

# Mixture distributions in a stochastic gene expression model with delayed feedback

Pavol Bokes<sup>1</sup>, Alessandro Borri<sup>2</sup>, Pasquale Palumbo<sup>2,3</sup>, and Abhyudai Singh<sup>4</sup>

<sup>1</sup>Comenius University, Bratislava, Slovakia

<sup>2</sup>Italian National Research Council (IASI-CNR), Roma, Italy

<sup>3</sup>University of Milano-Bicocca, Milano, Italy

<sup>4</sup>University of Delaware, Newark, Delaware, US

November 26, 2019

## Abstract

Noise in gene expression can be substantively affected by the presence of production delay. Here we consider a mathematical model with bursty production of protein, a one-step production delay (the passage of which activates the protein), and feedback in the frequency of bursts. We specifically focus on examining the steady-state behaviour of the model in the slow-activation (i.e. large-delay) regime. Using a quasi-steady-state (QSS) approximation, we derive an autonomous ordinary differential equation for the inactive protein that applies in the slow-activation regime. If the differential equation is monostable, the steady-state distribution of the inactive (active) protein is approximated by a single Gaussian (Poisson) mode located at the globally stable steady state of the differential equation. If the differential equation is bistable (due to cooperative positive feedback), the steady-state distribution of the inactive (active) protein is approximated by a mixture of Gaussian (Poisson) modes located at the stable steady states; the weights of the modes are determined from a WKB approximation to the stationary distribution. The asymptotic results are compared to numerical solutions of the chemical master equation.

**Keywords:** Stochastic gene expression; Bursting; Production delay; WKB approximation; Large deviations; Exponential asymptotics

**MSC 2010:** 92C40; 60J27; 41A60

## 1 Introduction

Gene expression in individual cells involves the interaction of molecules which are present at low copy numbers (Eldar and Elowitz, 2010; Munsky et al, 2012). The intrinsic noise generated by the low-copy-number reactions is passed down to the end product of gene expression, the protein, and results in temporal fluctuations and cell-to-cell heterogeneity of the protein copy number (Taniguchi et al, 2010; Suter et al, 2011). The production of proteins in bursts of multiple copies is one of the most important sources of protein variability (Singh

et al, 2010; Dar et al, 2012). Specifically, bursty production accounts for super-Poissonian variability observed in protein copy numbers (Thattai and van Oudenaarden, 2001).

Mathematical modelling has been proved useful in understanding the mechanisms of stochastic gene expression. The underlying probability distributions are typically defined as solutions to a specific master equation (Paulsson, 2005; Veerman et al, 2018; Albert, 2019). Explicit solutions to the master equation, especially at steady state, can be found for models with few components (Bokes et al, 2012; Zhou and Liu, 2015) and/or with special structural properties (Kumar et al, 2015; Anderson and Cotter, 2016). Generally, however, explicit solutions are unavailable or intractable and one resorts to stochastic simulation or seeks a numerical solution to a finite truncation of the master equation (Munsky and Khammash, 2006; Borri et al, 2016; Gupta et al, 2017). An alternative approach, which often provides useful qualitative insights into the model behaviour, is based on reduction techniques such as quasi-steady-state (Srivastava et al, 2011; Kim et al, 2014) and adiabatic reductions (Bruna et al, 2014; Popovic et al, 2016), piecewise-deterministic framework (Lin and Doering, 2016; Lin and Buchler, 2018), linear-noise approximation (Schnoerr et al, 2017; Modi et al, 2018), or moment closure (Singh and Hespanha, 2007; Andreychenko et al, 2017; Gast et al, 2019).

Production delay is an inevitable part of gene expression (Monk, 2003; Zavala and Marquez-Lago, 2014; Bokes et al, 2018). It can be caused by a number of mechanisms, e.g. transcriptional/translational elongation (Roussel and Zhu, 2006), post-translational modification (Gedeon and Bokes, 2012), or compartmental transport (Mor et al, 2010; Sturrock et al, 2017). The delay specifies the amount of time that needs to pass before a newly produced molecule can partake in its regulatory function (specifically in feedback). Delay can be fixed or randomly chosen from a distribution (Barrio et al, 2006; Lafuerza and Toral, 2011; Gupta et al, 2014). Exponentially distributed delays are the simplest among distributed delays as they are realised by the passage of a single memoryless step. Erlang and phase-type distributions provide a wider family of distributed delays which can be generated by a finite network of memoryless states (Soltani et al, 2016). Previous results indicate that large one-step (exponential) and multi-step (Erlang/phase-type) delays reduce the super-Poissonian noise in a bursty protein down to Poissonian levels (Singh and Bokes, 2012; Stoeger et al, 2016; Smith and Singh, 2019). This effect is also seen experimentally with buffered noise in cytoplasmic mRNA levels compared to nuclear mRNA levels due to transport delays (Battich et al, 2015). Additional effects of the inclusion of a delay are observed if the protein regulates, via transcriptional feedback, its burst frequency. In case of negative feedback, delays of moderate size lead to an increase, rather than decrease in protein noise (Smith and Singh, 2019). In case of non-cooperative positive feedback, noise-driven bimodal protein distributions, which are observed in the absence of delay, turn unimodal upon the inclusion of a distributed delay, and eventually converge to the Poissonian statistics as the delay increases (Borri et al, 2019). In case of cooperative positive feedback, the introduction of delay has been reported to enhance the stability of the modes of the protein distribution (Gupta et al, 2013; Feng et al, 2016; Kyrychko and Schwartz, 2018).

In the paper we focus, for its relative simplicity, on the case of exponential delay. We refer to the delay as activation and distinguish between the inactive

Name	Scheme	Rate	Reset map
Production	$\emptyset \rightarrow B \times X$	$\varepsilon^{-1} f_s$	$X \rightarrow X + B$
Activation	$X \rightarrow S$	$X$	$X \rightarrow X - 1,$ $s \rightarrow \min\{s + 1, s_{\max}\}$
Decay	$S \rightarrow \emptyset$	$\varepsilon^{-1} s$	$s \rightarrow s - 1$

Table 1: Reaction channels in the delayed feedback model. The copy number of the inactive protein  $X$  is denoted by  $X$  (in italic type) and that of the active protein  $S$  by  $s$ . The production burst rate  $f_s$  is a function of  $s$ , which implements the feedback. The production burst size  $B$  is in general drawn from a prescribed random distribution. The parameter  $\varepsilon \ll 1$  determines the discrepancy between the  $O(1)$  slow activation timescale and the  $O(\varepsilon)$  fast timescale of turnover of the active protein  $S$ . The reset map of the activation channel ensures that  $s$  never exceeds an upper bound  $s_{\max}$ .

and active protein species. We will argue that in the limit of slow activation rates the model behaviour becomes deterministic at the level of the inactive protein. Behaviour of stochastic models near a deterministic limit can be interpreted using the large-deviation theory (Tsimring and Pikovsky, 2001; Heymann and Vanden-Eijnden, 2008; Kumar and Kulkarni, 2019) and quantified by WKB asymptotic approximations (Schuss, 2009; Bressloff, 2014; Assaf and Meerson, 2017). The WKB approach has been successfully applied to stochastic reaction kinetics systems with large molecule copy numbers (Hinch and Chapman, 2005; Be’er and Assaf, 2016; Yin and Wen, 2019), fast switching of internal states (Newby, 2012; Lin and Galla, 2016), or a combination of both (Newby and Chapman, 2014, discussed at greater length in Section 7). Here we will use the WKB approximation to obtain reliable estimates of the stationary distribution of the active (and also the inactive) protein in the slow activation (large-delay) regime.

The outline of the paper is as follows. The stochastic model is introduced in Section 2 and reduced in Section 3 to a deterministic rate equation by means of a quasi-steady-state (QSS) reduction. Section 4 introduces the WKB solution and outlines the key aspects of the WKB analysis. Section 5 compares the WKB and numerical solutions to the master equation. The WKB analysis is performed in Section 6. Section 7 concludes the paper with a discussion.

## 2 Model formulation

The paper is concerned with a reaction system involving an inactive protein  $X$  and an active protein  $S$  which are subject to the production, activation, and decay reaction channels (Table 1). Each reaction is specified by its rate and reset map. The reaction rate, upon the multiplication by the length of an infinitesimally short time interval, gives the probability that the reaction will occur within the interval. The reset map determines the adjustment of the reaction species copy numbers resulting from an occurrence of the reaction.

We are specifically interested in studying the model in the regime of slow

activation. Making activation slow is equivalent to making all the remaining reactions fast: indeed, by Table 1, the activation rate is  $O(1)$ , whereas the production and decay rates are  $O(1/\varepsilon)$ , where  $\varepsilon \ll 1$  is a small dimensionless parameter. The aim of the paper is to find asymptotic approximations, valid for  $\varepsilon \ll 1$ , of the stationary behaviour of the model.

Let us talk through the specific forms of the reaction rates and reset maps of the individual reactions in Table 1. The production rate depends on the number  $s$  of active protein through a general (integer-valued) feedback response function  $f_s$ . The production reset map indicates that the number  $X$  of inactive protein is increased by the size  $B \geq 1$  of a production burst. Bursts sizes are drawn (independently of each other) from a distribution

$$b_j = P[B = j], \quad j = 1, 2, \dots \quad (1)$$

The activation rate is proportional to the number of inactive protein; the factor of proportionality is set to one without loss of generality. The activation reset map turns one inactive protein into an active protein if there is extra capacity for active protein ( $s < s_{\max}$ ); it removes an inactive protein without creating an active protein if there is no capacity ( $s = s_{\max}$ ). The importance of introducing an upper bound  $s_{\max}$  on the number of active protein is discussed at greater length in Section 7. The decay rate is proportional to the number of active protein; the decay reset map decreases the number of active protein by one.

### 3 QSS approximation

In the slow-activation regime ( $\varepsilon \ll 1$ ), the inactive protein  $X$  is present at a copy number  $X$  (written in *italics*) which is  $O(\varepsilon^{-1})$  large. In order to measure the abundance of species  $X$  on an  $O(1)$  scale, we define

$$x = \varepsilon X, \quad (2)$$

which we refer to as the concentration of the inactive protein  $X$ . In the limit of  $\varepsilon \rightarrow 0$ , the concentration  $x$  becomes a continuous quantity.

An additional effect of small values of  $\varepsilon$  is that the turnover of the active protein  $S$  becomes much faster than that of the inactive protein  $X$ . This difference in timescales leads to seeking a quasi-steady-state (QSS) reduction of the model. On the fast timescale of the turnover of  $S$ , we treat the concentration  $x$  of the slow species  $X$  as a fixed quantity. With this assumption, the dynamics of  $S$  is that of an  $M/M/s_{\max}/s_{\max}$  server with memoryless arrival and service times,  $s_{\max}$  servers, and no queue (Gross, 2008). This analogy identifies active protein molecules  $S$  with customers, protein activation with customer arrival (occurring with a constant rate  $\varepsilon^{-1}x$ ), and protein decay with customer service (with rate  $\varepsilon^{-1}$  per customer). A well-known result from the queuing theory implies that the QSS distribution of  $s$  is

$$\rho_s(x) = \mathcal{N}(x) \frac{x^s}{s!}, \quad s = 0, 1, \dots, s_{\max}, \quad (3)$$

where

$$\mathcal{N}(x) = \left( \sum_{s=0}^{s_{\max}} \frac{x^s}{s!} \right)^{-1} \quad (4)$$

is the normalisation constant. Thus, fixing the slow species  $X$  to a concentration  $x$ , the fast species copy number  $s$  obeys the (truncated) Poisson distribution (3) with location parameter  $x$ .

Given a concentration  $x$  of the inactive protein  $X$ , the effective production burst rate is equal to  $\varepsilon^{-1}\bar{f}(x)$ , where  $\bar{f}(x)$  is the expectation of  $f_s$  with respect to the QSS distribution (3), i.e.

$$\bar{f}(x) = \sum_{s=0}^{s_{\max}} f_s \rho_s(x). \quad (5)$$

The value of  $\bar{f}(x)$  is determined by all values of  $f_s$ : for example,  $\bar{f}(2)$  includes, with a small weight, the value  $f_{10}$ ; conversely,  $f_2$  contributes a little towards  $\bar{f}(10)$ . Nevertheless, the main contribution to  $\bar{f}(x)$  comes from the values of  $f_s$  whose argument  $s$  is close to  $x$ . Due to the contributions of neighbouring terms, any sharp features of  $f_s$  are “mollified” in the function  $\bar{f}(x)$ ; for example, a step function (Gedeon et al, 2017; Crawford-Kahrl et al, 2019)

$$f_s = \begin{cases} a_0 & \text{if } s < s_{\text{thresh}}, \\ a_1 & \text{if } s \geq s_{\text{thresh}}, \end{cases} \quad (6)$$

turns into a smooth sigmoid function by the application of (5); see Figure 1, top panels.

The concentration of protein produced per unit time is equal to the product of the effective burst rate  $\varepsilon^{-1}\bar{f}(x)$  and the mean burst size  $\varepsilon\langle B \rangle$  (in units of concentration), where

$$\langle B \rangle = \sum_{j=1}^{\infty} j b_j \quad (7)$$

is the mean burst size (in units of copy number). Subtracting from the product  $\langle B \rangle \bar{f}(x)$  the linear activation rate yields the limiting deterministic dynamics

$$\frac{dx}{dt} = \langle B \rangle \bar{f}(x) - x \quad (8)$$

for the inactive protein. Steady states of (8) are the intersections of  $\langle B \rangle \bar{f}(x)$  with the diagonal (i.e. fixed points), and we consider two cases separately:

**Monostable case.** Assume that (8) possesses a single globally stable steady state  $x_0$  (Figure 1, top left, blue diamond). After the elapse of an initial transient, the inactive protein concentration  $x$  is attracted to  $x_0$ , and the active protein copy number  $s$  follows the QSS distribution  $\rho_s(x_0)$ .

**Bistable case.** Assume that (8) possesses three steady states  $x_- < x_0 < x_+$ , of which  $x_-$  and  $x_+$  are stable and  $x_0$  is unstable (Figure 1, top right, blue diamonds). Since in the long run the inactive protein concentration will alternate between being very close to either of the two attractors, the steady-state distribution of the active protein is expected to take the form of a mixture of the QSS distributions  $\omega_- \rho_s(x_-) + \omega_+ \rho_s(x_+)$ , where  $\omega_- \geq 0$ ,  $\omega_+ \geq 0$ ,  $\omega_- + \omega_+ = 1$ , are the probabilistic weights of the stable steady states  $x_-$  and  $x_+$ . Thus, we expect the steady-state distributions of  $s$  to be a mixture of two Poissonian modes located at the two stable steady states of the ODE (8).

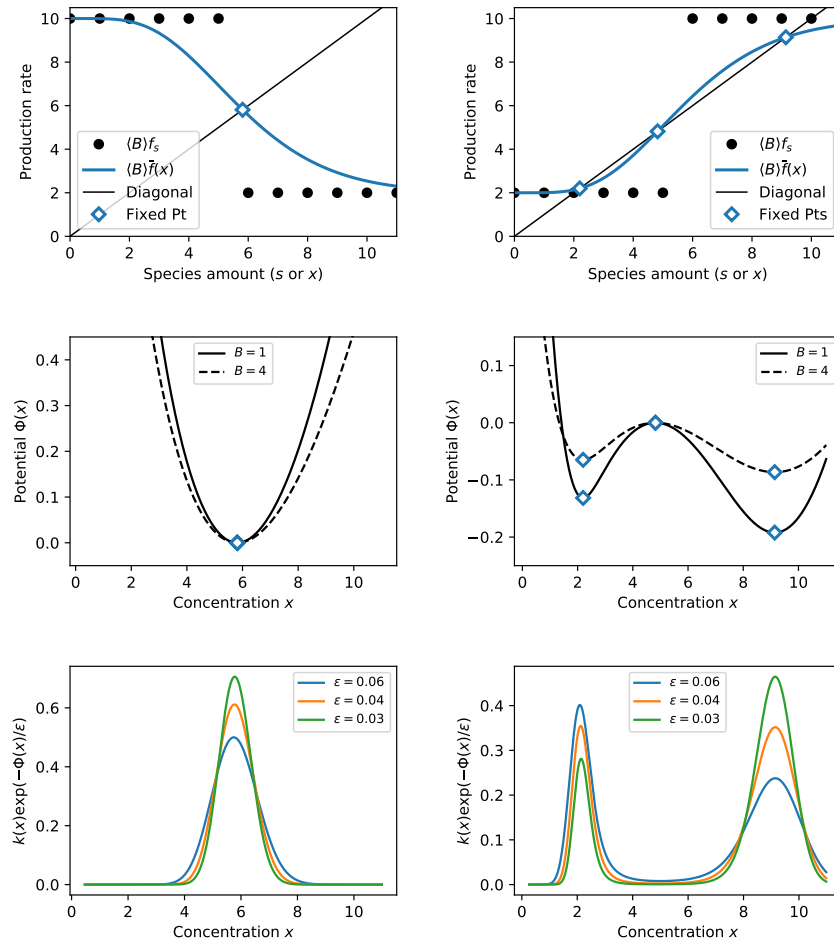


Figure 1: *Top*: The instantaneous production rate  $\langle B \rangle f_s$  (black circles) and the QSS-averaged production rate  $\langle B \rangle \bar{f}(x)$  (blue line) for negative (*left*) or positive (*right*) feedback. Blue diamond markers indicate the fixed points. *Middle*: The WKB potential  $\Phi(x)$ . Its local extrema collocate with the fixed points of the QSS-averaged production rate. The potential gets flatter as production becomes bursty (dashed line). *Bottom*: The WKB approximation of the probability density function of the inactive protein concentration  $x$ . The pdf is peaked around the potential minima for  $\epsilon \ll 1$ . *Parametric values*: The upper bound on  $S$  is  $s_{\max} = 20$ . The feedback threshold is  $s_{\text{thresh}} = 6$ . The burst size is fixed to  $B = 1$  except for the dashed line, middle panels, where it is fixed to  $B = 4$ . Unimodal examples use  $a_0 = 10$ ,  $a_1 = 2$ , except for the dashed line, middle panels, which uses  $a_0 = 2.5$ ,  $a_1 = 0.5$ . Bimodal examples use  $a_0 = 2$ ,  $a_1 = 10$ , except for the dashed line, middle panels, which uses  $a_0 = 0.5$ ,  $a_1 = 2.5$ .

## 4 WKB approximation: an overview

The approximation introduced in this Section reinforces the QSS-based heuristics with a systematic argumentation. Additionally, it provides asymptotic expansions as  $\varepsilon \rightarrow 0$  of the weights in the mixture distributions (which the QSS heuristics does not).

The WKB approximation of the probability mass function  $p(x, s; \varepsilon)$  of the inactive protein concentration  $x$  and the active protein copy number  $s$  is sought (and found in Section 6) in the form of

$$p(x, s; \varepsilon) \sim k(x)w_s(x)e^{-\frac{\Phi(x)}{\varepsilon}}, \quad \varepsilon \ll 1, \quad (9)$$

where  $k(x)$ ,  $w_s(x)$  and  $\Phi(x)$  are independent of  $\varepsilon$  and satisfy  $k(x) > 0$ ,  $w_s(x) > 0$ , and  $\sum_{s=0}^{s_{\max}} w_s(x) = 1$ . The variable  $s$  is placed in the subscript to emphasise its discreteness (contrasting it with the continuous nature of  $x$  in the limit of  $\varepsilon \rightarrow 0$ ). The ansatz (9) can be rephrased in terms of the marginal distribution of  $x$  and the conditional distribution of  $s$  as

$$p(x, \cdot; \varepsilon) = \sum_{s=0}^{s_{\max}} p(x, s; \varepsilon) \sim k(x)e^{-\frac{\Phi(x)}{\varepsilon}}, \quad (10)$$

$$p(s|x; \varepsilon) = \frac{p(x, s; \varepsilon)}{p(x, \cdot; \varepsilon)} \sim w_s(x). \quad (11)$$

We refer to  $\Phi(x)$  as the WKB potential, to  $k(x)$  as the WKB prefactor, and to  $w_s(x)$  as the WKB conditional distribution of  $s$ . In Section 6 we determine these terms from the underlying chemical master equation and establish a connection between the WKB and QSS approximations. At this stage, let us highlight the main outcomes of the forthcoming analysis.

We will show that the potential is a Lyapunov function of the limiting ODE model (8), possessing local minima (maxima) at the same points where (8) has stable (unstable) steady states (Figure 1, centre). Additionally, the potential carries information about the noise in the model that the ODE does not: specifically, the ODE depends only on the product of burst rate and burst frequency, remaining the same if the burst size is multiplied by the same factor as the burst frequency is divided by; the potential, on the other hand, becomes flatter as the system becomes more bursty (Figure 1, centre, dashed lines). This observation is consistent with an intuition that bursty production enhances noise and the chance to escape potential wells.

If the ODE is monostable, the potential has a single global minimum situated at the single global attractor  $x_0$  of the differential equation (Figure 1, centre left). For  $\varepsilon \ll 1$  the distribution (10) of  $x$  is sharply peaked around  $x_0$  (Figure 1, bottom left) so that it can be approximated by a delta peak

$$p(x, \cdot; \varepsilon) \sim \delta(x - x_0). \quad (12)$$

If the ODE is bistable,  $\Phi(x)$  is a double-well potential, with the wells at the stable steady states  $x_-$  and  $x_+$  being separated by a barrier at the unstable steady state  $x_0$  (Figure 1, centre right). For  $\varepsilon \ll 1$  the distribution (10) of  $x$  is sharply peaked around  $x_-$ ,  $x_+$  (Figure 1, bottom right) so that a mixture of delta peaks

$$p(x, \cdot; \varepsilon) \sim \omega_- \delta(x - x_-) + \omega_+ \delta(x - x_+) \quad (13)$$

can be used as a rough approximation.

We will show in Section 6 that at a steady state  $x_* \in \{x_-, x_0, x_+\}$  of (8) the WKB conditional distribution of  $s$  coincides with the QSS approximation,

$$w_s(x_*) = \rho_s(x_*), \quad 0 \leq s \leq s_{\max}, \quad (14)$$

where  $\rho_s(x)$  is defined by (3)–(4). Interestingly, conditioning on points  $x$  away from the steady states of (8) leads to WKB conditional distributions of  $s$  that differ from the QSS heuristics. Arguably, this disagreement arises because the simplifying QSS assumption of a fixed  $x$  is invalidated by making a large deviation from a steady state. Nevertheless, the contribution of non-QSS conditional distributions towards the total distribution of  $s$  turns out to be negligible.

The total distribution of  $s$  is given by

$$p(\cdot, s; \varepsilon) = \int_0^\infty p(s|x; \varepsilon) p(x, \cdot; \varepsilon) dx \sim \int_0^\infty w_s(x) p(x, \cdot; \varepsilon) dx. \quad (15)$$

Inserting the delta-peak singleton (12) or the delta-peak-mixture (13) approximations into (15) and using (14) we find a Poisson singleton

$$p(\cdot, s; \varepsilon) \sim \rho_s(x_0) \quad (16)$$

for the monostable case and a Poisson mixture

$$p(\cdot, s; \varepsilon) \sim \omega_- \rho_s(x_-) + \omega_+ \rho_s(x_+) \quad (17)$$

for the bistable case. These are the very approximations which were suggested by the QSS approach.

For the weights  $\omega_-$  and  $\omega_+$  we have

$$\frac{\omega_-}{\omega_+} \sim \frac{\int_{O(x_-)} k(x) e^{-\frac{\Phi(x)}{\varepsilon}} dx}{\int_{O(x_+)} k(x) e^{-\frac{\Phi(x)}{\varepsilon}} dx},$$

where  $O(x_-)$  and  $O(x_+)$  are neighbourhoods of  $x_-$  and  $x_+$ . Applying the Laplace method to approximate for small  $\varepsilon$  the integrals in the above equation, we find

$$\frac{\omega_-}{\omega_+} \sim \frac{k(x_-)}{k(x_+)} \sqrt{\frac{\Phi''(x_+)}{\Phi''(x_-)}} e^{-\frac{\Phi(x_-) - \Phi(x_+)}{\varepsilon}}, \quad (18)$$

which together with  $\omega_- + \omega_+ = 1$  determine the asymptotic behaviour of the weights. In particular, if the right well of the potential  $\Phi(x)$  is deeper than the right well, i.e.  $\Phi(x_-) > \Phi(x_+)$ , then (18) implies that  $(\omega_-, \omega_+) \rightarrow (0, 1)$  as  $\varepsilon \rightarrow 0$  (and  $(\omega_-, \omega_+) \rightarrow (1, 0)$  as  $\varepsilon \rightarrow 0$  if  $\Phi(x_-) < \Phi(x_+)$ ). However, if the two wells are finely balanced, the weights are comparable for a range of  $\varepsilon \ll 1$ , and both Poissons that constitute the steady-state distribution (17) of  $s$  are appreciable.

The approximation of the distribution of  $x$  by a single delta peak (monostable case) or a mixture of two delta peaks (bistable case) is sufficient for the purpose of deriving the leading-order approximation of the distribution of  $s$ . If one is interested in the distribution of  $x$  itself, a satisfactory result is obtained by

a parabolic approximation of the WKB potential around its minima. In the monostable case it is given by a Gaussian singleton

$$p(x, \cdot; \varepsilon) \sim \sqrt{\frac{\Phi''(x_0)}{2\pi\varepsilon}} \exp\left(-\frac{\Phi''(x_0)(x - x_0)^2}{2\varepsilon}\right), \quad (19)$$

and in the bistable case by a Gaussian mixture

$$p(x, \cdot; \varepsilon) \sim \omega_- \sqrt{\frac{\Phi''(x_-)}{2\pi\varepsilon}} \exp\left(-\frac{\Phi''(x_-)(x - x_-)^2}{2\varepsilon}\right) + \omega_+ \sqrt{\frac{\Phi''(x_+)}{2\pi\varepsilon}} \exp\left(-\frac{\Phi''(x_+)(x - x_+)^2}{2\varepsilon}\right). \quad (20)$$

The results (19) and (20) for the inactive protein distribution are expressed in terms of the concentration  $x$ . Expressions in terms of the copy number  $X = x/\varepsilon$  can be obtained using the well-known rule for a linear transformation of a normally distributed random variable.

For further background on this Section's results as well as practical recipes for the calculation of the WKB terms that appear in these results, we refer the reader to Section 6. In the next Section, we introduce the master equation, find its steady-state solution numerically, and compare it to the Poisson/Gaussian singleton/mixture approximations.

## 5 Master equation

The probability  $P = P(X, s, t)$  of having  $X$  molecules of inactive protein  $X$  and  $s$  molecules of active protein  $S$  (cf. Table 1) at time  $t$  satisfies the chemical master equation (CME)

$$\varepsilon \frac{\partial P}{\partial t} = f_s \left( \sum_{j=1}^{\infty} b_j \mathbb{E}_X^{-j} - 1 \right) P + \varepsilon (\mathbb{E}_X \mathbb{E}_s^{-1} - 1) X P + (\mathbb{E}_s - 1) s P, \quad (21)$$

in which  $\mathbb{E}_X$  and  $\mathbb{E}_s$  denote the van Kampen step operators (van Kampen, 2006) in variables  $X$  and  $s$  and  $\mathbb{E}_X^{-1}$  and  $\mathbb{E}_s^{-1}$  are their formal inverses. The master equation (21) applies in the unbounded case  $s_{\max} = \infty$ ; the bounded case will be dealt with in detail in Section 6. The principal interest of this paper is in steady-state solutions, which are obtained by setting the derivative in (21) to zero.

The first step in finding a numerical solution to a CME is its truncation to a finite number of equations. Following the approach illustrated e.g. by Borri et al (2016), we truncate the master equation to a finite lattice  $\{0, 1, \dots, s_{\max}\} \times \{0, 1, \dots, X_{\max}\}$ , and calculate the (unique) normalised steady-state solution; this amounts to finding a nullvector of a sparse square matrix of large order  $(s_{\max}+1)(X_{\max}+1)$ . The upper bound for the active protein is set to  $s_{\max} = 20$ , while the upper bound  $X_{\max}$  for the inactive protein is set to  $X_{\max} = 4\lceil x_+/\varepsilon \rceil$ , where  $x_+$  is the uppermost steady state of the ODE (8). Preliminary rounds of the Gillespie stochastic simulation algorithm (SSA) (Gillespie, 1976) show that sample trials (almost) never exceed  $X_{\max}$ , so that the truncated solution is expected to be a good approximation of the original one.

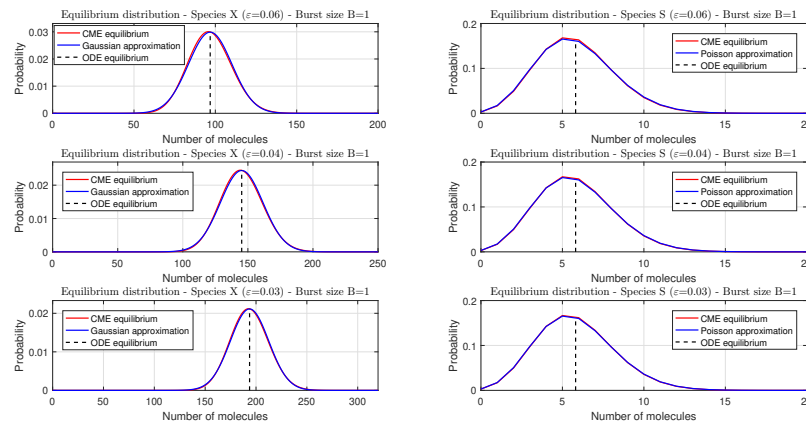


Figure 2: Steady-state distributions for negative feedback and fixed burst size  $B = 1$  obtained by numerical solution (red line) and asymptotic approximation (blue line). Panel columns refer to the two protein species; panel rows refer to distinct values of  $\varepsilon$ . The dashed lines indicate the locations of stable fixed points of the deterministic rate equation. The step function (6) parameters are  $s_{\text{thresh}} = 6$ ,  $a_0 = 10$ ,  $a_1 = 2$ .

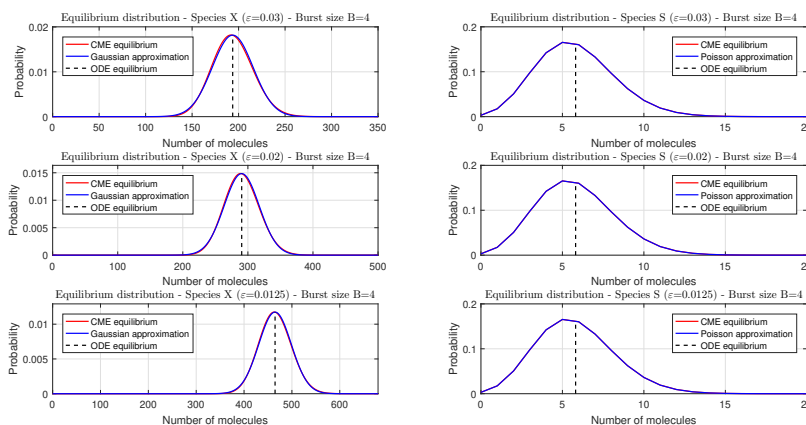


Figure 3: Steady-state distributions for negative feedback and fixed burst size  $B = 4$  obtained by numerical solution (red line) and asymptotic approximation (blue line). Panel columns refer to the two protein species; panel rows refer to distinct values of  $\varepsilon$ . The dashed lines indicate the locations of the stable fixed points of the deterministic equation. The step function (6) parameters are  $s_{\text{thresh}} = 6$ ,  $a_0 = 2.5$ ,  $a_1 = 0.5$ .

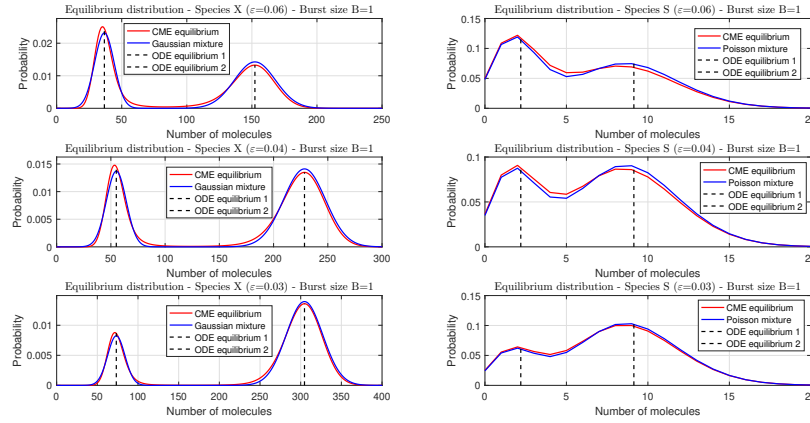


Figure 4: Steady-state distributions for the case of positive feedback and fixed burst size  $B = 1$  obtained by numerical solution (red line) and asymptotic approximation (blue line). Panel columns refer to the two protein species; panel rows refer to distinct values of  $\varepsilon$ . The dashed lines indicate the locations of the stable fixed points of the deterministic equation. The step function (6) parameters are  $s_{\text{thresh}} = 6$ ,  $a_0 = 2$ ,  $a_1 = 10$ .

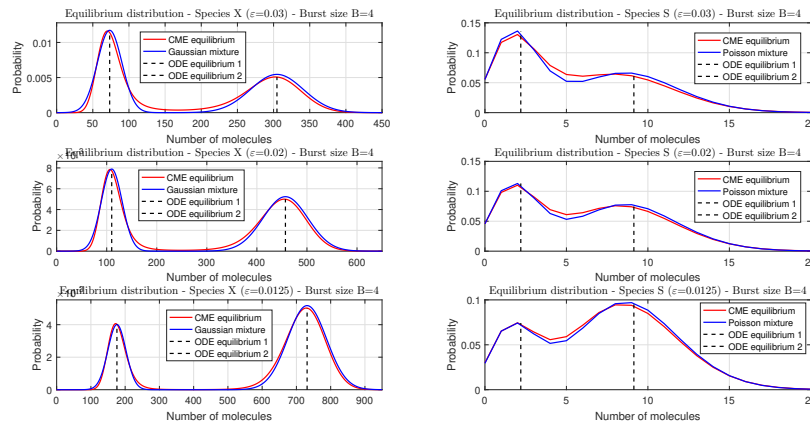


Figure 5: Steady-state distributions for the case of positive feedback and fixed burst size  $B = 4$  obtained by numerical solution (red line) and asymptotic approximation (blue line). Panel columns refer to the two protein species; panel rows refer to distinct values of  $\varepsilon$ . The dashed lines indicate the locations of the stable fixed points of the deterministic equation. The step function (6) parameters are  $s_{\text{thresh}} = 6$ ,  $a_0 = 0.5$ ,  $a_1 = 2.5$ .

Burst size	Feedback	$x_*$	$(\Phi''(x_*))^{-1}$	$k(x_-)/k(x_+)$	$\Phi(x_-) - \Phi(x_+)$
1	Negative	5.81	10.69	N/A	N/A
4	Negative	5.81	14.44	N/A	N/A
1	Positive	2.2	3.35	4.47	0.0606
		9.14	14.98		
4	Positive	2.2	8.67	4.42	0.0216
		9.14	40.17		

Table 2: Coefficients of the WKB approximation used in the numerical examples of Figures 2–5. The values of  $x_*$  give the location of the stable fixed points of ODE (8). The value of  $(\Phi''(x_*))^{-1}$  gives (up to a factor of  $1/\varepsilon$ ) the variance of the Gaussian modes. The prefactor ratio  $k(x_-)/k(x_+)$  and the potential difference  $\Phi(x_-) - \Phi(x_+)$  enter into the calculation of the mixture weights in the bistable case.

The numerical and the WKB-based asymptotic solutions to the CME are in a good agreement both for negative feedback (Figures 2–3) as well as for positive feedback (Figure 4–5). For each feedback type we consider two (fixed) burst sizes  $B = 1$  and  $B = 4$  ( $b_B = 1$ ,  $b_j = 0$  for  $j \neq B$ ). Panel columns in the figures refer to the two protein species; panel rows refer to the chosen values of  $\varepsilon$ . The parameters of the step function (6) are given in the figure captions. The asymptotic approximations (Figures 2–5, blue lines) are obtained by evaluating (16) and (19) in the monostable case (negative feedback) or (17) and (20) in the bistable case (positive feedback). The Poisson modes are evaluated from the formula (3)–(4). The weights in the mixture distributions are calculated from (18). The numerical values of the coefficients that enter these calculations are summarised in Table 2.

## 6 WKB approximation: analysis

We start our analysis by formulating a version of the master equation (21) that is applicable to the  $s_{\max} < \infty$  case (cf. Table 1). For this purpose it turns out to be helpful to use a modified version of the van Kampen step operators. For a sequence  $v_s$  indexed by an integer  $0 \leq s \leq s_{\max}$ , we define the left- and right-shift operators by

$$\mathcal{L}v_s = \begin{cases} v_{s+1} & \text{if } 0 \leq s < s_{\max}, \\ 0 & \text{if } s = s_{\max}, \end{cases} \quad \mathcal{R}v_s = \begin{cases} 0 & \text{if } s = 0, \\ v_{s-1} & \text{if } 0 < s < s_{\max}, \\ v_{s-1} + v_s & \text{if } s = s_{\max}. \end{cases} \quad (22)$$

Away from the boundary  $s = s_{\max}$ , the left- and right-shift operators  $\mathcal{L}$  and  $\mathcal{R}$  are equal to the van Kampen step operator and its formal inverse, respectively. Let us discuss the meaning of the modifications of the operators that are made on the boundary  $s = s_{\max}$ . The left-shift operator is used below to describe the transfer of probability mass due to protein decay. The modification of the left-shift operator at the boundary  $s = s_{\max}$  means there is no transfer of probability from the inadmissible state  $s_{\max} + 1$ . The right-shift operator is used below to

describe the transfer of probability mass due to protein activation. The reset map of the activation reaction channel (Table 1) implies that the states with  $s_{\max}$  as well as the states with  $s_{\max} - 1$  active protein molecules transfer into a state with  $s_{\max}$  molecules. Correspondingly, the right-shift operator returns at the boundary the sum of the ultimate and the penultimate terms of the original sequence.

The probability  $P(X, s, t)$  of at time  $t$  having  $X$  molecules of inactive protein  $X$  and  $s$  molecules of active protein  $S$  (cf. Table 1) satisfies

$$\begin{aligned} \varepsilon \frac{\partial}{\partial t} P(X, s, t) = f_s \sum_{j=1}^{\infty} b_j P(X - j, s, t) + \varepsilon(X + 1) \mathcal{R}P(X + 1, s, t) \\ + \mathcal{L}P(X, s, t) - (f_s + \varepsilon X + s)P(X, s, t), \end{aligned} \quad (23)$$

in which we use the operator formalism (22) in the variable  $s$  but the shifts in the variable  $X$  are made explicit. We thereby tacitly understand, as is customary in analyses of chemical master equation, that the probability of having a negative number of species  $X$  is equal to zero.

Inserting

$$X = \frac{x}{\varepsilon}, \quad P(X, s, t) = p(x, s, t)$$

into (23), we obtain

$$\begin{aligned} \varepsilon \frac{\partial}{\partial t} p(x, s, t) = f_s \sum_{j=1}^{\infty} b_j p(x - \varepsilon j, s, t) + (x + \varepsilon) \mathcal{R}p(x + \varepsilon, s, t) \\ + \mathcal{L}sp(x, s, t) - (f_s + x + s)p(x, s, t), \end{aligned} \quad (24)$$

which expresses the master equation in terms of the concentration of the inactive protein. Equating the derivative in (24) to zero, we arrive at a bivariate difference equation

$$0 = f_s \sum_{j=1}^{\infty} b_j p(x - \varepsilon j, s) + (x + \varepsilon) \mathcal{R}p(x + \varepsilon, s) + \mathcal{L}sp(x, s) - (f_s + x + s)p(x, s) \quad (25)$$

for the steady-state distribution  $p(x, s)$ . Below we seek an asymptotic approximation as  $\varepsilon \rightarrow 0$  to the (normalised) solution of (25).

## 6.1 Expansion

We seek a solution to (25) in the WKB form, cf. Equation (9),

$$p(x, s; \varepsilon) = (r_s^0(x) + \varepsilon r_s^1(x) + O(\varepsilon^2)) e^{-\frac{\Phi(x)}{\varepsilon}}, \quad (26)$$

where  $r_s^0(x) = k(x)w_s(x)$  and  $r_s^1(x)$  give the first two terms in the asymptotic expansion in the powers of  $\varepsilon$ . Below we develop, by means of (26), the individual terms of the difference equation (25) into asymptotic expansions of up to the second order. This is a mechanistic but laborious exercise. Therefore we suggest that, on first reading, the reader focus their attention on the leading-order terms in the expansions; these are sufficient for evaluating the potential  $\Phi(x)$  and  $w_s(x)$  (which play the central role in the analysis). The second-order terms, including

those that involve  $r_s^1(x)$ , come into play in Subsection 6.3 in the determination of the prefactor  $k(x)$ .

For the first term in equation (25) we find

$$\begin{aligned} \sum_{j=1}^{\infty} b_j p(x - \varepsilon j, s; \varepsilon) &= \sum_{j=1}^{\infty} b_j \left( r_s^0(x - \varepsilon j) + \varepsilon r_s^1(x - \varepsilon j) + O(\varepsilon^2) \right) e^{-\Phi(x - \varepsilon j)/\varepsilon} \\ &= e^{-\frac{\Phi(x)}{\varepsilon}} \sum_{j=1}^{\infty} b_j e^{j\theta} \left( r_s^0(x) - \varepsilon j r_s^{0'}(x) + \varepsilon r_s^1(x) + O(\varepsilon^2) \right) \left( 1 - \frac{\varepsilon j^2 \Phi''(x)}{2} + O(\varepsilon^2) \right) \\ &\sim e^{-\frac{\Phi(x)}{\varepsilon}} \left( r_s^0(x) M(\theta) + \varepsilon \left[ r_s^1(x) M(\theta) - r_s^{0'}(x) M'(\theta) - \frac{1}{2} \Phi''(x) r_s^0(x) M''(\theta) \right] \right), \end{aligned}$$

where

$$\theta = \Phi'(x), \quad M(\theta) = \sum_{j=1}^{\infty} b_j e^{j\theta}. \quad (27)$$

are the potential derivative and the moment generating function of the burst-size probability distribution (1), respectively.

The second term in equation (25) is developed into

$$\begin{aligned} (x + \varepsilon) \mathcal{R} p(x + \varepsilon, s; \varepsilon) &= (x + \varepsilon) \left( \mathcal{R} r_s^0(x + \varepsilon) + \varepsilon \mathcal{R} r_s^1(x + \varepsilon) + O(\varepsilon^2) \right) e^{-\Phi(x + \varepsilon)/\varepsilon} \\ &= e^{-\frac{\Phi(x)}{\varepsilon} - \theta} (x + \varepsilon) \left( \mathcal{R} r_s^0(x) + \varepsilon [\mathcal{R} r_s^{0'}(x) + \mathcal{R} r_s^1(x)] + O(\varepsilon^2) \right) \left( 1 - \frac{\varepsilon \Phi''(x)}{2} + O(\varepsilon^2) \right) \\ &\sim e^{-\frac{\Phi(x)}{\varepsilon} - \theta} \left( x \mathcal{R} r_s^0(x) + \varepsilon \left[ \left( 1 - \frac{\Phi''(x)x}{2} \right) \mathcal{R} r_s^0(x) + x (\mathcal{R} r_s^{0'}(x) + \mathcal{R} r_s^1(x)) \right] \right). \end{aligned}$$

The remaining terms in (25) are easy to expand.

We insert the WKB ansatz (26) into (25), expand the individual terms of the equation as suggested above, and collect terms of same order; at the leading order this yields

$$\mathcal{A} r_s^0(x) = 0, \quad (28)$$

where

$$\mathcal{A} v_s = e^{-\theta} x \mathcal{R} v_s + \mathcal{L} s v_s - (f_s(1 - M(\theta)) + x + s) v_s \quad (29)$$

is a linear operator acting on sequences  $v_s$  defined for  $0 \leq s \leq s_{\max}$ . Such sequences can be represented by  $(s_{\max} + 1)$ -dimensional column vectors  $\mathbf{v} = (v_0, v_1, \dots, v_{s_{\max}})^T$ , and the linear operator  $\mathcal{A}$  as a square matrix  $\mathbf{A}$  of order  $s_{\max} + 1$ . Here and throughout this Section, we will go back and forth between the operator-sequence and the matrix-vector notations, using that which expresses a given formula more succinctly.

The matrix  $\mathbf{A}$  is tridiagonal. On the diagonal it has the sequence  $-f_s(1 - M(\theta)) - x - s$ , where  $0 \leq s < s_{\max}$ , except for the last diagonal element, which is given by  $-f_{s_{\max}}(1 - M(\theta)) - x(1 - e^{-\theta}) - s_{\max}$ . On the upper diagonal it has the sequence  $1, 2, \dots, s_{\max}$ ; the elements of the lower diagonal are all equal to

$e^{-\theta}x$ . It looks like

$$\mathbf{A} = \begin{pmatrix} -f_0(1 - M(\theta)) - x & 1 & & 0 \\ e^{-\theta}x & \ddots & \ddots & \\ & \ddots & \ddots & s_{\max} \\ 0 & & e^{-\theta}x & -f_{s_{\max}}(1 - M(\theta)) - (1 - e^{-\theta})x - s_{\max} \end{pmatrix}. \quad (30)$$

The matrix  $\mathbf{A} = \mathbf{A}(x, \theta)$  — just like the associated operator  $\mathcal{A} = \mathcal{A}(x, \theta)$  — depends on the protein concentration  $x$  and the (yet unknown) potential derivative  $\theta = \Phi'(x)$ .

Equation (28) can be written in a matrix form as

$$\mathbf{A}(x, \theta) \mathbf{r}^0(x) = \mathbf{0}, \quad (31)$$

where  $\mathbf{r}^0(x) = (r_0^0(x), r_1^0(x), \dots, r_{s_{\max}}^0(x))^{\top}$  is a positive vector (i.e. a vector with only positive elements). The matrix  $\mathbf{A}$  does not have any negative off-diagonal elements. Denote by

$$H(x, \theta) = \lambda_1(\mathbf{A}(x, \theta)) \quad (32)$$

the eigenvalue with largest real part (which we refer to hereafter as the principal eigenvalue). The Perron–Frobenius theorem implies that the principal eigenvalue is real and that its right and left eigenvectors (referred to as principal eigenvectors) are real and positive. The right eigenvectors of non-principal eigenvalues, being orthogonal to the left principal eigenvector, cannot be positive. Therefore, equation (31) is solvable in  $\mathbf{r}^0(x) > 0$  if

$$H(x, \theta) = 0 \quad (33)$$

holds. Solving equation (33) in  $\theta$ , we obtain a functional dependence of the potential derivative  $\theta = \Phi'(x)$  on the concentration  $x$ . The potential itself can be obtained by (numerically) integrating its derivative.

Crucially, equation (33) does not define  $\theta = \theta(x)$  uniquely. First, there is a trivial solution  $\theta = 0$  for any value  $x > 0$ . Indeed,  $\theta = 0$  implies  $M(0) = 1$ , with which the matrix  $\mathbf{A}(x, 0)$  reduces to (the transpose of) the transition rate matrix of the QSS process (the  $M/M/s_{\max}/s_{\max}$  server; see Section 3). Therefore it is singular and its principal eigenvalue (32) is equal to zero. The trivial branch  $\theta \equiv 0$  of solutions to (33) cannot be used to construct the potential: we need to look for a different branch. To do so, it is useful to consider a Hamiltonian system of differential equations corresponding to the Hamiltonian (32).

## 6.2 The Hamiltonian system

Differentiating with respect to  $x$  equation (33) in which  $\theta = \theta(x)$  yields

$$\frac{\partial H}{\partial x} + \frac{\partial H}{\partial \theta} \frac{d\theta}{dx} = 0,$$

i.e.

$$\frac{d\theta}{dx} = -\frac{\frac{\partial H}{\partial x}(x, \theta)}{\frac{\partial H}{\partial \theta}(x, \theta)}. \quad (34)$$

The non-autonomous differential equation (34) is equivalent to the system of two autonomous differential equations

$$\dot{x} = \frac{\partial H}{\partial \theta}(x, \theta), \quad \dot{\theta} = -\frac{\partial H}{\partial x}(x, \theta), \quad (35)$$

in which the dot represents the time derivative. The system is Hamiltonian: its trajectories form the level sets of (32). We are specifically interested in the zero set (33). Borrowing terminology from the Hamiltonian formalism, we refer to the variable  $\theta$  as the conjugate momentum.

In order to solve system (35), we need to evaluate the right-hand sides. For this purpose it is useful to express the Hamiltonian (32) as

$$H = \mathbf{u}^\top \mathbf{A} \mathbf{v}, \quad (36)$$

where  $\mathbf{u} > 0$  and  $\mathbf{v} > 0$  are the left and right eigenvectors corresponding to the principal eigenvalue  $H$ , i.e.

$$\mathbf{u}^\top \mathbf{A} = H \mathbf{u}^\top, \quad \mathbf{A} \mathbf{v} = H \mathbf{v}, \quad (37)$$

which additionally satisfy

$$\mathbf{u}^\top \mathbf{v} = 1, \quad \sum_{s=0}^{s_{\max}} v_s = 1. \quad (38)$$

Conditions (38) can be met by a suitable choice of multiplicative constants.

The principal eigenvectors  $\mathbf{u}$  and  $\mathbf{v}$ , just like  $\mathbf{A}$  and the principal eigenvalue  $H$ , depend on the protein concentration  $x$  and the conjugate momentum  $\theta$ . For  $\theta = 0$  the matrix  $\mathbf{A}(x, 0)$  reduces to the transition rate matrix of the QSS process with principal eigenvalue  $H(x, 0) = 0$ ; the principal eigenvectors are given by the nullvectors of the QSS process

$$\mathbf{u}(x, 0) = \mathbf{1}, \quad \mathbf{v}(x, 0) = \boldsymbol{\rho}(x), \quad (39)$$

in which  $\mathbf{1} = (1, 1, \dots, 1)^\top$  is an  $(s_{\max} + 1)$ -dimensional vector of ones and  $\boldsymbol{\rho}(x) = (\rho_0(x), \rho_1(x), \dots, \rho_{s_{\max}}(x))^\top$  is the vector representation of the QSS stationary distribution (3).

The partial derivatives of the Hamiltonian (36) are given by

$$\begin{aligned} \frac{\partial H}{\partial x} &= \mathbf{u}^\top \frac{\partial \mathbf{A}}{\partial x} \mathbf{v} + \frac{\partial \mathbf{u}^\top}{\partial x} \mathbf{A} \mathbf{v} + \mathbf{u}^\top \mathbf{A} \frac{\partial \mathbf{v}}{\partial x} \\ &= \mathbf{u}^\top \frac{\partial \mathbf{A}}{\partial x} \mathbf{v} + H \frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial x} \\ &= \mathbf{u}^\top \frac{\partial \mathbf{A}}{\partial x} \mathbf{v}, \end{aligned} \quad (40)$$

$$\frac{\partial H}{\partial \theta} = \mathbf{u}^\top \frac{\partial \mathbf{A}}{\partial \theta} \mathbf{v}. \quad (41)$$

The derivatives of  $\mathbf{A}$  in (40)–(41) are matrix representations of the derivatives

$$\frac{\partial \mathcal{A}}{\partial x} v_s = e^{-\theta} \mathcal{R} v_s - v_s, \quad \frac{\partial \mathcal{A}}{\partial \theta} v_s = -e^{-\theta} x \mathcal{R} v_s + M'(\theta) f_s v_s \quad (42)$$

of the operator  $\mathcal{A}$  (29).

For  $\theta = 0$  the derivatives (40)–(41) of the Hamiltonian satisfy

$$\begin{aligned}\frac{\partial H}{\partial x}(x, 0) &= \mathbf{1}^\top \frac{\partial \mathbf{A}}{\partial x}(x, 0) \boldsymbol{\rho}(x) = \sum_{s=0}^{s_{\max}} \frac{\partial \mathcal{A}}{\partial x}(x, 0) \rho_s(x) \\ &= \sum_{s=0}^{s_{\max}} \mathcal{R} \rho_s(x) - \sum_{s=0}^{s_{\max}} \rho_s(x) = 0,\end{aligned}\quad (43)$$

$$\begin{aligned}\frac{\partial H}{\partial \theta}(x, 0) &= \mathbf{1}^\top \frac{\partial \mathbf{A}}{\partial \theta}(x, 0) \boldsymbol{\rho}(x) = \sum_{s=0}^{s_{\max}} \frac{\partial \mathcal{A}}{\partial \theta}(x, 0) \rho_s(x) \\ &= -x \sum_{s=0}^{s_{\max}} \mathcal{R} \rho_s(x) + M'(0) \sum_{s=0}^{s_{\max}} f_s \rho_s(x) \\ &= -x + \langle B \rangle \bar{f}(x).\end{aligned}\quad (44)$$

We note that in the equalities leading up to (43) and (44), we made use of the following properties: (i) the product of a row vector of ones and a column vector is equal to the sum of the column vector's elements; (ii) the right-shift operator  $\mathcal{R}$  (22) preserves the sum of vector elements; (iii) evaluating at zero the first derivative of the burst-size moment generating function (27) gives the mean burst size; we also used the definition of the QSS-averaged production rate (5).

Inserting  $\theta = 0$  into (35) and using (43)–(44), we find that on the  $x$ -axis the Hamiltonian system satisfies

$$\dot{x} = \langle B \rangle \bar{f}(x) - x, \quad \dot{\theta} = 0 \quad (\text{if } \theta = 0). \quad (45)$$

Equations (45) imply that (i) the  $x$ -axis is an invariant set of the Hamiltonian system and that (ii) the system reduces to the deterministic rate equation (8) when restricted to the invariant set. The Hamiltonian system (35) thus comprises, and extends by an additional dimension in  $\theta$ , the ODE dynamics established by the quasi-steady-state analysis of Section 3.

A point  $(x_*, 0)$ , where  $x_* \in \{x_-, x_0, x_+\}$  is any of the fixed points of (8), is also a steady state of the full Hamiltonian system (35). The linearisation matrix is given by

$$\mathbf{J} = \begin{pmatrix} \frac{\partial^2 H}{\partial x \partial \theta} & \frac{\partial^2 H}{\partial \theta^2} \\ 0 & -\frac{\partial^2 H}{\partial x \partial \theta} \end{pmatrix} \bigg|_{x=x_*, \theta=0}, \quad (46)$$

in which the second derivative of  $H$  with respect to  $x$  is immediately seen to be zero because of (43).

It follows that  $(x_*, 0)$  is a saddle of (35) with eigenvalues

$$\lambda_{1,2}(\mathbf{J}) = \pm \frac{\partial^2 H}{\partial x \partial \theta} \quad (47)$$

and eigenvectors

$$\text{Ker}(\mathbf{J} - \lambda_1(\mathbf{J})) \ni (1, 0)^\top, \quad \text{Ker}(\mathbf{J} - \lambda_2(\mathbf{J})) \ni \left( 1, -2 \left( \frac{\partial^2 H}{\partial \theta^2} \right)^{-1} \frac{\partial^2 H}{\partial x \partial \theta} \right)^\top, \quad (48)$$

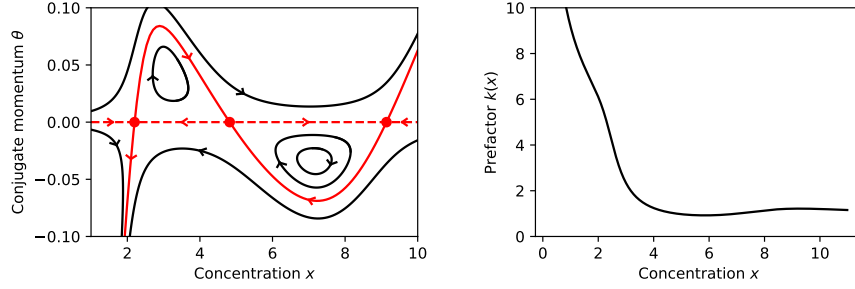


Figure 6: *Left*: Phase plane of the Hamiltonian system (35). The heteroclinic orbits (red colour) form the zero set of the Hamiltonian (32). The nontrivial portion of the zero set (full red line) defines the derivative of the WKB potential  $\theta = \Phi'(x)$ . The trivial portion of the zero set ( $\theta = 0$ ; dashed red line) is the phase line of the deterministic ODE (8). The red markers denote the steady states of (8), which are also saddles of the Hamiltonian system. *Right*: The WKB prefactor. The parameter values are: The upper bound on  $S$  is  $s_{\max} = 20$ . The burst size is fixed to  $B = 1$ . The step function (6) parameters are  $s_{\text{thresh}} = 6$ ,  $a_0 = 2$ ,  $a_1 = 10$ .

in which the derivatives of the Hamiltonian are evaluated at  $(x_*, 0)$ . One can show that

$$\frac{\partial^2 H}{\partial x \partial \theta}(x_*, 0) = -1 + \langle B \rangle \bar{f}'(x_*), \quad (49)$$

$$\frac{\partial^2 H}{\partial \theta^2}(x_*, 0) = x_* \left( 1 + \frac{\langle B^2 \rangle}{\langle B \rangle} \right) + 21^\top \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \tilde{\mathbf{v}}, \quad (50)$$

where  $\tilde{\mathbf{v}}$  is a solution to  $\mathbf{A}(x_*, 0) \tilde{\mathbf{v}} = -\frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \boldsymbol{\rho}(x_*)$ . Equation (49) is an immediate consequence of (44). Equation (50) is derived in Appendix A. The derivative of the effective production rate (5) satisfies

$$\bar{f}'(x) = \mathcal{N}(x) \left( (\bar{f}(x) - f_{s_{\max}}) \frac{x^{s_{\max}}}{s_{\max}!} + \sum_{s=0}^{s_{\max}-1} (f_{s+1} - f_s) \frac{x^s}{s!} \right). \quad (51)$$

Equations (49)–(51) provide a practical numerical recipe for calculating the nontrivial eigenvector (48) of the Hamiltonian system linearisation.

The trajectories emanating from a saddle  $(x_*, 0)$  along the direction of the eigenvector  $(1, 0)^\top$  form the trivial branch  $\theta = 0$  of the zero set (33). The trajectories emanating from the saddle along the nontrivial eigenvector (48) form the nontrivial branch of the zero set (33) (Figure 6, left). The nontrivial branch constitutes the sought-after WKB potential derivative  $\theta = \Phi'(x)$ . The derivative  $\Phi'(x)$  is equal to zero where the nontrivial  $\theta = \Phi'(x)$  and the trivial  $\theta = 0$  branches intersect, i.e. at the fixed points  $x = x_* \in \{x_-, x_0, x_+\}$  of the deterministic rate equation (8). This establishes the claim made in Section 4 that the potential extrema coincide with the fixed points of the deterministic model.

Formulae (18), (19), and (20) require us to evaluate the second derivative of the potential at fixed points. The second derivative of  $\Phi(x)$ , i.e. the first

derivative of  $\theta = \Phi'(x)$ , is equal at the fixed points to the second component of the nontrivial eigenvector (48) of the Hamiltonian system linearisation (cf. Figure 6, left panel). Away from fixed points, the derivative of  $\theta = \Phi'(x)$  can be evaluated by substituting into the right-hand side of (34).

In order to complete the WKB approximation (9) at the leading order, we express

$$\mathbf{r}^0(x) = k(x)\mathbf{w}(x), \quad \mathbf{w}(x) = \mathbf{v}(x, \Phi'(x)), \quad (52)$$

in which the  $l^1$ -normalised nullvector  $\mathbf{w}(x) = (w_0(x), \dots, w_{s_{\max}}(x))^T$  gives the distribution of  $s$  conditioned on a concentration  $x$ , and  $k(x)$  is a prefactor. In particular, if  $x = x_*$  is a fixed point, then we have  $\mathbf{w}(x_*) = \mathbf{v}(x_*, 0) = \boldsymbol{\rho}(x_*)$  (the Poissonian QSS distribution); this establishes the claim (14) made in Section 4. Away from the fixed points, the conditional probability mass function of  $s$  is in general different from the Poissonian QSS distribution.

Having determined the potential  $\Phi(x)$  and the probability mass function  $w_s(x)$ , the last remaining task is to calculate the prefactor  $k(x)$ . Doing so requires us to consult the second-order terms in the WKB expansion.

### 6.3 Calculating the prefactor

Having inserted the WKB ansatz (26) into the master equation (25) and having expanded the individual members up to the second order (Section 6.1), we now collect the second order terms and obtain

$$\begin{aligned} \mathcal{A}r_s^1(x) - f_s \left( M'(\theta)r_s^{0'}(x) + \frac{1}{2}\Phi''(x)M''(\theta)r_s^0(x) \right) \\ + e^{-\theta} \left[ \left( 1 - \frac{1}{2}\Phi''(x)x \right) \mathcal{R}r_s^0(x) + x\mathcal{R}r_s^{0'}(x) \right] = 0. \end{aligned}$$

Inserting the factorisation  $r_s^0(x) = k(x)w_s(x)$ , cf. (52), into the equation above yields

$$\mathcal{A}r_s^1(x) = k'(x)\alpha_s(x) + k(x)\beta_s(x), \quad (53)$$

where

$$\begin{aligned} \alpha_s(x) &= M'(\theta)f_s w_s(x) - e^{-\theta}x\mathcal{R}w_s(x), \\ \beta_s(x) &= M'(\theta)f_s w_s'(x) + \frac{1}{2}\Phi''(x)M''(\theta)f_s w_s(x) \\ &\quad - e^{-\theta} \left[ \left( 1 - \frac{1}{2}\Phi''(x)x \right) \mathcal{R}w_s(x) + x\mathcal{R}w_s'(x) \right]. \end{aligned}$$

In order that equation (53) be solvable in  $r_s^1(x)$ , its right-hand side must be orthogonal to the left nullvector  $l_s(x) = u_s(x, \theta(x))$  of  $\mathcal{A} = \mathcal{A}(x, \theta(x))$ , i.e.

$$k'(x) \sum_{s=0}^{s_{\max}} l_s(x)\alpha_s(x) + k(x) \sum_{s=0}^{s_{\max}} l_s(x)\beta_s(x) = 0,$$

Integrating the above linear homogeneous first-order equation in  $k(x)$  yields

$$k(x) = e^{-\psi(x)}, \quad \psi'(x) = \frac{\sum_{s=0}^{s_{\max}} l_s(x)\beta_s(x)}{\sum_{s=0}^{s_{\max}} l_s(x)\alpha_s(x)}.$$

The function  $\psi(x)$  is determined by numerically integrating  $\psi'(x)$ . The dependence of the prefactor  $k(x)$  on the protein concentration  $x$  is exemplified in Figure 6, right panel.

## 7 Discussion

We formulated and investigated a stochastic model for the production of a protein with delayed positive feedback. In the model, the protein is produced in bursts of multiple molecule copies. Newly produced protein molecules are inactive, and become activated by passing through a single activation step; biologically, the step can represent chemical modification, compartmental transport, or other scenarios. Active protein molecules regulate the frequency of bursty production of inactive protein. Such feedback can biologically be realised through transcriptional regulation.

The model incorporates an upper bound  $s_{\max}$  on the number of active protein. If  $s_{\max}$  active protein are already present, a new activation is allowed to occur, but is immediately followed by the removal of the activated molecule; consequently, the number of active protein molecules never exceeds  $s_{\max}$ . Thanks to the introduction of the upper bound, a number of crucial steps in the mathematical analysis involve finite, rather than infinite, calculation (e.g. the averaging (5) or the matrix (30)). Without an explicit upper bound in the model, each of these calculations would require an ad-hoc truncation; the explicit inclusion of the upper bound in the model guarantees a consistent use of truncation throughout the entire analysis. In the presented numerical examples, we choose  $s_{\max}$  large enough in order that the results be close to those expected without an upper bound.

We focused on examining the model behaviour in the regime of slow activation. The regime is characterised by activation rates of  $O(1)$  and production/decay rates of  $O(1/\varepsilon)$ , where  $\varepsilon \ll 1$ . Consequently, the inactive protein is present at  $O(1/\varepsilon)$  large copy numbers and fluctuates on a  $O(1)$  slow timescale, whereas the active protein is present at  $O(1)$  moderate copy numbers and fluctuates on a  $O(\varepsilon)$  fast timescale. Performing a quasi-steady-state (QSS) elimination of the fast active protein, we found that the QSS distribution of the active protein is a (truncated) Poisson, and that the inactive protein evolves on the  $O(1)$  timescale according to a deterministic rate equation (8).

We were specifically interested in situations where the rate equation has two stable steady states. Bistability occurs if the effective feedback response function, which quantifies the feedback of the inactive protein on itself, is sufficiently sigmoid so as to support multiple fixed points. The effective response function is equal to the average value, weighted by the QSS distribution of the active protein, of the original, active-protein-dependent, production rate. As a result of averaging by the noisy active protein, the effective response function is a smoothed-out (or “mollified”) version of the original response function; the requirement that the mollified function be sigmoid implies that the original function must be yet steeper. For simplicity, we used an (infinitely steep) step function in the examples of this paper. Biologically, highly sigmoid feedback responses can be maintained through cooperative binding of the protein to the regulatory DNA sequences.

If the model operates in the slow activation regime and if the limiting rate equation is monostable, then the steady-state distribution of the inactive (active) protein is nearly Gaussian (Poisson); the location of the Gaussian/Poisson mode is dictated by the unique fixed point of the rate equation. If the rate equation is bistable, the distribution of the inactive protein is approximated by a mixture of two small-noise Gaussians, and that of the active protein by a mix-

ture of two (moderate-noise) Poissons; the locations of the Gaussian/Poissonian modes are dictated by the fixed points of the rate equation. In order to obtain asymptotic approximations of the weights of the two modes, one needs to consult (and calculate) a WKB solution to the master equation; doing so was the concern of the bulk of the mathematical analysis presented in this paper. The asymptotic solution agrees well with a numerical solution to the master equation.

The principal step in the calculation of the asymptotic WKB solution is the determination of the WKB potential. The derivative of the potential is formed by the nontrivial heteroclinic connections between the steady states of a Hamiltonian system (Figure 6, full red line). The trivial heteroclinic connections that lie on the  $x$ -axis (Figure 6, dashed red line) satisfy the rate equation (the one previously established by the QSS analysis). The potential derivative and the deterministic rate have opposite signs: the potential has local minima/maxima where the rate equations has stable/unstable steady states; in other words, the WKB potential is the deterministic rate equation’s Lyapunov function.

Our asymptotic analysis stands on the shoulders of previous analyses (see Introduction for a limited review), and one in particular: Newby and Chapman (2014) study a stochastic gene expression model which is based on different biological assumptions than ours; the commonality is that it features two components with a similar pattern of time and abundance scales. The model of Newby and Chapman (2014) consists of an “internal” finite-state Markov chain (representing promoter states) coupled with an “external” birth and death process (representing protein). The coupling of the two components is in the dependence of transition rates for either component on the current state of both. The deterministic limit in protein dynamics is obtained by reducing both the internal and the external noise. The internal noise is reduced by speeding up the promoter transitions; the external noise is reduced by increasing the protein abundance. If both noise sources are reduced proportionally to each other (and to a small parameter  $\varepsilon$ ), the same configuration of time and abundance scales is achieved as in our model in the slow activation regime: the promoter state fluctuates at  $O(1)$  numbers on a  $O(\varepsilon)$  time scale and the protein at  $O(1/\varepsilon)$  numbers on a  $O(1)$  time scale. Like we did here for our model, Newby and Chapman (2014) used the WKB method to describe the large-time behaviour of their model in the  $\varepsilon \ll 1$  regime. There are some similarities, as well as differences, between the two models as well as in the methodologies of this paper and Newby and Chapman (2014). Our model features bursting and a stoichiometric connection between the two species that the model of Newby and Chapman (2014) does not. The WKB potential is determined, in both studies, from the condition that a matrix, here (30), be singular. Our matrix is large and sparse, whereas that of Newby and Chapman (2014) is dense and typically small (few gene states). Methodologically, we define the Hamiltonian as the principal eigenvalue (one with the largest real part), where Newby and Chapman (2014) use the determinant, of the matrix. The method of determining the prefactor (Section 6.3) from the higher-order terms of the master equation is the same as used in Newby and Chapman (2014).

In future work, we would like to look beyond the steady state and quantify the transition rates between the modes  $x_-$  and  $x_+$  of the mixture distributions identified in the present paper. It is expected that these rates are proportional to  $\exp(-(\Phi(x_0) - \Phi(x_{\pm}))/\varepsilon)$ , i.e. exponentially small as  $\varepsilon \rightarrow 0$ . The proportion-

ality constant will be determined by matching the WKB solution to a solution of a Fokker–Planck equation in the neighbourhood of the unstable steady state  $x_0$  (Hinch and Chapman, 2005; Bressloff, 2014). We would also like to see the current framework extended to more general distributed delays; we expect that delays composed of  $m$  slow memoryless steps will lead to a Hamiltonian system in the  $2m$ -dimensional Euclidean space.

In summary, we performed a detailed analysis of the steady-state distribution for an autoregulating protein with a large one-step production delay. Previous studies show that deterministically monostable positive feedbacks can exhibit bimodal distributions when formulated stochastically (Singh, 2012; Bokes and Singh, 2019). Our analysis shows that while both monostable and bistable feedbacks can exhibit bimodality at the single-cell level without any time delay, they will converge to different distributions with the inclusion of large delays, and hence provides a novel method to probe the structure of positive genetic feedback circuits.

## Appendix A: Linearisation of the Ham. system

Here we derive the expression (50) for the second  $\theta$ -derivative of the Hamiltonian (36). By doing so, we complete the linearisation analysis of the Hamiltonian system (32).

Differentiating  $\mathbf{A}\mathbf{v} = H\mathbf{v}$  with respect to  $\theta$  twice yields

$$\frac{\partial^2 \mathbf{A}}{\partial \theta^2} \mathbf{v} + 2 \frac{\partial \mathbf{A}}{\partial \theta} \frac{\partial \mathbf{v}}{\partial \theta} + \mathbf{A} \frac{\partial^2 \mathbf{v}}{\partial \theta^2} = \frac{\partial^2 H}{\partial \theta^2} \mathbf{v} + 2 \frac{\partial H}{\partial \theta} \frac{\partial \mathbf{v}}{\partial \theta} + H \frac{\partial^2 \mathbf{v}}{\partial \theta^2}.$$

At a saddle with coordinates  $x = x_* \in \{x_-, x_0, x_+\}$  and  $\theta = 0$ , we have  $H(x_*, 0) = \frac{\partial H}{\partial \theta}(x_*, 0) = 0$  and  $\mathbf{v}(x_*, 0) = \boldsymbol{\rho}(x_*)$ , so that

$$\begin{aligned} \frac{\partial^2 H}{\partial \theta^2}(x_*, 0) \boldsymbol{\rho}(x_*) &= \frac{\partial^2 \mathbf{A}}{\partial \theta^2}(x_*, 0) \boldsymbol{\rho}(x_*) + 2 \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \frac{\partial \mathbf{v}}{\partial \theta}(x_*, 0) \\ &\quad + \mathbf{A}(x_*, 0) \frac{\partial^2 \mathbf{v}}{\partial \theta^2}(x_*, 0). \end{aligned}$$

We multiply the above equation by  $\mathbf{u}^\top(x_*, 0) = \mathbf{1}^\top$  from the left; noting that  $\mathbf{1}^\top \boldsymbol{\rho}(x_*) = 1$  and  $\mathbf{1}^\top \mathbf{A}(x_*, 0) = \mathbf{0}^\top$ , we obtain

$$\frac{\partial^2 H}{\partial \theta^2}(x_*, 0) = \mathbf{1}^\top \frac{\partial^2 \mathbf{A}}{\partial \theta^2}(x_*, 0) \boldsymbol{\rho}(x_*) + 2 \mathbf{1}^\top \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \frac{\partial \mathbf{v}}{\partial \theta}(x_*, 0). \quad (\text{A1})$$

The second derivative of  $\mathbf{A}(x, \theta)$  is the matrix representation of the second derivative

$$\frac{\partial^2 \mathcal{A}}{\partial \theta^2}(x, \theta) v_s = e^{-\theta} x \mathcal{R} v_s + M''(\theta) f_s v_s$$

of the operator  $\mathcal{A}(x, \theta)$  (29). Therefore, the first term on the right-hand side of

(A1) satisfies

$$\begin{aligned} \mathbf{1}^\top \frac{\partial^2 \mathbf{A}}{\partial \theta^2}(x_*, 0) \boldsymbol{\rho}(x_*) &= \sum_{s=0}^{s_{\max}} \frac{\partial^2 \mathcal{A}}{\partial \theta^2}(x_*, 0) \rho_s(x_*) \\ &= x_* \sum_{s=0}^{s_{\max}} \mathcal{R} \rho_s(x_*) + M''(0) \sum_{s=0}^{s_{\max}} f_s \rho_s(x_*) \\ &= x_* + \langle B^2 \rangle \bar{f}(x_*) = x_* \left( 1 + \frac{\langle B^2 \rangle}{\langle B \rangle} \right), \end{aligned} \quad (\text{A2})$$

in which we utilised the fact that  $x_*$  is a fixed point of  $\langle B \rangle \bar{f}(x)$ .

The second term on the right-hand side of (A1) involves the derivative of the principal eigenvector with respect to  $\theta$ . Differentiating  $\mathbf{A}\mathbf{v} = H\mathbf{v}$  with respect to  $\theta$  and inserting  $x = x_*$  and  $\theta = 0$  yields

$$\mathbf{A}(x_*, 0) \frac{\partial \mathbf{v}}{\partial \theta}(x_*, 0) = -\frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \boldsymbol{\rho}(x_*), \quad (\text{A3})$$

Solving the inhomogeneous linear algebraic system (A3) yields

$$\frac{\partial \mathbf{v}}{\partial \theta}(x_*, 0) = \tilde{\mathbf{v}} + c \boldsymbol{\rho}(x_*),$$

where  $c \boldsymbol{\rho}(x_*)$  is a representant of the (right) kernel of  $\mathbf{A}(x_*, 0)$  and

$$\tilde{\mathbf{v}} = -[\mathbf{A}(x_*, 0)]^+ \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \boldsymbol{\rho}(x_*)$$

is a least-squares solution to (A3), with  $\mathbf{A}^+$  denoting the pseudo-inverse of  $\mathbf{A}$ .

The second term in (A1) therefore satisfies

$$\begin{aligned} \mathbf{1}^\top \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \frac{\partial \mathbf{v}}{\partial \theta}(x_*, 0) &= \mathbf{1}^\top \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \tilde{\mathbf{v}} + c \mathbf{1}^\top \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \boldsymbol{\rho}(x_*) \\ &= \mathbf{1}^\top \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \tilde{\mathbf{v}} + c \frac{\partial H}{\partial \theta}(x_*, 0) \\ &= \mathbf{1}^\top \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0) \tilde{\mathbf{v}}. \end{aligned} \quad (\text{A4})$$

Inserting (A2) and (A4) into (A1) recovers (50).

## References

- Albert J (2019) Path integral approach to generating functions for multistep post-transcription and post-translation processes and arbitrary initial conditions. *J Math Biol* DOI <https://doi.org/10.1007/s00285-019-01426-4>
- Anderson DF, Cotter SL (2016) Product-form stationary distributions for deficiency zero networks with non-mass action kinetics. *B Math Biol* 78(12):2390–2407
- Andreychenko A, Bortolussi L, Grima R, Thomas P, Wolf V (2017) Distribution approximations for the chemical master equation: comparison of the method of moments and the system size expansion. In: *Modeling Cellular Systems*, Springer, pp 39–66

- Assaf M, Meerson B (2017) WKB theory of large deviations in stochastic populations. *J Phys A: Math Theor* 50(26):263,001
- Barrio M, Burrage K, Leier A, Tian T (2006) Oscillatory regulation of *hes1*: discrete stochastic delay modelling and simulation. *PLoS Comput Biol* 2:e117
- Battich N, Stoeger T, Pelkmans L (2015) Control of transcript variability in single mammalian cells. *Cell* 163(7):1596–1610
- Be’er S, Assaf M (2016) Rare events in stochastic populations under bursty reproduction. *J Stat Mech Theory E* 2016:113,501
- Bokes P, Singh A (2019) Noise induced bimodality in genetic circuits with monostable positive feedback. In: 2019 18th European Control Conference (ECC), IEEE, pp 698–703
- Bokes P, King JR, Wood AT, Loose M (2012) Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *J Math Biol* 64:829–854
- Bokes P, Lin Y, Singh A (2018) High cooperativity in negative feedback can amplify noisy gene expression. *B Math Biol* 80:1871–1899, DOI <https://doi.org/10.1007/s11538-018-0438-y>
- Borri A, Carravetta F, Mavelli G, Palumbo P (2016) Block-tridiagonal state-space realization of chemical master equations: A tool to compute explicit solutions. *J Comput Appl Math* 296:410–426
- Borri A, Palumbo S, Singh A (2019) Time delays in a positive genetic feedback circuit. Submitted for publication
- Bressloff PC (2014) Stochastic processes in cell biology. Springer, New York
- Bruna M, Chapman SJ, Smith MJ (2014) Model reduction for slow–fast stochastic systems with metastable behaviour. *J Chem Phys* 140:174,107
- Crawford-Kahrl P, Cummins B, Gedeon T (2019) Comparison of combinatorial signatures of global network dynamics generated by two classes of ODE models. *Siam J Appl Dyn Syst* 18(1):418–457
- Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, Simpson ML, Weinberger LS (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. *P Natl Acad Sci USA* 109:17,454–17,459
- Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467(7312):167
- Feng J, Sevier SA, Huang B, Jia D, Levine H (2016) Modeling delayed processes in biological systems. *Phys Rev E* 94(3):032,408
- Gast N, Bortolussi L, Tribastone M (2019) Size expansions of mean field approximation: Transient and steady-state analysis. *Perform Evaluation* 129:60–80
- Gedeon T, Bokes P (2012) Delayed protein synthesis reduces the correlation between mRNA and protein fluctuations. *Biophys J* 103(3):377–385

- Gedeon T, Harker S, Kokubu H, Mischaikow K, Oka H (2017) Global dynamics for steep nonlinearities in two dimensions. *Physica D* 339:18–38
- Gillespie D (1976) A General method for numerically simulating stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22:403–34
- Gross D (2008) Fundamentals of queueing theory. John Wiley & Sons, Hoboken, N.J.
- Gupta A, Mikelson J, Khammash M (2017) A finite state projection algorithm for the stationary solution of the chemical master equation. *J Chem Phys* 147(15):154,101
- Gupta C, López JM, Ott W, Josić K, Bennett MR (2013) Transcriptional delay stabilizes bistable gene networks. *Phys Rev Lett* 111(5):058,104
- Gupta C, López JM, Azencott R, Bennett MR, Josić K, Ott W (2014) Modeling delay in genetic networks: From delay birth-death processes to delay stochastic differential equations. *J Chem Phys* 140(20):05B624\_1
- Heymann M, Vanden-Eijnden E (2008) The geometric minimum action method: A least action principle on the space of curves. *Commun Pur Appl Math* 61(8):1052–1117
- Hinch R, Chapman SJ (2005) Exponentially slow transitions on a markov chain: the frequency of calcium sparks. *Eur J Appl Math* 16(04):427–446
- van Kampen N (2006) Stochastic Processes in Physics and Chemistry. Elsevier, Amsterdam
- Kim JK, Josić K, Bennett MR (2014) The validity of quasi-steady-state approximations in discrete stochastic simulations. *Biophys J* 107(3):783–793
- Kumar N, Kulkarni RV (2019) A stochastic model for post-transcriptional regulation of rare events in gene expression. *Phys Biol*
- Kumar N, Singh A, Kulkarni RV (2015) Transcriptional bursting in gene expression: analytical results for general stochastic models. *Plos Comput Biol* 11(10):e1004,292
- Kyrychko Y, Schwartz I (2018) Enhancing noise-induced switching times in systems with distributed delays. *Chaos* 28(6):063,106
- Lafuerza L, Toral R (2011) Role of delay in the stochastic creation process. *Phys Rev E* 84:021,128
- Lin YT, Buchler NE (2018) Efficient analysis of stochastic gene dynamics in the non-adiabatic regime using piecewise deterministic markov processes. *J Roy Soc Interface* 15:20170,804
- Lin YT, Doering CR (2016) Gene expression dynamics with stochastic bursts: Construction and exact results for a coarse-grained model. *Phys Rev E* 93:022,409
- Lin YT, Galla T (2016) Bursting noise in gene expression dynamics: linking microscopic and mesoscopic models. *J Roy Soc Interface* 13:20150,772

- Modi S, Soltani M, Singh A (2018) Linear noise approximation for a class of piecewise deterministic Markov processes. In: 2018 Annual American Control Conference (ACC), IEEE, pp 1993–1998
- Monk N (2003) Oscillatory expression of Hes1, p53, and NF- $\kappa$ B driven by transcriptional time delays. *Curr Biol* 13:1409–1413
- Mor A, Suliman S, Ben-Yishay R, Yunger S, Brody Y, Shav-Tal Y (2010) Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. *Nat Cell Biol* 12(6):543
- Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys* 124:044,104
- Munsky B, Neuert G, Van Oudenaarden A (2012) Using gene expression noise to understand gene regulation. *Science* 336:183–187
- Newby J, Chapman SJ (2014) Metastable behavior in Markov processes with internal states. *J Math Biol* 69(4):941–976
- Newby JM (2012) Isolating intrinsic noise sources in a stochastic genetic switch. *Phys Biol* 9(2):026,002
- Paulsson J (2005) Models of stochastic gene expression. *Phys Life Rev* 2:157–175
- Popovic N, Marr C, Swain PS (2016) A geometric analysis of fast-slow models for stochastic gene expression. *J Math Biol* 72:87–122
- Roussel MR, Zhu R (2006) Stochastic kinetics description of a simple transcription model. *B Math Biol* 68(7):1681–1713
- Schnoerr D, Sanguinetti G, Grima R (2017) Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J Phys A: Math Theor* 50(9):093,001
- Schuss Z (2009) Theory and applications of stochastic processes: an analytical approach. Springer Science & Business Media, Berlin/Heidelberg
- Singh A (2012) Stochastic analysis of genetic feedback circuit controlling HIV cell-fate decision. In: Decision and Control (CDC), 2012 IEEE 51st Annual Conference on, IEEE, pp 4918–4923
- Singh A, Bokes P (2012) Consequences of mRNA transport on stochastic variability in protein levels. *Biophys J* 103:1087–1096
- Singh A, Hespanha J (2007) Stochastic analysis of gene regulatory networks using moment closure. In: Proceedings of the American Control Conference
- Singh A, Razooky B, Cox CD, Simpson ML, Weinberger LS (2010) Transcriptional bursting from the HIV-1 promoter is a significant source of stochastic noise in HIV-1 gene expression. *Biophys J* 98:L32–L34
- Smith M, Singh A (2019) Stochastic delays suppress noise in a genetic circuit with negative feedback. *bioRxiv* p 786491

- Soltani M, Vargas-Garcia CA, Antunes D, Singh A (2016) Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes. *Plos Comput Biol* 12(8):e1004972
- Srivastava R, Haseltine EL, Mastny E, Rawlings JB (2011) The stochastic quasi-steady-state assumption: Reducing the model but not the noise. *J Chem Phys* 134(15):154109
- Stoeger T, Battich N, Pelkmans L (2016) Passive noise filtering by cellular compartmentalization. *Cell* 164(6):1151–1161
- Sturrock M, Li S, Shahrezaei V (2017) The influence of nuclear compartmentalisation on stochastic dynamics of self-repressing gene expression. *J Theor Biol* 424:55–72
- Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F (2011) Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332:472–474
- Taniguchi Y, Choi P, Li G, Chen H, Babu M, Hearn J, Emili A, Xie X (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329:533–538
- Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *P Natl Acad Sci USA* 98:151588,598
- Tsimring L, Pikovsky A (2001) Noise-induced dynamics in bistable systems with delay. *Phys Rev Lett* 87(25):250602
- Veerman F, Marr C, Popović N (2018) Time-dependent propagators for stochastic models of gene expression: an analytical method. *J Math Biol* 77:261–312
- Yin H, Wen X (2019) First passage times and minimum actions for a stochastic minimal bistable system. *Chinese J Phys* 59:220–230
- Zavala E, Marquez-Lago TT (2014) Delays induce novel stochastic effects in negative feedback gene circuits. *Biophys J* 106(2):467–478
- Zhou T, Liu T (2015) Quantitative analysis of gene expression systems. *Quant Biol* 3(4):168–181