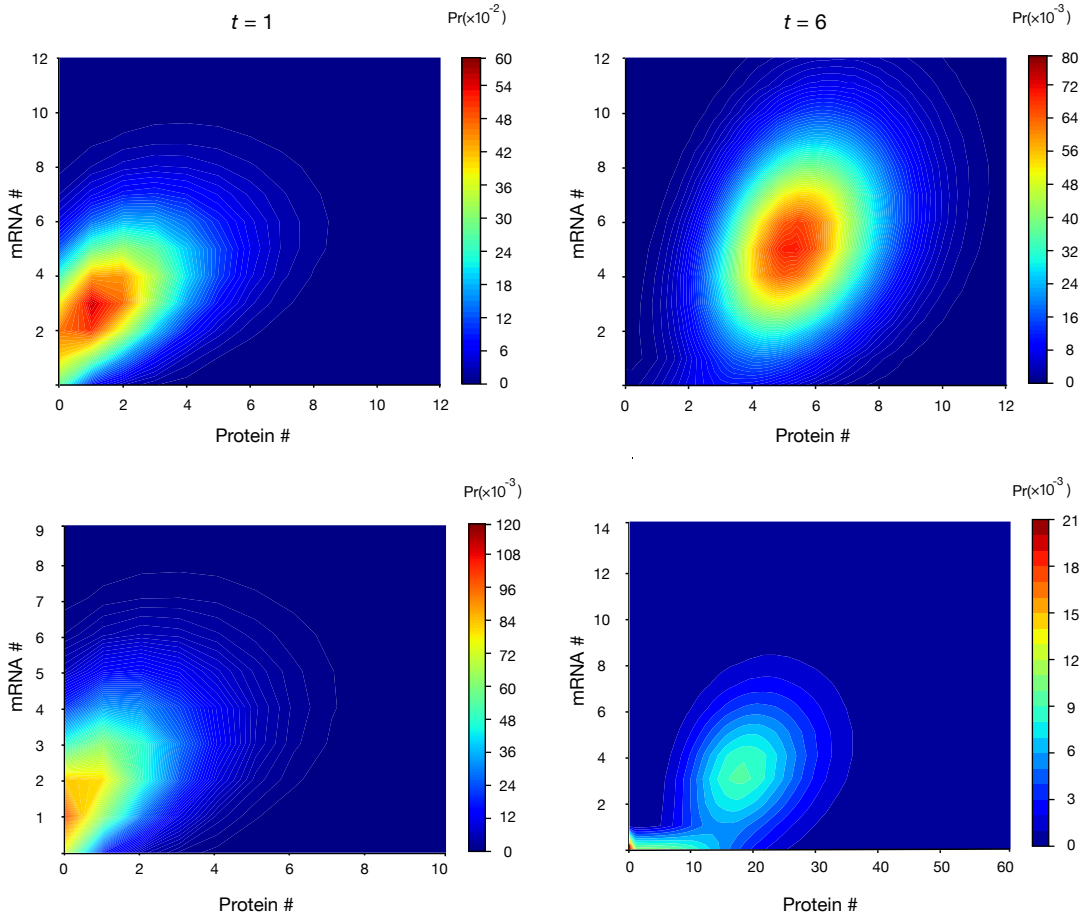
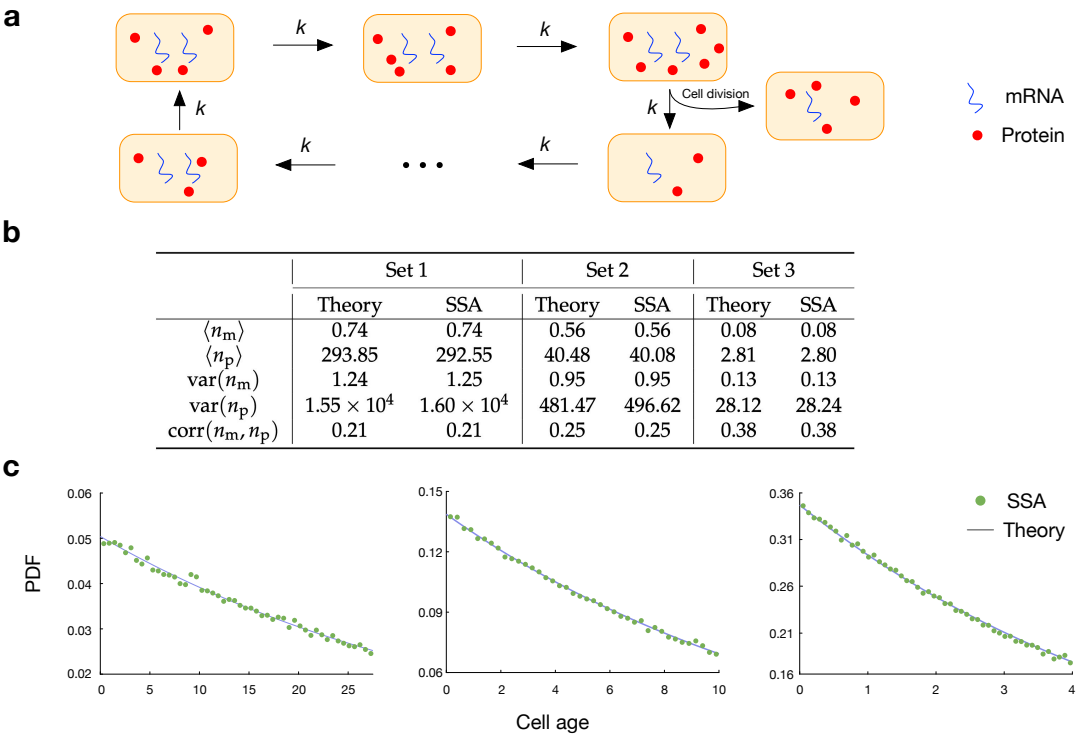


FIG. 5. Joint distributions of mRNA and protein numbers in Model II, as predicted by the SSA (10^7 realizations). The kinetic parameters and initial conditions are the same as those in Fig. 2.

(2012).
[32]H. Pendar, T. Platini, and R. V. Kulkarni, "Exact protein distributions for stochastic models of gene expression using partitioning of poisson processes," Phys. Rev. E **87**, 042720 (2013).
[33]J. Paulsson, "Models of stochastic gene expression," Phys. Life Rev. **2**, 157–175 (2005).
[34]N. Brenner, E. Braun, A. Yoney, L. Susman, J. Rotella, and H. Salman, "Single-cell protein dynamics reproduce universal fluctuations in cell populations," Eur. Phys. J. E **38**, 1–9 (2015).
[35]B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, "Global quantification of mammalian gene expression control," Nature **473**, 337–342 (2011).
[36]N. G. Van Kampen, *Stochastic processes in physics and chemistry*, Vol. 1 (Elsevier, 1992).
[37]Z. Cao and R. Grima, "Linear mapping approximation of gene regulatory networks with stochastic dynamics," Nat. Commun. **9**, 3305 (2018).
[38]P. Kurasov, A. Lück, D. Mugnolo, and V. Wolf, "Stochastic hybrid models of gene regulatory networks–a pde approach," Math. Biosci. **305**, 170–177 (2018).
[39]Y. Pinchover and J. Rubinstein, *An introduction to partial differential equations*, Vol. 10 (Cambridge university press, 2005).
[40]"Nist digital library of mathematical functions," `https://dlmf.nist.gov/`, Release 1.1.10 of 2023-06-15.
[41]D. T. Gillespie, "Stochastic simulation of chemical kinetics," Annu. Rev. Phys. Chem. **58**, 35–55 (2007).
[42]D. Schnoerr, G. Sanguinetti, and R. Grima, "Approximation and inference methods for stochastic biochemical kinetics—a tutorial review," J. Phys. A: Math. Theor. **50**, 093001 (2017).
[43]O. G. Berg, "A model for the statistical fluctuations of protein numbers in a microbial population," J. Theor. Biol. **71**, 587–603 (1978).
[44]Z. Cao and R. Grima, "Accuracy of parameter estimation for auto-regulatory transcriptional feedback loops from noisy data," J R Soc Interface **16**, 20180967 (2019).
[45]C. Albayrak, C. A. Jordi, C. Zechner, J. Lin, C. A. Bichsel, M. Khammash, and S. Tay, "Digital quantification of proteins and mRNA in single mammalian cells," Mol. Cell **61**, 914–924 (2016).
[46]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," Nat. Commun **12**, 2618 (2021).



	Set 1		Set 2		Set 3	
	Theory	SSA	Theory	SSA	Theory	SSA
$\langle n_{\mathrm{m}} \rangle$	0.74	0.74	0.56	0.56	0.08	0.08
$\langle n_{\mathrm{p}} \rangle$	293.85	292.55	40.48	40.08	2.81	2.80
$\mathrm{var}(n_{\mathrm{m}})$	1.24	1.25	0.95	0.95	0.13	0.13
$\mathrm{var}(n_{\mathrm{p}})$	1.55×10^4	1.60×10^4	481.47	496.62	28.12	28.24
$\mathrm{corr}(n_{\mathrm{m}}, n_{\mathrm{p}})$	0.21	0.21	0.25	0.25	0.38	0.38

FIG. 6. Comparison of the population snapshot predictions of Models III and IV. (a) Illustration of Model IV. This is described in detail in Appendix E. (b) The mean, variance and covariance of mRNA and protein numbers were calculated from SSA simulations of Model IV; specifically starting from a certain number of cells, the population grows exponentially and the moments are calculated after several generations when a steady-state distribution has been achieved. These moments are compared to those predicted by Model III using Eqs. (44)-(48) and excellent agreement is found between the two models for three sets of parameters. (c) From simulations of Model IV we also compute the cell-age distribution; the cell age for a cell with cell-cycle stage i is given by $t = iT/N$. The distribution from simulations is found to be in excellent agreement with the theoretical distribution $p(t) = 2^{1-t/T} \ln 2/T$ used in the analytical calculations for Model III. The kinetic parameters together with the initial number of cells n_{initial}, the total simulation time t_{max}, the number of phases N, and cell-cycle stage transition rate k used for SSA simulations are: (Set 1) $\sigma_{\mathrm{on}} = 0.1$, $\sigma_{\mathrm{off}} = 0.1$, $\rho = 3$, $\lambda = 10$, $d = 2$, $T = 27.5$, $N = 50$, $t_{\mathrm{max}} = 400$, $n_{\mathrm{initial}} = 5$, $k = 1.82$; (Set 2) $\sigma_{\mathrm{on}} = 0.5$, $\sigma_{\mathrm{off}} = 2$, $\rho = 3$, $\lambda = 5$, $d = 1$, $T = 10$, $N = 40$, $t_{\mathrm{max}} = 180$, $n_{\mathrm{initial}} = 1$, $k = 4$; (Set 3) $\sigma_{\mathrm{on}} = 0.1$, $\sigma_{\mathrm{off}} = 2$, $\rho = 2$, $\lambda = 6$, $d = 1$, $T = 4$, $N = 50$, $t_{\mathrm{max}} = 80$, $n_{\mathrm{initial}} = 1$, $k = 12.5$.