Postponing production exponentially enhances the molecular memory of a stochastic switch

Pavol Bokes*

Comenius University, Bratislava, Slovakia

June 19, 2020

Abstract

Delayed production can substantially alter the qualitative behaviour of feedback systems. Motivated by stochastic mechanisms in gene expression, we consider a protein molecule which is produced in randomly timed bursts, requires an exponentially distributed time to activate, and then partakes in positive regulation of its burst frequency. Asymptotically analysing the underlying master equation in the large-delay regime, we provide tractable approximations to time-dependent probability distributions of molecular copy numbers. Importantly, the presented analysis demonstrates that positive feedback systems with large production delays can constitute a stable toggle switch even if they operate with low copy numbers of active molecules.

Keywords: Stochastic gene expression; Bursting; Production delay; Chemical master equation; Linear noise approximation; System-size expansion; Fokker–Planck equation; WKB approximation; Large deviations; Exponential asymptotics; Asymptotic matching

1 Introduction

Protein products of a given gene can be present in single cells at low copy numbers, fluctuating in time due to random occurrences of production and degradation events [1–3]. The resulting protein probability distributions exhibit large ratios of variance to the mean, especially if the protein is produced in bursts [4–8]. Understanding how cells can maintain predictable behaviour in the presence of gene-expression noise poses an important challenge to mathematical and systems biology.

The completion of a functioning protein molecule involves a succession of processes such as transcriptional and translational elongation [9, 10], post-translational modification [11], and splicing and export to cytoplasm in case of eukaryotes [12–14]; any of these effectively postpones the production event. Delayed production can dramatically alter the protein dynamics, especially if the protein autoregulates its own production [15–17]. While delays in negative

*E-mail: pavol.bokes@fmph.uniba.sk

feedback systems can lead to destabilisation and noise amplification [18–20], delays have been reported to stabilise the behaviour of and reduce noise in positive feedback systems [21–24]. Long distributed delays buffer production bursts and reduce the large protein variance-to-mean ratios to the Poissonian value of one [25–28]. If a large exponentially distributed delay is added to a highly cooperative positive feedback system, the protein distribution is a mixture of multiple Poissonian modes [29]. In this paper, we extend the steady-state result of [29] and characterise the temporal evolution of the Poissonian modes.

The key insight of [29], on which we expand in this work, is that a stochastic autoregulatory circuit extended by a large delay exhibits a near-deterministic behaviour of the inactive protein (the production of which has been initiated but not yet completed). A classical means of eliminating noise in a stochastic model is to increase its system size/volume [30]; the near-deterministic behaviour in large-volume systems can be described using the linear noise approximation (LNA) [31–35] and the Wentzel–Kramers–Brillouin (WKB) approximation [36–40]. These two approaches are complementary: the LNA applies on finite temporal domains [41]; the WKB approximation covers slow metastable dynamics such as transitions between deterministically stable steady states or to a fixation/extinction point [42]. The LNA and WKB methodologies have been extended to large-volume systems which additionally include a fast-equilibrating low-copy number species [43–47]. While the assumptions of high abundance of slow species and fast equilibration of scarce species are both required for a deterministic limit, other studies considered the high-abundance limit without eliminating the fast species [48–51], or the fast equilibration limit without assuming high abundance in the slow species [52–54]. The lengthening of production delay in [29] has the combined effect of making the inactive protein highly abundant and the scarce active protein (relatively) fast; this renders large-delay systems tractable to LNA and WKB methods.

The outline of the paper is as follows. Section 2 introduces the delayed feedback model and its master equation. Sections 3, 4, and 5 develop the LNA, the WKB-based mixture approximation, and an initial-transient approximation to the solution of the master equation, respectively. Section 6 concludes the paper with a discussion.

2 Master equation

The numbers of inactive (X) and active (s) proteins are modelled by a bivariate Markov chain in continuous time with discrete state space $\{0, 1, \dots\} \times \{0, 1, \dots, s_{\max}\}$. Here we are making a modelling assumption that the number of active proteins s stays below an upper bound $s_{\max}$; this will enable us to use finite-dimensional techniques in our analysis later. The assumption is not restrictive, as a wide class of unbounded systems can be arbitrarily closely approximated by truncated systems [55–57].

The joint probability mass function (PMF) $P(X, s, t)$ of X and s satisfies

the Kolmogorov forward equation [58]

$$\begin{aligned} \varepsilon \partial_t P(X, s, t) = & \sum_{(\tilde{X}, \tilde{s}) \neq (X, s)} W(X, s | \tilde{X}, \tilde{s}) P(\tilde{X}, \tilde{s}, t) \\ & - \sum_{(\tilde{X}, \tilde{s}) \neq (X, s)} W(\tilde{X}, \tilde{s} | X, s) P(X, s, t), \end{aligned} \quad (1)$$

where $W(X, s | \tilde{X}, \tilde{s})$ gives the transition rates in the Markov chain; they are of the following types:

Production

$$W(X + j, s | X, s) = f_s b_j, \quad j = 1, 2, \dots, \quad (2)$$

in which $b_j \geq 0$, where $\sum_{j=0}^{\infty} b_j = 1$, gives the probability of a production burst of size $j = 0, 1, \dots$; the dependence of the production rate f_s on the active protein s implements the delayed feedback.

Activation

$$W(X - 1, s + 1 | X, s) = W(X - 1, s_{\max} | X, s_{\max}) = \varepsilon X, \quad s \neq s_{\max}; \quad (3)$$

the boundary modification of the activation transition guarantees that the number s of active proteins never exceeds the upper bound $s_{\max}$. The activation rate constant ε (which is the reciprocal of the mean delay) is assumed to be small.

Decay

$$W(X, s - 1 | X, s) = s. \quad (4)$$

The Kolmogorov forward equation (1) with transitions (2)–(4) reads

$$\begin{aligned} \varepsilon \partial_t P(X, s, t) = & f_s \sum_{j=0}^{\infty} b_j P(X - j, s, t) + \varepsilon(X + 1) P(X + 1, s - 1, t) \\ & + (s + 1) P(X, s + 1, t) - (f_s + \varepsilon X + s) P(X, s, t), \quad s \neq s_{\max}, \\ \varepsilon \partial_t P(X, s_{\max}, t) = & f_{s_{\max}} \sum_{j=0}^{\infty} b_j P(X - j, s_{\max}, t) \\ & + \varepsilon(X + 1) (P(X + 1, s_{\max} - 1, t) + P(X + 1, s_{\max}, t)) \\ & - (f_{s_{\max}} + \varepsilon X + s_{\max}) P(X, s_{\max}, t). \end{aligned} \quad (5)$$

The probability of having a negative amount of X or s is tacitly understood to be equal to zero in (5) and equations that follow. The aim of the paper is to provide asymptotic approximations as $\varepsilon \rightarrow 0$ for time-dependent solutions $P(X, s, t)$ to (5).

Expressing the shifts in the variable X using the exponential of the differential operator [36]

$$e^{\pm j \partial_X} v(X) = v(X \pm j) \text{ for any function } v \text{ of } X,$$

and defining the moment-generating function (MGF) of production burst size by

$$M(\theta) = \sum_{j=1}^{\infty} b_j e^{j\theta}, \quad (6)$$

we rewrite (5) as

$$\begin{aligned}\varepsilon \partial_t P(X, s, t) &= f_s M(-\partial_X) P(X, s, t) + \varepsilon e^{\partial_X} X P(X, s-1, t) \\ &\quad + (s+1) P(X, s+1, t) - (f_s + \varepsilon X + s) P(X, s, t), \quad s \neq s_{\max}, \\ \varepsilon \partial_t P(X, s_{\max}, t) &= f_{s_{\max}} M(-\partial_X) P(X, s_{\max}, t) \\ &\quad + \varepsilon e^{\partial_X} X (P(X, s_{\max}-1, t) + P(X, s_{\max}, t)) \\ &\quad - (f_{s_{\max}} + \varepsilon X + s_{\max}) P(X, s_{\max}, t).\end{aligned}\tag{7}$$

Since the inactive protein is produced at an $O(1)$ rate but is removed (by being activated) at an $O(\varepsilon)$ rate, its copy number X will accumulate to $O(1/\varepsilon)$ -large levels. In order to bring the inactive protein measurements back on the $O(1)$ scale, we introduce new variables

$$x = \varepsilon X, \quad p(x, s, t) = \varepsilon^{-1} P(\varepsilon^{-1} x, s, t). \tag{8}$$

We refer to x , which becomes a continuous quantity in the limit of $\varepsilon \rightarrow 0$, as the concentration of the inactive protein. The ε^{-1} prefactor is included in the second equality of (8) so that $p(x, s, t)$ represents the probability density function (PDF) of x , rather than its PMF. Inserting (8) into (7) gives

$$\begin{aligned}\varepsilon \partial_t p(x, s, t) &= f_s M(-\varepsilon \partial_x) p(x, s, t) + e^{\varepsilon \partial_x} x p(x, s-1, t) \\ &\quad + (s+1) p(x, s+1, t) - (f_s + x + s) p(x, s, t), \quad s \neq s_{\max}, \\ \varepsilon \partial_t p(x, s_{\max}, t) &= f_{s_{\max}} M(-\varepsilon \partial_x) p(x, s_{\max}, t) \\ &\quad + e^{\varepsilon \partial_x} x (p(x, s_{\max}-1, t) + p(x, s_{\max}, t)) \\ &\quad - (f_{s_{\max}} + x + s_{\max}) p(x, s_{\max}, t).\end{aligned}\tag{9}$$

Equation (9) can be vectorised into

$$\varepsilon \partial_t \mathbf{p} = \mathbf{A}(x, -\varepsilon \partial_x) \mathbf{p}, \tag{10}$$

where

$$\mathbf{p} = \mathbf{p}(x, t) = (p(x, 0, t), \dots, p(x, s_{\max}, t))^{\top} \tag{11}$$

and

$$\mathbf{A}(x, \theta) = \begin{pmatrix} d_0(\theta) & 1 & & 0 \\ e^{-\theta} x & d_1(\theta) & 2 & \\ & e^{-\theta} x & \ddots & \ddots \\ & & \ddots & s_{\max} \\ 0 & & & e^{-\theta} x & d_{s_{\max}}(\theta) \end{pmatrix} \tag{12}$$

is a tridiagonal matrix with main diagonal elements defined by

$$d_s(\theta) = -f_s(1 - M(\theta)) - (1 - \delta_{s, s_{\max}} e^{-\theta})x - s. \tag{13}$$

The insertion in (10) of the differential operator $-\varepsilon \partial_x$ in place of the parameter θ of the matrix (12)–(13) has one important subtlety: the composition of operators, unlike multiplication, is not commutative; in particular, $e^{-\theta} x \neq x e^{-\theta}$ in (12)–(13). Equation (10), being the focal point of much of the ensuing analysis, will be referred to as the master equation.

3 Linear noise approximation

We express the inactive protein concentration as the sum of a deterministic function of time and a small random correction

$$x = \xi(t) + \varepsilon^{1/2}y. \quad (14)$$

Denote by $\Pi(y, s, t)$ the joint probability distribution of the inactive protein fluctuation y and the active protein copy number s , and let

$$\mathbf{\Pi} = \mathbf{\Pi}(y, t) = (\Pi(y, 0, t), \dots, \Pi(y, s_{\max}, t))^T. \quad (15)$$

By the transformation rule, the distribution of y and that of x are related via

$$\mathbf{p}(x, t) = \varepsilon^{-1/2} \mathbf{\Pi} \left(\varepsilon^{-1/2}(x - \xi(t)), t \right). \quad (16)$$

The inclusion of the prefactor $\varepsilon^{-1/2}$ in (16) ensures that the distribution remains normalised after transformation. In the expansion procedure that follows, we determine $\xi(t)$ and, to the leading order, $\mathbf{\Pi}(y, t)$; equation (16) then reconstitutes the approximation in terms of the original joint distribution $\mathbf{p}(x, t)$.

3.1 Expansion

Inserting the transformation (14) and (16) into the master equation (10) and multiplying by $\varepsilon^{1/2}$ yield

$$-\varepsilon^{1/2}\xi'(t)\partial_y \mathbf{\Pi} + \varepsilon\partial_t \mathbf{\Pi} = \mathbf{A}(\xi(t) + \varepsilon^{1/2}y, -\varepsilon^{1/2}\partial_y) \mathbf{\Pi}. \quad (17)$$

Taylor-expanding the right-hand side of (17),

$$\begin{aligned} \varepsilon\partial_t \mathbf{\Pi} &= \mathbf{A}\mathbf{\Pi} + \varepsilon^{1/2}(y\mathbf{A}_x\mathbf{\Pi} + (\xi'(t) - \mathbf{A}_\theta)\partial_y \mathbf{\Pi}) \\ &\quad + \varepsilon \left(\frac{1}{2}\mathbf{A}_{\theta\theta}\partial_y^2 \mathbf{\Pi} - \mathbf{A}_{x\theta}\partial_y(y\mathbf{\Pi}) \right) + \dots \end{aligned} \quad (18)$$

The matrix $\mathbf{A}$ and its derivatives $\mathbf{A}_x$, $\mathbf{A}_\theta$, $\mathbf{A}_{\theta\theta}$, and $\mathbf{A}_{x\theta}$ are understood to be evaluated at $x = \xi(t)$, $\theta = 0$ in (18) and equations that follow; note that, since the matrix elements (12)–(13) depend linearly on x , we have $\mathbf{A}_{xx} = 0$.

We look for solutions to (18) in the form of a regular power series in $\varepsilon^{1/2}$, i.e.

$$\mathbf{\Pi}(y, t; \varepsilon) = \mathbf{\Pi}_0(y, t) + \varepsilon^{1/2}\mathbf{\Pi}_{1/2}(y, t) + \varepsilon\mathbf{\Pi}_1(y, t) + O(\varepsilon^{3/2}). \quad (19)$$

Inserting (19) into (18) and collecting terms of same order yield

$$O(1): \quad \mathbf{A}\mathbf{\Pi}_0 = \mathbf{0}, \quad (20)$$

$$O(\varepsilon^{1/2}): \quad \mathbf{A}\mathbf{\Pi}_{1/2} + y\mathbf{A}_x\mathbf{\Pi}_0 + (\xi'(t) - \mathbf{A}_\theta)\partial_y \mathbf{\Pi}_0 = \mathbf{0}, \quad (21)$$

$$\begin{aligned} O(\varepsilon): \quad \partial_t \mathbf{\Pi}_0 &= \mathbf{A}\mathbf{\Pi}_1 + y\mathbf{A}_x\mathbf{\Pi}_{1/2} + (\xi'(t) - \mathbf{A}_\theta)\partial_y \mathbf{\Pi}_{1/2} \\ &\quad + \frac{1}{2}\mathbf{A}_{\theta\theta}\partial_y^2 \mathbf{\Pi}_0 - \mathbf{A}_{x\theta}\partial_y(y\mathbf{\Pi}_0). \end{aligned} \quad (22)$$

In the next subsections we go through the consequences of equations (20)–(22), one by one.

Figure 1: The one-dimensional truncated birth-death process with constant birth rate x and linear death rate s . The discrete states correspond to the admissible copy numbers s of the active protein, births represent activation (here from an unchanging pool of inactive proteins), and deaths correspond to decay.

3.2 The active protein probability mass function

By (20), the leading-order approximation $\Pi_0 = \Pi_0(y, t)$ to the joint protein distribution belongs to the nullspace of the matrix $\mathbf{A} = \mathbf{A}(\xi(t), 0)$. Substituting $\theta = 0$ into (12)–(13), whilst noting that $M(0) = 1$, one realises that the matrix $\mathbf{A}(x, 0)$ is the Markovian generator matrix of a one-dimensional truncated birth-death process (Figure 1). The nullspace of $\mathbf{A}(x, 0)$ consists of scalar multiples of its stationary distribution $\boldsymbol{\rho}(x) = (\rho_0(x), \dots, \rho_{s_{\max}}(x))^T$, which is the truncated Poisson distribution

$$\rho_s(x) = \frac{\frac{x^s}{s!}}{1 + x + \frac{x^2}{2} + \dots + \frac{x^{s_{\max}}}{s_{\max}!}}, \quad s = 0, 1, \dots, s_{\max}. \quad (23)$$

The leading-order approximation to the joint protein distribution therefore satisfies

$$\Pi_0(y, t) = g(y, t)\boldsymbol{\rho}(\xi(t)). \quad (24)$$

By (24), the inactive protein concentration fluctuation y and the active protein copy number s are at any time t statistically independent; the probability mass function of the latter is given by (23); the probability density function of the former is given by the function $g(y, t)$, which will be determined in Subsection 3.4. Before doing that, let us identify in the next subsection the deterministic motion $\xi(t)$ of the inactive protein concentration.

3.3 The deterministic motion of the inactive protein

Inserting (24) into (21), one finds

$$\mathbf{A}\Pi_{1/2} = -yg\mathbf{A}_x\boldsymbol{\rho} + \partial_y g(\mathbf{A}_\theta - \xi'(t))\boldsymbol{\rho}. \quad (25)$$

Multiplying (25) from the left by the $(s_{\max} + 1)$ -dimensional row vector $\mathbf{1}^T = (1, \dots, 1)$ of ones, and noting that $\mathbf{1}^T\mathbf{A} = \mathbf{1}^T\mathbf{A}_x = \mathbf{0}$ and $\mathbf{1}^T\boldsymbol{\rho} = 1$, we realise that in order that (25) be solvable in $\Pi_{1/2}$, it is necessary that $\partial_y g(\mathbf{1}^T\mathbf{A}_\theta\boldsymbol{\rho} - \xi'(t)) = 0$ holds; and, as we do not allow g to be constant in y , this means that

$$\xi'(t) = \mathbf{1}^T\mathbf{A}_\theta(\xi(t), 0)\boldsymbol{\rho}(\xi(t)). \quad (26)$$

Next, we simplify the right-hand side of the ordinary differential equation (26). Differentiating the elements of the matrix (12) with respect to θ , then taking $\theta = 0$ whilst noting that $M'(0) = \langle B \rangle$, we find, after a little algebra, that

$$\mathbf{1}^T\mathbf{A}_\theta(x, 0)\boldsymbol{\rho}(x) = \langle B \rangle \bar{f}(x) - x, \quad (27)$$

where

$$\langle B \rangle = \sum_{j=1}^{\infty} j b_j \quad (28)$$

gives the mean burst size and

$$\bar{f}(x) = \sum_{s=0}^{s_{\max}} f_s \rho_s(x) \quad (29)$$

gives the effective production rate. Formula (29) calculates the average of the instantaneous production rate with respect to the Poisson distribution (23) of the active protein copy number.

Inserting (27) into (26), we find

$$\xi'(t) = \langle B \rangle \bar{f}(\xi(t)) - \xi(t), \quad (30)$$

meaning that the deterministic component $\xi(t)$ of the inactive protein concentration increases per unit time by the effective production rate and decreases by the linear activation rate.

3.4 The inactive protein fluctuation

Inserting (24) and (26) into (22) and multiplying from the left by $\mathbf{1}^\top$ yield

$$\partial_t g = \mathbf{1}^\top \mathbf{A}_\theta (\rho \mathbf{1}^\top - \mathbf{I}) \partial_y \Pi_{1/2} + \frac{\mathbf{1}^\top \mathbf{A}_{\theta\theta} \rho}{2} \partial_y^2 g - \mathbf{1}^\top \mathbf{A}_{x\theta} \rho \partial_y (yg). \quad (31)$$

In order to close (31), we need to express $\Pi_{1/2}$ in terms of the unknown g . The right-hand side of (25) is a linear combination whose coefficients depend on y but the vectors do not. We write the solution to (25) as

$$\Pi_{1/2} = c \rho + yg \rho_x + \partial_y g v, \quad (32)$$

where c is a scalar (possibly dependent on y and t), ρ is a homogeneous solution, and ρ_x and v solve the y -independent problems

$$\mathbf{A} \rho_x = -\mathbf{A}_x \rho, \quad \mathbf{A} v = (\mathbf{I} - \rho \mathbf{1}^\top) \mathbf{A}_\theta \rho. \quad (33)$$

That ρ_x is indeed a solution to the first of the inhomogeneous problems (33) can be established by differentiating $\mathbf{A} \rho = \mathbf{0}$.

Inserting (32) into (31) whilst noting that $\mathbf{1}^\top \rho_x = 1_x = 0$ yields a linear time-dependent Fokker–Planck equation

$$\partial_t g = -\alpha(\xi(t)) \partial_y (yg) + \beta(\xi(t)) \partial_y^2 g, \quad (34)$$

where

$$\alpha(x) = \mathbf{1}^\top \mathbf{A}_\theta \rho_x + \mathbf{1}^\top \mathbf{A}_{x\theta} \rho = (\mathbf{1}^\top \mathbf{A}_\theta \rho)_x = \langle B \rangle \bar{f}'(x) - 1, \quad (35)$$

$$\beta(x) = \mathbf{1}^\top \mathbf{A}_\theta (\rho \mathbf{1}^\top - \mathbf{I}) v + \frac{\mathbf{1}^\top \mathbf{A}_{\theta\theta} \rho}{2}. \quad (36)$$

The drift term $\alpha(x)$ is the linearisation of the deterministic motion (30). While no apparent simplification is available for the first term in (36), the second term satisfies $\mathbf{1}^\top \mathbf{A}_{\theta\theta} \rho = \langle B^2 \rangle \bar{f}(x) + x$.

Assuming without loss of generality that at the initial time $t = 0$ there is no deviation of the stochastic process from the initial condition $\xi(0) = x_{\text{init}}$ of the deterministic equation (30), we solve the Fokker–Planck equation (34) subject to the delta-function initial condition

$$g(y, 0) = \delta(y), \quad (37)$$

obtaining a Gaussian solution [32]

$$g(y, t) = \frac{1}{\sqrt{2\pi\sigma(t)}} \exp\left(-\frac{y^2}{2\sigma^2(t)}\right), \quad (38)$$

where the variance $\sigma^2(t)$ satisfies an inhomogeneous linear equation of the first order with variable coefficients

$$\frac{d\sigma^2}{dt} = 2\alpha(\xi(t))\sigma^2 + 2\beta(\xi(t)), \quad \sigma^2(0) = 0. \quad (39)$$

Equation (39) is in related contexts referred to as the fluctuation–dissipation relation [59]

3.5 The LNA and its small- t and large- t behaviour

Combining (16), (19), (24), and (38), we arrive at the final form of the linear noise approximation (LNA)

$$p(x, s, t) \sim \rho_s(\xi(t)) \times \frac{1}{\sqrt{2\pi\varepsilon\sigma(t)}} \exp\left(-\frac{(x - \xi(t))^2}{2\varepsilon\sigma^2(t)}\right), \quad \varepsilon \ll t, \quad \log t \ll \varepsilon^{-1} \quad (40)$$

for the joint distribution of the concentration x of the inactive and the copy number s of the active protein. The LNA (40) implies that: s follows the truncated Poisson distribution (23); x follows the Gaussian distribution with mean equal to the deterministic solution $\xi(t)$ to (30) and a small variance equal to $\varepsilon\sigma^2(t)$.

Although the analysis in the current section implicitly assumes $t = O(1)$, the validity of (40) extends to all times except for those that are $O(\varepsilon)$ small or whose logarithms are $O(1/\varepsilon)$ large (i.e. exponentially large times). For these two time domains different approximations will be developed in Sections 4 and 5. In preparation for these developments, note that the asymptotic matching principle [60] implies that the terminal behaviour of an early-time approximation has to coincide with the $t \ll 1$ behaviour of (40), and the $t \gg 1$ behaviour of (40) with the initial behaviour of a large-time approximation. The former is

$$p(x, s, t) \sim \rho_s(x_{\text{init}}) \times \delta(x - x_{\text{init}}), \quad \text{valid if } \varepsilon \ll t \ll 1, \quad (41)$$

as can be seen easily by taking the limit of $t \rightarrow 0$ of (40).

The $t \rightarrow \infty$ limit of (40) deserves a more detailed treatment especially if the deterministic motion (30) is bistable (which we assume to be the case henceforth). Bistability can occur if there is a positive feedback in protein production, and the effective production rate is sufficiently sigmoid (Figure 2, left). The deterministic model (30) then possesses three steady states $x_- < x_0 < x_+$, of which x_- and x_+ are stable and x_0 is unstable (Figure 2, left,

Figure 2: *Left*: The instantaneous production rate $\langle B \rangle f_s$ (dots) and the Poisson-averaged production rate $\langle B \rangle \bar{f}(x)$ (full line). As a result of the averaging (29), the sharp features of the former are smoothed out in the latter. Fixed points $x_- < x_0 < x_+$ (diamonds) of the deterministic model (30) occur where $\langle B \rangle \bar{f}(x)$ intersects the diagonal (dashed line). *Right*: The WKB potential possesses local extrema where (30) has fixed points. *Parameters for both panels*: the burst rate is $f_s = 0.5$ for $s \leq 5$; $f_s = 2.5$ for $s \geq 6$; the burst size is $B = 4$ ($b_j = \delta_{j,4}$).

diamonds). Depending on which side of the unstable fixed point x_0 the initial protein concentration $\xi(0) = x_{\text{init}}$ resides, the deterministic solution (30) and the variance coefficient (39) assume different large-time limits

$$\xi(t) \sim \begin{cases} x_-, & \sigma^2(t) \sim \begin{cases} \sigma_-^2 & \text{if } x_{\text{init}} < x_0, \\ \sigma_+^2 & \text{if } x_{\text{init}} > x_0, \end{cases} \end{cases} \quad \text{for } t \gg 1, \quad (42)$$

in which

$$\sigma_{\pm}^2 = -\frac{\beta(x_{\pm})}{\alpha(x_{\pm})} \quad (43)$$

are the two possible large-time limits of (39). With $\beta(x)$ being positive and $\alpha(x)$ negative at the stable steady states $x = x_{\pm}$, the variances (43) are positive as should be. Note that we discarded from our considerations the non-generic possibility $x_{\text{init}} = x_0$.

Inserting (42) into (40) yields

$$p(x, s, t) \sim \begin{cases} \rho_s(x_-) \times \frac{1}{\sqrt{2\pi\varepsilon\sigma_-}} \exp\left(-\frac{(x-x_-)^2}{2\varepsilon\sigma_-^2}\right) & \text{if } x_{\text{init}} < x_0, \\ \rho_s(x_+) \times \frac{1}{\sqrt{2\pi\varepsilon\sigma_+}} \exp\left(-\frac{(x-x_+)^2}{2\varepsilon\sigma_+^2}\right) & \text{if } x_{\text{init}} > x_0, \end{cases} \quad (44)$$

provided that $1 \gg t$ and $\log t \ll \varepsilon^{-1}$,

which can be expressed in terms of the Heaviside function in a single formula

$$p(x, s, t) \sim H(x_0 - x_{\text{init}}) \times \rho_s(x_-) \times \frac{1}{\sqrt{2\pi\varepsilon\sigma_-}} \exp\left(-\frac{(x-x_-)^2}{2\varepsilon\sigma_-^2}\right) \\ + H(x_{\text{init}} - x_0) \times \rho_s(x_+) \times \frac{1}{\sqrt{2\pi\varepsilon\sigma_+}} \exp\left(-\frac{(x-x_+)^2}{2\varepsilon\sigma_+^2}\right), \quad (45)$$

provided that $1 \gg t$ and $\log t \ll \varepsilon^{-1}$.

The resulting approximation (45) is a mixture of two Poisson/Gaussian modes located at the stable fixed points of the deterministic rate equation (30). The weights in (45) are binary: the whole probability mass is given to one or the other peak depending on the initial condition. On the exponentially large timescale, however, the weights of the peaks become probabilistic and slowly evolve in time, as described in the next section.

4 Mixture approximation

The large-timescale approximation to the joint probability distribution is obtained by replacing the binary weights in (45) by time-dependent ones,

$$p(x, s, t) \sim \omega_-(t) \times \rho_s(x_-) \times \frac{1}{\sqrt{2\pi\varepsilon\sigma_-}} \exp\left(-\frac{(x-x_-)^2}{2\varepsilon\sigma_-^2}\right) + \omega_+(t) \times \rho_s(x_+) \times \frac{1}{\sqrt{2\pi\varepsilon\sigma_+}} \exp\left(-\frac{(x-x_+)^2}{2\varepsilon\sigma_+^2}\right), \quad t \gg 1. \quad (46)$$

The weights are set to evolve according to a random telegraph process

$$\begin{aligned} \frac{d\omega_-}{dt} &= -\lambda_- \omega_- + \lambda_+ \omega_+, & \omega_-(0) &= H(x_0 - x_{\text{init}}), \\ \frac{d\omega_+}{dt} &= \lambda_- \omega_- - \lambda_+ \omega_+, & \omega_+(0) &= H(x_{\text{init}} - x_0), \end{aligned} \quad (47)$$

in which $\lambda_{\pm}$ are the metastable transition rates between the modes of the distributions. The weight values are initiated in (47) to the binary values that were identified by the large- t analysis of the LNA.

The explicit solution to the random-telegraph equations (47) is given by

$$\omega_-(t) = \frac{\lambda_+}{\lambda_- + \lambda_+} + \left(H(x_0 - x_{\text{init}}) - \frac{\lambda_+}{\lambda_- + \lambda_+} \right) e^{-(\lambda_- + \lambda_+)t}, \quad (48)$$

$$\omega_+(t) = \frac{\lambda_-}{\lambda_- + \lambda_+} + \left(H(x_{\text{init}} - x_0) - \frac{\lambda_-}{\lambda_- + \lambda_+} \right) e^{-(\lambda_- + \lambda_+)t}. \quad (49)$$

With (46) and (48)–(49) at hand, it remains to determine the metastable transition rates $\lambda_{\pm}$ to complete the large-time approximation to the joint protein distribution. This is done in the subsections that follow; at the present moment, we point out that the rates will be shown to be exponentially small,

$$\log \frac{1}{\lambda_{\pm}} = O(\varepsilon^{-1}). \quad (50)$$

It follows that unless time t is exponentially large, the product $(\lambda_+ + \lambda_-)t$ in the exponential in (48)–(49) is negligible, so that

$$\omega_-(t) \sim H(x_0 - x_{\text{init}}), \quad \omega_+(t) \sim H(x_{\text{init}} - x_0), \quad \log t \ll \varepsilon^{-1}. \quad (51)$$

Note that inserting (51) into (46) recovers (45). Thus, provided that time is large but not exponentially large, the mixture distribution (46) with random-telegraph weights and the LNA approximation (40) do indeed overlap.

The LNA (40) and the mixture approximation (46) to the probability distribution $p(x, s, t)$ are easily recast, by means of the transformation (8), into

Figure 3: Numeric (*dashed line*) and asymptotic (*full line*) approximations to the PMFs of the inactive (*left panel*) and active (*right panel*) protein. The markers indicate the discrete values of the active protein distribution. The LNA (40) is used for early, and the WKB-based approximation (46) for the later time-points; the last two time-points are set to $\log(2)/(\lambda_- + \lambda_+)$ and $10/(\lambda_- + \lambda_+)$, where $\lambda_{\pm}$ are the metastable transition rates (63). The model parameters are $f_s = 0.5$ for $s \leq 5$; $f_s = 2.5$ for $s \geq 6$; $B = 4$; $\varepsilon = 0.0125$; $s_{\max} = 20$. The initial values are $X_{\text{init}} = s_{\text{init}} = 0$. The upper bound for X required for the numerical solution is $X_{\max} = 1121$.

approximations for the joint probability mass function $P(X, s, t)$ of the copy numbers X and s . In order to validate the approximations, we also calculate the PMF numerically. This requires restricting the model in X like we did in s , i.e. to a finite state space $\{0, 1, \dots, X_{\max}\} \times \{0, 1, \dots, s_{\max}\}$. For the finite process, the bursty increases (2) apply only if the burst size satisfies $j < X_{\max} - X$, with larger bursts being redirected to the boundary as per $W(X_{\max}, s|X, s) = f_s \sum_{j=X_{\max}-X}^{\infty} b_j$. The Kolmogorov forward equation (1) is then a finite (albeit large) linear system of differential equations, and is solved by numerical integration.

Figure 3 shows the evolution of the marginals of a solution which is initialised to zero copy numbers, i.e. $P(X, s, 0) = \delta_{X,0} \delta_{s,0}$. The zero initial state lies in the region of attraction of the lower stable fixed point x_- . The early dynamics is characterised by a steady shift of the distributions to the right until they are located at x_- . This initial dynamics is closely approximated by the LNA (40) (Figure 3, first three timepoints). At much greater times, the weight of the mode around x_- begins to decrease, and a new mode around the upper stable fixed point x_+ arises. This later dynamics is closely approximated by the mixture approximation (46) (Figure 3, last three timepoints); for reasons that will become apparent in the next section we refer to the mixture as the WKB-based approximation. At the final timepoint, a steady state is reached, which is characterised by a balance of probability transfer between the two modes.

4.1 WKB approximation

The determination of $\lambda_{\pm}$ rests on a uniformly valid approximation to the stationary distribution that has been provided in an earlier work [29] in the WKB

form

$$\mathbf{p}(x; \varepsilon) \sim C(\varepsilon)k(x) \exp\left(-\frac{\Phi(x)}{\varepsilon}\right) \mathbf{w}(x). \quad (52)$$

For a detailed exposition how the individual terms in (52) can be calculated we refer the reader to the original paper. Here we merely summarise the most important properties of these terms. The function $\Phi(x)$, which is referred to as the potential, plays a key role in the approximation. Small differences in the potential are exponentially exacerbated in (52); it follows that the main contribution toward the total probability mass comes from around the local minima of the potential. Importantly, [29] showed that $\Phi(x)$ has local minima at the stable fixed points $x_{\pm}$ of the differential equation (30), separated by a local maximum at the unstable fixed point x_0 (Figure 2, right panel). The vector $\mathbf{w}(x) > 0$ in (52) satisfies $\mathbf{1}^T \mathbf{w}(x) = 1$ and its components provide an approximation to the stationary distribution of the active protein copy number conditioned on a given concentration x of the inactive protein. The function $k(x) > 0$ is a prefactor, and the coefficient $C(\varepsilon) > 0$ is a normalisation constant.

Assume for concreteness that the protein concentration is initialised at $t = 0$ to a value $x_{\text{init}} < x_0$. After the elapse of an initial $O(1)$ temporal transient, the concentration x equilibrates to a quasi-stationary distribution located around the lower fixed point x_- . The quasi-stationary distribution coincides with the stationary distribution (52) for $x < x_0$, and is equal to zero for $x > x_0$; the normalisation constant $C(\varepsilon)$ is therefore calculated by integrating (52) over $0 < x < x_0$ only, i.e.

$$C(\varepsilon) = \left(\int_0^{x_0} \mathbf{1}^T \mathbf{p}(x; \varepsilon) dx \right)^{-1} \sim \frac{1}{k(x_-)} \sqrt{\frac{\Phi''(x_-)}{2\pi\varepsilon}} \exp\left(\frac{\Phi(x_-)}{\varepsilon}\right), \quad (53)$$

in which the approximation is due to the Laplace method. The truncation of (52) to zero for $x > x_0$ introduces a discontinuity at x_0 in the quasi-steady-state distribution. The discontinuity can be resolved by constructing in a small neighbourhood of x_0 an inner approximation to the quasi-stationary distribution that asymptotically matches (52) to its left and zero to its right. Importantly, the construction of the inner solution and the matching procedure will help quantify the rate of probability leakage λ_- across the potential barrier x_0 . This is done in the next two subsections.

The quasi-stationary distribution can be alternatively approximated by the large- t limit of the LNA (44). Indeed, comparing the behaviour of (44) and (52) around the stable fixed points, one derives consistency relations

$$\mathbf{w}(x_{\pm}) = \boldsymbol{\rho}(x_{\pm}), \quad \Phi''(x_{\pm}) = \frac{1}{\sigma_{\pm}^2}, \quad (54)$$

which have previously been obtained by [29] by other means. The first equality of (54) means that the WKB and LNA approaches consistently predict the same Poisson distributions of active protein if conditioned on a fixed-point concentration of the inactive protein; the second equality of (54) says that the linearised noise around a fixed point increases as the WKB potential well flattens. While the LNA and the WKB approximation are both appropriate around the fixed points, only the WKB approximation provides an adequate approximation at the tails of the distribution away from the fixed points. Since the exchange of

probability mass between the fixed points occur through the tails, it is necessary to use the uniformly valid WKB approximation to obtain the correct rate of leakage. Nevertheless, the LNA methodology, aside from providing time-dependent approximations for $t = O(1)$, plays a key role in the construction of the inner approximation to the quasi-stationary distribution, as explained below.

4.2 Inner approximation

Applying the transformation (14) with the deterministic motion set constantly to the unstable fixed point $\xi(t) \equiv x_0$, and carrying through with the expansion procedure, we approximate via (16), (19), and (24) the quasi-stationary distribution by

$$\mathbf{p}(x; \varepsilon) \sim \varepsilon^{-1/2} g\left(\varepsilon^{-1/2}(x - x_0)\right) \rho(x_0), \quad (55)$$

where $g = g(y)$ is a solution to the time-independent version of the Fokker-Planck equation (34),

$$\alpha(yg)' - \beta g'' = 0 \quad (56)$$

with $\alpha = \alpha(x_0)$ and $\beta = \beta(x_0)$ given by (35)–(36). Reducing the order of (56) gives

$$\alpha yg - \beta g' = \lambda_-, \quad (57)$$

where λ_- is a constant of integration. Solving (57) in g subject to the terminal condition $g(\infty) = 0$ yields

$$g(y) = \frac{\lambda_-}{\beta} \exp\left(\frac{\alpha y^2}{2\beta}\right) \int_y^\infty \exp\left(-\frac{\alpha \zeta^2}{2\beta}\right) d\zeta. \quad (58)$$

The integration constant λ_- is identified with the metastable transition rate as it gives the (drift-diffusion) probability current — the amount of probability that is being transferred per unit time through the unstable fixed point in the direction of increasing x . The value of the probability current is determined in the next subsection by an asymptotic matching of (58) to the WKB formula for the quasi-stationary distribution.

4.3 Asymptotic matching

The WKB approximation (52) and (53) to the quasi-stationary distribution applies for values of x that lie to the left of the unstable fixed point x_0 . On the other hand, the LNA approximation (55) and (58) applies for x in the vicinity of x_0 . Asymptotic matching principle states that the two approximations should be equivalent in an overlap region, which is here given by

$$\varepsilon^{1/2} \ll x_0 - x \ll 1. \quad (59)$$

The LNA approximation (55) and (58) simplifies in the overlap region (59) to

$$\mathbf{p}(x; \varepsilon) \sim \lambda_- \sqrt{\frac{2\pi}{\varepsilon\alpha\beta}} \exp\left(\frac{\alpha(x - x_0)^2}{2\varepsilon\beta}\right) \rho(x_0) \quad (60)$$

and the WKB approximation (52) and (53) to

$$\mathbf{p}(x; \varepsilon) \sim \frac{k(x_0)}{k(x_-)} \sqrt{\frac{\Phi''(x_-)}{2\pi\varepsilon}} \exp\left(-\frac{\Phi(x_0) - \Phi(x_-)}{\varepsilon} - \frac{\Phi''(x_0)(x - x_0)^2}{2\varepsilon}\right) \mathbf{w}(x_0). \quad (61)$$

Comparing (60) and (61) yields

$$\mathbf{w}(x_0) = \boldsymbol{\rho}(x_0), \quad \frac{\alpha(x_0)}{\beta(x_0)} = -\Phi''(x_0), \quad (62)$$

$$\lambda_- = \frac{\alpha(x_0)}{2\pi} \frac{k(x_0)}{k(x_-)} \sqrt{-\frac{\Phi''(x_-)}{\Phi''(x_0)}} \exp\left(-\frac{\Phi(x_0) - \Phi(x_-)}{\varepsilon}\right). \quad (63)$$

Since $\Phi(x_0) > \Phi(x_-)$, the current (63) is exponentially small as $\varepsilon \rightarrow 0$ as advertised in (50). The WKB terms $k(x_-)$, $k(x_0)$, and $\Phi(x_0) - \Phi(x_-)$ can be numerically determined as described by [29]. The formula for λ_+ is obtained by replacing the symbol ‘−’ by ‘+’ wherever it appears in formula (63).

5 Early-time approximation

For completeness, we provide an approximation to the joint protein distribution that applies during the small initial transient $t = O(\varepsilon)$. First, we introduce the fast time variable by

$$t = \varepsilon\tau. \quad (64)$$

Inserting (64) into the master equation (10) yields

$$\partial_\tau \mathbf{p} = \mathbf{A}(x, -\varepsilon \partial_x) \mathbf{p}. \quad (65)$$

Taking $\varepsilon \rightarrow 0$ in (65), and solving the limiting equation yields an approximation

$$\mathbf{p}(x, \tau) \sim \exp(\mathbf{A}(x, 0)\tau) \mathbf{p}(x, 0). \quad (66)$$

Expressing (66) in terms of the original time variable for deterministic initial conditions, we arrive at

$$p(x, s, t) \sim \delta(x - x_{\text{init}}) [\exp(\mathbf{A}(x_{\text{init}}, 0)t/\varepsilon)]_{s, s_{\text{init}}}, \quad t \ll 1. \quad (67)$$

The exponential of the birth–death process’s (Figure 1) Markovian generator matrix in (67) can for $\varepsilon \ll t$ be approximated by the its stationary distribution $\rho_s(x_{\text{init}})$; making this approximation recovers (41). Thus, provided that time is small but sufficiently larger than ε , the birth–death process dynamics of (67) and the LNA approximation (40) do indeed overlap.

6 Discussion

The paper describes the stochastic dynamics of a bursty protein with delayed feedback in the large-delay regime. A key observation is that long delays lead to a large build-up of protein molecules which have already been produced but not yet activated. The reservoir of inactive protein molecules confers stability to random perturbations on the inactive protein. Indeed, in the limit of very

large delays, the inactive protein evolves deterministically according to an autonomous ordinary differential equation (30). Meanwhile, the number of active protein fluctuates fast, being “slaved” — borrowing the metaphor from [61] — by the inactive protein. Provided that the protein sustains its expression via a sufficiently cooperative positive feedback loop, the limiting differential equation possesses two stable fixed points, a lower one of basal expression and a higher one of up-regulated expression. Bistability generated by a positive feedback can implement molecular memory in living cells and serve to maintain cell-fate decisions [62–65].

The delayed regulatory circuit is a multiscale problem with three distinct timescales: the fast timescale of the active protein turnover, the slow timescale of the inactive protein turnover, and the very slow timescale of metastable transitions between the fixed points of the deterministic motion. As the main mathematical result of the paper, we develop on each of the three timescales an approximation to the joint probability distribution of inactive and active protein abundances. On the fast timescale, the inactive protein stays constant, while the active protein equilibrates to the Poisson distribution (67). On the slow timescale, the inactive and active protein species follow respectively the Gaussian and the Poisson distributions (40) located around the deterministic motion. On the very slow timescale, the inactive protein distribution is a mixture of Gaussians, and the active protein distribution is a mixture of Poisson modes (46) located at the fixed points of the deterministic motion. The weights of the modes evolve in time according to a very slow two-state switch (a random telegraph process), the holding times of which exponentially lengthen as the delay increases. Therefore, even moderately large delays can implement an extremely stable switching mechanism that can function as a basic molecular memory unit in a living cell.

As part of our asymptotic analysis, we demonstrated that the successive timescales share a common overlap domain on which the respective approximations provide consistent results. On the one hand, the existence of the overlap demonstrates the completeness of our approximation scheme in the sense of covering the whole temporal frame; on the other hand, the requirement that the terminal behaviour of a preceding approximation is continued by the successive approximation determines the initial condition for the latter. Asymptotic matching is used not only in the temporal but also in the protein domain, where it serves to determine the rates of metastable escape from the basins of attraction of the deterministic fixed points. The procedure, which has previously been applied to large-volume [36, 37, 39] and fast-switching systems [44, 45], involves outer and inner approximations to the quasi-stationary protein distribution. The outer approximation, which applies within the basin, is calculated by the WKB method following [29]. The inner approximation, which is valid in a narrow neighbourhood of the unstable fixed point that separates the basins, is found as a constant-flux solution to a Fokker–Planck equation. The probability flux of the inner solution is determined from matching the inner to the outer approximation; the value of the flux provides the desired metastable rate of escape from the basin.

In summary, our study demonstrates that extending a noisy autoregulatory loop by a sufficiently large production delay can form a stable toggle switch. Using a matched asymptotic approach, our study quantifies the effect and connects the metastable switching to shorter-term dynamics. Finally, we expect that the

phenomenon is present not only in case of a single gene, but also appears for mutually repressive genes, and other manifestations of positive feedback systems.

Acknowledgement

The author thanks Abhyudai Singh, Pasquale Palumbo, and Alessandro Borri for discussions which influenced the paper.

References

- [1] L. Cai, N. Friedman, and X. Xie, “Stochastic protein expression in individual cells at the single molecule level,” *Nature*, vol. 440, pp. 358–362, 2006.
- [2] Y. Taniguchi, P. Choi, G. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. Xie, “Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells,” *Science*, vol. 329, pp. 533–538, 2010.
- [3] G.-W. Li and X. S. Xie, “Central dogma at the single-molecule level in living cells,” *Nature*, vol. 475, no. 7356, pp. 308–315, 2011.
- [4] M. Thattai and A. van Oudenaarden, “Intrinsic noise in gene regulatory networks,” *P. Natl. Acad. Sci. USA*, vol. 98, p. 151588598, 2001.
- [5] B. Munsky, G. Neuert, and A. Van Oudenaarden, “Using gene expression noise to understand gene regulation,” *Science*, vol. 336, pp. 183–187, 2012.
- [6] N. Kumar and R. V. Kulkarni, “Constraining the complexity of promoter dynamics using fluctuations in gene expression,” *Phys. Biol.*, vol. 17, no. 1, p. 015001, 2019.
- [7] J. Rodriguez and D. R. Larson, “Transcription in living cells: Molecular mechanisms of bursting,” *Annu. Rev. Biochem.*, vol. 89, 2020.
- [8] A. Giaretta, G. M. Toffolo, and T. C. Elston, “Stochastic modeling of human papillomavirusearly promoter gene regulation,” *J. Theor. Biol.*, vol. 486, p. 110057, 2020.
- [9] M. R. Roussel and R. Zhu, “Stochastic kinetics description of a simple transcription model,” *B. Math. Biol.*, vol. 68, no. 7, pp. 1681–1713, 2006.
- [10] V. Belitsky and G. Schütz, “Rna polymerase interactions and elongation rate,” *J. Theor. Biol.*, vol. 462, pp. 370–380, 2019.
- [11] M. Gallego and D. M. Virshup, “Post-translational modifications regulate the ticking of the circadian clock,” *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 2, pp. 139–148, 2007.
- [12] A. Mor, S. Suliman, R. Ben-Yishay, S. Yunger, Y. Brody, and Y. Shav-Tal, “Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells,” *Nat. Cell Biol.*, vol. 12, no. 6, p. 543, 2010.

- [13] N. Battich, T. Stoeger, and L. Pelkmans, “Control of transcript variability in single mammalian cells,” *Cell*, vol. 163, no. 7, pp. 1596–1610, 2015.
- [14] M. Sturrock, S. Li, and V. Shahrezaei, “The influence of nuclear compartmentalisation on stochastic dynamics of self-repressing gene expression,” *J. Theor. Biol.*, vol. 424, pp. 55–72, 2017.
- [15] M. Barrio, K. Burrage, A. Leier, and T. Tian, “Oscillatory regulation of hes1: discrete stochastic delay modelling and simulation,” *PLoS Comput. Biol.*, vol. 2, p. e117, 2006.
- [16] E. Zavala and T. T. Marquez-Lago, “Delays induce novel stochastic effects in negative feedback gene circuits,” *Biophys. J.*, vol. 106, no. 2, pp. 467–478, 2014.
- [17] K. Parmar, K. B. Blyuss, Y. N. Kyrychko, and S. J. Hogan, “Time-delayed models of gene regulatory networks,” *Comput. Math. Method. M.*, vol. 2015, 2015.
- [18] R. Martinez-Corral, E. Raimundez, Y. Lin, M. B. Elowitz, and J. Garcia-Ojalvo, “Self-amplifying pulsatile protein dynamics without positive feedback,” *Cell Syst.*, vol. 7, no. 4, pp. 453–462, 2018.
- [19] J. Wang, M. Lefranc, and Q. Thommen, “Stochastic oscillations induced by intrinsic fluctuations in a self-repressing gene,” *Biophys. J.*, vol. 107, no. 10, pp. 2403–2416, 2014.
- [20] R. E. Baker and G. Röst, “Global dynamics of a novel delayed logistic equation arising from cell biology,” *J. Nonlinear Sci.*, vol. 30, no. 1, pp. 397–418, 2020.
- [21] L. Lafuerza and R. Toral, “Role of delay in the stochastic creation process,” *Phys. Rev. E*, vol. 84, p. 021128, 2011.
- [22] C. Gupta, J. M. López, W. Ott, K. Josić, and M. R. Bennett, “Transcriptional delay stabilizes bistable gene networks,” *Phys. Rev. Lett.*, vol. 111, no. 5, p. 058104, 2013.
- [23] J. Feng, S. A. Sevier, B. Huang, D. Jia, and H. Levine, “Modeling delayed processes in biological systems,” *Phys. Rev. E*, vol. 94, no. 3, p. 032408, 2016.
- [24] Y. Kyrychko and I. Schwartz, “Enhancing noise-induced switching times in systems with distributed delays,” *Chaos*, vol. 28, no. 6, p. 063106, 2018.
- [25] A. Singh and P. Bokes, “Consequences of mRNA transport on stochastic variability in protein levels,” *Biophys. J.*, vol. 103, pp. 1087–1096, 2012.
- [26] T. Stoeger, N. Battich, and L. Pelkmans, “Passive noise filtering by cellular compartmentalization,” *Cell*, vol. 164, no. 6, pp. 1151–1161, 2016.
- [27] A. Borri, P. Palumbo, and A. Singh, “Time delays in a genetic positive-feedback circuit,” *IEEE Control Syst. Lett.*, vol. 4, no. 1, pp. 163–168, 2019.

- [28] M. Smith and A. Singh, “Stochastic delays suppress noise in a genetic circuit with negative feedback,” *bioRxiv*, p. 786491, 2019.
- [29] P. Bokes, A. Borri, P. Palumbo, and A. Singh, “Mixture distributions in a stochastic gene expression model with delayed feedback: a WKB approximation approach,” *J. Math. Biol. (to appear in)*, preprint: <https://doi.org/10.1101/855783> (2019).
- [30] T. G. Kurtz, “The relationship between stochastic and deterministic models for chemical reactions,” *J. Chem. Phys.*, vol. 57, no. 7, pp. 2976–2978, 1972.
- [31] A. Andreychenko, L. Bortolussi, R. Grima, P. Thomas, and V. Wolf, “Distribution approximations for the chemical master equation: comparison of the method of moments and the system size expansion,” in *Modeling Cellular Systems*, pp. 39–66, Springer, 2017.
- [32] N. van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, 2006.
- [33] P. Thomas and R. Grima, “Approximate probability distributions of the master equation,” *Phys. Rev. E*, vol. 92, p. 012120, 2015.
- [34] P. Thomas, R. Grima, and A. V. Straube, “Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators,” *Phys. Rev. E*, vol. 86, no. 4, p. 041110, 2012.
- [35] N. Herath and D. Del Vecchio, “Reduced linear noise approximation for biochemical reaction networks with time-scale separation: The stochastic tqssa+,” *J. Chem. Phys.*, vol. 148, no. 9, p. 094108, 2018.
- [36] R. Hinch and S. J. Chapman, “Exponentially slow transitions on a Markov chain: the frequency of calcium sparks,” *Eur. J. Appl. Math.*, vol. 16, no. 04, pp. 427–446, 2005.
- [37] C. Escudero and A. Kamenev, “Switching rates of multistep reactions,” *Phys. Rev. E*, vol. 79, no. 4, p. 041149, 2009.
- [38] M. Assaf, E. Roberts, and Z. Luthey-Schulten, “Determining the stability of genetic switches: explicitly accounting for mRNA noise,” *Phys. Rev. Lett.*, vol. 106, no. 24, p. 248102, 2011.
- [39] P. C. Bressloff, *Stochastic processes in cell biology*. Springer, New York, 2014.
- [40] C. Çelik, P. Bokes, and A. Singh, “Stationary distributions and metastable behaviour for self-regulating proteins with general lifetime distributions,” *bioRxiv*, 2020.
- [41] E. Wallace, D. Gillespie, K. R. Sanft, and L. Petzold, “Linear noise approximation is valid over limited times for any chemical system that is sufficiently large,” *IET Syst. Biol.*, vol. 6, no. 4, pp. 102–115, 2012.
- [42] M. Assaf and B. Meerson, “WKB theory of large deviations in stochastic populations,” *J. Phys. A: Math. Theor.*, vol. 50, no. 26, p. 263001, 2017.

- [43] T. Kepler and T. Elston, “Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations,” *Biophys. J.*, vol. 81, pp. 3116–3136, 2001.
- [44] J. Newby and S. J. Chapman, “Metastable behavior in Markov processes with internal states,” *J. Math. Biol.*, vol. 69, no. 4, pp. 941–976, 2014.
- [45] J. Newby, “Bistable switching asymptotics for the self regulating gene,” *J. Phys. A-math. Gen.*, vol. 48, p. 185001, 2015.
- [46] P. G. Hufton, Y. T. Lin, and T. Galla, “Classical stochastic systems with fast-switching environments: Reduced master equations, their interpretation, and limits of validity,” *Phys. Rev. E*, vol. 99, no. 3, p. 032121, 2019.
- [47] P. G. Hufton, Y. T. Lin, and T. Galla, “Model reduction methods for population dynamics with fast-switching environments: Reduced master equations, stochastic differential equations, and applications,” *Phys. Rev. E*, vol. 99, no. 3, p. 032122, 2019.
- [48] T. Lipniacki, P. Paszek, A. Marciniak-Czochra, A. Brasier, and M. Kimmel, “Transcriptional stochasticity in gene expression,” *J. Theor. Biol.*, vol. 238, pp. 348–67, 2006.
- [49] S. Smith, C. Cianci, and R. Grima, “Model reduction for stochastic chemical systems with abundant species,” *J. Chem. Phys.*, vol. 143, no. 21, p. 12B615.1, 2015.
- [50] P. G. Hufton, Y. T. Lin, T. Galla, and A. J. McKane, “Intrinsic noise in systems with switching environments,” *Phys. Rev. E*, vol. 93, p. 052119, 2016.
- [51] P. Kurasov, A. Lück, D. Mugnolo, and V. Wolf, “Stochastic hybrid models of gene regulatory networks — a PDE approach,” *Math. Biosci.*, vol. 305, pp. 170–177, 2018.
- [52] C. Jia and R. Grima, “Dynamical phase diagram of an auto-regulating gene in fast switching conditions,” *J. Chem. Phys.*, vol. 152, no. 17, p. 174110, 2020.
- [53] T. Plesa, R. Erban, and H. G. Othmer, “Noise-induced mixing and multi-modality in reaction networks,” *Eur. J. Appl. Math.*, vol. 30, no. 5, pp. 887–911, 2019.
- [54] M. A. Al-Radhawi, D. Del Vecchio, and E. D. Sontag, “Multi-modality in gene regulatory networks with slow promoter kinetics,” *Plos Comput. Biol.*, vol. 15, no. 2, p. e1006784, 2019.
- [55] A. G. Hart and R. L. Tweedie, “Convergence of invariant measures of truncation approximations to markov processes,” *Appl. Math.*, vol. 3, p. 26113, 2012.
- [56] A. Gupta, J. Mikelson, and M. Khammash, “A finite state projection algorithm for the stationary solution of the chemical master equation,” *J. Chem. Phys.*, vol. 147, no. 15, p. 154101, 2017.

- [57] Y. Satin, A. Zeifman, A. Korotysheva, and K. Kiseleva, “Two-sided truncations for a class of continuous-time Markov chains,” in *International Conference on Information Technologies and Mathematical Modelling* (A. Dudin *et al.*, eds.), pp. 312–323, Springer, 2017.
- [58] D. J. Warne, R. E. Baker, and M. J. Simpson, “Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art,” *J. Roy. Soc. Interface*, vol. 16, no. 151, p. 20180943, 2019.
- [59] I. Lestas, J. Paulsson, N. Ross, and G. Vinnicombe, “Noise in gene regulatory networks,” *IEEE T. Circuits-I*, vol. 53, no. 1, pp. 189–200, 2008.
- [60] J. Kevorkian and J. Cole, *Perturbation Methods in Applied Mathematics*. Springer, 1981.
- [61] H. Haken, “Synergetics,” in *Self-Organizing Systems*, pp. 417–434, Springer, 1987.
- [62] A. Becskei, B. Séraphin, and L. Serrano, “Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion,” *EMBO J.*, vol. 20, pp. 2528–35, 2001.
- [63] R. Kerr, S. Jabbari, and I. G. Johnston, “Intracellular energy variability modulates cellular decision-making capacity,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.
- [64] K. H. Chong, S. Samarasinghe, D. Kulasiri, and J. Zheng, “Mathematical modelling of core regulatory mechanism in p53 protein that activates apoptotic switch,” *J. Theor. Biol.*, vol. 462, pp. 134–147, 2019.
- [65] Y. Xin, B. Cummins, and T. Gedeon, “Multistability in the epithelial-mesenchymal transition network,” *BMC Bioinformatics*, vol. 21, no. 1, p. 71, 2020.