

# Stationary distributions and metastable behaviour for self-regulating proteins with general lifetime distributions

Candan Çelik<sup>1</sup>, Pavol Bokes<sup>1,2</sup>, and Abhyudai Singh<sup>3</sup>

<sup>1</sup> Department of Applied Mathematics and Statistics, Comenius University,  
Bratislava 84248, Slovakia

`candan.celik@fmph.uniba.sk`

<sup>2</sup> Mathematical Institute, Slovak Academy of Sciences, Bratislava 81473, Slovakia  
`pavol.bokes@fmph.uniba.sk`

<sup>3</sup> Department of Electrical and Computer Engineering, University of Delaware,  
Newark, Delaware 19716, USA  
`absingh@udel.edu`

**Abstract.** Regulatory molecules such as transcription factors are often present at relatively small copy numbers in living cells. The copy number of a particular molecule fluctuates in time due to the random occurrence of production and degradation reactions. Here we consider a stochastic model for a self-regulating transcription factor whose lifespan (or time till degradation) follows a general distribution modelled as per a multi-dimensional phase-type process. We show that at steady state the protein copy-number distribution is the same as in a one-dimensional model with exponentially distributed lifetimes. This invariance result holds only if molecules are produced one at a time: we provide explicit counterexamples in the bursty production regime. Additionally, we consider the case of a bistable genetic switch constituted by a positively autoregulating transcription factor. The switch alternately resides in states of up- and downregulation and generates bimodal protein distributions. In the context of our invariance result, we investigate how the choice of lifetime distribution affects the rates of metastable transitions between the two modes of the distribution. The phase-type model, being non-linear and multi-dimensional whilst possessing an explicit stationary distribution, provides a valuable test example for exploring dynamics in complex biological systems.

**Keywords:** stochastic gene expression · master equation · stationary distribution · metastable systems

## 1 Introduction

Biochemical processes at the single-cell level involve molecules such as transcription factors that are present at low copy numbers [6, 46]. The dynamics of these processes is therefore well described by stochastic Markov processes in

continuous time with discrete state space [15, 22, 42]. While few-component or linear-kinetics systems [16] allow for exact analysis, in more complex system one often uses approximative methods [12], such as moment closure [4], linear-noise approximation [3, 9], hybrid formulations [25, 26, 33], and multi-scale techniques [38, 39].

In simplest Markovian formulations, the lifetime of a regulatory molecule is memoryless, i.e. exponentially distributed [10, 47]. However, transcription factors are complex macromolecules, which can be present in various molecular conformations, and whose removal can require a complex interplay of multiple pathways [13, 37]. Therefore, their lifetime distributions can assume far more complex forms than the simple exponential.

In Section 2, we formulate, both in the deterministic and stochastic settings, a one-dimensional model for the abundance of a transcription factor with a memoryless lifetime. Since many transcription factors regulate their own gene expression [2], we allow the production rate to vary with the copy number. We show that the deterministic solutions tend to the fixed points of the feedback response function; in the stochastic framework, we provide the stationary distribution of the protein copy number.

In Section 3, we proceed to characterise the steady-state behaviour of a structured model that accounts for complex lifetime pathways. The model is multidimensional, each dimension corresponding to a different class and stage of a molecule’s lifetime; the chosen structure accounts for a wide class of phase-type lifetime distributions [34, 45]. We demonstrate that the deterministic fixed points and the stochastic stationary distribution that were found for the one-dimensional framework remain valid for the total protein amount in the multi-dimensional setting.

We emphasise that the distribution invariance result rests on the assumption of non-bursty production of protein. The case of bursty production is briefly discussed in Section 4, where explicit counter-examples are constructed by means of referring to explicit mean and variance formulae available from literature for systems without feedback [28, 36].

In the final Section 5, we approximate the stochastic protein distribution by a mixture of Gaussians with means at deterministic fixed points and variances given by the linear-noise approximation [8, 30]. Additionally, we study the rates of metastable transitions [40, 43] between the Gaussian modes in the one-dimensional and structured settings.

## 2 One-dimensional model

*Deterministic framework.* The dynamics of the abundance of protein  $X$  at time  $t$  can be modelled deterministically by an ordinary differential equation

$$\frac{dX}{dt} = \tau^{-1} (f(X) - X), \quad (1)$$

which states that the rate of change in  $X$  is equal to the difference of production and decay rates. The decay rate is proportional to  $X$ ; the factor of proportion-

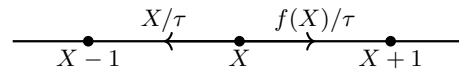


Fig. 1: A diagram of the one-dimensional model. The number of molecules  $X$  can decrease by one or increase by one. The stochastic rates (or propensities) of these transitions are indicated above the transition edges.

ality is the reciprocal of the expected lifetime  $\tau$ . The rate of production per unit protein lifetime is denoted by  $f(X)$  in (1); the dependence of the production rate on the protein amount  $X$  implements the feedback in the model. Equating the right-hand side of (1) to zero yields

$$f(X) = X, \quad (2)$$

meaning that steady states of (1) are given by the fixed point of the production response function  $f(X)$ .

*Stochastic framework.* The stochastic counterpart of (1) is the Markov process with discrete states  $X \in \mathbb{N}_0$  in continuous time with transitions  $X \rightarrow X - 1$  or  $X \rightarrow X + 1$ , occurring with rates  $X/\tau$  and  $f(X)/\tau$  respectively (see the schematic in Figure 1). Note that in case of a constant production rate, i.e.  $f(X) \equiv \lambda$ , the model turns into the immigration-and-death process [32]; in queueing theory this is also known as  $M/M/\infty$  queue [21]. The stationary distribution of the immigration-death process is known to be Poissonian with mean equal to  $\lambda$  [32].

For a system with feedback, the probability  $P(X, t)$  of having  $X$  molecules at time  $t$  satisfies the master equation

$$\frac{dP(X, t)}{dt} = \tau^{-1} (\mathbb{E}^{-1} - 1) f(X) P(X, t) + \tau^{-1} (\mathbb{E} - 1) X P(X, t), \quad (3)$$

in which  $\mathbb{E}$  is the van-Kampen step operator [30]. Inserting  $P(X, t) = \pi(X)$  into (3) and solving the resulting difference equation, one finds a steady-state distribution in the explicit form

$$\pi(X) = \pi(0) \frac{\prod_{k=0}^{X-1} f(k)}{X!}. \quad (4)$$

The probability  $\pi(0)$  of having zero molecules plays the role of the normalisation constant in (4), which can be uniquely determined by imposing the normalisation condition  $\pi(0) + \pi(1) + \dots = 1$ . Note that inserting  $f(X) \equiv \lambda$  into (4) results in the aforementioned Poissonian distribution with  $\pi(0) = e^{-\lambda}$ .

### 3 Multiclass-multistage model

In this section, we introduce a structured multiclass-multistage model which is an extension of one-dimensional model introduced in the previous section. The

fundamentals of the multidimensional model are as shown in Figure 2. A newly produced molecule is assigned into one of  $K$  distinct classes. Which class is selected is chosen randomly according to a discrete distribution  $p_1, \dots, p_K$ . The lifetime of a molecule in the  $i$ -th class consists of  $S_i$  stages. The holding time in any of these stages is memoryless (exponential), and parametrised by its mean  $\tau_{ij}$ , where  $i$  indicates which class and  $j$  indicates which stage. Note that

$$\tau = \sum_{i=1}^K \sum_{j=1}^{S_i} p_i \tau_{ij} \quad (5)$$

gives the expected lifetime of a newly produced molecule. After the last ( $S_i$ -th) stage, the molecule is degraded. The total distribution of a molecule lifetime is a mixture, with weights  $p_i$ , of the lifetime distributions of the individual classes, each of which is a convolution of exponential distributions of the durations of the individual stages; such distributions are referred to as phase-type distribution and provide a wide family of distribution to approximate practically any distribution of a positive random variable [45].

We denote by  $X_{ij}$  the number of molecules in the  $i$ -th class and the  $j$ -th stage of their lifetime, by

$$X = (X_{11}, \dots, X_{1S_1}, X_{21}, \dots, X_{2S_2}, \dots, X_{K1}, \dots, X_{KS_K})$$

the  $\sum_{i=1}^K S_i$ -dimensional copy-number vector, and by

$$\|X\| = \sum_{i=1}^K \sum_{j=1}^{S_i} X_{ij} \quad (6)$$

the total number of molecules across all classes and stages.

*Deterministic framework.* The deterministic description of the structured model is given by a system of coupled ordinary differential equations

$$\frac{dX_{i1}}{dt} = \frac{p_i f(\|X\|)}{\tau} - \frac{X_{i1}}{\tau_{i1}}, \quad i = 1, \dots, K, \quad (7)$$

$$\frac{dX_{ij}}{dt} = \frac{X_{ij-1}}{\tau_{ij-1}} - \frac{X_{ij}}{\tau_{ij}}, \quad i = 1, \dots, K \text{ and } j = 2, \dots, S_i. \quad (8)$$

The right-hand sides of (7)–(8) are each equal to the difference of appropriate arrival and departure rates at/from a particular compartment of the structured model. The departure rates are proportional to the number of molecules in the compartment, with the reciprocal of the holding time giving the factor of proportionality. The arrival rate takes a different form for the first stages (7) and for the other stages (8). For the first stage, the arrival is obtained by the product of the production rate  $f(\|X\|)/\tau$  and the probability  $p_i$  of selecting the  $i$ -th class. For the latter stages, the arrival rate is equal to the departure rate of the previous stage.

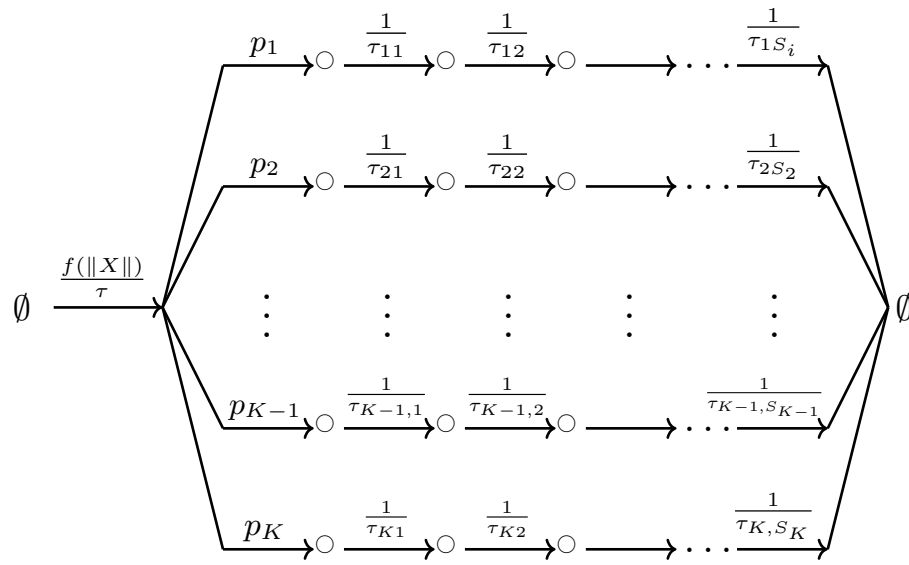


Fig. 2: A schematic representation of multiclass-multistage model. A newly produced molecule is randomly assigned, according to a prescribed distribution  $p_1, \dots, p_K$ , into one of  $K$  distinct classes. The lifetime of a molecule in the  $i$ -th class consists of  $S_i$  consecutive memoryless stages, and ends in the degradation of the molecule. The expected holding time in the  $j$ -th stage of the  $i$ -th class is  $\tau_{ij}$ . The production rate is a function of the total number  $\|X\|$  of molecules across all stages and classes.

Equating (7)–(8) to zero, we find that

$$\frac{p_i f(\|X\|)}{\tau} = \frac{X_{i1}}{\tau_{i1}} = \frac{X_{i2}}{\tau_{i2}} = \dots = \frac{X_{ij}}{\tau_{ij}} \quad (9)$$

hold at steady state, from which it follows that

$$X_{ij} = \frac{p_i \tau_{ij} f(\|X\|)}{\tau}. \quad (10)$$

Summing (10) over  $i = 1, \dots, K$  and  $j = 2, \dots, S_i$ , and using (5) and (6), yield

$$\|X\| = f(\|X\|) \quad (11)$$

for the total protein amount (6). Thus, the protein amount at steady state is obtained, like in the one-dimensional model, by calculating the fixed points of the feedback response function.

Combining (11) and (9) we find

$$X_{ij} = \frac{p_i \tau_{ij} \|X\|}{\tau}, \quad (12)$$

which means that at steady state the total protein amount is distributed among the compartments proportionally to the product of class assignment probability and the mean holding time of the particular compartment.

*Stochastic framework.* Having shown that the stationary behaviour of the one-dimensional and the structured multi-dimensional models are the same in the deterministic framework, we next aim to demonstrate that the same is also true in the stochastic context. Prior to turning our attention to the feedback system, it is again instructive to discuss the case without regulation, i.e.  $f(\|X\|) \equiv \lambda$ ; the new molecule arrivals are then exponentially distributed. In the language of queueing theory, the process can be reinterpreted as the  $M/G/\infty$  queue with exponential arrivals of customers, a general phase-type distribution of service times, and an infinite number of servers. It is well known that the steady-state distribution of an  $M/G/\infty$  queue is Poisson with mean equal to  $\lambda$  [41]. Thus, without feedback, we obtain the very same Poisson( $\lambda$ ) distribution that applies in the one-dimensional case.

In the feedback case, the probability  $P(X, t)$  of having  $X = (X_{11}, \dots, X_{K, S_K})$  copy numbers in the individual compartments at any time  $t$  satisfies the master equation

$$\frac{dP(X, t)}{dt} = \tau^{-1} \sum_{i=1}^K p_i (\mathbb{E}_{i1}^{-1} - 1) f(\|X\|) P(X, t) \quad (13)$$

$$+ \sum_{i=1}^K \sum_{j=1}^{S_i-1} \tau_{ij}^{-1} (\mathbb{E}_{ij} \mathbb{E}_{ij+1}^{-1} - 1) X_{ij} P(X, t) \quad (14)$$

$$+ \sum_{i=1}^K \tau_{iS_i}^{-1} (\mathbb{E}_{iS_i} - 1) X_{iS_i} P(X, t). \quad (15)$$

The right-hand-side terms (13), (14), and (15) stand for the change in probability mass function due to the production, moving to next stage, and decay reactions, respectively. Note that  $\mathbb{E}_{ij}$  is a step operator which increases the copy number of molecules in the  $i$ -th class at the  $j$ -th stage by one [30]. Likewise,  $\mathbb{E}_{ij}^{-1}$  decreases the same copy number by one. Rearrangement of terms in the master equation yields

$$\begin{aligned} \frac{dP(X, t)}{dt} = & \sum_{i=1}^K (\tau^{-1} p_i \mathbb{E}_{i1}^{-1} f(\|X\|) P(X, t) - \tau_{i1}^{-1} X_{i1} P(X, t)) \\ & + \sum_{i=1}^K \sum_{j=1}^{S_i-1} (\tau_{ij}^{-1} \mathbb{E}_{ij} \mathbb{E}_{ij+1}^{-1} X_{ij} P(X, t) - \tau_{ij+1}^{-1} X_{ij+1} P(X, t)) \\ & + \sum_{i=1}^K \tau_{iS_i}^{-1} \mathbb{E}_{iS_i} X_{iS_i} P(X, t) - \tau^{-1} f(\|X\|) P(X, t). \end{aligned}$$

Equating the derivative to zero, we derive for the stationary distribution  $\pi(X)$  an algebraic system

$$\begin{aligned} 0 = & \sum_{i=1}^K (\tau^{-1} p_i \mathbb{E}_{i1}^{-1} f(\|X\|) \pi(X) - \tau_{i1}^{-1} X_{i1} \pi(X)) \\ & + \sum_{i=1}^K \sum_{j=1}^{S_i-1} (\tau_{ij}^{-1} \mathbb{E}_{ij} \mathbb{E}_{ij+1}^{-1} X_{ij} \pi(X) - \tau_{ij+1}^{-1} X_{ij+1} \pi(X)) \\ & + \sum_{i=1}^K \tau_{iS_i}^{-1} \mathbb{E}_{iS_i} X_{iS_i} \pi(X) - \tau^{-1} f(\|X\|) \pi(X). \end{aligned} \quad (16)$$

Clearly, it is sufficient that

$$\begin{aligned} \tau^{-1} p_i \mathbb{E}_{i1}^{-1} f(\|X\|) \pi(X) &= \tau_{i1}^{-1} X_{i1} \pi(X), \\ \tau_{ij}^{-1} \mathbb{E}_{ij} \mathbb{E}_{ij+1}^{-1} X_{ij} \pi(X) &= \tau_{ij+1}^{-1} X_{ij+1} \pi(X), \\ \sum_{i=1}^K \tau_{iS_i}^{-1} \mathbb{E}_{iS_i} X_{iS_i} \pi(X) &= \tau^{-1} f(\|X\|) \pi(X) \end{aligned} \quad (17)$$

hold for  $\pi(X)$  in order that (16) be satisfied. One checks by direct substitution that

$$\pi(X) \propto \prod_{k=0}^{\|X\|-1} f(k) \times \prod_{i=1}^K \prod_{j=1}^{S_i} \frac{(p_i \tau_{ij} / \tau)^{X_{ij}}}{X_{ij}!} \quad (18)$$

satisfies (17); therefore, (18) represents the stationary distribution of the structured model. In order to interpret (18), we condition the joint distribution on the total protein copy number, writing

$$\pi(X) = \pi_{\text{cond}}(X \mid \|X\|) \pi_{\text{tot}}(\|X\|), \quad (19)$$

in which the conditional distribution is recognised as the multinomial [29]

$$\pi_{\text{cond}}(X \mid \|X\|) = \binom{\|X\|}{X} \prod_{i=1}^K \prod_{j=1}^{S_i} (p_i \tau_{ij} / \tau)^{X_{ij}}, \quad (20)$$

and the total copy number distribution is given by

$$\pi_{\text{tot}}(\|X\|) = \pi_{\text{tot}}(0) \frac{\prod_{k=0}^{\|X\|-1} f(k)}{\|X\|!}. \quad (21)$$

By (20), the conditional means of  $X_{ij}$  coincide with the deterministic partitioning of the total copy number (12). Importantly, comparing (21) to (4), we conclude that the one-dimensional and multi-dimensional models generate the same (total) copy number distributions.

## 4 Bursting

The independence of stationary distribution on the lifetime distribution relies on the assumption of non-bursty production of protein that has implicitly been made in our model. In this section, we allow for the synthesis of protein in bursts of multiple molecules at a single time [14, 17]. Referring to previously published results [28, 36], we provide an counterexample that demonstrates that in the bursty case different protein lifetime distributions can lead to different stationary copy-number distributions. The counterexample can be found even in the absence of feedback.

Let  $G(t) = \text{Prob}[T > t]$  denote the survival function of an individual molecule and let

$$\tau = - \int_0^\infty t G'(t) dt = \int_0^\infty G(t) dt \quad (22)$$

be the mean lifetime. Bursty production means that the number of molecules can change within an infinitesimally small time interval of length  $dt$  from  $X$  to  $X + j$ , where  $j \geq 1$ , with probability  $\lambda \tau^{-1} b_j dt$ , in which  $\lambda$  is the burst frequency and  $b_j = \text{Prob}[B = j]$  is the probability mass function of the random burst size  $B$  variable.

In queueing theory, bursty increases in the state variable are referred to as batch customer arrivals. Specifically, a bursty gene-expression model without feedback and with general lifetime distribution corresponds to the  $M^X/G/\infty$  queue with memoryless (exponential) batch arrivals, general service distribution, and an infinite number of servers.

Previous analyses [28, 36] show that the steady-state protein mean  $\langle X \rangle$  and the Fano factor  $F = \text{Var}(X)/\langle X \rangle$  are given by

$$\langle X \rangle = \lambda \langle B \rangle, \quad F = 1 + K_s \left( \frac{\langle B^2 \rangle}{\langle B \rangle} - 1 \right), \quad (23)$$



where

$$K_s = \frac{\int_0^\infty G^2(t)dt}{\tau} \quad (24)$$

is referred to as the senescence factor. Elementary calculation shows that  $K_s = 1/2$  if the lifetime distribution is exponential with survival function  $G(t) = e^{-t/\tau}$  and that  $K_s = 1$  if the lifetime distribution is deterministic with survival function  $G(t) = 1$  for  $t < \tau$  and  $G(t) = 0$  for  $t \geq \tau$ . Thus, although two lifetime distributions result in the same value of the stationary mean protein copy number, they give a different value of the noise (the Fano factor); therefore the copy-number distributions are different.

## 5 Metastable transitioning

Transcription factors that self-sustain their gene expression by means of a positive feedback loop can act as a simple genetic switch [5, 20]. A positive-feedback switch can be in two states, one in which the gene is fully activated through its feedback loop, while in the other the gene is expressed at a basal level. The switch serves as a basic memory unit, retaining the information on its initial state on long timescales, and very slowly relaxing towards an equilibrium distribution. It is therefore important to investigate not only the stationary, but also transient distributions, which are generated by a positively autoregulating transcription factor.

Following previous studies [7, 11, 19], we model positive feedback by the Hill function response curve

$$f(X) = \Omega \left( a_0 + \frac{a_1 X^H}{\Omega^H + X^H} \right), \quad (25)$$

in which  $a_0$  and  $a_1$  represent the basal and regulable production rates,  $H$  is the cooperativity coefficient, and  $\Omega$  gives the critical amount of protein required for half-stimulation of feedback. Provided that  $H > 1$ , one can find  $a_0$  and  $a_1$  such that (25) possesses three distinct fixed points  $X_- < X_0 < X_+$ , of which the central is unstable and the other two are stable (Figure 3, left). The two stable fixed points provide alternative large-time outcomes of the deterministic models (1) and (7)–(8).

Bistability of deterministic models translates into bimodal distributions in the stochastic framework. For large values of  $\Omega$ , the bimodal protein distribution can be approximated by a mixture of Gaussian modes which are located at the stable fixed points  $X_\pm$  (see Figure 4),

$$P(X, t) \sim p_-(t) \frac{e^{-\frac{(X-X_-)^2}{2\sigma_-^2}}}{\sqrt{2\pi}\sigma_-} + p_+(t) \frac{e^{-\frac{(X-X_+)^2}{2\sigma_+^2}}}{\sqrt{2\pi}\sigma_+}. \quad (26)$$

The mixture approximation (26) is determined not only by the locations  $X_\pm$ , but also on the variances  $\sigma_\pm^2$  and the weights  $p_\pm(t)$  of the two modes (which

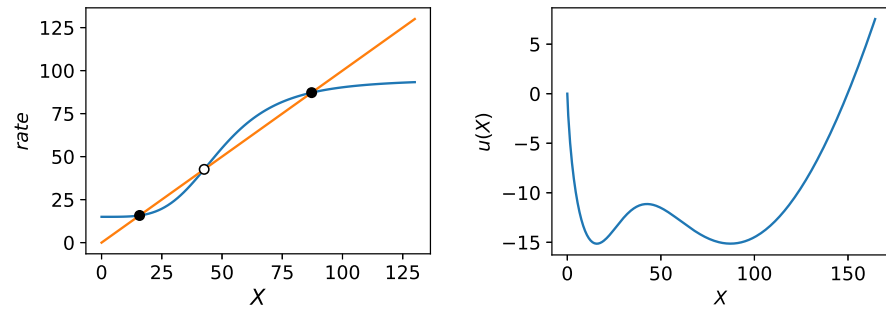


Fig. 3: *Left*: A sigmoid feedback response function (blue curve) intersects the diagonal (orange line) in multiple fixed points. Ones that are stable to the rate equation (1) (full circles) are interspersed by unstable ones (empty circle). *Right*: The potential  $u(x)$ , defined by (33), is a Lyapunov function of the rate equation (1). The local minima, or the troughs/wells, of the potential are situated at its stable fixed points; the local maximum, or the barrier, of the potential coincides with the unstable fixed point. *Parameter values for both panels*: We use the Hill-type response (25) with  $a_0 = 0.3$ ,  $a_1 = 1.6$ ,  $H = 4$ ,  $\Omega = 50$ .

are given below). The weights in (26) are allowed to vary with time in order to account for the slow, metastable transitions that occur between the distribution modes.

The invariance result for stationary distributions derived in the preceding sections implies that, in the limit of  $t \rightarrow \infty$ , the protein distribution (26) becomes independent of the choice of the protein lifetime distribution. In particular, the same variances  $\sigma_{\pm}^2$  and the same limit values  $p_{\pm}(\infty)$  of the weights will apply for exponentially distributed and phase-type decay processes. In what follows, we first consult literature to provide results  $\sigma_{\pm}^2$  and  $p_{\pm}(t)$  that apply for the one-dimensional model with exponential decay. Next, we use stochastic simulation to investigate the effect of phase-type lifespan distributions on metastable dynamics.

The variances of the modes are obtained by the linear-noise approximation [35, 44] of the master equation (3), which yields

$$\sigma_{\pm}^2 = \frac{X_{\pm}}{1 - f'(X_{\pm})}; \quad (27)$$

the right-hand side of (27) is equal to the ratio of a fluctuation term (equal to the number of molecules) to a dissipation term (obtained by linearising the rate equation (1) around a stable fixed point).

The metastable transitions between the distribution modes can be described by a random telegraph process (cf. Figure 5, left)

$$\ominus \xrightleftharpoons[1/T_+]{1/T_-} \oplus, \quad (28)$$

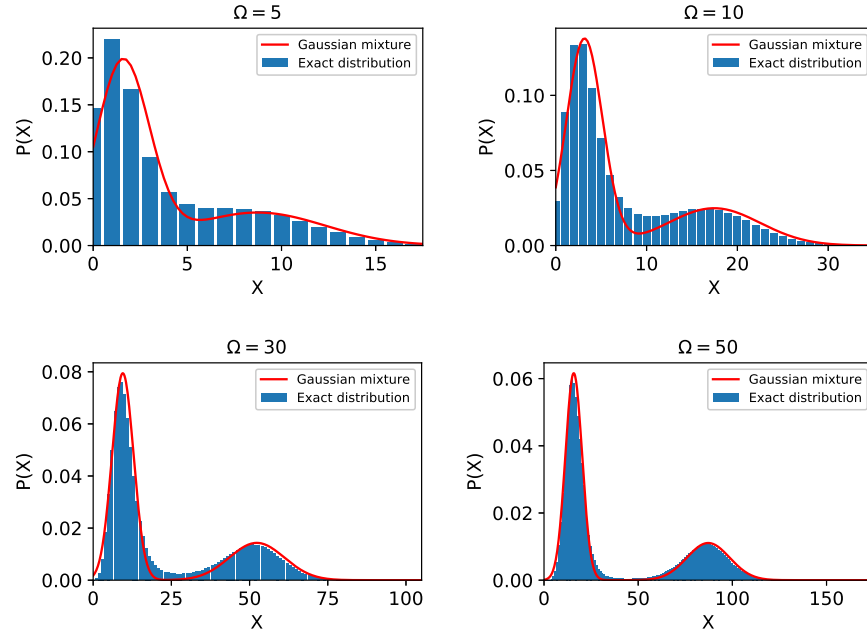


Fig. 4: Exact stationary protein distribution (4) and the Gaussian-mixture approximation (26) in varying system-size conditions. The means of the Gaussians are given by the stable fixed points of  $f(X)$ ; the variances are given by linear-noise approximation (27). The mixture weights are given by  $p_+(\infty) = T_+/(T_+ + T_-)$ ,  $p_-(\infty) = T_-/(T_+ + T_-)$ , where the residence times are given by the Arrhenius-type formula (32). We use a Hill-type response (25) with  $a_0 = 0.3$ ,  $a_1 = 1.6$ ,  $H = 4$ , and  $\Omega$  shown in panel captions.

in which the lumped states  $\ominus$  and  $\oplus$  correspond to the basins of attractions of the two stable fixed points;  $T_-$  and  $T_+$  are the respective residence times. The mixture weights  $p_-(t)$  and  $p_+(t)$  in (26) are identified with the probabilities of the lumped states in (28); these satisfy the Chapman–Kolmogorov equations

$$\frac{dp_-}{dt} = -\frac{p_-}{T_-} + \frac{p_+}{T_+}, \quad \frac{dp_+}{dt} = \frac{p_-}{T_-} - \frac{p_+}{T_+}, \quad (29)$$

which admit an explicit solution

$$p_+(t) = \frac{T_+}{T_+ + T_-} + \left( p_+(0) - \frac{T_+}{T_+ + T_-} \right) \exp \left( - \left( \frac{1}{T_+} + \frac{1}{T_-} \right) t \right), \quad (30)$$

$$p_-(t) = \frac{T_-}{T_+ + T_-} + \left( p_-(0) - \frac{T_-}{T_+ + T_-} \right) \exp \left( - \left( \frac{1}{T_+} + \frac{1}{T_-} \right) t \right). \quad (31)$$

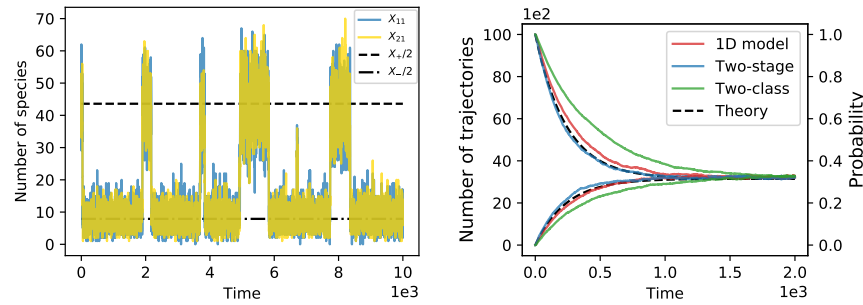


Fig. 5: *Left*: Large-time stochastic trajectories of a structured two-class model with parameters as given below. The horizontal lines represent deterministic fixed points as given by (11)–(12). *Right*: The number of trajectories, out of  $10^4$  simulation repeats, that reside in the basin of attraction of the upper stable fixed point as function of time. Simulation is initiated at the upper stable fixed point (the decreasing function) or at the lower stable fixed point (the increasing function). The dashed black curve gives the theoretical probability (30) with initial condition  $p_+(0) = 1$  (the decreasing solution) or  $p_+(0) = 0$  (the increasing solution). *Parameter values*: The Hill-function parameters are:  $\Omega = 50$ ,  $H = 4$ ,  $a_0 = 0.3$ ,  $a_1 = 1.6$ . The mean lifetime is  $\tau = 1$ . The two-stage model parameters are:  $K = 1$ ,  $p_1 = 1$ ,  $S_1 = 2$ ,  $\tau_{11} = \tau_{12} = 0.5$ . The two-class model parameters are:  $K = 2$ ,  $p_1 = 1/6$ ,  $p_2 = 5/6$ ,  $S_1 = S_2 = 1$ ,  $\tau_{11} = 3$ ,  $\tau_{21} = 3/5$ .

The initial probability  $p_+(0) = 1 - p_-(0)$  is set to one or zero in (30)–(31) depending on whether the model is initialised in the neighbourhood of the upper or the lower stable fixed point.

With (30)–(31) at hand, the problem of determining the mixture weights in (26) is reduced to that of determining the residence times  $T_{\pm}$ . Previous large-deviation and WKB analyses of the one-dimensional model [18, 23, 24] provide an Arrhenius-type formula

$$T_{\pm} = 2\pi\tau X_{\pm}^{-1}\sigma_{\pm}\sqrt{-\sigma_0^2}\exp(u(X_0) - u(X_{\pm})). \quad (32)$$

Formula (32) features, on top of the familiar symbols (the mean lifetime  $\tau$ , fixed points  $X_{\pm}$  and  $X_0$ , linearised variances  $\sigma_{\pm}$ , and the Ludolph-van-Ceulen constant  $\pi$ ), two new symbols: a value  $\sigma_0^2$  and a function  $u(X)$ . The value  $\sigma_0^2$  is readily calculated by inserting 0 instead of  $\pm$  into the fluctuation–dissipation relation (27); note that for the unstable fixed point  $X_0$ , the denominator in (27) is negative (cf. Figure 3, left), which renders the whole fraction also negative.

In analogy with the Arrhenius law, the function  $u(X)$  represents an “energy” of state  $X$ , and is given here explicitly by an indefinite integral [18, 23, 24]

$$u(X) = \int \ln\left(\frac{X}{f(X)}\right) dX. \quad (33)$$

Note that the derivative of (33),

$$u'(X) = \ln \left( \frac{X}{f(X)} \right), \quad (34)$$

is zero if  $f(X) = X$ , i.e. at the fixed points of the feedback response function, is negative if  $f(X) > X$  and positive if  $f(X) < X$ . Substituting into (33) the solution  $X = X(t)$  to the deterministic rate equation (1) and evaluating the time derivative, we find

$$\frac{du(X(t))}{dt} = u'(X(t)) \frac{dX(t)}{dt} = \tau^{-1} (f(X) - X) \ln \left( \frac{X}{f(X)} \right) \Big|_{X=X(t)} \leq 0, \quad (35)$$

with equality in (35) holding if and only if  $X$  is a fixed point of the feedback response function  $f(X)$ . Therefore, the energy function  $u(X)$  is a Lyapunov function of the ordinary differential equation (1) (Figure 3, right). The exponentiation in (32) dramatically amplifies the potential difference between the stable and the unstable fixed points. For example, a moderately large potential barrier, say 5 (which is about the height of the potential barrier in Figure 3, right), introduces a large factor  $e^5 \approx 150$  in (32). This confirms an intuition that metastable transitions between the distribution modes are very (exponentially) slow.

The random telegraph solution (30) is compared in Figure 5 to the residence of stochastically generated trajectories in the basin of attraction of the upper fixed point. The agreement is close for simulations of the one-dimensional model (with an exponential lifetime) and for a structured model with one class and two stages (with an Erlangian lifetime). For a two-class model (with an exponential mixture lifetime), the transitioning also occurs on the exponentially slow timescale, but is perceptibly slower. Sample trajectories were generated in Python's package for stochastic simulation of biochemical systems GillesPy2 [1]. The one-dimensional model was initiated with  $\lfloor X_+ \rfloor$  molecules. The two species in the two-stage and two-class models were initiated to  $S$  and  $\lfloor X_+ \rfloor - S$ , where  $S$  was drawn from the binomial distribution  $\text{Binom}(\lfloor X_+ \rfloor, 0.5)$ .

## 6 Discussion

In this paper we studied a stochastic chemical reaction system for a self-regulating protein molecule with exponential and phase-type lifetimes. We demonstrated that the exponential and phase-type models support the same stationary distribution of the protein copy number. While stationary distributions of similar forms have previously been formulated in the context of queueing theory [21, 27, 31], our paper provides a self-contained and concise treatment of the one-dimensional model and the multi-dimensional structured model that is specifically tailored for applications in systems biology.

We showed that the invariance result rests on the assumption of non-bursty production of protein. We demonstrated that, in the presence of bursts, exponential and deterministic lifetimes generate stationary protein-level distributions with different variances.

Deterministic modelling approaches are used in systems biology as widely as stochastic ones. Therefore, we complemented the stationary analysis of the stochastic Markov-chain models by a fixed-point analysis of deterministic models based on differential equations. The result is that, irrespective of lifetime distribution, the deterministic protein level is attracted, for large times, to the stable fixed points of the feedback response function. Connecting the stochastic and deterministic frameworks, we demonstrated that the stationary distribution of the Markovian model is sharply peaked around the fixed points of the deterministic equation. We showed that the distribution can be approximated by a mixture of Gaussian modes with means given by the deterministic fixed points and variances that are consistent with the traditional linear-noise analysis results.

Next, we focused on the transitions between the distribution modes. These occur rarely with rates that are exponentially small. We compared an asymptotic result, derived in previous literature for the one-dimensional model, to stochastic simulation results of the one-dimensional model and two specific structured models: we chose a model with one class and two stages and a model with two classes each with one stage. The simulation results of the one-dimensional and two-stage models agreed closely to the theoretical prediction; intriguingly, the agreement with theory was closer for the two-stage model. On the other hand, a two-class model showed slower transitioning rates. The theoretical asymptotic results have been derived in [18, 23, 24] only for the one-dimensional model. Large deviations in multi-dimensional models are much harder to quantify than one-variable ones. We believe that the current model, being multi-dimensional while possessing a tractable steady-state distribution, provides a convenient framework on which such methodologies can be developed.

In summary, our study provides an invariance-on-lifetime-distribution result in the deterministic and stochastic contexts for a non-bursty regulatory protein. While the main results concern the stationary behaviour, our study also performs simulation, and opens avenue for future enquiries, into the transient transitioning dynamics.

## Bibliography

- [1] Abel, J.H., Drawert, B., Hellander, A., Petzold, L.R.: Gillespy: A python package for stochastic model building and simulation. *IEEE Life Sci. Lett.* **2**(3), 35–38 (2017)
- [2] Alon, U.: *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC (2007)
- [3] Andreychenko, A., Bortolussi, L., Grima, R., Thomas, P., Wolf, V.: Distribution approximations for the chemical master equation: comparison of the method of moments and the system size expansion. In: *Modeling Cellular Systems*, pp. 39–66. Springer (2017)
- [4] Backenköhler, M., Bortolussi, L., Wolf, V.: Control variates for stochastic simulation of chemical reaction networks. In: *International Conference on Computational Methods in Systems Biology*. pp. 42–59. Springer (2019)
- [5] Becskei, A., S  raphin, B., Serrano, L.: Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–35 (2001)
- [6] Blake, W., Kaern, M., Cantor, C., Collins, J.: Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003)
- [7] Bokes, P., Lin, Y., Singh, A.: High cooperativity in negative feedback can amplify noisy gene expression. *B. Math. Biol.* **80**, 1871–1899 (2018)
- [8] Bokes, P., Borri, A., Palumbo, P., Singh, A.: Mixture distributions in a stochastic gene expression model with delayed feedback. *bioRxiv* p. 855783 (2019)
- [9] Bokes, P., Hojcka, M., Singh, A.: Buffering gene expression noise by microRNA based feedforward regulation. In:   eska, M.,   afr  nek, D. (eds.) *Computational Methods in Systems Biology, 16th International Conference, CMSB 2018. Lect. N. Bioinformat.*, vol. 11095, pp. 129–145 (2018)
- [10] Bokes, P., King, J.R., Wood, A.T., Loose, M.: Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *J. Math. Biol.* **64**, 829–854 (2012)
- [11] Bokes, P., Singh, A.: Controlling noisy expression through auto regulation of burst frequency and protein stability. In:   eska M., Paoletti N. (eds) *Hybrid Systems Biology. HSB 2019. Lecture Notes in Computer Science*, vol 11705. Springer, Cham (2019)
- [12] Bortolussi, L., Lanciani, R., Nenzi, L.: Model checking Markov population models by stochastic approximations. *Inform. Comput.* **262**, 189–220 (2018)
- [13] Bratsun, D., Volfson, D., Tsimring, L.S., Hasty, J.: Delay-induced stochastic oscillations in gene regulation. *P. Natl. Acad. Sci. USA* **102**(41), 14593–14598 (2005)
- [14] Cai, L., Friedman, N., Xie, X.: Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006)
- [15] Cinquemani, E.: Identifiability and reconstruction of biochemical reaction networks from population snapshot data. *Processes* **6**(9), 136 (2018)

- [16] Cinquemani, E.: Stochastic reaction networks with input processes: Analysis and application to gene expression inference. *Automatica* **101**, 150–156 (2019)
- [17] Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., Weinberger, L.S.: Transcriptional burst frequency and burst size are equally modulated across the human genome. *P. Natl. Acad. Sci. USA* **109**, 17454–17459 (2012)
- [18] Escudero, C., Kamenev, A.: Switching rates of multistep reactions. *Phys. Rev. E* **79**(4), 041149 (2009)
- [19] Friedman, N., Cai, L., Xie, X.: Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302 (2006)
- [20] Griffith, J.: Mathematics of cellular control processes II. positive feedback to one gene. *J. Theor. Biol.* **20**(2), 209–216 (1968)
- [21] Gross, D.: Fundamentals of queueing theory. John Wiley & Sons, Hoboken, N.J. (2008)
- [22] Guet, C., Henzinger, T.A., Igler, C., Petrov, T., Sezgin, A.: Transient memory in gene regulation. In: International Conference on Computational Methods in Systems Biology. pp. 155–187. Springer (2019)
- [23] Hanggi, P., Grabert, H., Talkner, P., Thomas, H.: Bistable systems: Master equation versus Fokker–Planck modeling. *Phys. Rev. A* **29**(1), 371 (1984)
- [24] Hinch, R., Chapman, S.J.: Exponentially slow transitions on a Markov chain: the frequency of calcium sparks. *Eur. J. Appl. Math.* **16**(04), 427–446 (2005)
- [25] Innocentini, G.C., Antoneli, F., Hodgkinson, A., Radulescu, O.: Effective computational methods for hybrid stochastic gene networks. In: International Conference on Computational Methods in Systems Biology. pp. 60–77. Springer (2019)
- [26] Innocentini, G.C., Hodgkinson, A., Radulescu, O.: Time dependent stochastic mrna and protein synthesis in piecewise-deterministic models of gene networks. *Front. Phys.* **6**, 46 (2018)
- [27] Jackson, J.R.: Jobshop-like queueing systems. *Manage. Sci.* **10**(1), 131–142 (1963)
- [28] Jia, T., Kulkarni, R.: Intrinsic noise in stochastic models of gene expression with molecular memory and bursting. *Phys. Rev. Lett.* **106**(5), 58102 (2011)
- [29] Johnson, N., Kotz, S., Kemp, A.: Univariate Discrete Distributions, 3rd ed. Wiley-Interscience (2005)
- [30] van Kampen, N.: Stochastic Processes in Physics and Chemistry. Elsevier, Amsterdam (2006)
- [31] Kelly, F.P.: Reversibility and stochastic networks. Cambridge University Press (2011)
- [32] Kendall, D.: Stochastic processes and population growth. *J. Roy. Stat. Soc. B* **11**, 230–82 (1949)
- [33] Kurasov, P., Lück, A., Mugnolo, D., Wolf, V.: Stochastic hybrid models of gene regulatory networks — a PDE approach. *Math. Biosci.* **305**, 170–177 (2018)



- [34] Lagershausen, S.: Performance analysis of closed queueing networks, vol. 663. Springer Science & Business Media, Heidelberg (2012)
- [35] Lestas, I., Paulsson, J., Ross, N., Vinnicombe, G.: Noise in gene regulatory networks. *IEEE T. Circuits-I* **53**(1), 189–200 (2008)
- [36] Liu, L., Kashyap, B., Templeton, J.: On the  $GI^X/G/\infty$  system. *J. Appl. Probab.* **27**(3), 671–683 (1990)
- [37] Miękisz, J., Poleszczuk, J., Bodnar, M., Foryś, U.: Stochastic models of gene expression with delayed degradation. *B. Math. Biol.* **73**(9), 2231–2247 (2011)
- [38] Michaelides, M., Hillston, J., Sanguinetti, G.: Geometric fluid approximation for general continuous-time Markov chains. *Proc. Roy. Soc. A* **475**(2229), 20190100 (2019)
- [39] Michaelides, M., Hillston, J., Sanguinetti, G.: Statistical abstraction for multi-scale spatio-temporal systems. *ACM T. Model. Comput. S.* **29**(4), 1–29 (2019)
- [40] Newby, J., Chapman, S.J.: Metastable behavior in Markov processes with internal states. *J. Math. Biol.* **69**(4), 941–976 (2014)
- [41] Norris, J.R.: Markov chains. Cambridge Univ Press, Cambridge, UK (1998)
- [42] Prajapat, M.K., Ribeiro, A.S.: Added value of autoregulation and multi-step kinetics of transcription initiation. *R. Soc. Open Sci.* **5**(11), 181170 (2018)
- [43] Samal, S.S., Krishnan, J., Esfahani, A.H., Lüders, C., Weber, A., Radulescu, O.: Metastable regimes and tipping points of biochemical networks with potential applications in precision medicine. In: *Automated Reasoning for Systems Biology and Medicine*, pp. 269–295. Springer (2019)
- [44] Schnoerr, D., Sanguinetti, G., Grima, R.: Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J. Phys. A: Math. Theor.* **50**(9), 093001 (2017)
- [45] Soltani, M., Vargas-Garcia, C.A., Antunes, D., Singh, A.: Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes. *Plos Comput. Biol.* **12**(8), e1004972 (2016)
- [46] Taniguchi, Y., Choi, P., Li, G., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.: Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010)
- [47] Thattai, M., van Oudenaarden, A.: Intrinsic noise in gene regulatory networks. *P. Natl. Acad. Sci. USA* **98**, 151588598 (2001)