Mathematical Biology

Mixture distributions in a stochastic gene expression model with delayed feedback: a WKB approximation approach

Pavol Bokes[1,2] · Alessandro Borri[3] · Pasquale Palumbo[3,4] · Abhyudai Singh[5]

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 formal asymptotic approach, 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 fixed point 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 fixed points; 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.

✉ Pavol Bokes
   pavol.bokes@fmph.uniba.sk

[1] Comenius University, Bratislava, Slovakia

[2] Mathematical Institute, Slovak Academy of Sciences,, Bratislava, Slovakia

[3] Italian National Research Council (IASI-CNR), Rome, Italy

[4] University of Milano-Bicocca, Milan, Italy

[5] University of Delaware, Newark, DE, USA

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; Plesa et al. 2019) 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; Qiu et al. 2020). 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 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 Sect. 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 paper is structured as follows. Section 2 introduces the stochastic model and its chemical master equation (CME). Section 3 seeks a WKB approximation to the stationary solution of the CME and formulates a specific algebraic problem that needs to be solved to determine the approximation. Section 4 solves the algebraic problem by analysing the phase plane of an associated dynamical system; a specific restriction of the dynamical system is thereby interpreted in terms of the emergent deterministic dynamics of the model in the slow-activation limit. Section 5 completes the calculation of the terms required for the WKB approximation. Section 6 develops the WKB approximation into tractable Gaussian/Poisson singleton/mixture approximations and compares them to the numerical solution of the CME. Section 7 concludes the paper with a discussion.

2 Master equation

The paper is concerned with a reaction system involving the inactive and active protein species X and 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 is applied on the copy numbers of the two molecular species each time the reaction occurs.

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

Table 1 Reaction channels of the delayed feedback model with reaction species X (inactive protein) and S (active protein)

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

The copy number of the active protein is denoted by s and that of the inactive protein by X (we reserve the lower case for its concentration $x = \varepsilon X$, which is of the same order as 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 $O(1)$ long timescale of slow activation and the $O(\varepsilon)$ short timescale of fast turnover of the active protein. The reset map of the activation channel ensures that s never exceeds an upper bound $s_{\max}$

$O(1/\varepsilon)$, 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 proteins is increased by the size $B \geq 0$ of a production burst. Bursts sizes are drawn (independently of each other) from a distribution

$$P[B = j] = b_j, \quad j = 0, 1, 2, \ldots \tag{1}$$

The activation rate is proportional to the number of inactive proteins; 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 motivation for including the upper bound is technical: it allows for the use of finite-dimensional mathematical techniques. We will set $s_{\max}$ to a value at which the decay rate exceeds, by a factor of two, the maximal rate of production; such a limit is reached rarely and its introduction affects the system dynamics negligibly. The decay rate is proportional to the number of active proteins; the decay reset map decreases the number of active proteins by one.

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

$$\varepsilon \frac{\partial P}{\partial t} = f_s \left(\sum_{j=0}^{\infty} b_j \mathbb{E}_X^{-j} - 1 \right) P + \varepsilon \left(\mathbb{E}_X \mathbb{E}_s^{-1} - 1 \right) X P + (\mathbb{E}_s - 1) s P, \tag{2}$$

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 (2) applies in the unbounded case $s_{\max} = \infty$; in order to formulate one that is applicable to the $s_{\max} < \infty$ case, it turns out to be helpful to use modified versions 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

$$
\mathcal{L}v_s = \begin{cases} v_{s+1} & \text{if } 0 \leq s < s_{\max}, \\ 0 & \text{if } s = s_{\max}, \end{cases} \qquad \mathcal{R}v_s = \begin{cases} 0 & \text{if } s = 0, \\ v_{s-1} & \text{if } 0 < s < s_{\max}, \\ v_{s-1} + v_s & \text{if } s = s_{\max}. \end{cases} \tag{3}
$$

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 states with $s_{\max} - 1$ as well as $s_{\max}$ 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.

In the bounded case $s_{\max} < \infty$, the CME can be written as

$$
\varepsilon \frac{\partial}{\partial t} P(X, s, t) = f_s \sum_{j=0}^{\infty} b_j P(X - j, s, t) + \varepsilon (X + 1) \mathcal{R} P(X + 1, s, t)
$$
$$
+ \mathcal{L}s P(X, s, t) - (f_s + \varepsilon X + s) P(X, s, t), \tag{4}
$$

in which we use the operator formalism (3) in the variable s, but the shifts in the variable X are made explicit. We thereby tacitly understand, as is customary in analyses of CMEs, that the probability of having a negative number of species X is equal to zero.

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

$$
X = \frac{x}{\varepsilon}, \quad P(X, s, t) = p(x, s, t). \tag{5}
$$

We refer to the new variable x, which becomes in the limit of $\varepsilon \to 0$ a continuous quantity, as the concentration of the inactive protein. Inserting (5) into (4), we obtain

$$
\varepsilon \frac{\partial}{\partial t} p(x, s, t) = f_s \sum_{j=0}^{\infty} b_j p(x - \varepsilon j, s, t) + (x + \varepsilon) \mathcal{R} p(x + \varepsilon, s, t)
$$
$$
+ \mathcal{L}s p(x, s, t) - (f_s + x + s) p(x, s, t). \tag{6}
$$

Equating the derivative in (6) to zero, we arrive at a bivariate difference equation

$$0 = f_s \sum_{j=0}^{\infty} b_j p(x - \varepsilon j, s) + (x + \varepsilon) \mathcal{R} p(x + \varepsilon, s) + \mathcal{L} s p(x, s) - (f_s + x + s) p(x, s) \quad (7)$$

for the steady-state distribution $p(x, s)$. Below we seek an asymptotic approximation as $\varepsilon \to 0$ to the (normalised) solution of (7).

3 Expansion

We seek a solution to (7) in the WKB form

$$p(x, s; \varepsilon) = \left(r_s^0(x) + \varepsilon r_s^1(x) + O(\varepsilon^2) \right) e^{-\frac{\Phi(x)}{\varepsilon}}, \quad (8)$$

where $r_s^0(x) > 0$ and $r_s^1(x)$ give the first two terms in the asymptotic expansion in the powers of ε. The variable s is placed in the subscript to emphasise its discreteness (contrasting it with the continuous nature of x in the limit of $\varepsilon \to 0$). The function $\Phi(x)$ in the exponent of (8) is referred to as the WKB potential.

Below we develop, by means of (8), the individual terms of the difference Eq. (7) 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 calculating the potential $\Phi(x)$, which plays the central role in the analysis.

For the first term in Eq. (7) we find

$$\sum_{j=0}^{\infty} b_j p(x - \varepsilon j, s; \varepsilon)$$

$$= \sum_{j=0}^{\infty} b_j \left(r_s^0(x - \varepsilon j) + \varepsilon r_s^1(x - \varepsilon j) + O(\varepsilon^2) \right) e^{-\Phi(x - j\varepsilon)/\varepsilon}$$

$$= e^{-\frac{\Phi(x)}{\varepsilon}} \sum_{j=0}^{\infty} b_j e^{j\theta} \left(r_s^0(x) - \varepsilon j r_s^{0\prime}(x) + \varepsilon r_s^1(x) + O(\varepsilon^2) \right) \left(1 - \frac{\varepsilon j^2 \Phi''(x)}{2} + O(\varepsilon^2) \right)$$

$$\sim e^{-\frac{\Phi(x)}{\varepsilon}} \left(r_s^0(x) M(\theta) + \varepsilon \left[r_s^1(x) M(\theta) - r_s^{0\prime}(x) M'(\theta) - \frac{1}{2} \Phi''(x) r_s^0(x) M''(\theta) \right] \right),$$

where

$$\theta = \Phi'(x), \quad M(\theta) = \sum_{j=0}^{\infty} b_j e^{j\theta} \quad (9)$$

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

The second term in Eq. (7) is developed into

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

The remaining terms in (7) are easy to expand.

We insert the WKB ansatz (8) into (7), expand the individual terms of the equation as done above, and collect terms of same order; at the leading order this yields

$$\mathcal{A}r_s^0(x) = 0, \tag{10}$$

where

$$\mathcal{A}v_s = e^{-\theta}x\mathcal{R}v_s + \mathcal{L}sv_s - (f_s(1 - M(\theta)) + x + s)v_s \tag{11}$$

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 $\boldsymbol{v} = (v_0, v_1, \ldots, v_{s_{\max}})^{\mathsf{T}}$, and the linear operator $\mathcal{A}$ as a square matrix $\boldsymbol{A}$ of order $s_{\max} + 1$. Here and below, 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 $\boldsymbol{A}$ is tridiagonal. On the main 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 $1, 2, \ldots, s_{\max}$; the elements of the lower diagonal are all equal to $e^{-\theta}x$. It looks like

$$\boldsymbol{A} = \begin{pmatrix} -f_0(1 - M(\theta)) - x & 1 & & & 0 \\ & e^{-\theta}x & \ddots & \ddots & \\ & & \ddots & \ddots & s_{\max} \\ 0 & & & e^{-\theta}x & \begin{matrix} -f_{s_{\max}}(1 - M(\theta)) \\ -(1 - e^{-\theta})x - s_{\max} \end{matrix} \end{pmatrix}. \tag{12}$$

The matrix $\boldsymbol{A} = \boldsymbol{A}(x, \theta)$ — 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 (10) can be written in a matrix form as

$$\boldsymbol{A}(x, \theta)\boldsymbol{r}^0(x) = \boldsymbol{0}, \tag{13}$$

where $\boldsymbol{r}^0(x) = (r_0^0(x), r_1^0(x), \ldots, r_{s_{\max}}^0(x))^{\mathsf{T}}$ is required to be a positive vector (i.e. a vector with only positive elements). Denote by

$$H(x, \theta) = \lambda_1(\boldsymbol{A}(x, \theta)) \tag{14}$$

the eigenvalue with largest real part (Bressloff and Newby [2014]), which we refer to hereafter as the principal eigenvalue. The matrix A does not have any negative off-diagonal elements; the Perron–Frobenius theorem implies that the principal eigenvalue is real, and 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, Eq. (13) is solvable in $r^0(x) > 0$ if and only if

$$H(x, \theta) = 0 \tag{15}$$

holds. In what follows, we show that the zero-level set (15) includes a graph of a function $\theta = \theta(x)$, which forms the derivative (9) of the desired potential. In order to characterise the level sets of the function $H(x, \theta)$, it is useful to consider the associated Hamiltonian system of differential equations.

4 Hamiltonian system

Differentiating with respect to x Eq. (15) in which $\theta = \theta(x)$ yields

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

i.e.

$$\frac{\mathrm{d}\theta}{\mathrm{d}x} = -\frac{\frac{\partial H}{\partial x}(x, \theta)}{\frac{\partial H}{\partial \theta}(x, \theta)}. \tag{16}$$

The non-autonomous differential Eq. (16) is equivalent to the system of two autonomous differential equations

$$\dot{x} = \frac{\partial H}{\partial \theta}(x, \theta), \quad \dot{\theta} = -\frac{\partial H}{\partial x}(x, \theta), \tag{17}$$

in which the dot represents the time derivative. The system is Hamiltonian: its trajectories form the level sets of (14). We are specifically interested in the zero set (15). Borrowing terminology from the Hamiltonian formalism, we refer to the variable θ as the conjugate momentum.

In order to solve system (17), we need to evaluate the right-hand sides. For this purpose it is useful to express the Hamiltonian (14) as

$$H = u^\mathsf{T} A v, \tag{18}$$

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

$$u^\mathsf{T} A = H u^\mathsf{T}, \quad A v = H v, \tag{19}$$

which additionally satisfy

$$\boldsymbol{u}^\mathsf{T}\boldsymbol{v} = 1, \quad \sum_{s=0}^{s_{\max}} v_s = 1. \tag{20}$$

Conditions (20) can be met by a suitable choice of multiplicative constants.

The principal eigenvectors $\boldsymbol{u}$ and $\boldsymbol{v}$, just like $\boldsymbol{A}$ and the principal eigenvalue H, depend on the protein concentration x and the conjugate momentum θ. Differentiating (18) with respect to x yields

$$\begin{aligned}
\frac{\partial H}{\partial x} &= \boldsymbol{u}^\mathsf{T}\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{v} + \frac{\partial \boldsymbol{u}^\mathsf{T}}{\partial x}\boldsymbol{A}\boldsymbol{v} + \boldsymbol{u}^\mathsf{T}\boldsymbol{A}\frac{\partial \boldsymbol{v}}{\partial x} \\
&= \boldsymbol{u}^\mathsf{T}\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{v} + H\frac{\partial \boldsymbol{u}^\mathsf{T}\boldsymbol{v}}{\partial x} \\
&= \boldsymbol{u}^\mathsf{T}\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{v},
\end{aligned} \tag{21}$$

where the first equality applies the product rule, the second follows by (19), while the third is due to (20). Differentiating (18) with respect to θ and simplifying as before gives

$$\frac{\partial H}{\partial \theta} = \boldsymbol{u}^\mathsf{T}\frac{\partial \boldsymbol{A}}{\partial \theta}\boldsymbol{v}. \tag{22}$$

The derivatives of $\boldsymbol{A}$ in (21)–(22) are matrix representations of the derivatives

$$\frac{\partial \mathcal{A}}{\partial x}v_s = \mathrm{e}^{-\theta}\mathcal{R}v_s - v_s, \quad \frac{\partial \mathcal{A}}{\partial \theta}v_s = M'(\theta)f_s v_s - \mathrm{e}^{-\theta}x\mathcal{R}v_s \tag{23}$$

of the operator $\mathcal{A}$ (11). Before analysing the entire phase plane of system (17), we first characterise its flow on the x-axis and interpret it in terms of the model dynamics.

4.1 Deterministic rate equation

The matrix $\boldsymbol{A}(x, 0)$ simplifies, owing to the relation $M(0) = 1$, to a Markovian transition matrix of a birth–death process with state space $\{0, 1, \ldots, s_{\max}\}$, a constant birth rate x (as long as $s < s_{\max}$), and a linear death rate s. In our context, births correspond to activation (from a constant source of inactive protein), and deaths correspond to decay of the active protein. In queueing theory, which identifies births and deaths with customer arrivals and departures, such a process is referred to as the $M/M/s_{\max}/s_{\max}$ server with memoryless arrival and service times, $s_{\max}$ servers, and no queue (Gross 2008).

The transition matrix $\boldsymbol{A}(x, 0)$ of the $M/M/s_{\max}/s_{\max}$ server has the principal eigenvalue and eigenvectors given by

$$H(x, 0) = 0, \quad \boldsymbol{u}(x, 0) = \boldsymbol{1}, \quad \boldsymbol{v}(x, 0) = \boldsymbol{\rho}(x), \tag{24}$$

in which $\boldsymbol{1} = (1, 1, \ldots, 1)^\mathsf{T}$ is an $(s_{\max} + 1)$-dimensional vector of ones and $\boldsymbol{\rho}(x) = (\rho_0(x), \rho_1(x), \ldots, \rho_{s_{\max}}(x))^\mathsf{T}$ is the process's stationary distribution; this is given by

the truncated Poisson distribution with location parameter x (Gross 2008),

$$\rho_s(x) = \mathcal{N}(x)\frac{x^s}{s!}, \quad s = 0, 1, \ldots, s_{\max}, \tag{25}$$

where

$$\mathcal{N}(x) = \left(\sum_{s=0}^{s_{\max}} \frac{x^s}{s!}\right)^{-1} \tag{26}$$

is the normalisation constant. We remark that the probability of the server running at full capacity, which is returned by (25) at $s = s_{\max}$, goes under the name of Erlang's loss formula (Gross 2008).

Inserting (24) and (23) into (21) yields

$$\frac{\partial H}{\partial x}(x, 0) = \mathbf{1}^{\mathsf{T}}\frac{\partial A}{\partial x}(x, 0)\boldsymbol{\rho}(x) = \sum_{s=0}^{s_{\max}} \frac{\partial \mathcal{A}}{\partial x}(x, 0)\rho_s(x)$$
$$= \sum_{s=0}^{s_{\max}} \mathcal{R}\rho_s(x) - \sum_{s=0}^{s_{\max}} \rho_s(x) = 0, \tag{27}$$

in which we used that the product of a row vector of ones and a column vector is equal to the sum of the column vector's elements and that the right-shift operator $\mathcal{R}$ (3) preserves the sum of vector elements. Inserting (24) and (23) into (22) we find, after similar simplifications, that

$$\frac{\partial H}{\partial \theta}(x, 0) = \mathbf{1}^{\mathsf{T}}\frac{\partial A}{\partial \theta}(x, 0)\boldsymbol{\rho}(x) = \sum_{s=0}^{s_{\max}} \frac{\partial \mathcal{A}}{\partial \theta}(x, 0)\rho_s(x)$$
$$= M'(0)\sum_{s=0}^{s_{\max}} f_s\rho_s(x) - x\sum_{s=0}^{s_{\max}} \mathcal{R}\rho_s(x)$$
$$= \langle B \rangle \bar{f}(x) - x, \tag{28}$$

where

$$\langle B \rangle = M'(0) = \sum_{j=0}^{\infty} jb_j \tag{29}$$

is the mean burst size and

$$\bar{f}(x) = \sum_{s=0}^{s_{\max}} f_s\rho_s(x) \tag{30}$$

is the expectation of f_s with respect to the truncated Poisson distribution (25).

Inserting (27) into (17), we find that $\dot{\theta} = 0$ if $\theta = 0$: the x-axis is an invariant set of the Hamiltonian system (17). Inserting (28) into (17), we find that on the invariant set $\theta = 0$ the Hamiltonian system reduces to the rate equation

$$\dot{x} = \langle B \rangle \bar{f}(x) - x. \tag{31}$$

The Hamiltonian system (17) thus comprises, and extends by an additional dimension in θ, the concentration dynamics given by (31). The rate equation (31) represents a deterministic reduction of the delayed feedback model. This can be understood intuitively by invoking a quasi-steady-state (QSS) approximation (Rao and Arkin 2003). On $O(\varepsilon)$-short timescales, the inactive protein varies little and slowly, and its concentration x is nearly constant; on the other hand, the active protein is noisy and fast, and its copy number s evolves like an $M/M/s_{\max}/s_{\max}$ server, equilibrating to the QSS distribution (25). On $O(1)$-long timescales, the inactive protein is produced with an effective burst rate (30) which is obtained by averaging the instantaneous burst rate with respect to the QSS distribution. Multiplying the effective burst rate by the expected burst size and subtracting the linear activation rate yields the emergent deterministic dynamics (31).

Although the main contribution to $\bar{f}(x)$ comes from the values of f_s whose argument s is close to x, the value of $\bar{f}(x)$ is determined by all values of f_s. Due to the contributions of neighbouring terms, any sharp features of f_s are smoothed out, or "mollified", in the function $\bar{f}(x)$; for example, a step function (Gedeon et al. 2017; Crawford-Kahrl et al. 2019)

$$f_s = \begin{cases} a_0 & \text{if } s < s_{\text{thresh}}, \\ a_1 & \text{if } s \geq s_{\text{thresh}} \end{cases} \tag{32}$$

turns into a smooth sigmoid function by the application of (30); see Fig. 1, top panels.

By (24), the x-axis belongs to the zero-level set (15). However, since $\theta = 0$ cannot serve as the derivative of an appropriate potential, we need to look for a different branch $\theta = \Phi'(x)$ of solutions to (15). The nontrivial branch is found by linearising the Hamiltonian system around its steady states on the x-axis.

4.2 Phase-plane analysis

A point $(x_*, 0)$, where x_* is any of the fixed points of (31), is also a steady state of the full Hamiltonian system (17). The linearisation matrix is given by

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial^2 H}{\partial x \partial \theta} & \frac{\partial^2 H}{\partial \theta^2} \\ 0 & -\frac{\partial^2 H}{\partial x \partial \theta} \end{pmatrix} \Bigg|_{x=x_*, \theta=0,} \tag{33}$$

in which the second derivative of H with respect to x is immediately seen to be zero because of (27).

It follows that $(x_*, 0)$ is a saddle of (17) with eigenvalues

$$\lambda_{1,2}(\boldsymbol{J}) = \pm \frac{\partial^2 H}{\partial x \partial \theta} \tag{34}$$
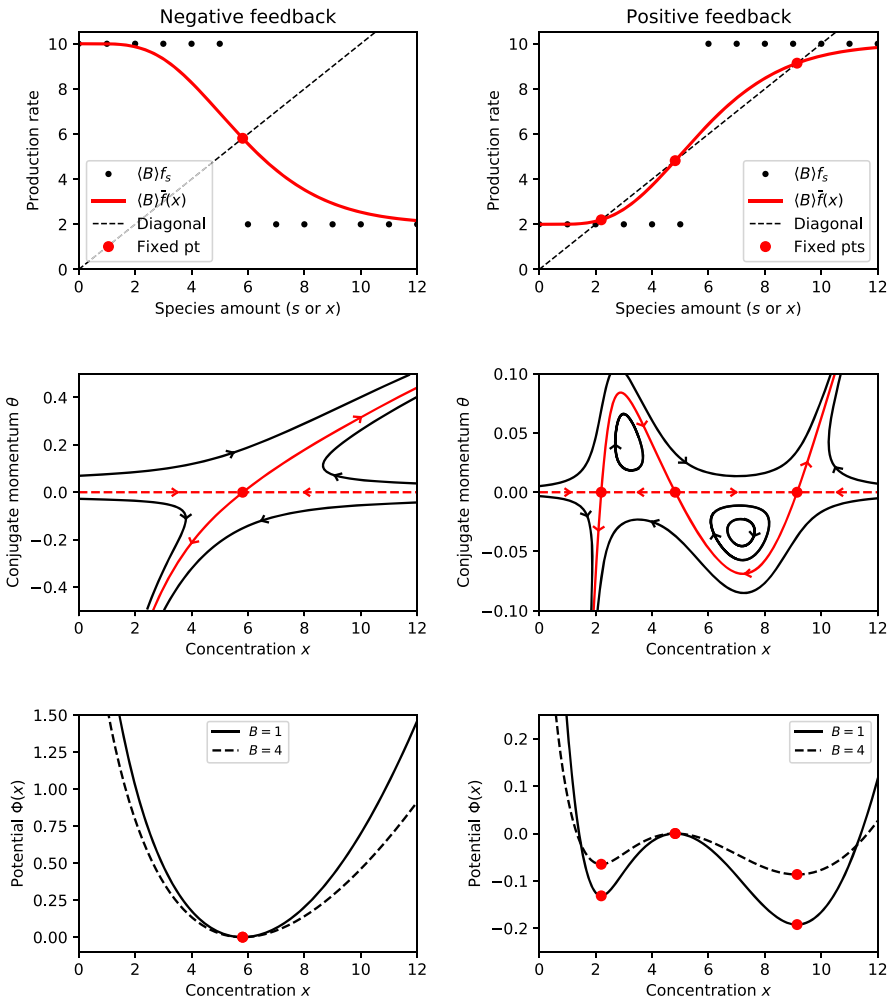
Fig. 1 Top: The instantaneous production rate $\langle B \rangle f_s$ (black dots) and the QSS-averaged production rate $\langle B \rangle \bar{f}(x)$ (red curve). Centre: Phase plane of the Hamiltonian system (17). The heteroclinic orbits (shown in red) form the zero set (15). The nontrivial portion of the zero set (solid red curve) 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 rate equation (31). Bottom: The WKB potential $\Phi(x)$. Red circles in all panels: The fixed points of the QSS-averaged production rate, the saddles of the Hamiltonian system, and the extrema of the potential coincide. 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 curve, bottom panels, where it is fixed to $B = 4$. Negative feedback examples use $a_0 = 10$, $a_1 = 2$, except for the dashed curve, bottom left panel, which uses $a_0 = 2.5$, $a_1 = 0.5$. Positive feedback examples use $a_0 = 2$, $a_1 = 10$, except for the dashed curve, bottom right panel, which uses $a_0 = 0.5$, $a_1 = 2.5$ (color figure online)

and eigenvectors

$$\text{Ker}(\boldsymbol{J} - \lambda_1(\boldsymbol{J})) \ni (1, 0)^\mathsf{T}, \quad \text{Ker}(\boldsymbol{J} - \lambda_2(\boldsymbol{J})) \ni \left(1, -2\left(\frac{\partial^2 H}{\partial\theta^2}\right)^{-1}\frac{\partial^2 H}{\partial x \partial\theta}\right)^\mathsf{T}, \quad (35)$$

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

$$\frac{\partial^2 H}{\partial x \partial\theta}(x_*, 0) = \langle B\rangle \bar{f}'(x_*) - 1, \tag{36}$$

$$\frac{\partial^2 H}{\partial\theta^2}(x_*, 0) = x_*\left(1 + \frac{\langle B^2\rangle}{\langle B\rangle}\right) + 2\mathbf{1}^\mathsf{T}\frac{\partial \boldsymbol{A}}{\partial\theta}(x_*, 0)\tilde{\boldsymbol{v}}, \tag{37}$$

where $\tilde{\boldsymbol{v}}$ is a solution to $\boldsymbol{A}(x_*, 0)\tilde{\boldsymbol{v}} = -\frac{\partial \boldsymbol{A}}{\partial\theta}(x_*, 0)\boldsymbol{\rho}(x_*)$. Equation (36) is an immediate consequence of (28). Equation (37) is derived in the "Appendix". The derivative of the effective production rate (30) satisfies

$$\bar{f}'(x) = \sum_{s=0}^{s_{\max}-1} (f_{s+1} - f_s)\rho_s(x) + \left(\bar{f}(x) - f_{s_{\max}}\right)\rho_{s_{\max}}(x). \tag{38}$$

Equations (36)–(38) provide a practical numerical recipe for calculating the nontrivial eigenvector (35) of the Hamiltonian system linearisation.

The trajectories emanating from a saddle $(x_*, 0)$ along the direction of the eigenvector $(1, 0)^\mathsf{T}$ form the trivial branch $\theta = 0$ of the zero set (15) (Fig. 1, central panels, dashed red line). The trajectories emanating from the saddle along the nontrivial eigenvector (35) form the nontrivial branch of the zero set (15) (Fig. 1, central panels, solid red curve). The nontrivial branch constitutes the sought-after WKB potential derivative $\theta = \Phi'(x)$. Given that the potential derivative has the opposite sign to the deterministic flow on the x-axis, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(x(t)) = \Phi'(x(t))\dot{x}(t) < 0 \tag{39}$$

for non-stationary solutions to (31). Therefore, the potential $\Phi(x)$ is a Lyapunov function of the rate equation (31), possessing local minima (maxima) where (31) has stable (unstable) fixed points (Fig. 1, bottom panels).

The potential carries additional information about the noise in the model that the rate equation does not: specifically, the rate equation depends only on the product of burst size 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 (Fig. 1, bottom panels, dashed curves). This observation is consistent with an intuition that bursty production enhances noise and the chance to escape potential wells.
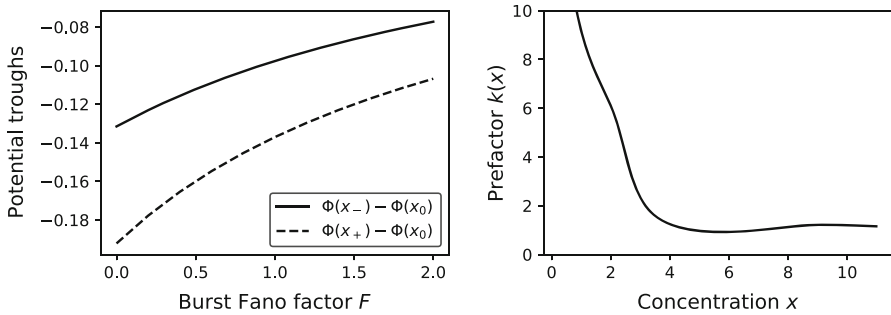
Fig. 2 Left: The depths of potential troughs as functions of the Fano factor of binomially-distributed burst sizes with mean $\langle B \rangle = 1$. Right: The WKB prefactor in the non-bursty case ($B = 1$). Parameter values for both panels: $s_{\max} = 20$; $s_{\text{thresh}} = 6$; $a_0 = 2$; $a_1 = 10$

In order to investigate the impact of distributed burst sizes, we consider burst distributions with moment generating functions (MGFs)

$$M(\theta) = \left(F + (1 - F)e^{\theta}\right)^{\frac{\langle B \rangle}{1-F}} \tag{40}$$

parametrised by the mean $\langle B \rangle > 0$ and the Fano factor $F \geq 0$ (the variance-to-mean ratio). For $F = 0$, definition (40) simplifies to $M(\theta) = e^{\langle B \rangle \theta}$, which is the MGF of a fixed burst size $B = \langle B \rangle$ (assuming that $\langle B \rangle$ is an integer). For $0 < F < 1$, formula (40) gives the MGF of a binomial distribution. As $F \to 1$, the right-hand side of (40) tends to $e^{\langle B \rangle (e^{\theta} - 1)}$, which gives the MGF of the Poisson distribution. For $F > 1$, expression (40) is the MGF of the negative binomial distribution; specifically, for $F = 1 + \langle B \rangle$, it reduces to the MGF $M(\theta) = (1 + \langle B \rangle (1 - e^{\theta}))^{-1}$ of the widely used geometric distribution of burst sizes (McAdams and Arkin 1997). Fixing the burst size mean $\langle B \rangle$ and varying the Fano factor F, the troughs of a double-well potential become shallower (Fig. 2, left).

5 Prefactor

In order to complete the WKB approximation (8) at the leading order, we express the solution to the linear Eq. (10) in terms of the l^1-normalised nullvector (19) as

$$r_s^0(x) = k(x)w_s(x), \quad w_s(x) = v_s(x, \Phi'(x)), \tag{41}$$

where $k(x)$ is an s-independent prefactor.

The calculation of the prefactor requires to consult the second-order terms in the WKB expansion of the master equation. Recall that in Sect. 3 we inserted the WKB ansatz (8) into the master equation (7), expanded (the individual terms of the equation) up to the second order, and collected the first-order terms. Collecting the second-order

terms yields

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

Inserting the factorisation (41) into the above equation yields

$$\mathcal{A}r_s^1(x) = k'(x)\alpha_s(x) + k(x)\beta_s(x), \tag{42}$$

where

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

In order that Eq. (42) be solvable in $r_s^1(x)$, its right-hand side must be orthogonal to the left nullvector $l_s(x) = u_s(x, \Phi'(x))$ of $\mathcal{A} = \mathcal{A}(x, \Phi'(x))$, 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

$$k(x) = \exp\left(-\int \frac{\sum_{s=0}^{s_{\max}} l_s(x)\beta_s(x)}{\sum_{s=0}^{s_{\max}} l_s(x)\alpha_s(x)}\mathrm{d}x \right).$$

The dependence of the prefactor $k(x)$ on the protein concentration x is exemplified in Fig. 2, right.

6 Mixture approximations

Combining (8) and (41), we express the WKB approximation to the joint distribution $p(x, s; \varepsilon)$ of the inactive protein concentration x and the active protein copy number s in the form of

$$p(x, s; \varepsilon) \sim k(x)w_s(x)e^{-\frac{\Phi(x)}{\varepsilon}}, \quad \varepsilon \ll 1, \tag{43}$$

where $k(x)$, $w_s(x)$ and $\Phi(x)$ are independent of ε and satisfy $k(x) > 0$, $w_s(x) > 0$, and $\sum_{s=0}^{s_{\max}} w_s(x) = 1$.

Expressing the marginal distribution of x and the conditional distribution of s as

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

we refer to $w_s(x)$ as the WKB conditional distribution of s, and note that the potential $\Phi(x)$ and the prefactor $k(x)$ constitute the WKB marginal distribution of x.

Dominant contributions to the protein distribution (43) come from the neighbourhoods of the minima of the potential $\Phi(x)$, where the potential can be approximated by parabolas. In the monostable regime of the rate equation (31), when the potential has a global minimum at x_0, doing so leads to

$$p(x, s; \varepsilon) \sim k(x_0) w_s(x_0) \exp\left(-\frac{\Phi(x_0)}{\varepsilon} - \frac{\Phi''(x_0)(x - x_0)^2}{2\varepsilon}\right), \qquad (44)$$

whereas in the bistable regime with two minima x_- and x_+ we obtain

$$\begin{aligned}
p(x, s; \varepsilon) \sim{}& k(x_-) w_s(x_-) \exp\left(-\frac{\Phi(x_-)}{\varepsilon} - \frac{\Phi''(x_-)(x - x_-)^2}{2\varepsilon}\right) \\
&+ k(x_+) w_s(x_+) \exp\left(-\frac{\Phi(x_+)}{\varepsilon} - \frac{\Phi''(x_+)(x - x_+)^2}{2\varepsilon}\right).
\end{aligned} \qquad (45)$$

Combining (41) and (24), we find that at critical points of the WKB potential the WKB conditional distribution of s satisfies

$$w_s(x_*) = v_s(x_*, 0) = \rho_s(x_*), \quad 0 \le s \le s_{\max}, \quad x_* \in \{x_-, x_0, x_+\}, \qquad (46)$$

where $\rho_s(x)$ is the QSS Poisson distribution defined by (25)–(26). Interestingly, conditioning on non-critical points of the potential leads to WKB conditional distributions of s that differ from the QSS ones. Arguably, this disagreement arises because the simplifying QSS assumption of a fixed inactive protein concentration is invalidated by making a large and improbable deviation from a fixed point. Nevertheless, the contribution of non-QSS conditional distributions towards the total distribution of s is seen to be negligible by the application of the parabolic approximations (44)–(45).

Inserting (46) into (44) and replacing (x, s)-independent factors by an appropriate normalisation constant yields

$$p(x, s; \varepsilon) \sim \rho_s(x_0) \times \sqrt{\frac{\Phi''(x_0)}{2\pi\varepsilon}} \exp\left(-\frac{\Phi''(x_0)(x - x_0)^2}{2\varepsilon}\right). \qquad (47)$$

Integrating (47) over x implies that s follows a truncated Poisson distribution with location parameter x_0 (Fig. 3, right panels); summing (47) over s implies that x follows a Gaussian distribution with mean x_0 and variance $\varepsilon/\Phi''(x_0)$. Linearly transforming concentration into copy number by means of $X = x/\varepsilon$, we find that the latter also
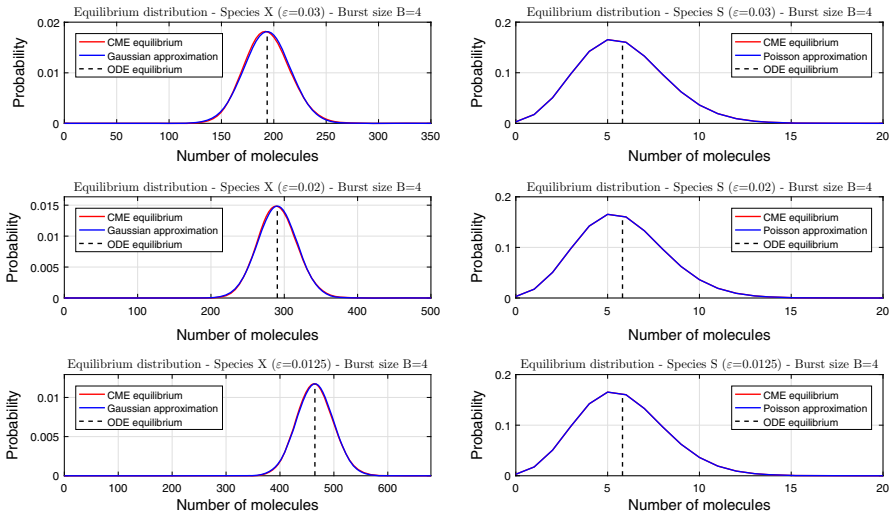
Fig. 3 Steady-state distributions for negative feedback and fixed burst size $B = 4$ obtained by numerical solution (red) and asymptotic approximation (blue). Panel columns refer to the two protein species; panel rows refer to distinct values of ε. The dashed lines indicate the locations of the stable fixed points (in units of molecules in the left column, and units of concentration in the right column) of the deterministic equation (31). The step function (32) parameters are $s_{\text{thresh}} = 6$, $a_0 = 2.5$, $a_1 = 0.5$ (color figure online)

follows a Gaussian distribution with mean x_0/ε and variance $1/\varepsilon \Phi''(x_0)$ (Fig. 3, left panels).

Inserting (46) into (45) leads to

$$
\begin{aligned}
p(x, s; \varepsilon) \sim &\ \omega_- \times \rho_s(x_-) \times \sqrt{\frac{\Phi''(x_-)}{2\pi\varepsilon}} \exp\left(-\frac{\Phi''(x_-)(x - x_-)^2}{2\varepsilon}\right) + \\
&\ \omega_+ \times \rho_s(x_+) \times \sqrt{\frac{\Phi''(x_+)}{2\pi\varepsilon}} \exp\left(-\frac{\Phi''(x_+)(x - x_+)^2}{2\varepsilon}\right)
\end{aligned}
\tag{48}
$$

with weights given by

$$
\omega_\pm = \frac{Ck(x_\pm)}{\sqrt{\Phi''(x_\pm)}} \exp\left(-\frac{\Phi(x_\pm)}{\varepsilon}\right),
\tag{49}
$$

in which the constant C is determined from the normalisation condition $\omega_- + \omega_+ = 1$. Integrating (48) over x implies that the marginal distribution of s is a mixture, with weights ω_- and ω_+, of two truncated Poisson distributions with location parameters x_- and x_+ (Fig. 4, right panels). Summing (48) over s implies that the marginal distribution of x is a mixture, with the same weights, of two Gaussians with means $x_\pm$ and variances $\varepsilon/\Phi''(x_\pm)$, which transform to $x_\pm/\varepsilon$ and $1/\varepsilon\Phi''(x_\pm)$ in units of molecules $X = x/\varepsilon$ (Fig. 4, left panels).

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 (35) of the
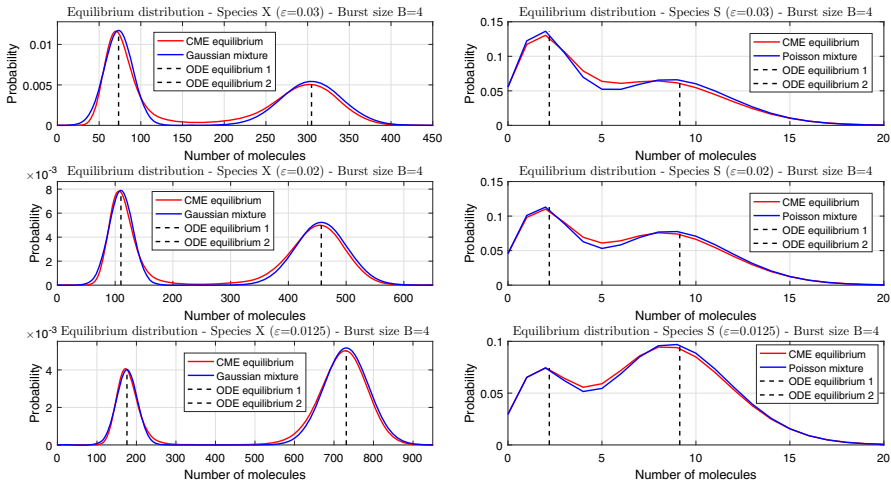
Fig. 4 Steady-state distributions for the case of positive feedback and fixed burst size $B = 4$ obtained by numerical solution (red) and asymptotic approximation (blue). Panel columns refer to the two protein species; panel rows refer to distinct values of ε. The dashed lines indicate the locations of the stable fixed points (in units of molecules in the left column, and units of concentration in the right column) of the deterministic equation (31). The step function (32) parameters are $s_{\text{thresh}} = 6$, $a_0 = 0.5$, $a_1 = 2.5$ (color figure online)

Hamiltonian system linearisation (cf. Fig. 1, central panels). Away from the fixed points, the derivative of $\theta = \Phi'(x)$ can be evaluated by substituting into the right-hand side of (16).

In the chosen example $\Phi(x_-) > \Phi(x_+)$, i.e. the right well of the double-well potential is deeper than the left well. Equation (49) then implies that $(\omega_-, \omega_+) \rightarrow (0, 1)$ as $\varepsilon \rightarrow 0$. However, the two potential wells are finely balanced, so that the weights are comparable for a range of $\varepsilon \ll 1$, and the two Poissons/Gaussians that constitute the steady-state distribution (48) both contribute.

Figures 3 and 4 demonstrate that the WKB-based Gaussian/Poisson singleton and mixture approximations are in a close agreement with numerical solutions to the chemical master equation; the latter are calculated as follows. First, the CME (4) is truncated 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 $\{0, 1, ..., s_{\max}\} \times \{0, 1, ..., X_{\max}\}$, 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 rate equation (31). The truncated solution is expected to be a close approximation to the original one because sample trials produced by the stochastic simulation algorithm (Gillespie 1976) (almost) never exceed the upper bound.

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 (30) or the matrix (12)). Without an explicit upper bound in the model, each of these calculations would require a numerical 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 $\varepsilon \ll 1$ regime of slow activation. The regime is characterised by an $O(1)$-slow activation rate and $O(1/\varepsilon)$-fast production and decay rates. Consequently, the inactive protein is present at $O(1/\varepsilon)$-large copy numbers and fluctuates slowly on $O(1)$-long timescales, whereas the active protein is present at $O(1)$-low copy numbers and fluctuates fast on $O(\varepsilon)$-short timescales. Neglecting the slow fluctuations in the inactive protein, we found that the active protein obeys a one-dimensional birth–death process which equilibrates to a (truncated) Poisson quasi-steady-state (QSS) distribution. On the slow timescale, the inactive protein is produced with a self-dependent rate that is obtained by averaging the instantaneous production rate with respect to the QSS distribution of the active protein. Depending on whether this effective feedback response function has a single or multiple fixed points, the limiting deterministic dynamics of the inactive protein is monostable or bistable.

Bistability occurs if the effective feedback response function is sufficiently sigmoid. As a result of averaging by the noisy active protein, the effective response function smooths out, or "mollifies", any sharp features of the instantaneous 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 the limiting deterministic 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 mixture 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 approximate solution closely agrees 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 (Fig. 1, central panels, solid red curve). The trivial heteroclinic connections that lie on the x-axis satisfy the limiting deterministic rate equation (Fig. 1, central panels, dashed red line). The potential derivative and the deterministic rate have opposite signs: the potential has local minima/maxima where the rate equations has stable/unstable fixed points; 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 ε), 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 an $O(\varepsilon)$ timescale and the protein at $O(1/\varepsilon)$ numbers on an $O(1)$ timescale. 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 those of 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 (12), 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 (Sect. 5) from the higher-order terms of the master equation is the same as used by Newby and Chapman (2014).

In future work, it will be interesting 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 $\varepsilon \to 0$. The proportionality constant will be determined by matching the WKB solution to a solution of a Fokker–Planck equation in the neighbourhood of the unstable fixed point x_0 (Hinch and Chapman 2005; Bressloff 2014). We also expect that the current framework can be extended to more general distributed delays. Any non-negative distribution can be arbitrarily closely approximated by a phase-type distribution (Lagershausen 2012). Large deviations driven by a phase-type delay composed of m slow memoryless steps will be characterised by a Hamiltonian system with m degrees of freedom. On the other hand, a fixed delay does not buffer bursts and is conjectured to generate super-Poissonian distributions at the active protein stage. We expect that the current framework can also be extended to account for multiple protein species. Specifically, in case of a co-repressive toggle switch, we imagine that the relative stabilities of its fixed points can be modulated by the relative lengths of the two delays.

In summary, we performed a detailed analysis of the steady-state distribution for an autoregulating protein with a large one-step production delay. While both monostable and bistable feedbacks can exhibit bimodality at the single-cell level without time delay (Singh 2012; Bokes and Singh 2019), the current results imply that with the inclusion of large delays they will generate qualitatively different distributions. Our analysis thus provides a novel method to probe the structure of positive genetic feedback circuits.

Appendix: Linearisation of Hamiltonian system

Here we derive expression (37) for the second θ-derivative of the Hamiltonian (18). By doing so, we complete the linearisation analysis of the Hamiltonian system (14).

Differentiating $\boldsymbol{A}\boldsymbol{v} = H\boldsymbol{v}$ with respect to θ twice yields

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

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 $\boldsymbol{v}(x_*, 0) = \boldsymbol{\rho}(x_*)$, so that

$$\frac{\partial^2 H}{\partial \theta^2}(x_*, 0)\boldsymbol{\rho}(x_*) = \frac{\partial^2 \boldsymbol{A}}{\partial \theta^2}(x_*, 0)\boldsymbol{\rho}(x_*) + 2\frac{\partial \boldsymbol{A}}{\partial \theta}(x_*, 0)\frac{\partial \boldsymbol{v}}{\partial \theta}(x_*, 0)$$

$$+ A(x_*, 0) \frac{\partial^2 \boldsymbol{v}}{\partial \theta^2}(x_*, 0).$$

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

$$\frac{\partial^2 H}{\partial \theta^2}(x_*, 0) = \mathbf{1}^{\mathsf{T}} \frac{\partial^2 A}{\partial \theta^2}(x_*, 0)\boldsymbol{\rho}(x_*) + 2\mathbf{1}^{\mathsf{T}} \frac{\partial A}{\partial \theta}(x_*, 0) \frac{\partial \boldsymbol{v}}{\partial \theta}(x_*, 0). \qquad (50)$$

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

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

of the operator $\mathcal{A}(x, \theta)$ (11). Therefore, the first term on the right-hand side of (50) satisfies

$$\begin{aligned} \mathbf{1}^{\mathsf{T}} \frac{\partial^2 A}{\partial \theta^2}(x_*, 0)\boldsymbol{\rho}(x_*) &= \sum_{s=0}^{s_{\max}} \frac{\partial^2 \mathcal{A}}{\partial \theta^2}(x_*, 0)\rho_s(x_*) \\ &= x_* \sum_{s=0}^{s_{\max}} \mathcal{R}\rho_s(x_*) + M''(0) \sum_{s=0}^{s_{\max}} f_s \rho_s(x_*) \\ &= x_* + \langle B^2 \rangle \bar{f}(x_*) = x_* \left(1 + \frac{\langle B^2 \rangle}{\langle B \rangle} \right), \end{aligned} \qquad (51)$$

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 (50) involves the derivative of the principal eigenvector with respect to θ. Differentiating $A\boldsymbol{v} = H\boldsymbol{v}$ with respect to θ and inserting $x = x_*$ and $\theta = 0$ yields

$$A(x_*, 0) \frac{\partial \boldsymbol{v}}{\partial \theta}(x_*, 0) = -\frac{\partial A}{\partial \theta}(x_*, 0)\boldsymbol{\rho}(x_*). \qquad (52)$$

Solving the inhomogeneous linear algebraic system (52) yields

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

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

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

is a least-squares solution to (52), with A^+ denoting the pseudo-inverse of A.

The second term in (50) therefore satisfies

$$\mathbf{1}^{\mathsf{T}} \frac{\partial A}{\partial \theta}(x_*, 0) \frac{\partial \boldsymbol{v}}{\partial \theta}(x_*, 0) = \mathbf{1}^{\mathsf{T}} \frac{\partial A}{\partial \theta}(x_*, 0)\tilde{\boldsymbol{v}} + c\mathbf{1}^{\mathsf{T}} \frac{\partial A}{\partial \theta}(x_*, 0)\boldsymbol{\rho}(x_*)$$

$$= \mathbf{1}^{\mathsf{T}} \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0)\tilde{\mathbf{v}} + c\frac{\partial H}{\partial \theta}(x_*, 0)$$

$$= \mathbf{1}^{\mathsf{T}} \frac{\partial \mathbf{A}}{\partial \theta}(x_*, 0)\tilde{\mathbf{v}}. \tag{53}$$

Inserting (51) and (53) into (50) recovers (37).

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 79(6–7):2211–2236

Anderson DF, Cotter SL (2016) Product-form stationary distributions for deficiency zero networks with non-mass action kinetics. Bull 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: Graw F, Matthaeus F, Pahle J (eds) 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. Bull Math Biol 80:1871–1899. 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 P, Singh A (2019) Time delays in a genetic positive-feedback circuit. IEEE Control Syst Lett 4(1):163–168

Bressloff PC (2014) Stochastic processes in cell biology. Springer, New York

Bressloff PC, Newby JM (2014) Path integrals and large deviations in stochastic hybrid systems. Phys Rev E 89(4):042,701

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. Proc 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 Eval 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. Wiley, Hoboken

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 16:045,003

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

Lagershausen S (2012) Performance analysis of closed queueing networks, vol 663. Springer, Heidelberg

Lin YT, Buchler NE (2018) Efficient analysis of stochastic gene dynamics in the non-adiabatic regime using piecewise deterministic markov processes. J R 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 R Soc Interface 13:20150,772

McAdams H, Arkin A (1997) Stochastic mechanisms in gene expression. Proc Natl Acad Sci USA 94:814–819

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-κB 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

Plesa T, Erban R, Othmer HG (2019) Noise-induced mixing and multimodality in reaction networks. Eur J Appl Math 30(5):887–911

Popovic N, Marr C, Swain PS (2016) A geometric analysis of fast-slow models for stochastic gene expression. J Math Biol 72:87–122

Qiu H, Zhang B, Zhou T (2020) Explicit effect of stochastic reaction delay on gene expression. Phys Rev E 101(1):012,405

Rao C, Arkin A (2003) Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. J Chem Phys 118:4999–5010

Roussel MR, Zhu R (2006) Stochastic kinetics description of a simple transcription model. Bull 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, Berlin

Singh A (2012) Stochastic analysis of genetic feedback circuit controlling HIV cell-fate decision. In: 2012 IEEE 51st annual conference on decision and control (CDC). 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 JP (2007) Stochastic analysis of gene regulatory networks using moment closure. In: 2007 American control conference. IEEE, pp 1299–1304

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):e1004,972

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):154,109

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. Proc 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):250,602

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. Chin 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